Article

Neuron

Task-Dependent Changes in the Large-Scale Dynamics and Necessity of Cortical Regions

Highlights

- Mice navigated in virtual reality while performing one of three related tasks
- All dorsal cortex contributed to memory-dependent but not visually guided navigation
- Higher task demands induced decorrelation of whole-cortex Ca^{2+} activity patterns
- A modular RNN model suggested that differences in computations can explain results

Authors

Lucas Pinto, Kanaka Rajan, Brian DePasquale, Stephan Y. Thiberge, David W. Tank, Carlos D. Brody

Correspondence

dwtank@princeton.edu (D.W.T.), brody@princeton.edu (C.D.B.)

In Brief

Pinto et al. show that more complex perceptual decisions engage more diverse and spatially distributed computations across the cortex than simpler decisions, even when sensory stimuli and motor output are held constant.

Pinto et al., 2019, Neuron *104*, 810–824
November 20, 2019
https://doi.org/10.1016/j.neuron.2019.08.025

CellPress

CellPress

Neuron
Article

Task-Dependent Changes in the Large-Scale Dynamics and Necessity of Cortical Regions

Lucas Pinto,[1] Kanaka Rajan,[2,6] Brian DePasquale,[1] Stephan Y. Thiberge,[3] David W. Tank,[1,3,4,7,*] and Carlos D. Brody[1,4,5,7,8,*]

[1]Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA
[2]Joseph Henry Laboratories of Physics and Lewis–Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA
[3]Bezos Center for Neural Dynamics, Princeton University, Princeton, NJ 08544, USA
[4]Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA
[5]Howard Hughes Medical Institute, Princeton University, Princeton, NJ 08544, USA
[6]Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10014, USA
[7]Senior Author
[8]Lead Contact
*Correspondence: dwtank@princeton.edu (D.W.T.), brody@princeton.edu (C.D.B.)
https://doi.org/10.1016/j.neuron.2019.08.025

SUMMARY

Neural activity throughout the cortex is correlated with perceptual decisions, but inactivation studies suggest that only a small number of areas are necessary for these behaviors. Here we show that the number of required cortical areas and their dynamics vary across related tasks with different cognitive computations. In a visually guided virtual T-maze task, bilateral inactivation of only a few dorsal cortical regions impaired performance. In contrast, in tasks requiring evidence accumulation and/or post-stimulus memory, performance was impaired by inactivation of widespread cortical areas with diverse patterns of behavioral deficits across areas and tasks. Widefield imaging revealed widespread ramps of Ca^{2+} activity during the accumulation and visually guided tasks. Additionally, during accumulation, different regions had more diverse activity profiles, leading to reduced inter-area correlations. Using a modular recurrent neural network model trained to perform analogous tasks, we argue that differences in computational strategies alone could explain these findings.

INTRODUCTION

Making decisions based on sensory stimuli is a crucial cognitive ability, and its neural underpinnings have been the subject of intense scrutiny for the past decades. A particular puzzle has been the observation of task-related activity across the cortex, often with similar response patterns (Allen et al., 2017; Brody and Hanks, 2016; Dotson et al., 2018; Gold and Shadlen, 2007; Hernández et al., 2010; Koay et al., 2019; Scott et al., 2017; Siegel et al., 2015; Steinmetz et al., 2018). At face value, these findings suggest the existence of distributed cortical computations. However, systematic optogenetic inactivation studies have generally found very localized behavioral effects, even when widespread activity was observed under the same experimental conditions (Allen et al., 2017; Guo et al., 2014; Zatka-Haas et al., 2019). This highlights the possibility that such apparently distributed computations might serve other roles not relevant to the decision on individual trials. However, perceptual decisions can vary in difficulty, stimulus complexity, and cognitive computation requirements. Decision-making tasks can consist of fixed one-to-one sensory-motor mappings (Allen et al., 2017; Goard et al., 2016; Guo et al., 2014; Harvey et al., 2012; Musall et al., 2018; Pinto and Dan, 2015) or have multidimensional stimulus sets, trial-varying difficulty, or sensory-motor transformations that rely on stimulus working memory (Akrami et al., 2018; Brunton et al., 2013; Gold and Shadlen, 2007; Morcos and Harvey, 2016; Odoemene et al., 2018; Pinto et al., 2018). Additionally, tasks may vary widely in terms of the timescales over which decisions unfold.

The aforementioned inactivation findings were obtained while mice performed discriminations between two stimuli over relatively short timescales, so it remains unclear whether they are generalizable to other kinds of perceptual decisions. Indeed, it has been proposed that the degree to which different brain areas are causally involved in a behavior varies with the cognitive requirements of the task (Fuster, 1997; Lashley, 1931). However, direct evidence supporting this idea is lacking. Here we used large-scale optical perturbation and recording methods to perform an unbiased survey of both the Ca^{2+} activity patterns and perturbation effects across the entire dorsal cortex of mice performing, in the same virtual reality (VR) environment, three related tasks with different cognitive requirements. We show that, in contrast to what has been observed previously, distributed dorsal cortical areas appear to contribute to performance of an evidence-accumulation task and a memory-guided task but not a visually guided task. The nature of the behavioral effects depended on both the inactivated region and the task. Moreover, the magnitude of performance drops was correlated

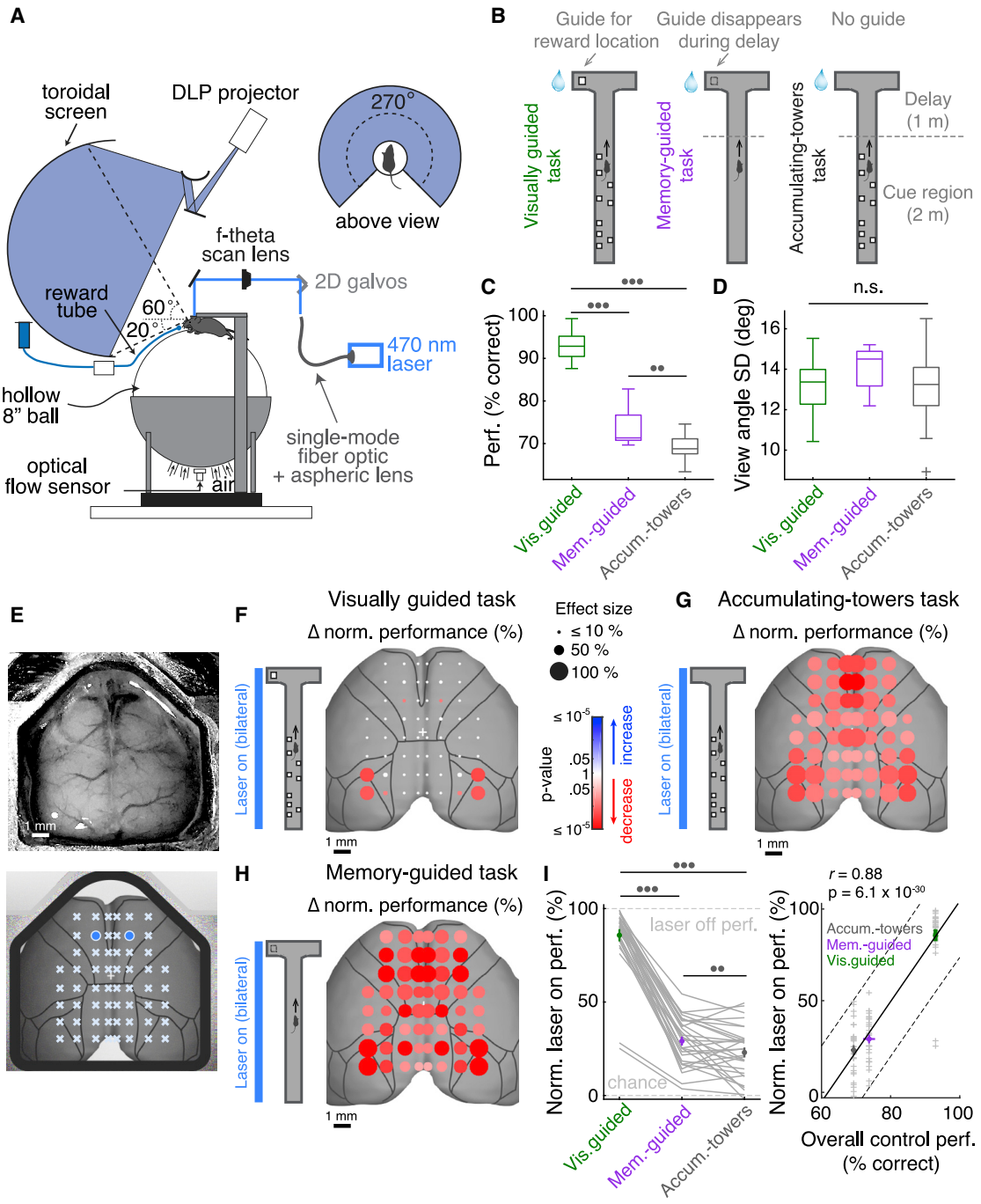

Figure 1. Widespread Cortical Involvement in Evidence- and Memory-Based Navigation

(A) Schematic of the VR setup integrated with a scanning laser for optogenetic inactivation.

(B) Schematic of the three tasks. Right: accumulating-towers task. The animal navigates a 3-m virtual T-maze to retrieve a reward in the arm corresponding to the side with the majority of towers, which appear transiently on either side anywhere in the first 2 m. Left: visually guided task. Maze and tower stimuli are the same, but, additionally, a tall visual guide visible from the beginning of the maze indicates the reward location. Center: memory-guided task. The maze is the same, but there are no tower stimuli, just a distal visual guide that disappears during the final 1 m.

(C) Overall performance of the three tasks (n = 36, 31, and 8 mice with at least 100 trials for the accumulating-towers, visually guided, and memory-guided tasks, respectively). Circles indicate statistical significance (two, $p < 0.01$; three, $p < 0.001$; Tukey's post hoc test).

(D) Average SD of view angle trajectories across spatial positions for the three tasks, illustrating similar motor behavior.

(E) Top: VGAT-ChR2-EYFP mouse with the cleared skull preparation. Bottom: schematic showing targeting of 29 bilateral cortical patches spanning many areas (crosses), overlaid on a reference from the Allen Brain Atlas. Circles illustrate bilateral targeting of a pair of homotopic areas.

(legend continued on next page)

with how demanding the tasks were. Although both the accumulation and the visually guided tasks were associated with global Ca^{2+} activity, we observed differences in cortical dynamics that appeared to be related to the different underlying cognitive computations. These data suggest that performance of tasks with more complex cognitive demands results in engagement of diverse computations across widespread regions of the cortex. Finally, we trained a recurrent neural network (RNN) model with a cortex-inspired modular architecture to perform both the visually guided and the accumulation task with fixed synaptic connectivity. Based on this model, we argue that differences in the underlying computations without changes in connectivity across different tasks could themselves explain our imaging and inactivation findings.

RESULTS

Widespread Effects of Cortical Inactivation in Memory-Dependent Navigation Tasks

We trained head-fixed mice to perform three tasks, all of which happened in the same VR T-maze (Figures 1A and 1B; Video S1). One of the tasks was our previously developed "accumulating-towers task" (Pinto et al., 2018; Figure 1B). As mice navigated through the initial 2 m (~4 s) of the T-maze, salient objects (towers) appeared transiently on either side. After a 1-m (~2-s) delay, the mice turned to the arm corresponding to the side with the highest tower count to retrieve a reward. The towers occurred at random positions and counts in each trial so that the mice experienced a wide range of magnitudes of sensory evidence, defined as # right − # left (Δ) towers. Compatible with previous behavioral analyses (Pinto et al., 2018), performance was modulated by the amount of evidence and was sensitive to single-tower count differences (Figures 1C and S1A; overall performance: 69.3% ± 0.4% correct, mean ± SEM, n = 36 mice). Moreover, a logistic regression analysis suggested spatially uniform evidence integration (Figure S1B). The second task, dubbed "visually guided," maintained all maze features, with the addition of a tall visual guide in the rewarded arm, visible from the beginning of the maze (Figure 1B). Thus, although sensory and motor features were highly similar between the two tasks (Figures 1C, 1D, and S1), only the accumulating-towers task required gradual evidence accumulation and working memory. Consistent with lower cognitive demands, performance was much higher in the visually guided task and had very little sensitivity to Δ towers (Figures 1C and S1C; overall performance: 92.7% ± 0.6% correct, mean ± SEM, n = 31). The third task, called "memory-guided," was designed to control for the long post-stimulus delay in the accumulating-towers task. This task happened in the same maze but did not contain towers, just a distal reward-indicating visual guide, which disappeared in the last 1 m of the maze. Thus, the task required no evidence accumulation, like the visually guided task, but contained a post-stimulus delay, like the accumulating-towers task. However, the absence of proximal towers resulted in less salient visual stimuli. Indeed, despite not requiring accumulation, the behavioral difficulty of the memory-guided task appeared closer to the accumulating-towers than to the visually guided task, although performance was still significantly higher (Figures 1C and S1C; 73.2% ± 1.4% correct, mean ± SEM, n = 8, $p_{task} = 1.9 \times 10^{-44}$, one-way ANOVA, accumulating towers versus memory-guided: p = 0.007, Tukey's post hoc test).

We first asked which dorsal cortical regions are required for the mice to perform the tasks. To answer this, we trained mice expressing Channelrhodopsin-2 (ChR2) transgenically in inhibitory interneurons (VGAT-ChR2-EYFP) and rendered their intact skulls optically transparent (Guo et al., 2014). This allowed us to scan a blue laser over the dorsal cortex, randomly targeting 29 bilateral homotopic locations spanning sensory, motor, and association cortices, in a subset of behavioral trials (Figure 1E). We confirmed electrophysiologically that this approach robustly inactivated discrete patches of the cortex (~1.5–2-mm radius) in a laser power-dependent fashion with little activity rebound (Figure S2). First we performed whole-trial inactivation experiments ($0 \leq y \leq 300$ cm) as mice performed the visually guided task. Significant inactivation effects were restricted to visual cortical areas and, to a lesser extent, the medial premotor cortex (Figure 1F; $p < 0.01$, bootstrapping). These results are unlikely to be due to inadvertent inhibition of underlying subcortical regions such as the striatum, where ChR2 is not expressed (Guo et al., 2014). However, given the inactivation spreads, we cannot rule out that the effects attributed to the premotor cortex are partly due to inactivation of the underlying medial prefrontal cortex, which is required for perceptual discrimination (Pinto and Dan, 2015). Regardless, our results are broadly compatible with previous reports that, in the dorsal cortex, only premotor and sensory areas are required for simple perceptual decisions (Allen et al., 2017; Guo et al., 2014; Zatka-Haas et al., 2019), highlighting the fact that the VR-navigation component of our tasks does not qualitatively change our findings.

(F) Effects of whole-trial inactivation of 29 bilateral cortical patches on performance of the visually guided task (n = 5 mice, 8,140 "laser on" trials; mean, 271.3 trials/location). The size of each circle indicates the size of the effect (caption on the right), given by the normalized performance drop between "laser on" and "laser off" trials (STAR Methods). The color indicates the sign of the effect (red, decreased performance), and the saturation is proportional to the p value (color bar), thresholded so that non-significant effects appear white.

(G) Same as (F) for the accumulating-towers task (n = 11 mice, 9,443 "laser on" trials; mean, 314.8 trials/location).

(H) Same as (F) and (G) for the memory-guided task (n = 10 mice, 3,750 "laser on" trials; mean, 129.3 trials/location).

(I) The magnitude of behavioral deficit is related to baseline task difficulty. Left: comparison of normalized performance during inactivation across tasks. Lines: each of the 29 bilateral locations. Two circles, $p < 0.01$; three circles, $p < 0.001$; one-sided paired t test (the effect was larger for 19 of 29 regions in the accumulation versus the memory-guided task). Colored circles and error bars, mean ± SEM across locations. Right: significant relationship between overall performance (data in C) and normalized "laser on" performance (left). Crosses, average "laser on" performance for each of the 29 bilateral locations; solid line, best linear fit (the slope is significantly different from zero, 95% confidence interval); dashed lines, 95% confidence interval; colored circles and error bars, mean ± SEM.

See also Figures S1–S3 and Video S1.

Next we repeated the experiments as mice performed the accumulating-towers task. Surprisingly, in contrast with the visually guided task, inactivating each of the 29 bilateral regions resulted in significant decreases in overall performance (Figure 1G; $p < 0.01$, bootstrapping). This result was not due to non-specific light effects because we observed no significant effects in control mice not expressing ChR2 (Figures S2M–S2S; ChR2+ effects were significantly larger than controls for 27 of 29 locations). Furthermore, although possibly an overestimate of the amount of required cortical territory, the result cannot be exclusively explained by large inactivation spreads encompassing multiple cortical areas. Repeating the experiments with much smaller laser powers and, thus, more spatially restricted inactivation yielded qualitatively similar results with widespread effects even at powers likely resulting in incomplete local silencing (Figures S2K and S2L). Moreover, even at the lowest tested laser power, the spatial map of inactivation effects did not resemble that of the visually guided task, arguing against the possibility that both tasks require the same set of cortical areas but that sensitivity to inactivation is higher for the accumulating-towers task. However, within visual cortical areas (required for both tasks), the accumulating-towers task was associated with more sensitivity to the magnitude of perturbation than the visually guided task (Figure S2J).

The accumulating-towers task is different from the visually guided task in two ways. First, the visually guided task does not require evidence accumulation; second, it does not contain a post-stimulus delay. To identify which of the two features contributed most to the difference between the tasks, we also performed inactivations during the memory-guided task, which contains a delay but no evidence accumulation. Interestingly, as in the accumulating-towers task, the inactivation of all bilateral locations resulted in significant performance drops (Figure 1H; $p < 0.01$, bootstrapping). To directly compare the effect magnitudes between the different tasks, we normalized inactivation-induced performance decreases so that 0 reflected chance performance and 100 reflected control (laser off) performance (STAR Methods). Effect sizes varied significantly as a function of tasks and the location of inactivation (Figure 1I; $p_{task} = 5.1 \times 10^{-32}$, $p_{location} = 9.7 \times 10^{-7}$, 2-way ANOVA), indicating that the requirements of different regions depend both quantitatively and qualitatively on the tasks' computational demands. On average, the magnitude of the effect was significantly higher in the accumulating-towers compared with the memory-guided task ($p = 0.006$, one-sided paired t test, 19 of 29 regions had a larger effect in the accumulation task). In addition, the magnitude of the effects was significantly correlated with how demanding the tasks were, as indexed by overall baseline performance (Figure 1I, $r = 0.88$, $p = 6.1 \times 10^{-30}$, Pearson correlation). These findings indicate that, although post-stimulus memory is a key contributor to the widespread cortical requirement, the evidence accumulation process is also important because it is associated with significantly larger deficits. Further evidence for this comes from similarly widespread effects upon selective cue region inactivations during the accumulating-towers task (Figure S3A). Thus, tasks with different underlying cognitive computations are associated with different sensitivities to, and spatial effect distributions upon, cortical inactivations.

Inactivation Effects Are Task and Region Dependent

We next wondered whether the nature of the behavioral effects of inactivation varied across the targeted regions. First we analyzed the effect of inactivation on other behavioral measures beyond overall performance and observed differences between cortical regions (Figures 2A, 2B, and S3). For instance, during the accumulating-towers task, we noted a large effect on side biases upon bilateral inactivation of frontal, but to a much lesser extent, posterior areas, although the preferred side was idiosyncratic (Figure 2A). Note that a trivial explanation such as side differences in laser power cannot account for these results; hardware issues would affect all mice equally, and implant inhomogeneities are unlikely to specifically affect the same regions across animals. Moreover, no significant biases occurred during the visually guided task (Figure S3A). We also observed different effects on running speeds and trajectory lengths that depended on both the region and the task the mice were performing (Figures 2B, S3B, and S3C). Importantly, however, these effects were unlikely to reflect gross motor deficits because the frequency of motor errors was largely unaffected (Figure S4C). Cortical regions also differed according to the behavioral effects of their inactivation during the memory-guided task, but with a different spatial distribution (Figures 2C and S3D). For instance, we again observed idiosyncratic increases in side bias, but with the inactivation of more posteromedial regions resulting in larger biases instead (Figure 2C).

To better quantify these differences, we combined these various measures of behavioral deficits during the accumulating-towers or the memory-guided task and grouped them using hierarchical clustering (Figures S3E–S3G; STAR Methods). During both tasks, the different areas could be grouped into three clusters, which significantly differed in the magnitude of the effect on all behavioral indicators (one-way ANOVA, $p \leq 0.01$). Compatible with our previous analyses, however, the clusters contained different cortical regions in the two tasks (Figures 2D and 2E). Overall, our results suggest the existence of spatially distributed processes underlying more demanding perceptual decisions, but with different regions contributing to different aspects of the behavior, contrary to previous conclusions (Lashley, 1931). Further, they support the notion that different tasks with different underlying computations, even when resulting in similar overall demands, are associated with distinct whole-cortex activity patterns.

Cortical Dynamics Are Task Specific

The inactivation results thus suggest differences in the underlying large-scale cortical activity, even between related tasks. To observe this directly, we performed wide-field Ca^{2+} imaging at mesoscale resolution ($\sim$70 μm) over the dorsal surface of the cortex, taking advantage of the same cleared skull preparation applied to mice expressing GCaMP6f in cortical excitatory neurons (Emx1-Ai93, n = 6; Figures 3A, 3B, S4, and S5). Note that, despite the incidence of ictal events in some mice from this line (Steinmetz et al., 2017), we have shown previously that it has statistically indistinguishable behavior (Pinto et al., 2018) and Ca^{2+} dynamics (Koay et al., 2019) from other mouse lines.

The wide-field signal is composed of both somatic and neuropil activity, primarily from superficial layers (Allen et al., 2017;

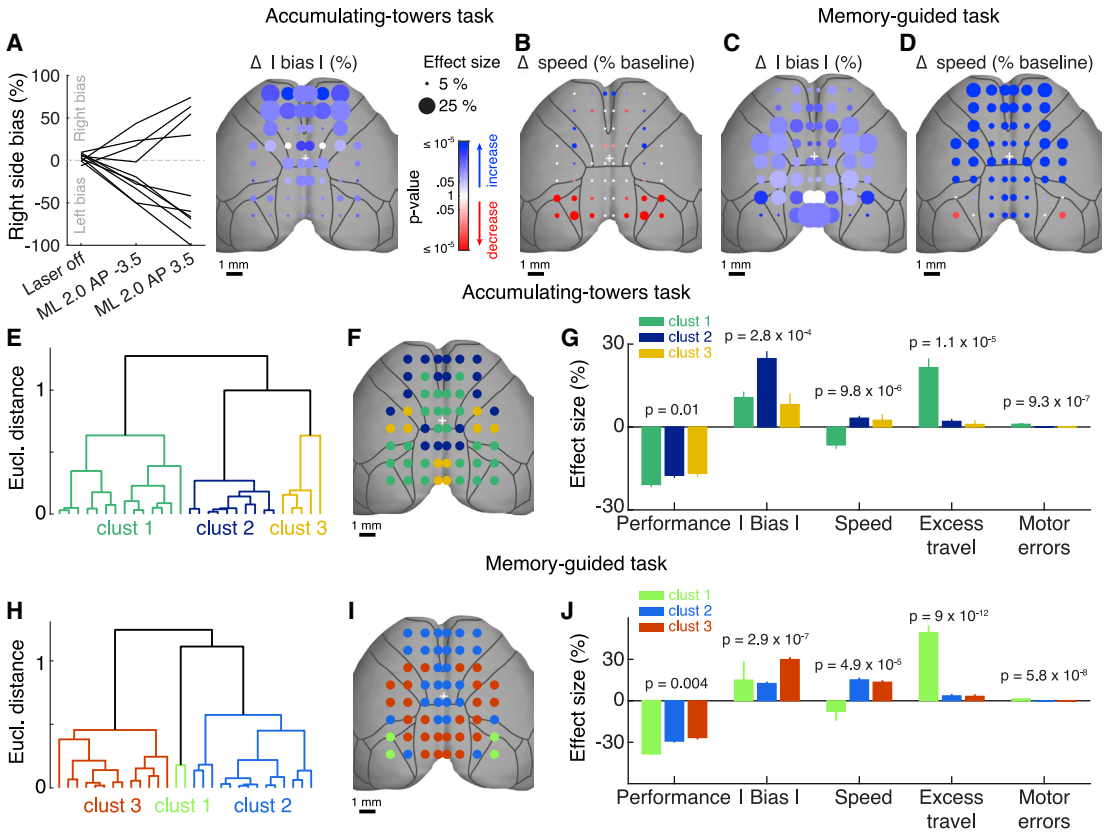

Figure 2. Location- and Task-Specific Spatial Patterns of Behavioral Deficits

(A) Bilateral whole-trial inactivation of frontal regions during the accumulating-towers task led to large idiosyncratic side biases. Left: magnitude of side biases for individual mice (n = 11, black lines). Shown are biases in "laser off" and "laser on" trials for a sample posterior and a sample frontal region. Right: effect of inactivating each region on absolute side bias. The size of each circle indicates the size of the effect (caption on the right). The color indicates the sign of the effect (red, decrease; blue, increase), and the saturation is proportional to the p value (color bar; STAR Methods), thresholded so that non-significant effects appear white.

(B) Different effects of inactivating bilateral regions on running speed during the accumulating-towers task. Conventions as in (A).

(C and D) Inactivation effects on absolute bias (C) and speed (D) for the memory-guide task. Conventions as in (A) and (B). Note the different spatial pattern.

(E) Dendrogram showing that cortical regions cluster into three groups according to the pattern of behavioral deficits in the accumulating-towers task.

(F) Each of the 29 bilateral cortical regions is color-coded according to the cluster to which they belong (colors in E).

(G) Significantly different effect sizes for five behavioral indicators across the three clusters (color code as in E and F). The p values on top are from one-way ANOVAs performed separately for each indicator. Error bars, ± SEM.

(H–J) Same as (E)–(G) for the memory-guided task.

See also Figure S3.

Clancy et al., 2019; Ma et al., 2016a; Makino et al., 2017; Xiao et al., 2017), so that it likely reflects the combined activity of superficial somata and neuropil and distal axons arriving through layer 1. This signal has been shown in a variety of GCaMP6-expressing mouse lines to be correlated with local spiking activity (Allen et al., 2017; Clancy et al., 2019; Ma et al., 2016b; Makino et al., 2017; Xiao et al., 2017) and to be largely dependent on local synaptic transmission (Makino et al., 2017), but comparison with cellular-resolution imaging suggested higher correlation with layer 1 than layer 2/3 signals (Allen et al., 2017). Despite not allowing us to isolate local computations definitively, it allowed us to probe mesoscale dynamics simultaneously from the same regions we inactivated and, crucially, to compare dynamics across tasks.

We measured Ca^{2+} dynamics during both the accumulating-towers and the visually guided tasks from blocks of trials within the same behavioral sessions (n = 25 from 6 mice, 2–5 sessions/mouse). We focused on these two tasks because they are directly comparable in terms of sensory stimuli and motor output (Figure S1). Similar to recent reports (Allen et al., 2017; Musall et al., 2018; Orsolic et al., 2019; Zatka-Haas et al., 2019), both tasks were associated with widespread Ca^{2+} activity (Figures 3B–3F and S6; Video S2). In particular, we observed prominent activity ramps across the dorsal cortex, leading to the reward location during both tasks (Figures 3E and 3F). These correlated ramps, observed here for the first time in VR navigation, have been reported previously in other tasks (Allen et al., 2017; Musall et al., 2018; Orsolic et al., 2019), suggesting that

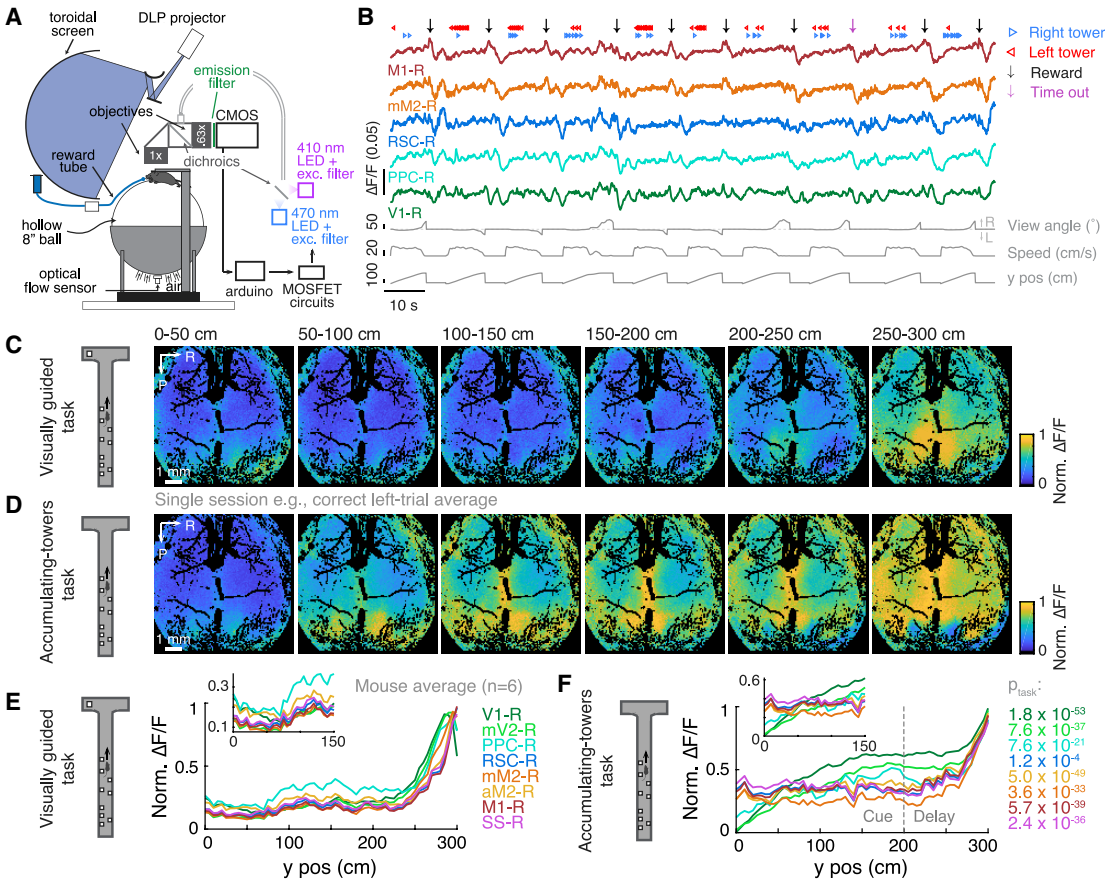

Figure 3. Task-Dependent Cortical Dynamics

(A) Schematics of the VR setup integrated with a wide-field macroscope.

(B) Example ΔF/F traces extracted for five ROIs (labels on the left), with behavioral events for ten consecutive trials of the accumulating-towers task (captions on the right).

(C) Pixel-wise Ca^{2+} activity maps during the visually guided task, averaged over correct left-choice trials and maze regions as labeled on top. Black pixels correspond to vasculature and headplate mask and were not included in the analyses.

(D) Same as (C) for accumulating-towers task trials during the same behavioral session.

(E) Spatially downsampled activity for all contralateral ROIs (i.e., right hemisphere) during the visually guided task, averaged over mice. The inset is a magnification of the same data, focused on the first half of the maze. Error bars are omitted for clarity.

(F) Same as (E) for the accumulating-towers task. The p values on the right are for the task factor in a 2-way repeated measures ANOVA (factors task and position). In (C)–(F), ΔF/F was normalized separately for each task and ROI to emphasize relative activity timing. See also Figures S4–S6 and Video S2.

they might be a general phenomenon of decision-making. Additionally, extending these findings, and compatible with our inactivation results, we observed differences in the large-scale activity patterns between the two tasks, primarily in the evidence accumulation period (Figure 3F). To summarize these effects, we averaged cortical activity over anatomically defined regions of interest (ROIs; Figure S5). The temporal profiles of Ca^{2+} activity of all ROIs were significantly different between the two tasks (Figures 3E, 3F, S6A, and S6B; $p < 0.001$, n = 6 mice, two-way ANOVAs with repeated measures). Similar conclusions could be drawn from pixel-wise analyses (Figures S6C–S6E).

We also observed task differences in the large-scale correlations between cortical regions, which were also modulated by whether the mice were engaged in a task (Figure 4A). Specifically, although the general structure of inter-ROI correlations was roughly preserved across tasks and spontaneous running behavior (being largely explained by the physical inter-ROI distance; Figure S7A), we found that the overall magnitude of the correlations decreased significantly when the mice were engaged in a task and were smallest during the accumulating-towers task (Figure 4B; $p = 3.1 \times 10^{-23}$, one-way ANOVA with repeated measures; $p < 10^{-4}$ for all pairwise comparisons, Tukey's post hoc test, n = 16 ROIs). Thus, because running patterns were indistinguishable between the two tasks (Figures 1D and S1), different cognitive computations themselves, not motor activity, appeared to modulate inter-ROI correlations.

Further inspection of the data suggested that areas could be grouped based on their correlations and that decreases in correlation happened primarily across rather than within these groups (i.e., note the darker regions in the off-diagonals; Figure 4A). To test this directly, we used hierarchical clustering to group the ROIs (STAR Methods), obtaining four clusters: one with visual

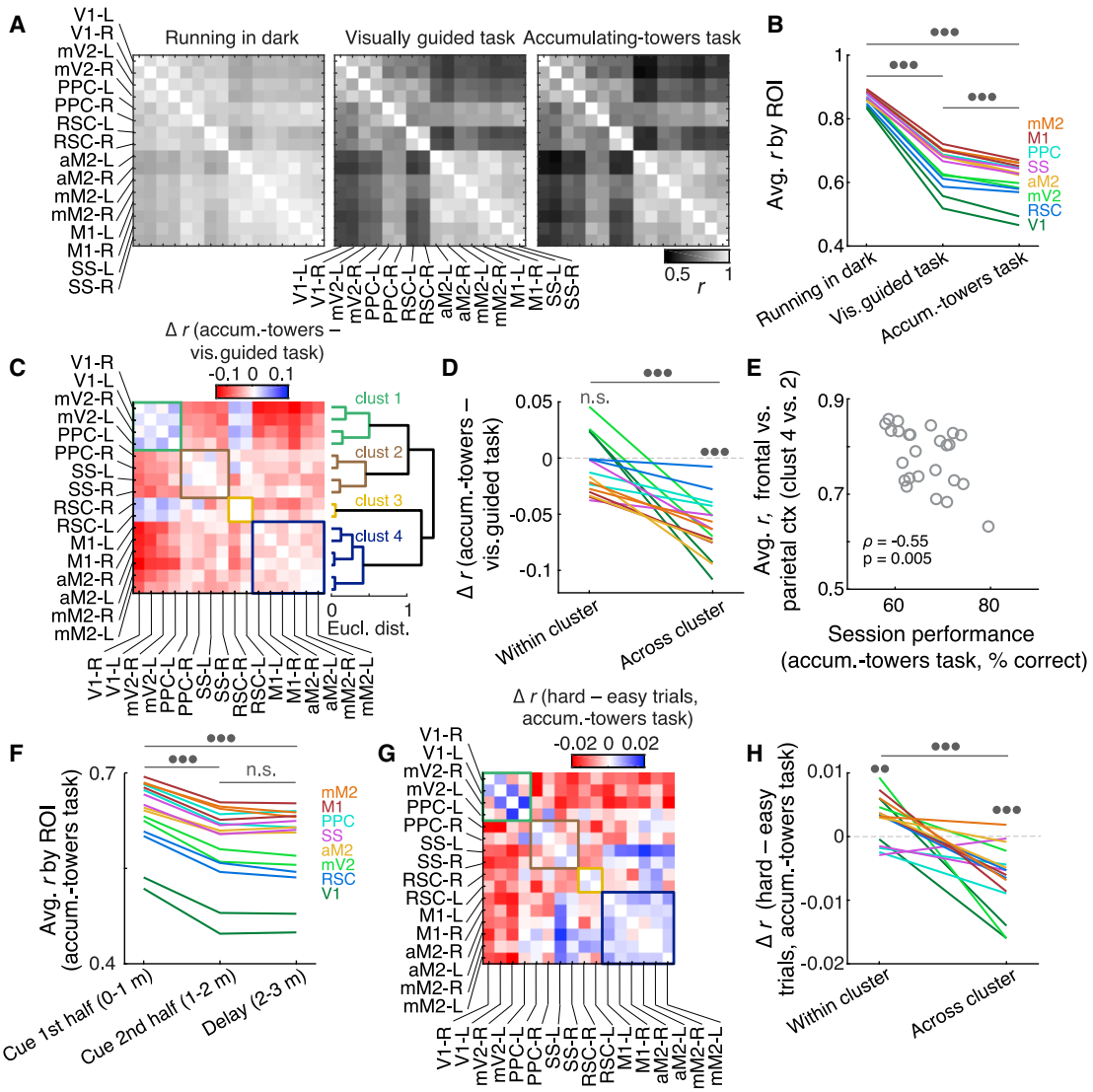

Figure 4. Correlations between ROIs Are High but Modulated by Behavioral Context

(A) Average pairwise correlation matrices between all 16 ROIs for different conditions labeled on top.

(B) Average correlation between each ROI and the other 15 as a function of behavioral context. Circles, $p < 0.001$.

(C) Left: difference in correlation matrices between the accumulating-towers and visually guided tasks, with ROIs sorted according to their cluster membership (right). Right: dendrogram showing hierarchical clustering of ROIs into 4 groups.

(D) For each ROI, we compared the average task-related change (Δr) in correlation with ROIs within the same cluster or outside of it. Color code as in (B). One circle, $p < 0.05$; three circles, $p < 0.001$.

(E) Relationship between session performance (n = 25 from 6 mice) and average correlation between the ROIs in the frontal cortex versus parietal cortices (clusters 4 and 2 in C).

(F) Average correlation between each ROI and the other 15 as a function of maze region in the accumulating-towers task. Circles, $p < 0.001$. n.s., not significant.

(G) Difference in correlation matrices between hard and easy trials in the accumulating-towers task.

(H) Average trial difficulty-related change in correlation with ROIs within the same cluster or outside of it (accumulating-towers task). Color code as in (F). Two circles, $p < 0.01$; three circles, $p < 0.001$.

See also Figure S7.

areas, one with parietal areas, one with retrosplenial areas, and one with frontal areas (Figure 4C). We then calculated the change in correlation in the accumulating-towers versus the visually guided task and compared, for each ROI, the amount of change within and across its cluster. Confirming our visual impression, correlations during the accumulating-towers task significantly decreased only across clusters (Figure 4D; $p = 8.0 \times 10^{-8}$ and 0.26, respectively, for across and within; paired t test, n = 16 ROIs; differences between specific cluster pairs were not significant, p = 0.09, one-way ANOVA with repeated measures). Next

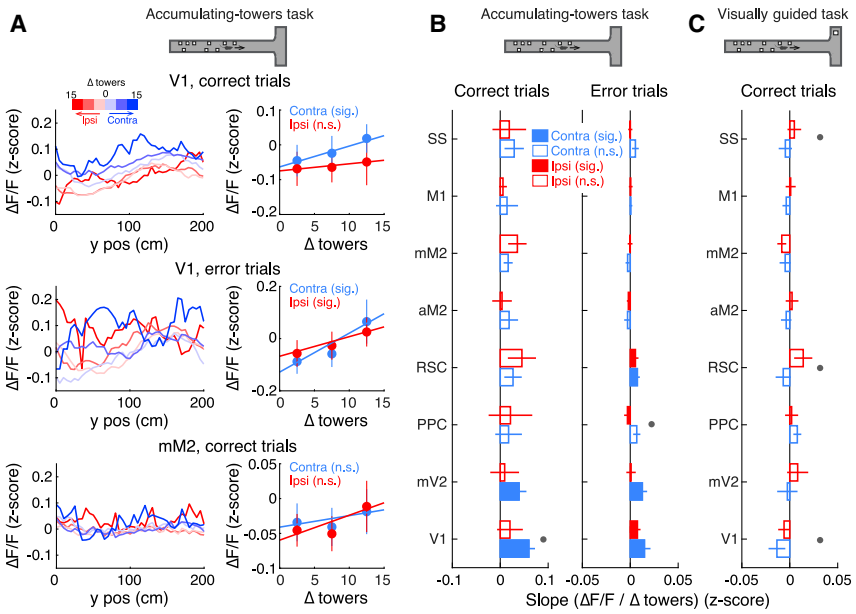

Figure 5. Large-Scale Ca^{2+} Activity Is Sensitive to Sensory Evidence Only during the Accumulating-Towers Task

(A) Left: ROI examples (labels on top) of average *Z*-scored ΔF/F as a function of the final amount of sensory evidence (binned, color bar on top) during the accumulating-towers task. Lines: mouse averages (n = 6). Error bars are omitted for clarity. Right: linear fits to quantify the amount of evidence tuning, corresponding to the examples on the left. Data points, mouse averages; error bars, ± SEM. Lines are average best fits from 50 bootstrapping iterations. Sig., the slope is significantly different from 0; n.s., the slope is not significantly different from 0.

(B) Average slopes of linear fits for each ROI during correct (left) and error (right) trials for the accumulating-towers task. Error bars, SD from bootstrapping (n = 50 iterations). Filled bars indicate slopes that are significantly different from 0. Circles indicate significant differences between contra- and ipsilateral slopes for each ROI.

(C) Same as (B) for correct trials in the visually guided task. Note no significant tuning to sensory evidence.

we wondered whether these changes in correlation, in addition to reflecting changes in task demands, were also related to performance of the task. Interestingly, although the general decrease in correlations did not itself appear to be correlated with performance (data not shown), the amount of decorrelation between parietal and frontal cortices (cluster 2 versus cluster 4) was significantly related to how well the animals performed the accumulating-towers task on a session-by-session basis (Figure 4E; $\rho = -0.55$, p = 0.005, Spearman correlation, corrected for multiple comparisons; this appeared to be largely driven by decorrelation between the somatosensory cortex and frontal areas, $\rho = -0.56$, p = 0.004). There were no significant relationships between overall ROI Ca^{2+} activity levels and session performance (data not shown), indicating that this finding is specific to the inter-ROI correlations. This result shows that even apparently small changes in correlation (~0.1) can be associated with sizable behavioral effects (~10% correct). It is unclear how these results relate to findings that the mammalian cortex has antagonistic networks with anti-correlated blood-oxygen-level-dependent (BOLD) signals during resting states and task execution (Raichle, 2015; Wig, 2017) because our mice were running under all conditions, and running itself can modulate large-scale cortical correlations (Clancy et al., 2019).

To further explore the relationship between task demands and large-scale correlational structures, we asked whether they are also modulated by within-task variations in cognitive load. We reasoned that, on average, both memory load and stimulus complexity increased with trial progression during the accumulating-towers task, which should, in turn, decrease correlations. Indeed, we observed a significant decrease in correlations during later trial epochs (Figure 4F; $p = 1.2 \times 10^{-17}$ across epochs, n = 16 ROIs, one-way ANOVA with repeated-measures), which was not observed in the visually guided task (Figure S7B). Along the same lines, we reasoned that trial difficulty should also modulate correlations. Across-cluster correlations were, in fact, significantly lower in hard compared with easy trials (Figures 4G and 4H; $p = 1.7 \times 10^{-4}$, two-sided paired t test). Similar results were obtained at finer spatial resolutions (Figures S7C–S7F). These findings further lend support to the notion that cognitive load modulates correlations, even when task rules are held constant. Together, our correlation results suggest that different cortical areas form functional modules that perform more independent computations when task demands increase.

Distributed Representations of Task Variables throughout the Dorsal Cortex

Next we asked which aspects of the accumulating-towers task are encoded in dorsal cortical Ca^{2+} activity. We first wondered whether the differences observed between the two tasks during the nominal cue region (Figure 3) were, in fact, related to the evidence accumulation computation. We plotted the average activity of each ROI separately for different amounts of final ipsi- and contralateral sensory evidence. Posterior ROIs, in particular V1 (Figure 5A) and mV2, appeared to have activity that increased with the amount of contralateral evidence, unlike frontal ROIs (e.g., mM2; Figure 5A). To better quantify this, we fit lines to the cue-period-averaged activity of each ROI as a function of contra- or ipsilateral evidence. During correct trials, V1 and mV2 activity was significantly tuned to the amount of contra- but not ipsilateral evidence (Figure 5B). Interestingly, during error trials, significant tuning to ipsilateral evidence was present in several ROIs (Figure 5B). In contrast, we observed no systematic relationship between activity and sensory evidence during the visually guided task (Figure 5C), consistent with our behavioral data (Figures 1C and S1D).

The analysis above, however, assumes that tuning is static throughout the maze, and it does not take into account finer-grained spatial patterns of activity that may be present in the data. To address this, we built separate linear decoders for accumulated evidence, choice, and previous choice at each position

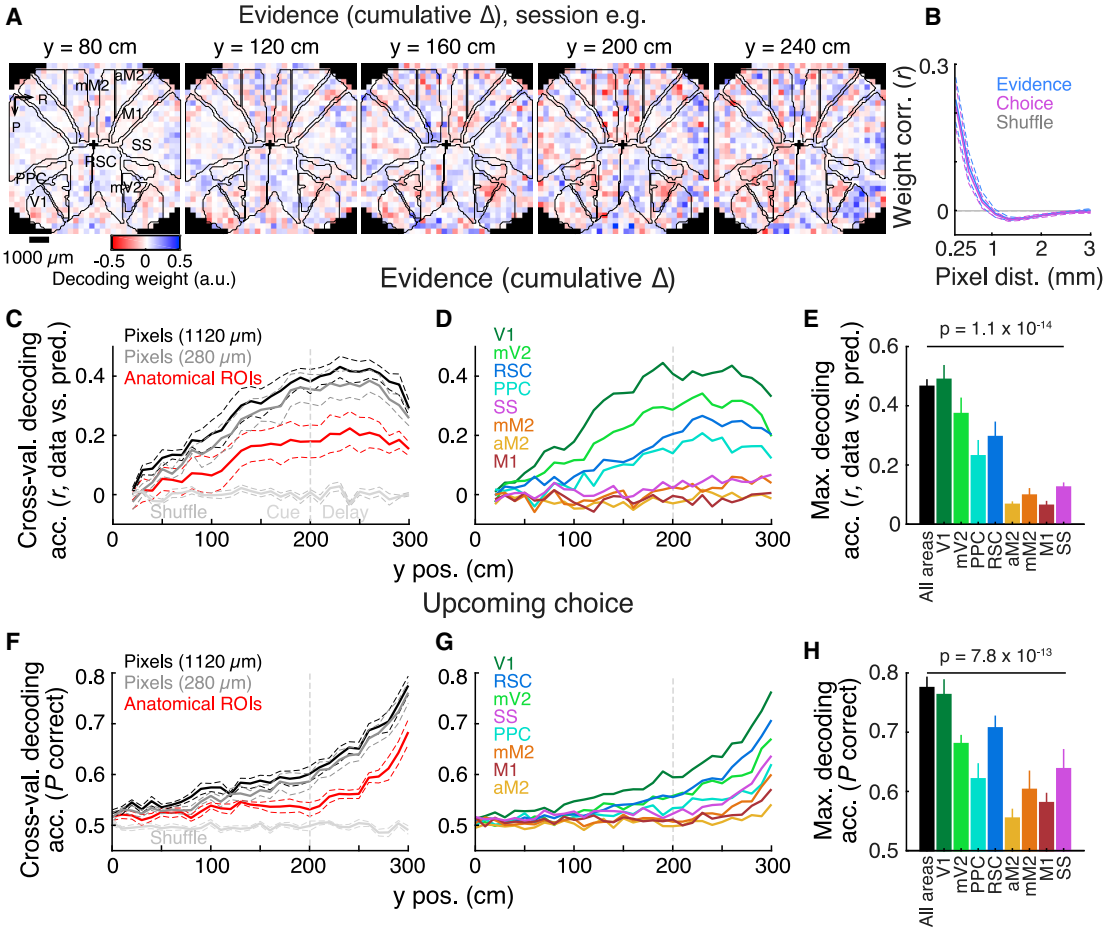

Figure 6. Distributed Coding of Task Variables during the Accumulating-Towers Task

(A) Evidence decoding weights plotted in pixel space (~280 × 280 μm) at different maze positions, from an example session. Outlines show anatomical ROIs (labels on the left); crosses indicate bregma location.

(B) Average Pearson correlation of evidence and choice decoding weights as a function of the distance between pixels. Dashed lines, ± SEM across mice (n = 6). Shuffles were obtained by randomizing the pixel coordinates 50 times.

(C) Cross-validated performance of linear decoders of cumulative evidence (Δ towers) across maze y positions, using activity averaged within anatomically defined ROIs or smaller pixels. Dashed lines, ± SEM (n = 6 mice).

(D) Accuracy of decoders employing pixels from only pairs of homotopic ROIs across maze y positions. Error bars are omitted for clarity.

(E) Maximal accuracy for decoders using pixels from the whole cortex (black, 1,120 × 1,120 μm) or only from homotopic ROI pairs. Error bars, ± SEM.

(F–H) Same as (C)–(E) for upcoming choice. (F) Decoding accuracy of pixel- or ROI-based decoders. (G) Accuracy of decoders using pixels from pairs of homotopic ROIs. (H) Maximal decoding accuracy using pixels from the whole cortex or only pairs of ROIs.

See also Figures S8 and S9.

in the maze. We used Ca^{2+} activity that was either averaged within each anatomically defined ROI or within finer-grained pixels (~280 × 280 μm; Figure 6A). We could decode the amount of accumulated evidence above chance levels from the start of the cue period, but with higher and stable accuracy after the second half of the maze, using either ROIs or pixels (Figure 6C). However, cross-validated decoding accuracy was significantly higher when pixels were used (Figures 6C and S8A; p = 0.02, two-sided paired t test), indicating that more information is present on meso than on large scales. Moreover, the decoding accuracy was comparable for pixel sizes as large as ~1120 × 1120 μm and lower for larger pixels (Figures S8A and S8B; average ROI size, ~2,500 × 2,500 μm), compatible with recent cellular-resolution findings in the posterior cortex (Minderer et al., 2019). In pixel-based models, we typically observed that the decoding weights were distributed broadly over the dorsal cortex (Figure 6A). To quantify the spatial distribution of weights, we calculated, for each pixel, its pairwise Pearson correlation to all the other pixels and plotted that as a function of their pairwise distance. Weight correlations fell to chance levels at ~1 mm and, interestingly, were below chance between ~2 and ~3 mm (Figure 6B). This spatial scale agrees with our previous analysis (Figures S8A and S8B).

Next, to test whether there are differences in the amount of information about accumulated evidence across different anatomically defined areas, we fit separate decoders using

only pixels from a pair of homotopic ROIs at a time. We observed significant differences in decoding accuracy across cortical regions that followed a general posterior-to-frontal gradient so that visual areas had the highest accuracy and frontal areas the lowest (Figures 6D and 6E; $p = 1.1 \times 10^{-14}$ across decoders, one-way ANOVA with repeated measures). This result is not likely to be due to inter-ROI differences in the magnitude of responses to individual towers interacting with slow GCaMP6f dynamics because we observed no significant differences in tower-evoked activity at short lags (<0.5 s; Figures S9A and S9D; $p = 0.24$, one-way ANOVA with repeated measures). Moreover, if the decoding results were strictly an artifact of this interaction, then one would expect that, unlike our observations, the decoding accuracy would fall off during the delay, which is much longer than the time constant of GCaMP6f (Chen et al., 2013). Note, however, that the lower decoding accuracies from frontal regions could be an artifact of spatially averaging intermingled neuronal populations. Future cellular-resolution studies should clarify this.

The accuracy of choice decoding was also significantly higher than chance throughout the maze (Figure 6F), possibly related to trial history effects (Koay et al., 2019; Pinto et al., 2018). However, it remained low until the end of the cue region, growing thereafter, compatible with our behavioral analysis showing that the mice integrate evidence from the whole cue period to make a decision (Figure S1B). Conversely, the decoding accuracy of previous choice, although also significant throughout the maze, was highest in the beginning of the trial (Figure S8E). As with accumulated evidence, pixel-based choice decoders outperformed anatomical ROI-based ones (Figures 6F and S8B; $p = 0.02$, two-sided paired t test). Likewise, significant differences were observed across ROIs (Figures 5G and 5H; $p = 7.8 \times 10^{-13}$, one-way ANOVA with repeated measures). Because view angle is itself correlated with choice (Pinto et al., 2018) and may be encoded in the activity of many cortical areas (Krumin et al., 2018), we devised a stringent analysis to control for potential view angle confounds (STAR Methods). The results from this analysis supported most of our conclusions (Figures S8C–S8H).

As a complementary analysis, we calculated average responses to the presentation of each tower as well as activity preceding and immediately following turns. These behavioral events evoked activity throughout the dorsal cortex, often with similar time courses across areas (albeit with different magnitudes; Figures S9A–S9F), consistent with previous reports on other tasks and sensory modalities (Allen et al., 2017; Ferezou et al., 2007; Orsolic et al., 2019; Scott et al., 2017; Sreenivasan et al., 2017). To better quantify the contribution of different behavioral variables and account for their correlations, we built linear encoding models of the Ca^{2+} activity of each ROI using different task events at different spatial lags as predictors (Pinto and Dan, 2015; Scott et al., 2017). We observed similar levels of encoding of both sensory and motor events across cortical regions (Figures S9G–S9J).

Together, the analyses above suggest distributed encoding of task variables in the accumulation task but with differences across cortical regions, in overall agreement with the inactivation findings.

More Complex Computations Diversify Dynamics and Increase Sensitivity to Perturbations in a Modular RNN

Compared to the visually guided task, the evidence accumulation task was associated with more decorrelated dorsal cortical dynamics, particularly across cortical area clusters, as well as more widespread effects of inactivation. We next asked whether differences in the underlying computations can explain such dynamic changes in a network without changing its underlying connectivity. To test this, we built an RNN model and trained it to perform both tasks, allowing us to gain insight into the possible computational and wiring constraints behind our findings (Li et al., 2016; Machens et al., 2005; Mante et al., 2013; Rajan et al., 2016). Thus, the model was intended as a conceptual exercise and not meant to reproduce every aspect of our multi-area dataset.

The design of the RNN (Figure 7A) was inspired by our behavioral and imaging results. It consisted of two modules ($N = 500$ units in each): one that received sensory input and another one from which the "behavioral" response (i.e., RNN output) was read out. For simplicity, we refer to them as the "posterior" and the "frontal" modules, respectively. We used the full-FORCE algorithm (DePasquale et al., 2018) to train the RNN. full-FORCE uses a second, "teacher" RNN to identify a set of desired ("target") activity patterns for performing the task. In addition, external inputs ("task hints") can be applied to the teacher network to influence the details of the target activity, allowing us to implement specific computational hypotheses based on our experimental observations. Inspired by anatomical data (Gămănuț et al., 2018) and the large-scale correlational structure we observed (Figure 4), we constrained the teacher network to have dense connectivity within modules and sparse connections between them (Figure S10A). Furthermore, because sensory evidence can be decoded with much higher accuracy from the posterior than from the frontal cortex (Figure 6C), only the posterior teacher module received the accumulated evidence as a task hint (Figure S10B). Finally, because both behavior and Ca^{2+} activity in the visually guided task are insensitive to the amount of sensory evidence (Figures S1C and 5C), the teacher network was provided no tower-dependent hint during the visually guided task, leading it to not accumulate the pulsed inputs (Figure S10B). Importantly, no hints specifying the correlational structure of network activity were provided. After learning, the teacher network was removed, and we analyzed the learner network operating on its own.

The modular RNN found a solution for both tasks using the same set of recurrent weights (Figures 7E and S10A), and qualitatively replicated several key features of our results. We observed much more diverse activity patterns among the units during accumulation (Figure 7B). To quantify this, we calculated pairwise Pearson correlations between the rates of all units during each task. Unlike in our imaging data, in which there is a substantial positive baseline correlation common to all regions (Figure 4; $r \sim 0.5$), potentially related to the presence of global ramps (Figure 3), correlation distributions in the RNN were symmetrically distributed around zero. However, they differed across tasks in terms of their widths and degree of unimodality (Figure S10E). To compare directly with imaging

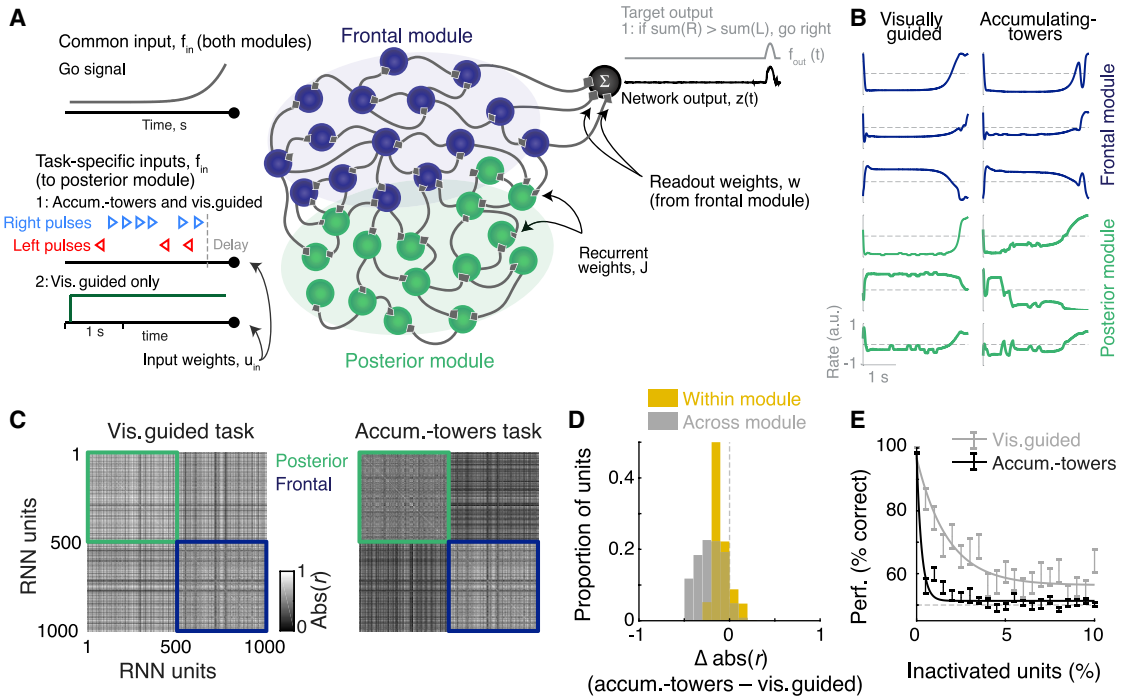

Figure 7. A Modular RNN Model Recapitulates Key Features of the Data

(A) Schematics of the modular multi-task RNN.

(B) Single-trial examples of activity from the same 6 units from both modules while the RNN is performing either task.

(C) Pairwise rate correlation matrices for both tasks.

(D) Histogram of changes in rate correlations (accumulating-towers – visually guided task) within and across modules.

(E) Effects of silencing different fractions of randomly selected units from the RNN on task performance. Error bars, ± SEM across inactivation runs (n = 50). Lines, best-fitting exponential functions to average data.

See also Figure S10.

data, we displayed the absolute values of the correlations, highlighting the degree of similarity in temporal activity profiles of the network units regardless of sign (Figure 7B; adding a surrogate global signal to RNN units results in similar distributions to the data; Figure S10E). During both tasks, correlations were generally lower within than across modules, much like our data (compare Figure 7C with Figure 4A). Moreover, the magnitude of correlations was significantly decreased during the accumulation task (Figure 7C; $p = 1.3 \times 10^{-144}$, signed-rank test, N = 1,000 units). Interestingly, and again consistent with our data (Figure 4C, D), correlations decreased significantly more across than within modules (Figure 7D; $p = 3.7 \times 10^{-40}$, signed-rank test, N = 1,000 units), even though the underlying synaptic connectivity was held constant between tasks. This result was qualitatively similar across a range of inter-module connectivity levels in the teacher network (Figure S10C).

Thus, like the dorsal cortex, the RNN displayed more correlated dynamics during the visually guided task. We hypothesized that this amount of redundancy would result in lower sensitivity to perturbations compared with the accumulation task. We tested this by randomly silencing various fractions of the RNN units, selected at random across both modules. Similar to increasing laser power during inactivations (Figures S2I–S2L), increasing the fraction of silenced units led to a progressive drop in performance during either task. Additionally, the network was indeed more sensitive to perturbation during the accumulation task (Figures 7E and S10D; $p_{task} = 4.1 \times 10^{-37}$, two-way ANOVA with factor inactivation levels and task; $p_{decay\ rate} < 0.01$, bootstrap, taken from exponential fits to the data). These results were analogous to the V1 inactivation data at various laser powers (Figure S2J). Thus, the modular RNN model qualitatively recapitulated key aspects of dorsal cortical dynamics and inactivation effects, suggesting that differences in task-related computations per se could be sufficient to drive the complexification of dynamics we observed.

DISCUSSION

Here we show that cortical dynamics during cognitive behavior are rearranged according to the computational requirements of the task so that mesoscale Ca^{2+} activity becomes more decorrelated across different regions, and the effects of perturbing cortical dynamics become more widespread and of higher magnitude, when computations become more complex. Moreover, different cortical areas contribute to distinct but overlapping aspects of the computations, but their contributions are rearranged in a task-dependent fashion.

Global Signals and Localized Inactivation Effects during Simple Decisions

Our results are broadly consistent with recent reports that simple two-stimulus discrimination is associated with correlated, large-scale ramping up of wide-field Ca^{2+} activity in dorsal cortical areas (Allen et al., 2017; Musall et al., 2018), although performance is only affected by the inactivation of a small subset of these areas (Allen et al., 2017; Guo et al., 2014; Zatka-Haas et al., 2019). Because the wide-field Ca^{2+} signal probably includes layer 1 activity (Allen et al., 2017; Ma et al., 2016a) and, thus, non-local input, these widespread ramps could a reflect a top-down input signal. Indeed, M2 inactivation has been shown to abolish global cortical activity ramps (Allen et al., 2017) and to disrupt the relative timing of cortical activity (Makino et al., 2017). Further supporting this notion, recent results from our group show that the ramps are far less prevalent in the somatic Ca^{2+} activity of individual layer 2/3 and layer 5 posterior cortical neurons, which instead show choice-specific sequences of activity that tile the trial during the accumulating-towers task (Koay et al., 2019). The exact nature and mechanisms of these ramping signals will be an interesting future research avenue. Possible candidates include motor preparation and/or execution (Allen et al., 2017; Gilad et al., 2018; Inagaki et al., 2019; Makino et al., 2017; Musall et al., 2018; Orsolic et al., 2019; Stringer et al., 2019), reward expectation (Inagaki et al., 2019; Shuler and Bear, 2006), or learned task timing (Makino and Komiyama, 2015; Orsolic et al., 2019).

Widespread Requirements during Demanding Decisions

In contrast to the visually guided task, we show, for the first time, that inactivation of widespread regions of the dorsal cortex leads to behavioral deficits during more demanding memory-dependent tasks (Figure 1). Importantly, the patterns and magnitude of behavioral effects vary across cortical regions and tasks (Figures 1 and 2), which generally agrees with our analyses of wide-field data suggesting functional gradients across the cortex. Further, this diversity of impairments argues against non-specific inactivation effects related to aversion to or distraction by silencing otherwise irrelevant regions. Of course, inferring function from localized inactivations is fraught with difficulties stemming from possible off-target effects (Hong et al., 2018; Wolff and Ölveczky, 2018; Young et al., 2000). Further work with cellular-resolution signals and simultaneous inactivation and imaging will be needed to understand the exact nature of the contributions of different areas.

Another crucial question for future studies is precisely which differences between the tasks drive the changes in the requirement and large-scale cortical dynamics we observed. Because our tasks were nearly identical in terms of sensory and motor features (Figures 1 and S1), trivial differences in sensory or motor-related activity (Gilad et al., 2018; Musall et al., 2018) cannot easily explain our results. Instead, we suggest that differences in the underlying computations themselves appear to be the origin of the findings, a conclusion supported by our RNN model (Figure 7). The more complex computations in the accumulating-towers task primarily stem from four features: its dependence on short-term memory of the stimulus stream, the need to constantly update this memory (i.e., evidence accrual), the relatively long accumulation period (~4 s), and the post-stimulus delay (~2 s). The memory-guided task controls primarily for the latter feature, and the finding of similarly widespread inactivation effects in this task suggests that memory load is an important contributor to the tasks' computational demands. On the other hand, the larger deficits and similarly widespread effects during cue-period inactivation in the accumulating-towers task show that evidence accumulation is also important (Figures 1 and S3). Another possibility highlighted by our findings is that even small changes in underlying task computations may lead to a qualitative large-scale reorganization of cortical activity. Further support for this idea comes from recent findings that different behavioral strategies within the same task are associated with different cortical dynamics and perturbation effects (Gilad et al., 2018).

The long timescale of evidence accumulation could also explain why we observed behavioral deficits upon posterior parietal cortex (PPC) inactivation, unlike previous tasks with much shorter accumulation periods (Erlich et al., 2015; Katz et al., 2016). Indeed, non-accumulation tasks with long stimulus durations require PPC activity (Driscoll et al., 2017; Goard et al., 2016; Harvey et al., 2012). Of course, given the small size of the mouse PPC and the spread of inactivations, we cannot rule out that our results are not due to inadvertent inactivation of nearby regions. Future work should clarify this issue. Along similar lines, a longer post-stimulus delay could explain the differences between our findings in the memory-guided task and the more localized effects in a similar task (Guo et al., 2014), although differences in sensory modality or the presence of virtual navigation could also underlie the discrepancies.

Distributed Representations of Task Variables

Our imaging results (Figures 6 and S9) agree with many reports of similar response properties and representations of multiple task features in any given cortical area, both during evidence accumulation (Brody and Hanks, 2016; Gold and Shadlen, 2007; Koay et al., 2019; Scott et al., 2017; Siegel et al., 2015) and other types of memory-dependent decision-making (Dotson et al., 2018; Hernández et al., 2010; Siegel et al., 2015). More specifically, direct comparison of different cortical areas during the same behavioral tasks in monkeys has shown that task information is widespread, with stimulus signals following a general sensory-to-motor cortex progression and choice information appearing near-simultaneously throughout the network (Hernández et al., 2010; Siegel et al., 2015). The same overall pattern was present in our data but with low accumulated evidence information in frontal regions (although we did observe responses to towers; Figure S9), which could be due to the spatial averaging of the widefield signal. Our main conclusions are unlikely to be influenced by this technical aspect, however. First, mesoscale wide-field activity has been shown to be generally consistent with local cellular activity (Allen et al., 2017; Chen et al., 2017; Clancy et al., 2019; Gilad et al., 2018; Musall et al., 2018). Moreover, recent findings from our group show that coding similarities are also present in single neurons from across the posterior cortex during the accumulating-towers task (Koay et al., 2019). Taken together, our imaging and inactivation results point to a similar picture of distributed processing, which could be a general hallmark of decision-making (Yoo and Hayden, 2018).

Our findings of widespread task-related Ca^{2+} activity are also supported by recent anatomical findings of very dense cortico-cortical connectivity in the mouse (Gămănuţ et al., 2018; Oh et al., 2014; Zhang et al., 2016; Zingg et al., 2014). Given this amount of long-range recurrence, it is perhaps not surprising that we observed correlated activity between different areas. General mesoscale correlational structures do, in fact, appear to be related to underlying anatomical connectivity (Mohajerani et al., 2013; Stafford et al., 2014), but our results extend these findings to show that they can also be dynamically modulated by task features, in agreement with human fMRI data (Gonzalez-Castillo et al., 2015). Specifically, the evidence accumulation task induces cortex-wide decorrelation (Figure 4), and the results of our RNN model suggest that this can be driven primarily by changes in the computational strategies of each task because network connectivity is fixed (Figure 7). Similar to theoretical conclusions on neuronal populations, such decreases in correlations may increase coding capacity (Kohn et al., 2016). Moreover, the correlations were more decreased between than within area clusters; this is a signature of network modularity, which can itself generate more complex dynamics (Sporns et al., 2000). Interestingly, short-term memory capacity in humans has been linked to the amount of modularity across the cortex (Stevens et al., 2012). However, caution should be exerted when interpreting our correlation results, both because of the potentially nonlinear nature of the GCaMP6f signals (Chen et al., 2013) and because the exact relationship between mesoscale and cellular-resolution activity is presently unclear. A more mechanistic understanding of these results will require follow-up studies using cellular-resolution imaging and electrophysiology.

Conclusions

At first glance, our results seem reminiscent of near-century-old findings that complex-maze learning and retrieval is affected solely by the size, and not the location, of cortical lesions (Lashley, 1931). However, our conclusions are fundamentally different. Specifically, careful quantitative assessment of behavior shows that cortical regions differ in terms of the effects of their inactivation in a task-specific fashion. Thus, our work reconciles apparently disparate accounts of localization and integration of cortical function by showing that these are complementary concepts and that the degree of localization or integration depends critically on behavioral complexity. Many challenges remain before we can obtain a complete understanding of the circuit mechanisms underlying perceptual decision-making. Our findings emphasize the fact that, in addition to causal requirements and local circuitry, the answer will necessarily involve an account of how different brain areas interact dynamically and at large scales to produce complex behavior.

STAR★METHODS

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

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Behavior
  - Scanning-laser optogenetic inactivation
  - Electrophysiological measurement of inactivation
  - Wide-field Ca^{2+} imaging
  - Sensory cortex mapping
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data selection
  - Behavior
  - Statistics of inactivation effects
  - Electrophysiology
  - Wide-field image preprocessing
  - Isosbestic correction of hemodynamic artifacts and $\Delta F/F$ calculation
  - Sensory maps
  - Atlas registration and ROI selection
  - Quantification of average time course
  - Linear fits for evidence tuning
  - ROI correlations
  - Decoding models
  - Event-triggered averages
  - Clustering
  - General statistics
  - Recurrent Neural Network Model
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.neuron.2019.08.025.

ACKNOWLEDGMENTS

We thank S.A. Koay for discussions and visual mapping software; B.B. Scott, A. Piet, and E.S. Carter for discussions; S. Stein and S. Baptista for technical support; and H.K. Inagaki and K. Svoboda for advice regarding the cleared skull preparation. We also thank P.D. Rich, E. Dennis, A. Riordan, and E. Nieh for helpful comments on the manuscript. The LED strobing Arduino software was modified from code provided by M. Shaik and E.M.C. Hillman. B.E. Engelhard built some of the VR training setups. This work was supported by NIH grants U01NS090541, U19NS104648, and F32NS101871 (to L.P.) and K99MH120047 (to L.P.).

AUTHOR CONTRIBUTIONS

L.P. performed the experiments and analyzed the data. K.R. and B.D. developed the RNN models. S.Y.T. and D.W.T. designed the VR and optical setups. L.P. wrote the manuscript with input from C.D.B. and D.W.T. L.P., C.D.B., and D.W.T. conceived the project.

DECLARATION OF INTERESTS

The authors declare no competing financial interests.

Received: November 2, 2018
Revised: June 18, 2019
Accepted: August 13, 2019
Published: September 26, 2019

REFERENCES

Akrami, A., Kopec, C.D., Diamond, M.E., and Brody, C.D. (2018). Posterior parietal cortex represents sensory history and mediates its effects on behaviour. Nature *554*, 368–372.

Allen, W.E., Kauvar, I.V., Chen, M.Z., Richman, E.B., Yang, S.J., Chan, K., Gradinaru, V., Deverman, B.E., Luo, L., and Deisseroth, K. (2017). Global representations of goal-directed behavior in distinct cell types of mouse neocortex. Neuron *94*, 891–907.e6.

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.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B Stat. Methodol. *57*, 289–300.

Brody, C.D., and Hanks, T.D. (2016). Neural underpinnings of the evidence accumulator. Curr. Opin. Neurobiol. *37*, 149–157.

Brunton, B.W., Botvinick, M.M., and Brody, C.D. (2013). Rats and humans can optimally accumulate evidence for decision-making. Science *340*, 95–98.

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.

Chen, T.-W., Li, N., Daie, K., and Svoboda, K. (2017). A map of anticipatory activity in mouse motor cortex. Neuron *94*, 866–879.e4.

Clancy, K.B., Orsolic, I., and Mrsic-Flogel, T.D. (2019). Locomotion-dependent remapping of distributed cortical networks. Nat. Neurosci. *22*, 778–786.

DePasquale, B., Cueva, C.J., Rajan, K., Escola, G.S., and Abbott, L.F. (2018). full-FORCE: A target-based method for training recurrent networks. PLoS ONE *13*, e0191527.

Dotson, N.M., Hoffman, S.J., Goodell, B., and Gray, C.M. (2018). Feature-based visual short-term memory is widely distributed and hierarchically organized. Neuron *99*, 215–226.e4.

Driscoll, L.N., Pettit, N.L., Minderer, M., Chettih, S.N., and Harvey, C.D. (2017). Dynamic reorganization of neuronal activity patterns in parietal cortex. Cell *170*, 986–999.e16.

Erlich, J.C., Brunton, B.W., Duan, C.A., Hanks, T.D., and Brody, C.D. (2015). Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat. eLife *4*, e05457.

Ferezou, I., Haiss, F., Gentet, L.J., Aronoff, R., Weber, B., and Petersen, C.C.H. (2007). Spatiotemporal dynamics of cortical sensorimotor integration in behaving mice. Neuron *56*, 907–923.

Fuster, J.M. (1997). Network memory. Trends Neurosci. *20*, 451–459.

Gămănuț, R., Kennedy, H., Toroczkai, Z., Ercsey-Ravasz, M., Van Essen, D.C., Knoblauch, K., and Burkhalter, A. (2018). The mouse cortical connectome, characterized by an ultra-dense cortical graph, maintains specificity by distinct connectivity profiles. Neuron *97*, 698–715.e10.

Garrett, M.E., Nauhaus, I., Marshel, J.H., and Callaway, E.M. (2014). Topography and areal organization of mouse visual cortex. J. Neurosci. *34*, 12587–12600.

Gilad, A., Gallero-Salas, Y., Groos, D., and Helmchen, F. (2018). Behavioral strategy determines frontal or posterior location of short-term memory in neocortex. Neuron *99*, 814–828.e7.

Goard, 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*, e13764.

Gold, J.I., and Shadlen, M.N. (2007). The neural basis of decision making. Annu. Rev. Neurosci. *30*, 535–574.

Gonzalez-Castillo, J., Hoy, C.W., Handwerker, D.A., Robinson, M.E., Buchanan, L.C., Saad, Z.S., and Bandettini, P.A. (2015). Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns. Proc. Natl. Acad. Sci. USA *112*, 8762–8767.

Guo, Z.V., Li, N., Huber, D., Ophir, E., Gutnisky, D., Ting, J.T., Feng, G., and Svoboda, K. (2014). Flow of cortical activity underlying a tactile decision in mice. Neuron *81*, 179–194.

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.

Haykin, S. (1986). Adaptive filter theory (Englewood Cliffs, NJ: Prentice-Hall).

Hernández, A., Nácher, V., Luna, R., Zainos, A., Lemus, L., Alvarez, M., Vázquez, Y., Camarillo, L., and Romo, R. (2010). Decoding a perceptual decision process across cortex. Neuron *66*, 300–314.

Hong, Y.K., Lacefield, C.O., Rodgers, C.C., and Bruno, R.M. (2018). Sensation, movement and learning in the absence of barrel cortex. Nature *561*, 542–546.

Inagaki, H.K., Fontolan, L., Romani, S., and Svoboda, K. (2019). Discrete attractor dynamics underlies persistent activity in the frontal cortex. Nature *566*, 212–217.

Kalatsky, V.A., and Stryker, M.P. (2003). New paradigm for optical imaging: temporally encoded maps of intrinsic signal. Neuron *38*, 529–545.

Katz, L.N., Yates, J.L., Pillow, J.W., and Huk, A.C. (2016). Dissociated functional significance of decision-related activity in the primate dorsal stream. Nature *535*, 285–288.

Koay, S.A., Thiberge, S.Y., Brody, C.D., and Tank, D.W. (2019). Neural correlates of cognition in primary visual versus neighboring posterior cortices during visual evidence-accumulation-based navigation. bioRxiv. https://doi.org/10.1101/568766.

Kohn, A., Coen-Cagli, R., Kanitscheider, I., and Pouget, A. (2016). Correlations and Neuronal Population Information. Annu. Rev. Neurosci. *39*, 237–256.

Krumin, M., Lee, J.J., Harris, K.D., and Carandini, M. (2018). Decision and navigation in mouse parietal cortex. eLife *7*, e42583.

Lashley, K.S. (1931). Mass action in cerebral function. Science *73*, 245–254.

Lerner, T.N., Shilyansky, C., Davidson, T.J., Evans, K.E., Beier, K.T., Zalocusky, K.A., Crow, A.K., Malenka, R.C., Luo, L., Tomer, R., and Deisseroth, K. (2015). Intact-brain analyses reveal distinct information carried by SNc dopamine subcircuits. Cell *162*, 635–647.

Li, N., Daie, K., Svoboda, K., and Druckmann, S. (2016). Robust neuronal dynamics in premotor cortex during motor planning. Nature *532*, 459–464.

Ma, Y., Shaik, M.A., Kim, S.H., Kozberg, M.G., Thibodeaux, D.N., Zhao, H.T., Yu, H., and Hillman, E.M.C. (2016a). Wide-field optical mapping of neural activity and brain haemodynamics: considerations and novel approaches. Philos. Trans. R. Soc. Lond. B Biol. Sci. *371*, 20150360.

Ma, Y., Shaik, M.A., Kozberg, M.G., Kim, S.H., Portes, J.P., Timerman, D., and Hillman, E.M.C. (2016b). Resting-state hemodynamics are spatiotemporally coupled to synchronized and symmetric neural activity in excitatory neurons. Proc. Natl. Acad. Sci. USA *113*, E8463–E8471.

Machens, C.K., Romo, R., and Brody, C.D. (2005). Flexible control of mutual inhibition: a neural model of two-interval discrimination. Science *307*, 1121–1124.

Makino, H., and Komiyama, T. (2015). Learning enhances the relative impact of top-down processing in the visual cortex. Nat. Neurosci. *18*, 1116–1122.

Makino, H., Ren, C., Liu, H., Kim, A.N., Kondapaneni, N., Liu, X., Kuzum, D., and Komiyama, T. (2017). Transformation of cortex-wide emergent properties during motor learning. Neuron *94*, 880–890.e8.

Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature *503*, 78–84.

Minderer, M., Brown, K.D., and Harvey, C.D. (2019). The spatial structure of neural encoding in mouse posterior cortex during navigation. Neuron *102*, 232–248.e11.

Mohajerani, M.H., Chan, A.W., Mohsenvand, M., LeDue, J., Liu, R., McVea, D.A., Boyd, J.D., Wang, Y.T., Reimers, M., and Murphy, T.H. (2013). Spontaneous cortical activity alternates between motifs defined by regional axonal projections. Nat. Neurosci. *16*, 1426–1435.

Morcos, A.S., and Harvey, C.D. (2016). History-dependent variability in population dynamics during evidence accumulation in cortex. Nat. Neurosci. *19*, 1672–1681.

Musall, S., Kaufman, M.T., Gluf, S., and Churchland, A. (2018). Movement-related activity dominates cortex during sensory-guided decision making. bioRxiv. https://doi.org/10.1101/308288.

Odoemene, O., Pisupati, S., Nguyen, H., and Churchland, A.K. (2018). Visual evidence accumulation guides decision-making in unrestrained mice. J. Neurosci. *38*, 10143–10155.

Oh, S.W., Harris, J.A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A.M., et al. (2014). A mesoscale connectome of the mouse brain. Nature *508*, 207–214.

Orsolic, I., Rio, M., Mrsic-Flogel, T.D., and Znamenskiy, P. (2019). Mesoscale cortical dynamics reflect the interaction of sensory evidence and temporal expectation during perceptual decision-making. bioRxiv. https://doi.org/10.1101/552026.

Pinto, L., and Dan, Y. (2015). Cell-type-specific activity in prefrontal cortex during goal-directed behavior. Neuron *87*, 437–450.

Pinto, L., Koay, S.A., Engelhard, B., Yoon, A.M., Deverett, B., Thiberge, S.Y., Witten, I.B., Tank, D.W., and Brody, C.D. (2018). An accumulation-of-evidence task using visual pulses for mice navigating in virtual reality. Front. Behav. Neurosci. *12*, 36.

Raichle, M.E. (2015). The brain's default mode network. Annu. Rev. Neurosci. *38*, 433–447.

Rajan, K., Harvey, C.D., and Tank, D.W. (2016). Recurrent network models of sequence generation and memory. Neuron *90*, 128–142.

Rossant, C., Kadir, S.N., Goodman, D.F.M., Schulman, J., Hunter, M.L.D., Saleem, A.B., Grosmark, A., Belluscio, M., Denfield, G.H., Ecker, A.S., et al. (2016). Spike sorting for large, dense electrode arrays. Nat. Neurosci. *19*, 634–641.

Scott, B.B., Constantinople, C.M., Akrami, A., Hanks, T.D., Brody, C.D., and Tank, D.W. (2017). Fronto-parietal cortical circuits encode accumulated evidence with a diversity of timescales. Neuron *95*, 385–398.e5.

Shuler, M.G., and Bear, M.F. (2006). Reward timing in the primary visual cortex. Science *311*, 1606–1609.

Siegel, M., Buschman, T.J., and Miller, E.K. (2015). Cortical information flow during flexible sensorimotor decisions. Science *348*, 1352–1355.

Sporns, O., Tononi, G., and Edelman, G.M. (2000). Theoretical neuroanatomy: relating anatomical and functional connectivity in graphs and cortical connection matrices. Cereb. Cortex *10*, 127–141.

Sreenivasan, V., Kyriakatos, A., Mateo, C., Jaeger, D., and Petersen, C.C.H. (2017). Parallel pathways from whisker and visual sensory cortices to distinct frontal regions of mouse neocortex. Neurophotonics *4*, 031203.

Stafford, J.M., Jarrett, B.R., Miranda-Dominguez, O., Mills, B.D., Cain, N., Mihalas, S., Lahvis, G.P., Lattal, K.M., Mitchell, S.H., David, S.V., et al. (2014). Large-scale topology and the default mode network in the mouse connectome. Proc. Natl. Acad. Sci. USA *111*, 18745–18750.

Steinmetz, N.A., Buetfering, C., Lecoq, J., Lee, C.R., Peters, A.J., Jacobs, E.A.K., Coen, P., Ollerenshaw, D.R., Valley, M.T., de Vries, S.E.J., et al. (2017). Aberrant cortical activity in multiple GCaMP6-expressing transgenic mouse lines. eNeuro *4*, ENEURO.0207-17.2017.

Steinmetz, N., Zatka-Haas, P., Carandini, M., and Harris, K. (2018). Distributed correlates of visually-guided behavior across the mouse brain. bioRxiv. https://doi.org/10.1101/474437.

Stevens, A.A., Tappon, S.C., Garg, A., and Fair, D.A. (2012). Functional brain network modularity captures inter- and intra-individual variation in working memory capacity. PLoS ONE *7*, e30468.

Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C.B., Carandini, M., and Harris, K.D. (2019). Spontaneous behaviors drive multidimensional, brainwide activity. Science *364*, 255–258.

Sussillo, D., and Abbott, L.F. (2009). Generating coherent patterns of activity from chaotic neural networks. Neuron *63*, 544–557.

Tian, L., Hires, S.A., Mao, T., Huber, D., Chiappe, M.E., Chalasani, S.H., Petreanu, L., Akerboom, J., McKinney, S.A., Schreiter, E.R., et al. (2009). Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. Nat. Methods *6*, 875–881.

Wig, G.S. (2017). Segregated systems of human brain networks. Trends Cogn. Sci. *21*, 981–996.

Wolff, S.B., and Ölveczky, B.P. (2018). The promise and perils of causal circuit manipulations. Curr. Opin. Neurobiol. *49*, 84–94.

Xiao, D., Vanni, M.P., Mitelut, C.C., Chan, A.W., LeDue, J.M., Xie, Y., Chen, A.C., Swindale, N.V., and Murphy, T.H. (2017). Mapping cortical mesoscopic networks of single spiking cortical or sub-cortical neurons. eLife *6*, e19976.

Yang, G.R., Joglekar, M.R., Song, H.F., Newsome, W.T., and Wang, X.-J. (2019). Task representations in neural networks trained to perform many cognitive tasks. Nat. Neurosci. *22*, 297–306.

Yoo, S.B.M., and Hayden, B.Y. (2018). Economic choice as an untangling of options into actions. Neuron *99*, 434–447.

Young, M.P., Hilgetag, C.-C., and Scannell, J.W. (2000). On imputing function to structure from the behavioural effects of brain lesions. Philos. Trans. R. Soc. Lond. B Biol. Sci. *355*, 147–161.

Zatka-Haas, P., Steinmetz, N.A., Carandini, M., and Harris, K.D. (2019). Distinct contributions of mouse cortical areas to visual discrimination. bioRxiv. https://doi.org/10.1101/501627.

Zhang, S., Xu, M., Chang, W.-C., Ma, C., Hoang Do, J.P., Jeong, D., Lei, T., Fan, J.L., and Dan, Y. (2016). Organization of long-range inputs and outputs of frontal cortex for top-down control. Nat. Neurosci. *19*, 1733–1742.

Zhuang, J., Ng, L., Williams, D., Valley, M., Li, Y., Garrett, M., and Waters, J. (2017). An extended retinotopic map of mouse cortex. eLife *6*, e18372.

Zingg, B., Hintiryan, H., Gou, L., Song, M.Y., Bay, M., Bienkowski, M.S., Foster, N.N., Yamashita, S., Bowman, I., Toga, A.W., and Dong, H.W. (2014). Neural networks of the mouse neocortex. Cell *156*, 1096–1111.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Organisms/Strains		
Mouse: C57BL6/J	The Jackson Laboratory	JAX: 000664
Mouse: B6.Cg-Tg(Slc32a1-COP4*H134R/EYFP) 8Gfng/J	The Jackson Laboratory	JAX: 014548
Mouse: IgS6$^{tm93.1(tetO-GCaMP6f)Hze}$ Tg(Camk2a-tTA) 1Mmay/J	The Jackson Laboratory	JAX: 024108
Mouse: B6.129S2-Emx1$^{tm1(cre)Krj}$/J	The Jackson Laboratory	JAX: 005628
Mouse: B6.Cg-Tg(Thy1-YFP)HJrs/J	The Jackson Laboratory	JAX: 003782
Software and Algorithms		
MATLAB 2015b, 2016b, 2017b	Mathworks	https://www.mathworks.com/products/matlab.html
ViRMEn	Aronov and Tank, 2014	https://pni.princeton.edu/pni-software-tools/virmen
NI DAQmx 9.5.1	National Instruments	https://www.ni.com/en-us/support/downloads/drivers/download.ni-daqmx.html
HCImage	Hamamatsu	https://hcimage.com
Klusta	Rossant et al., 2016	https://github.com/kwikteam/klusta
Scanning laser control software	This paper	https://github.com/BrainCOGS/laserGalvoControl
Other		
Virtual reality setup tools	Pinto et al., 2018	https://github.com/sakoay/AccumTowersTools

LEAD CONTACT AND MATERIALS AVAILABILITY

This study did not generate new unique reagents. Further information requests may be directed to the Lead Contact, Dr. Carlos D. Brody (brody@princeton.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All procedures were approved by the Institutional Animal Care and Use Committee at Princeton University and were performed in accordance with the Guide for the Care and Use of Laboratory Animals. We used both male and female mice aged 2 – 16 months, from the following strains: wild-types used in control experiments for cortical inactivations (n = 3, C57BL6/J, Jackson Laboratories, stock # 000664); VGAT-ChR2-EYFP for cortical inactivations [n = 28, B6.Cg-Tg(Slc32a1-COP4*H134R/EYFP)8Gfng/J, Jackson Laboratories, stock # 014548]; triple transgenic crosses expressing GCaMP6f under the CaMKIIα promoter, for widefield Ca^{2+} imaging (n = 6), from the following two lines: Ai93-D;CaMKIIα-tTA [IgS6$^{tm93.1(tetO-GCaMP6f)Hze}$ Tg(Camk2a-tTA)1Mmay/J, Jackson Laboratories, stock # 024108] and Emx1-IRES-Cre [B6.129S2-Emx1$^{tm1(cre)Krj}$/J, Jackson Laboratories, stock # 005628]; Thy1-YFP-H for imaging controls [n = 2, B6.Cg-Tg(Thy1-YFP)HJrs/J, Jackson Laboratories, stock # 003782]. Note that, despite the reported tendency for a subpopulation of the Emx1-Ai93 triple transgenics to display ictal events (Steinmetz et al., 2017), we have previously shown that this line has statistically indistinguishable behavior from other transgenic mouse lines (Pinto et al., 2018). Moreover, cellular-resolution data from these mice during the accumulating-towers task did not differ from Thy1-GCaMP6f transgenics in a recent study from our group (Koay et al., 2019).

The mice underwent sterile stereotaxic surgery to implant a custom lightweight titanium headplate (~1 g) and optically clear their intact skulls, under isoflurane anesthesia (2.5% for induction, 1.5% for maintenance). The clear skull procedure was similar to previously described (Guo et al., 2014). After asepsis the skull was exposed, and the periosteum removed using a bonn micro probe (Fine Science Tools). The incision margins were covered with veterinary cyanoacrylate glue (surgi-lock 2-oc, Meridian), and a coat of commercial cyanoacrylate glue (krazy glue, Elmers) was applied over the skull. Before the glue dried, a layer of diluted clear metabond (Parkell) was applied (~2-mm thick, 2 scoops powder:7 drops of quick base:2 drops catalyzer) and allowed to cure for ~10 minutes. The surface of the cured metabond was then polished using a cement polishing kit (Pearson dental, going progressively from coarse to fine). The metal headplate was attached to the cleared skull using metabond. After the metabond was cured, a ~2-mm layer of

transparent nail polish (Electron Microscopy Sciences) was applied and allowed to cure for ~15 min. The implant was typically covered using a custom 3D-printed headplate cap. Analgesia consisted of two doses of meloxicam (1 mg/kg I.P or S.C.), given pre-operatively and 24 h later. Mice also received peri-operative I.P. injections of 37°C saline to maintain hydration. Body temperature was maintained constant using a homeothermic control system (Harvard Apparatus). The mice were allowed to recover for 5 days before starting water restriction for behavioral training. They were then restricted to an allotted water volume of 1 – 2 mL per day, typically entirely delivered during behavioral sessions, always ensuring that no clinical signs of dehydration were present and body mass was at least 80% of the initial value. If any of these conditions were not met, the mice received supplemental water until recovering. The mice were extensively handled and had access to enriched environments as described elsewhere (Pinto et al., 2018).

METHOD DETAILS

Behavior

The mice were trained to perform the accumulating-towers task in a virtual reality (VR) environment (Figure 1A), as previously described in detail (Pinto et al., 2018). Briefly, the head-fixed mice sat on an 8-inch Styrofoam® ball suspended by compressed air (~60 p.s.i), and ball movements were measured with optical flow sensors (ADNS-3080 APM2.6) connected to an Arduino Due running custom code to transform ball rotations into virtual-world velocity (https://github.com/sakoay/AccumTowersTools/tree/master/OpticalSensorPackage). The VR environment was projected onto a custom-built Styrofoam® toroidal screen spanning a visual field of ~270° of horizontally and ~80° altitude using a DLP projector (Optoma HD141X) with a refresh rate of 120 Hz, a resolution of 1024 × 768 pixels and an RGB color balance of 0, 0.4 and 0.5, respectively. The set-up was enclosed in a custom-designed sound-attenuating chamber (8020.inc), and VR was controlled using the MATLAB (Mathworks)-based package ViRMEn (Aronov and Tank, 2014) running on a PC.

The accumulating-towers task took place in a 3.3-m virtual T-maze comprised of a 30-cm start region, a 2-m cue region and a 1-m delay region (Figure 1B). As the mice navigated the stem of the maze, they saw tall, high-contrast visual cues (towers, 6-cm tall and 2-cm wide) along either wall during the cue region, and were rewarded for turning, after the delay, into the arm where the largest number of towers occurred (4–8 μL of 10% v/v sweet condensed milk). Rewarded trials were followed by a 3 s ITI and error trials were followed by a sound and a 12 s ITI. Tower positions were drawn randomly from two Poisson distributions with rates of 7.7 and 2.3 m^{-1} on the rewarded and non-reward side, respectively, with a 12-cm refractory period (or 8.0:1.6 m^{-1} in some optogenetic inactivation sessions). Towers appeared when the animals were 10 cm away from their drawn location and disappeared 200 ms later. To discourage side biases, we used a debiasing algorithm described elsewhere (Pinto et al., 2018). Each session started with at least 10 warm-up trials of the visually-guided task (see below, trials continued until mice reached at least 85% correct performance), followed by the accumulating-towers task. If overall performance fell below 55% calculated over a 40-trial running window, animals were transitioned to an easy 10-trial block of a maze without any distractor towers, to increase motivation (Pinto et al., 2018). These trials were not included in any of the analyses. Behavioral sessions lasted for ~1h, and typically consisted of 200 – 250 trials.

The visually-guided task not requiring evidence accumulation happened in the same virtual maze as the accumulating-towers task, but, in addition to the towers, reward location was indicated using a tall visual guide located in the corresponding reward arm (30-cm tall, 4-cm wide), visible from 10 cm onward. No other explicit task cue beyond the presence (or absence) of the guide was provided to the animal to indicate task switches. For about half the inactivation experiment sessions (53/98), the towers in this visually-guided task were displayed only on the rewarded side. Similarly, for 15/25 imaging sessions visually-guided task trials consisted of only towers on one side. The mice achieved equivalent levels of performance with or without distractor towers, suggesting that they use exclusively the visual guide (average performance of 94.1 and 97.3% correct for towers on one or both sides, respectively, $p = 0.21$, two-sided t test, $n = 5$ mice who ran this version of the visually-guided task). Thus, these trials were combined for data analysis, unless stated otherwise.

The memory-guided task also happened in the same maze but contained no towers. Instead, a visual guide (60-cm tall, 8-cm wide) was visible from the start of the maze ($y = 10$ cm) and disappeared at $y = 200$ cm, matching the post-stimulus delay of the towers task. Each session started with at least 10 warm-up trials of the visually-guided task (as in the accumulating-towers task). If overall performance fell below 60% calculated over a 40-trial running window, animals were transitioned to an easy 10-trial block of the visually-guided task. Reward sizes and ITI lengths were the same for all tasks.

Scanning-laser optogenetic inactivation

We used a 473-nm laser (OBIS, Coherent) combined with 2D galvanometers (3-mm 6210H, class 0, with individual 671XX controllers, Cambridge Technologies) to allow scanning over the cortical surface. The laser was located outside the behavioral setup. The beam was directed to the galvanometers inside the set-up using a single-mode fiber optic (460HP, 125 μm diameter, Thorlabs) coupled to an aspheric lens fiber port for collimation (PAF-X-5-A, f = 4.6 mm, Thorlabs) and then passed through an f-theta scan lens (f = 160 mm, LINOS) and a pellicle beamsplitter (CM1-BP108, Thorlabs) before reaching the cortical surface. The beamsplitter directed 8% of the light to a camera (grasshopper3, Point Grey), allowing remote visualization of beam location. The beam had a Gaussian profile with a diameter of 65 μm measured as full width at half height (115 μm at $1/e^2$), and the system had a position accuracy of ~15 μm (measured as the standard deviation of the distribution of the differences between target and actual beam location for 70 X-Y combinations).

Reported laser powers were measured at the plane corresponding to the mice's skull position. For all experiments, we used a laser square wave at 40 Hz (80% duty cycle) with a linear 100-ms down ramp at laser offset, in order to minimize rebound activity (Guo et al., 2014; Figure S2). Because laser offset was determined by the animals' spatial position in the maze, ramp onset time was determined by estimating the desired offset time based on the animals' current position and velocity. For bilateral inactivation, the galvanometers alternated between the two hemispheres at 200 Hz (20-mm travel time: ~250 μs).

The galvanometers and the laser were controlled with analog voltage generated using a National Instruments DAQ card on a PC running custom MATLAB code. This PC was slave to another PC running ViRMEn, which sent laser on, laser off, and galvanometer position commands through digital lines. For whole-trial inactivation experiments (Figures 1, 2, and S1–S3), all bilateral grid locations were inactivated in the same session, each with a probability of 0.01 (total 'laser on' probability of 0.29), drawn pseudorandomly. For sub-trial inactivations of a subset of the areas (Figures S3A), between 1 and 5 areas were targeted in each session, with a probability of 0.05 – 0.15 per area, depending on the experiment. Unless otherwise stated, laser power was 6 mW.

Inactivation coordinates were defined in relation to bregma, set manually by the experimenter. To ensure consistency, in each session images of the implant acquired using the Point Grey camera were aligned to a reference using rigid transformations. These images were acquired under green LED (532 nm) illumination to enhance vascular contrast. For the full-grid experiments, grid positions went from –3.5 to +3.5 AP in 1 mm increments, and 0.25, 1, 2 mm ML between +1.5 and +3.5 AP, and also 3 mm in the other AP coordinates. For the other experiments, we defined the following coordinates V1: –3.5 AP, 3 ML; RSC: –2.5 AP, 0.5 ML ; PPC: –2 AP, 1.75 ML ; mM2: 0.0 AP, 0.5 ML ; aM2: +3 AP, 1 ML.

Electrophysiological measurement of inactivation

We drilled a small (~300 μm diameter) craniotomy using a carbide burr, covered it with a silicone elastomer (Kwik-cast, World Precision Instruments) and allowed the mice (n = 3 VGAT-ChR2-EYFP) to recover from the procedure for 24h. These mice had previously undergone the headplate implant and cleared skull preparation, and anesthesia and analgesia procedures were followed as described above. On the following day, the awake mice were head-fixed on a plastic running dish, the silicone plug was removed, and a 32-channel, single-shank silicon probe spanning the depth of the cortex (Poly2, Neuronexus) was inserted using a micromanipulator under visual guidance, and allowed to stabilize for 30 min. Signals were acquired at 20 kHz and high-pass filtered at 250 Hz using a digital amplifier board (RHD2132, Intan Technologies). While recording, we used the laser set-up to inactivate locations centered at various distances from the recording site ([0 0.25 0.5 1 2] mm), at various powers ([.5 1 2 4 6 8] mW). Laser pulses were 1.5 s long, and other parameters were as described above. Each of the 30 conditions had 10 repeats, drawn pseudorandomly, with a 5 s inter-repeat interval. Voltage traces controlling the laser were also acquired with the amplifier board for synchronization.

Wide-field Ca^{2+} imaging

We used a custom-built, tandem-lens wide-field macroscope integrated with the VR setup to image GCaMP6f fluorescence at mesoscale resolution. The objective close to the sample was a planapo 1x (Leica, M series) and the one close to sensor was a planapo 0.63x (Leica, M series). Images were acquired using an sCMOS (OrcaFlash4.0, Hamamatsu), at 20 frames per second with alternating blue (470-nm) and violet (410-nm) illumination (~0.2 mW/mm^2), and a size of 512 × 512 pixels (pixel size of ~17 μm). Both blue and violet LEDs (Luxeon star) were coupled to clean-up bandpass filters centered at 482 and 405 nm, respectively (FF02-482/18, FF01-405/10, Semrock), and collimators (COP1-A, Thorlabs). After passing through a dichroic beamsplitter (FF458-Di02, Semrock), they were directed to a flexible fiber optic light guide (1/2'' diameter, Edmond Optics) for epifluorescence illumination, with approximately equal path lengths for both wavelengths. Green emission fluorescence passed through a dichroic beamsplitter between the two objectives (FF495-Di03, Semrock) and and was bandpass-filtered at 525 nm (FF01-525/45, Semrock) before reaching the sCMOS. The LEDs were connected to MOSFET circuits controlled by custom-written code running on an Arduino Due to turn them on in alternating imaging frames. The arduino received analog frame exposure voltage traces from the sensor. The imaging plane was focused on the superficial vasculature and the space between the headplate and the objective was covered using a custom 3D-printed cone to avoid light contamination from the projector. Images were acquired with HCImage (Hamamatsu) running on a PC. Image acquisition was triggered by a TTL pulse from another PC running ViRMEn, and analog frame exposure voltage traces were acquired through a DAQ card (National Instruments) and saved in the behavioral log file for synchronization.

Sensory cortex mapping

Mapping experiments were carried out in the same imaging setup described above but without violet-LED illumination since hemodynamic contamination is unlikely to have a large impact on the fast component of airpuff stimulus-locked responses (Figure S5), and is accounted for in the visual mapping method (Kalatsky and Stryker, 2003). In both cases, stimuli were delivered using a separate PC, which also controlled the blue LED, and blank (LED off) frames were used to synchronize the stimulation and image sequences. For somatosensory responses, we delivered 30 200-ms airpuffs (15 p.s.i.) to the right whisker pad of the mice, with 8 s intervals, by sending a TTL pulse to a solenoid valve (NReseach). For visual area mapping, we used the method described in Zhuang et al. (2017), adapted for bilateral stimulation in the toroidal screen. Briefly, a 20°-wide bar with a full-contrast checkerboard texture (25°) changing polarity at 12 Hz drifted through the toroidal screen at 9°/s, in each of the four cardinal directions. For vertical directions, a single bar spanning the horizontal width of the toroidal screen drifted between 60° and −20° of elevation, and for horizontal directions, two bars drifted between 0° and 120° (one on each half of the screen). Each direction was repeated 25 times.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data selection

Because trials occurred in a block structure, for the accumulating-towers and the memory-guided tasks we selected sessions with at least one 40-trial block with overall performance above 60% correct and analyzed trials only from these blocks. Using these criteria, we selected 25 imaging sessions from 6 mice and 984 inactivation sessions from 31 mice for the accumulating-towers task, and 97 inactivation sessions from 10 mice for the memory-guided task. The visually-guided task did not contain blocks, so we applied a session-wide performance criterion of 80% to include the data. This yielded 95 sessions from 5 mice for the inactivation experiments in Figure 1F. For all tasks, we also excluded timed-out trials (> 60 s) and trials in which the animals traveled more than 110% of nominal maze length (except for analyses measuring the effects on inactivations on these parameters).

Behavior

Overall percent correct performance was defined as the percentage of trials in which the mice chose the arm corresponding to the largest number of towers, and side bias was defined as the difference between percent correct performance in right- and left-choice trials, such that negative values indicate a left bias. Excess travel was defined as the distance traveled in excess of the nominal maze length, in percentage of maze length. Speed was calculated using the total x-y displacement in the stem of the maze ($0 < y < 300$ cm), and acceleration was defined as the average derivative of speed for a given maze region. Motor errors were defined as the combined frequency (% trials) of the following events: large-magnitude view angles during the cue period (> 60°), trials with early turns (i.e., a turn immediately before the arm, resulting in a wall collision), trials in which the mouse first entered the opposite arm to its final choice, trials with speeds below the 10th percentile (from the combined distribution of laser and control trials), and trials with at least 10% excess travel. Psychometric curves were computed by plotting the percentage of right-choice trials as a function of the difference in the number of right and left towers (#R – #L, or Δ). Δ was binned in increments of 5 and its value defined as the average Δ weighted by the number of trials. We fitted the psychometric curves using a 4-parameter sigmoid:

$$p_R = 1 + \frac{a}{1 + \exp[-(\Delta - \Delta_0)/\lambda]}$$

To assess how mice weighted sensory evidence from different segments of the cue region, we performed a logistic regression analysis in which the probability of a right choice was predicted from a logistic function of the weighted sum of the net amount of sensory evidence per each of four segments, equally spaced between 10 (the earliest possible tower) and 200 cm:

$$p_R = 1 + \frac{1}{1 + \exp\left[-\left(\beta_0 + \sum_{i=1}^{4} \beta_i \Delta_i\right)\right]}$$

To analyze virtual view angle (θ) trajectories, we defined the view angle at a particular Y position, $\theta(Y)$, as the value of θ at the first time point t at which $y(t) \geq Y$. We then performed a linear choice decoding analysis based on the distributions of right- and left-choice view angles for each y position, with 5-cm steps, by defining the optimal choice decoding boundary $\theta_{cd}(y)$ that most equally separated the two distributions. Decoding accuracy was defined as the percent of right-choice trials with with $\theta(y) > \theta_{cd}(y)$ (Pinto et al., 2018). For the tower-triggered view angle analysis (Figure S1), we first subtracted from each trial the average view angle trajectory separately for left- and right choice trials. We then averaged the view angle aligned by the presentation of each left or right tower between 0 and 50 cm from the point the tower appeared, and subtracted the average view angle over the 20 cm preceding tower presentation. The position at which the mouse turned was defined for each trial as the first point at which the derivative of the view angle trajectory exceeded 3 standard deviations of the baseline derivative (calculated over 0 – 75 cm) for at least 4 consecutive 1-cm spatial bins (as defined above).

Normalized inactivation-induced performance decreases (Figures 1F–1I) were calculated as *100 – [(laser – control) / (control – chance performance)]*, where chance was determined empirically from the data as the average of the proportion of right (left) trials (whichever was lowest) for each mouse, weighted by the number of trials, and calculated only for control (laser off) trials. This was necessary because of the debiasing algorithm, which may result in unbalanced trial draws (Pinto et al., 2018).

Trial difficulty in the accumulating-towers task was defined as $|\#R - \#L| / (\#R + \#L)$, because performance depends both on the difference between the sides and the total tower counts (Pinto et al., 2018). For the analysis in Figures 4G and 4H, we divided trials into two groups using the median difficulty as the threshold.

Statistics of inactivation effects

The measures described above were calculated separately for 'laser on' and 'laser off' trials, and effect sizes were calculated as the difference between the two. For measures that are not in percentage units (e.g., speed), effect size was calculated as percentage change in relation to control data [i.e., 100 x (laser – control) / control)]. p values for the effects were calculated following the bootstrapping procedure described in Guo et al. (2014). Data from different mice were combined, and in each of 10,000 iterations, subjects, sessions and trials were sampled with replacement in this order, and effects were calculated as above. The p value was the proportion of iterations in which the sign of the effect was different than the sign calculated using all the data. Because this is a

two-sided test, significance levels were set to 0.025, and corrected for multiple comparisons using a false discovery rate procedure (FDR, see General statistics). Error estimates on the behavioral indicators were calculated as the standard deviation across the bootstrapping iterations. Effect sizes between different inactivation locations or groups (e.g., ChR2+ versus ChR2– controls) were statistically compared using the same bootstrapping method.

Electrophysiology

Semi-automatic spike sorting was done using Klusta (Rossant et al., 2016), and manually curated by merging or excluding clusters based on waveform similarity and time course of principal components. Clusters were considered to be single units if fewer than 1% of spikes violated a 2-ms refractory period and were classified as multi-unit otherwise. Single and multi-units were analyzed together for the effects of laser inactivation. For each unit, waveforms were extracted for the channel with the largest amplitude, which also determined the cortical recording depth assigned to it. We classified units as putative fast spiking (FS) cells if their peak-to-trough width was smaller than 0.5 ms. Peristimuls time histograms were built with a bin width of 100 ms, and the amount of suppression by the laser in % was calculated as $100 - 100 \times FR_{laser\ on}/FR_{baseline}$, where *FR* is the firing rate, *laser on* is the full 1.5 s laser on period and *baseline* is the 1.5 s period preceding laser onset. For putative FS cells, % excitation was calculated as $100 \times FR_{laser\ on}/FR_{baseline}$. Laser offset rebound (%) was calculated as $100 \times (FR_{laser\ off} - FR_{baseline})/FR_{baseline}$, where *laser off* is the 0.5 s period following laser offset.

Wide-field image preprocessing

Images in the sequence were corrected for motion artifacts by applying the x-y shift that maximized the cross-correlation between each frame and a reference frame given by the mean across all frames. This average image was also used to detect large surface blood vessels using a median filter iteratively applied at several spatial scales. The output of the filter was then thresholded to determine pixels corresponding to vasculature. The parameters of the algorithm were adjusted manually for each mouse. The final output was a vasculature and off-headplate mask, which was applied to the entire image sequence to convert fluorescence values for pixels in the mask to NaN. Note that we chose to mask large vessels because they have different hemodynamics than parenchymal vessels (Ma et al., 2016a), such that the assumptions underlying our correction procedures (below) may not hold. Following this masking step, we spatially binned the images to 128 × 128 pixels (pixel size of ~68 μm) and separated interleaved 470- and 410-nm-illumination frames into two separate stacks for hemodynamic contamination correction and ΔF/F calculation as described below.

Isosbestic correction of hemodynamic artifacts and ΔF/F calculation

We used an approach similar to one described previously (Allen et al., 2017), with a few crucial differences. The approach relies on the fact that when 410-nm (the isosbestic) wavelength is used to excite GCaMP, the indicator produces green wavelength emissions that do not depend on intracellular Ca^{2+} concentration (Tian et al., 2009), and can thus be used to estimate Ca^{2+}-independent changes in measured fluorescence (Lerner et al., 2015), which presumably largely come from hemodynamic contamination (Ma et al., 2016a). An advantage of this method is that it does not rely on reflectance measures, such that the photon path length of the two wavelengths is much more similar, potentially resulting is less approximation error. For each pixel, we first smoothed the 410-excitation trace, F_v, using a running Gaussian window (σ = 400 ms) to get rid of high-frequency noise. We then used least-squares regression to find a scaling factor and an offset so as to minimize the distance between F_v and 470-excitation fluorescence, F_b, thus obtaining a corrected version of F_v, F_v'. We found this step to be necessary because we determined that ~20% of the 410-nm signal comes from autofluorescence of the metabond and/or nail polish, such that uncorrected F_v resulted in underestimation of fractional fluorescence changes. The heuristically determined correction factors agreed well with independent autofluorescence estimates (not shown). After applying this correction, we calculated fractional fluorescence changes as F/F_0, separately for F_b and F_v', where F_0 for each channel was calculated as the mode of all F values over a 30 s sliding window with single-frame steps. We thus obtained pixel-wise traces of fractional fluorescence with either 470- or 410-excitation, R_b and R_v, respectively. Because hemodynamic contamination is multiplicative, we then applied a divisive correction: $\Delta F/F = R_b / R_v - 1$ (Ma et al., 2016a). For ROI analyses, we first averaged F_b and F_v across all pixels belonging to the ROI, and calculated ΔF/F as described above.

Sensory maps

Retinotopic maps were computed following a previously described method (Garrett et al., 2014; Kalatsky and Stryker, 2003; Zhuang et al., 2017). Briefly, ΔF/F was calculated for each pixel (where F_0 was defined as above) and averaged over repeats for each of the four cardinal directions. Averages were Fourier decomposed and retinotopic positions were extracted from the phase of the spectral peak corresponding to the temporal frequency of the stimulus drift. Azimuth and altitude maps were combined into a field sign map by taking the sine of the angle between the local gradients of the two. V1 borders were manually drawn on the field sign map. Airpuff response maps were constructed by averaging response z-scores over trials, calculated as ΔF/F responses following 400 ms of the airpuff subtracted by the baseline ΔF/F (400 ms preceding the airpuff) and divided by the baseline standard deviation. The final map was obtained by thresholding pixels contralateral to the stimulus with a z-score higher than 1 for at least 100 ms. To calculate average maps across mice, we chose a reference visual map and registered the others to it using rigid transformations. These transformations were then applied to the airpuff maps.

Atlas registration and ROI selection

Data were downloaded from the Allen Brain Atlas API Mouse Brain (ccv3) at 50-μm resolution and Z-projected to create a flat map of the following cortical areas (Allen abbreviation in parentheses): V1 (primary visual, VISp), mV2 (medial secondary visual, VISam and VISpm), PPC (posterior parietal cortex, overlapping with anterior visual area, VISa), RSC (retrosplenial cortex, RSP), SS (somatosensory cortex, SS), M1 (primary motor cortex, MOp), M2 (secondary motor cortex, MOs). Because we observed different inactivation effects (Figures 1, 2, S3), and because functional gradients have been previously observed (Chen et al., 2017), M2 was then subdivided into a medial and an anterior portion according to an anteroposterior line 1.25 mm lateral to bregma. The flat map was then registered to each animal by creating an image containing only V1 pixels from the atlas and another with V1 pixels obtained from visual area mapping. These two images were registered using affine transformations, which were then applied to the atlas coordinates to extract pixels belonging to each ROI for a given imaging session. Different sessions from the same animal were aligned using rigid transformations. For the analysis in Figure S7, we defined ROIs based on the coordinates used for whole-trial inactivation (Figure 1E), but separately for each hemisphere. We thus selected 58 non-overlapping, square-shaped ROIs of side 0.5 mm, centered at the coordinates as described above.

Quantification of average time course

Unless otherwise stated, analyses were based on spatially downsampled activity, in 5-cm steps for the average time course (Figures 3 and S6). For each trial and y position in the maze, we sampled ΔF/F from a single frame corresponding to the first point in time that the mouse visited that position. For the analysis in Figure S6, for each behavioral session we calculated the C.O.M. for each pixel in the image as described above and created a vector of pixel C.O.M.s for each task (accumulating-towers and control) and trial type (left- and right-choice), and computed linear Pearson correlation coefficients between different combinations of these vectors. For task comparisons, correlation coefficients were first calculated separately for right- and left-trials and then averaged.

Linear fits for evidence tuning

For the analysis in Figure 5, average anatomical ROI ΔF/F was z-scored individually for each ROI and session. We then calculated the average z-scored activity during the cue period, $0 \leq y \leq 200$ cm, separately for each final value of sensory evidence (Δ towers), divided into 6 equally sized bins, between 0 and 15 Δ towers, and converted into contra- or ipsilateral counts (from #R – #L) for averaging across homotopic ROIs. These values were then averaged across sessions for each mouse (n = 6). We used the ΔF/F values per evidence bin to fit a line separately for ipsi- and contralateral evidence, repeating the procedure 50 times, resampling the mouse-wide averages with replacement. This procedure was done separately for each ROI, task and correct or error trials. The slope of the evidence-tuning line was considered statistically significant if the average slope was beyond (or below) zero by at least two standard deviations of the distribution given by the 50 bootstrapping iterations.

ROI correlations

Spontaneous (running in dark) correlations (Figure 4A) were calculated for a period of 10 minutes when the mice ran inside the VR setup, with the projector turned off. Task correlations were calculated for spatially downsampled activity as described above, using 10-cm bins, only for correct trials. Adding error trials or using all imaging frames while the animals performed either task yielded very similar results (not shown). For the analysis in Figure 4E, correlations were calculated using the whole activity traces (i.e., not downsampling or selecting trials), and behavioral performance was calculated for the entire session, including trial blocks that may have been excluded from other analyses for not meeting the performance threshold (60% correct overall).

Decoding models

We trained linear decoders for upcoming choice, accumulated evidence and previous choice based on the simultaneous Ca^{2+} activity of all the 16 ROIs, spatially binned pixels for whole-cortex decoders, or all pixels within each pair of homotopic ROIs. To decorrelate evidence from choice, we used both correct and error trials when fitting the models. The models were fit with 50 runs of 3-fold cross-validation using L2 penalty (ridge regression). For each spatial position (0 – 300 cm, with activity downsampled in 10-cm steps as described previously) and each of 20 values of the penalty term λ, we fitted a linear regression model using $^2/_3$ of the data, from both correct and error trials:

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \lambda \| \vec{B} \|$$

where y = cumulative #R – #L towers for that position in the case of evidence, or ± 1 for right and left choices, respectively, in the case of upcoming and previous choice; β_i is the decoding weight for each pixel (ROI), β_0 is an offset term, n is the number of pixels (ROIs), X is z-scored ΔF/F for each pixel (ROI), and $\| \vec{B} \|$ is the L2 norm of the vector of decoding weights. For each fit, we calculated decoding accuracy on the $^1/_3$ of the data not used to fit the model, using the best value of λ determined from training data. For evidence decoding, accuracy was defined as the Pearson correlation coefficient between actual and decoded evidence across trials, for each spatial bin. For choice and previous choice, we picked an optimal decoding boundary that most equally separated the left- and right- choice trial distributions on the training set (see behavioral view angle analysis above), and applied that boundary to classify trials in the test

set. Accuracy was defined as the proportion of trials in the test set with correct classification. We picked the value of λ that maximized accuracy, and used median accuracy and weight values across all 50 × 3 runs for that λ. Shuffles were constructed by randomly permuting the y values in the training set and calculating accuracy on non-shuffled test sets, performed 10 times with 3-fold cross-validation (i.e., 150 times) for the best λ value only. For the view angle control analysis (Figure S8), we first fit the view angle θ to the activity of each pixel, X, i.e., $X = a\ \theta + b$, and used the residuals of this regression as the inputs to the decoder instead of X, and used the same fitting procedure as described above.

Event-triggered averages

Tower, post- and pre-turn triggered averages (Figures S9A–S9F) were calculated in time for each pixel (ROI) by aligning all trials to event onset and z-scoring post-event activity (0 – 2 s) to a 500-ms baseline preceding the event. In the case of pre-turn averages, activity was defined as the 2 s preceding turns, and baseline was taken to be the 500 ms following the turn.

Linear encoding model (GLM)

We fitted the Ca^{2+} activity of each ROI with a generalized linear model (GLM) (Pinto and Dan, 2015; Scott et al., 2017), where z-scored ΔF/F for each spatial position y (0 – 300 cm, in 1-cm steps) was described as a linear combination of task events with different spatial lags, with L2 penalty and 3-fold cross-validation:

$$\Delta F/F(y) = \beta_0 + \sum_{i=1}^{100} \beta_i^{tR} E_{y-i}^{tR} + \sum_{i=1}^{100} \beta_i^{tL} E_{y-i}^{tL} + \sum_{i=1}^{50} \beta_i^{\Delta} E_{y-i}^{\Delta} + \sum_{i=-30}^{30} \beta_i^{d\theta/dt} E_{y-i}^{d\theta/dt} +$$
$$\sum_{i=-30}^{30} \beta_i^{sp} E_{y-i}^{sp} + \beta^{\theta}\theta(y) + \beta^{ch} ch + \beta^{pch} pch + \beta^{prw} prw + \lambda \| \overrightarrow{B} \|$$

where β_i^x is the encoding weight for predictor x at spatial lag i, β_0 is an offset term, E_y^x is a delta function indicating the occurrence of event x at position y, λ is the penalty term and $\| \overrightarrow{B} \|$ is the L2 norm of the weight vector. *tR* indicates the occurrence of a right tower, *tL* of a left tower, Δ = cumulative #R – #L towers, *dθ/dt* is virtual view angle velocity, *sp* is running speed, θ is view angle (no lags), y is spatial position in the maze stem (no lags), and *ch*, *pch* and *prw* are constant offsets for a given trial, indicating upcoming choice, previous choice (+1 for right and –1 for left) and previous reward (1 for reward and –1 otherwise), respectively. The exact combination of model parameters and number of lags for each parameter was chosen because it maximized cross-validated accuracy in an initial model comparison. Cross-validation was performed as described for the decoding models, selecting $^2/_3$ – $^1/_3$ of trials (both correct and error) instead of individual data points (i.e., points from the same trial were always in the same set). Model accuracy was defined as the Pearson correlation between actual ΔF/F and that predicted by the model.

Clustering

We performed hierarchical clustering on correlation matrices as follows. For Figure 2, a 29 × 29 cortical region correlation matrix was calculated from 5-element vectors given by inactivation effect sizes on performance, absolute side bias, speed, excess travel and rate of motor errors. For Figure 4, we used the 16 × 16 ROI correlation matrix from correct trials in the accumulating-towers task, although using data from the other conditions, or including error trials, yielded nearly identical results. For Figure S7, we used correct trials from the towers task to compute two separate 29 × 29 inactivation-grid-based ROI correlation matrices, one for each hemisphere, and averaged the two. We did so in order to be able to directly compare between bilateral inactivation and imaging results. We then performed principal components analysis (PCA) on the correlation matrices and selected the top components that accounted for at least 90% of variance – yielding 3 principal components (PCs) for Figures 2 and 4 (99.8% and 92.1% variance explained for inactivation and activity correlation, respectively) and 4 PCs for Figure S8 (90.8% of variance) – thus obtaining an n area x n PC matrix. This matrix was then used to calculate Euclidean distances between the data points, providing the input to the hierarchical clustering algorithm. We tested between 2 and 5 clusters, and the optimal number of clusters, k, was given by the k that maximized clustering quality as measured by the Calinski-Harabasz index. For ease of comparison, GLM data (Figure S9) were divided according to the clusters obtained from the correlation analysis (Figure 4), but clustering these data separately yielded very similar ROI clusters (data not shown).

General statistics

For non-bootstrapping-based analyses, we first tested datasets for normality using the Lilliefors' modification of the Kolmogorov-Smirnov test. We used two-sided t tests to compare normally distributed datasets and the Wilcoxon sign rank test otherwise (or their paired test counterparts where appropriate). We corrected for multiple comparisons using FDR (Benjamini and Hochberg, 1995; Guo et al., 2014). Briefly, p values were ranked in ascending order, and the ith ranked p-value, P_i, was deemed significant if it satisfied $P_i \leq (\alpha i)/n$, where n is the number of comparisons and α is the significance level. For comparisons among multiple groups, we performed one- or two-way ANOVAs (with repeated-measures if group data were not independent), followed by Tukey's post hoc tests where appropriate. Binomial confidence intervals were calculated as 1-standard-deviation intervals using Jeffrey's method. Unless otherwise stated, imaging data were averaged over sessions for each animal first, and central tendency and error estimates were calculated across mice (n = 6).

Recurrent Neural Network Model

Network elements

We consider a network of N interconnected model neurons ('units') described by a standard firing rate model. Each model neuron is characterized by an activation variable x_i for $i = 1, 2, 3, \ldots, N$ (N = 1000) and a nonlinear response function, $\phi(x) = \tanh(x)$. The model neurons' connections are determined by a recurrent synaptic weight matrix **J** with element J_{ij} representing the strength of the connection from presynaptic neuron *j* to postsynaptic neuron *i* (Figures 7A and S10A). Individual synaptic weights are changed by applying the full-FORCE learning algorithm (DePasquale et al., 2018). The activation variable for each neuron $x_i$is determined by:

$$\tau \frac{dx_i}{dt} = -x_i + \sum_j^N J_{ij}\phi(x_j) + h_i$$

In the above equation, τ = 10 ms is the time constant of each unit in the network, which sets the timescale of network dynamics. h_i is the external input to the unit *i*. The network equations are integrated using Euler method with an integration time step, dt = 0.1ms.

Design of inputs and network output

The input $h_i = u_{in} f_{in}$ is provided to 50% of the network neurons (''posterior module'' indicated in green in Figure 7A) through the vector of input weights, u_{in}. To simulate the experimental design for the accumulating-towers task, f_{in} takes the form of two trains of square waves or ''pulses,'' one corresponding to the left and the other, to the right input (bottom left of Figure 7A). The pulses are generated independently from a Poisson distribution and, in each trial, the network experiences a different number of pulses (up to 10 on each side). For the visually-guided task, f_{in} consists of pulses as explained above, and an additional constant input applied for the duration of the cue period (Figure 7A, bottom left). For both tasks, the cue period is 2.5 s long.

After training, the network should respond to the pair of pulse trains in the accumulation task by producing a target output ''bump'' ($f_{out}(t)$; gray trace on the top right, Figure 7A) of duration 0.5 s at the end of the cue period. The output should be upward when the number of pulses on the right exceed the number of pulses on the left, and downward otherwise. In the visually-guided task, the upward bump indicates the presence of a constant input indicating a right choice and a downward bump indicates a constant input indicating a left choice. The actual output of the network, $z(t)$, is a linear combination of the activity of the 50% of the network neurons that are not directly input-driven ('frontal module', Figure 7A): $z(t) = \sum_j w_j \phi(x_j)$. Successful performance of the task is achieved by matching $z(t)$ to the target output $f_{out}(t)$, up to a desired degree of accuracy. The readout weights **w** are plastic and modifiable by recursive least-squares (Haykin, 1986; see also Mante et al., 2013; Rajan et al., 2016; Sussillo and Abbott, 2009). A 'go' signal was applied to instruct the network when to report a choice. The go signal was shaped as a quadratic function (top left of Figure 7A) to mimic the shape of the global ramps we observed in Ca^{2+} activity (Figure 3).

Network training and performance evaluation

The full-FORCE (DePasquale et al., 2018) algorithm modifies the recurrent synaptic weights J by comparing the input into each unit with a 'target' function. These target functions are the activity of a second, 'teacher' network (N = 1000) in which there are no plastic synapses. This network obeys the same dynamical equations as described above and is driven by the same inputs as the network it is meant to train. However, it has random, fixed recurrent connections (Gaussian distributed with zero mean and variance g^2 / N), and receives an extra ''hint'' input that enforces a specific computational strategy or rule for performing the task. For the accumulating-towers task, the hint is the instantaneous integral of the difference in pulse counts, $\int (f_{in}^R - f_{in}^L) dt$ (Figure S10B). There is no hint input for the visually-guided task.

We extend the above technique in two important ways. First, like the learner network, the teacher network is constructed to have two equally-sized modules, with strong, fully-reciprocal connectivity *within* each module (g = 1.5, connection probability, P_{conn} = 1) and weaker, sparse connectivity *between* the two modules (g = 0.75, P_{conn} = 0.05, but a range of values did not affect our main conclusions, Figure S10). This modularity gives rise to specific features in its activity that the learner network must adopt. Second, the trained network operates in a *multi-task mode* (Yang et al., 2019), in which it switches randomly between performing the accumulation or guided task.

During training, the recurrent weights are modified by a recursive least-squares algorithm (Haykin, 1986), which compares the activity of the teacher and learner network (DePasquale et al., 2018; Rajan et al., 2016; Sussillo and Abbott, 2009). The convergence of the training procedure was assayed by calculating the squared error between the individual network neurons and their target functions. We trained the network for 100 learning steps and computed performance after training based on an additional 100 steps. A ''step'' is defined as one run of the program for the duration of the relevant trial (T = 3 s).

The behavioral performance of the trained network was computed as the percentage of trials in which the network produced a correct response relative to the total number of trials the trained network was tested on. A response was considered correct if the sign of its integral during the response period matched that of the target output. A typical trained network was able to perform the task to a percent correct level of 100% (Figure 7E). Performance was evaluated in the same way after silencing neurons, either picked at random (Figure 7E), or from specific modules (Figure S10E). Each data point in silencing experiments consisted of 50 runs, each with different sets of random units being silenced.

DATA AND CODE AVAILABILITY

Data analysis code is available at https://github.com/BrainCOGS/widefieldImaging and https://github.com/BrainCOGS/behavioralAnalysis. Additionally, code for laser-scanning control software is available at https://github.com/BrainCOGS/laserGalvoControl. The curated raw datasets are still being documented and present technical difficulties for being deposited in a public repository due to their large size. We are attempting to resolve these difficulties, but in the meantime the data are available from the corresponding authors upon request. MATLAB source data for the main figures in the paper are available at Mendeley Data, https://doi.org/10.17632/d2dkk9647b.1.

Neuron, Volume *104*

Supplemental Information

Task-Dependent Changes in the Large-Scale Dynamics and Necessity of Cortical Regions

Lucas Pinto, Kanaka Rajan, Brian DePasquale, Stephan Y. Thiberge, David W. Tank, and Carlos D. Brody

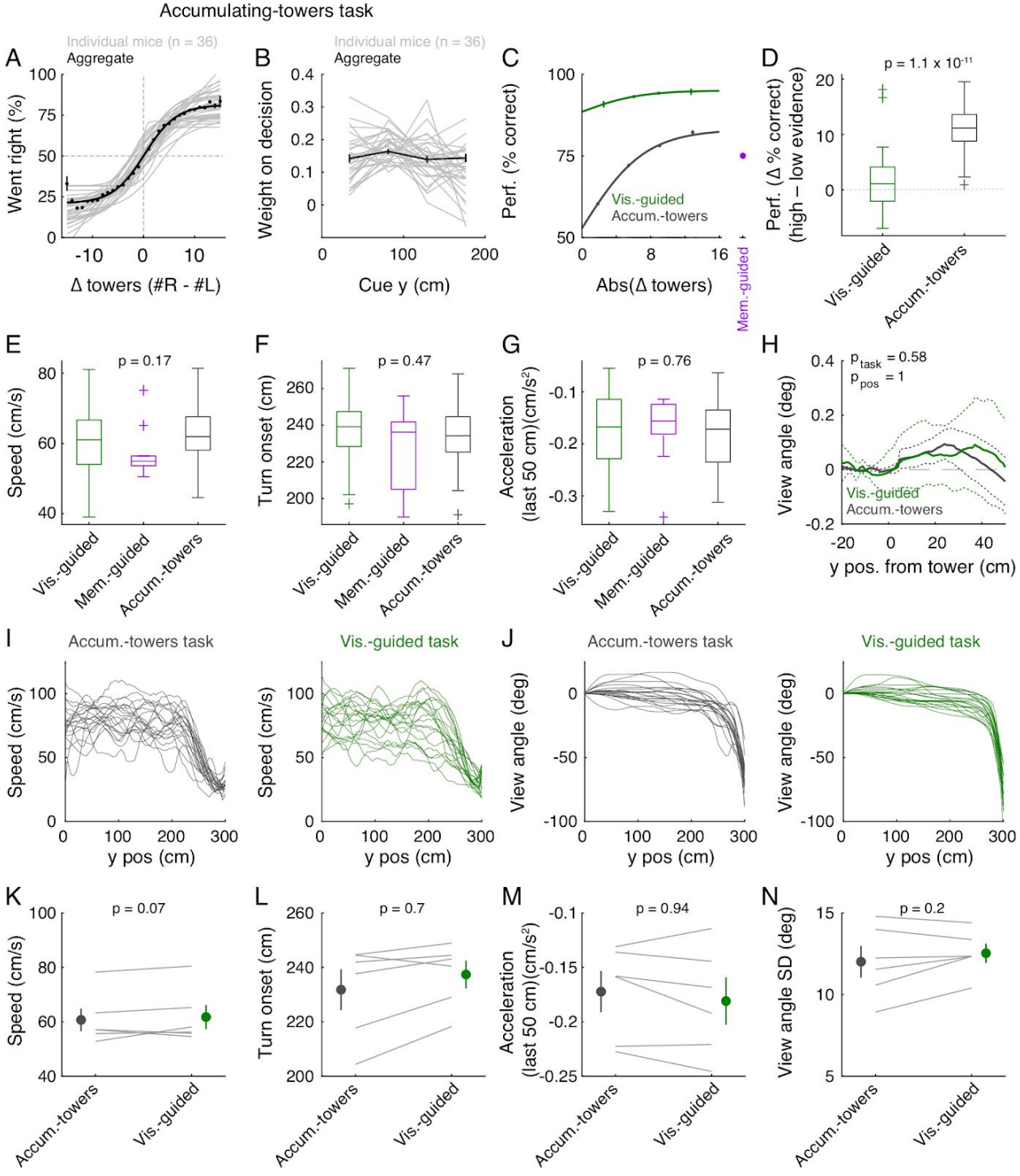
Accumulating-towers task
A
Individual mice (n = 36)
Aggregate
Went right (%)
Δ towers (#R - #L)
B
Individual mice (n = 36)
Aggregate
Weight on decision
Cue y (cm)
C
Perf. (% correct)
Vis.-guided
Accum.-towers
Abs(Δ towers)
Mem.-guided
D
p = 1.1 x 10-11
Perf. (Δ % correct)
(high – low evidence)
Vis.-guided
Accum.-towers
E
p = 0.17
Speed (cm/s)
Vis.-guided
Mem.-guided
Accum.-towers
F
p = 0.47
Turn onset (cm)
Vis.-guided
Mem.-guided
Accum.-towers
G
p = 0.76
Acceleration
(last 50 cm)(cm/s2)
Vis.-guided
Mem.-guided
Accum.-towers
H
ptask = 0.58
ppos = 1
View angle (deg)
Vis.-guided
Accum.-towers
y pos. from tower (cm)
I
Accum.-towers task
Vis.-guided task
Speed (cm/s)
y pos (cm)
J
Accum.-towers task
Vis.-guided task
View angle (deg)
y pos (cm)
K
p = 0.07
Speed (cm/s)
Accum.-towers
Vis.-guided
L
p = 0.7
Turn onset (cm)
Accum.-towers
Vis.-guided
M
p = 0.94
Acceleration
(last 50 cm)(cm/s2)
Accum.-towers
Vis.-guided
N
p = 0.2
View angle SD (deg)
Accum.-towers
Vis.-guided

Figure S1. Task performance comparison, related to Figure 1.

(A) Performance of the accumulating-towers task by each mouse in this study, as assessed by a psychometric functions for % right choice as a function of the Δ number of towers, defined as #R – #L towers. Thin gray lines: best-fitting psychometric functions for each mouse (n = 36). Thick black line: aggregate data (n = 194,934 trials). Error bars: binomial confidence intervals. (B) Weights of a logistic regression model in which choice is predicted by Δ towers from 4 equally-spaced segments of the cue region, showing that the mice on average weight evidence fairly evenly across different portions of the maze. Thin gray lines: individual mice, thick black line: average across mice. Error bars: ± SEM. (C) Comparison of aggregate psychometric functions between the accumulating-towers and visually-guided tasks (n = 36 and 37 mice, and 194,934 and 28,759 trials, for towers and guided task, respectively). Memory-guided task performance is added for comparison. Error bars: binomial confidence intervals, circles: average, lines: best-fitting psychometric functions. (D) Boxplot quantifying the modulation of performance by the amount of evidence in the accumulating-towers and the visually-guided tasks, given by the difference in performance between trials in top half vs. the bottom half of the Δ towers distribution. p-value is from a *t* test. (E) – (G) Comparison of motor indicators across the three tasks taking into account all mice, respectively for average running speed, estimated point of turn onset, and average acceleration during the last 50 cm of the maze stem. p-values are from a one-way ANOVA. (H) Average tower-triggered changes in view angle for the accumulating-towers and the visually-guided tasks, for mice used in imaging experiments (n = 6). Positive values indicate turns towards the side where the tower appeared. P-values are from a repeated-measures two-way ANOVA with factors task and position. Dotted lines: ± SEM across mice. (I) Speed traces from example left-turn trials (n = 20) from one of the mice used in imaging experiments, performing the accumulating-towers and the visually-guided tasks within the same sessions. Left: accumulating-towers task, right: visually-guided task. Speed was downsampled to 5-cm spatial bins and smoothed with a 25-cm running window for display only. (J) Same as I, for view angle. (K) – (N) Paired comparison of motor parameters for the mice used in imaging experiments, performing the accumulating-towers and the visually-guided tasks within the same sessions. P-values are from paired *t* tests. Lines: individual mice. Circles: population average (n = 6). Error bars: : ± SEM.

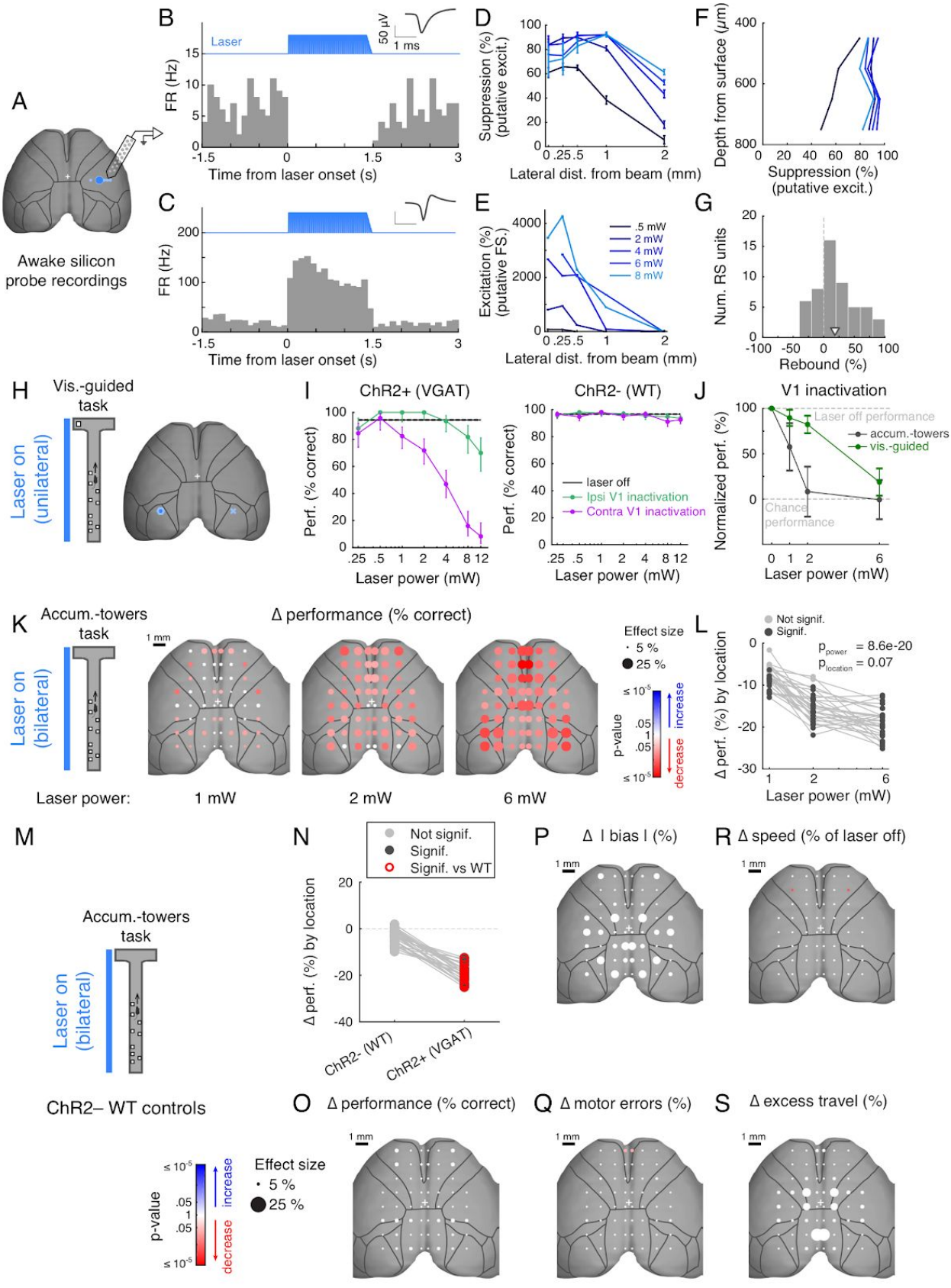
A
Awake silicon probe recordings
B
Laser
50 µV
1 ms
FR (Hz)
Time from laser onset (s)
C
FR (Hz)
Time from laser onset (s)
D
Suppression (%) (putative excit.)
Lateral dist. from beam (mm)
E
Excitation (%) (putative FS.)
.5 mW
2 mW
4 mW
6 mW
8 mW
Lateral dist. from beam (mm)
F
Depth from surface (µm)
Suppression (%) (putative excit.)
G
Num. RS units
Rebound (%)
H
Vis.-guided task
Laser on (unilateral)
I
ChR2+ (VGAT)
ChR2- (WT)
Perf. (% correct)
laser off
Ipsi V1 inactivation
Contra V1 inactivation
Laser power (mW)
J
V1 inactivation
Laser off performance
accum.-towers
vis.-guided
Chance performance
Normalized perf. (%)
Laser power (mW)
K
Accum.-towers task
Laser on (bilateral)
Δ performance (% correct)
1 mm
Effect size
5 %
25 %
p-value
increase
decrease
Laser power:
1 mW
2 mW
6 mW
L
Not signif.
Signif.
p_power = 8.6e-20
p_location = 0.07
Δ perf. (%) by location
Laser power (mW)
M
Accum.-towers task
Laser on (bilateral)
ChR2– WT controls
N
Not signif.
Signif.
Signif. vs WT
Δ perf. (%) by location
ChR2- (WT)
ChR2+ (VGAT)
P
Δ | bias | (%)
R
Δ speed (% of laser off)
O
Δ performance (% correct)
Q
Δ motor errors (%)
S
Δ excess travel (%)

Figure S2. Cortical spread of inactivation and ChR2– controls, related to Figure 1.
(A) We recorded single (n = 29) and multi-unit (n = 45) activity from 3 awake VGAT-ChR2-EYFP mice using a silicon probe spanning all layers while inactivating the cortex at various lateral distances from the probe and with varying laser powers. (B) Example peristimulus time histogram (PSTH) of a putative excitatory unit (inset: average waveform) being silenced upon laser onset (blue trace, n = 10 trials at 6 mW and 0.5 mm of lateral distance from the beam). (C) Example PSTH of a putative fast-spiking inhibitory unit increasing its firing rate upon laser onset. Conventions as in B. (D) Average suppression of activity as a function of lateral distance from the laser beam and laser power (legend in E) for putative excitatory units (n = 55). Error bars, ± SEM. (E) Average increase in firing rate for putative fast spiking inhibitory neurons (n = 19). Conventions as in D. (F) Average suppression of activity as a function of depth from cortical surface and laser power (legend in E) for putative excitatory units. (G) Distribution of post-laser offset rebound in the firing rate of putative excitatory units relative to pre-laser baseline, measured during the 500 ms following laser offset. Arrowhead: mean. (H) We performed whole-trial, unilateral inactivation of the visual cortex with varying laser powers, as VGAT-ChR2-EYFP (n = 2, mean of 52.6 ‘laser on’ trials/power/location) or ChR2– wild type controls (n = 3, mean of 238.3 ‘laser on’ trials/power/location) performed the visually-guided task. (I) Unilateral visual cortical inactivation led to disruption of performance of trials where reward was contralateral to the inactivated hemisphere, in a laser power-dependent fashion, specifically in animals expressing ChR2. Error bars: binomial confidence intervals. (J) We directly compared the behavioral effect of varying laser power in V1 inactivations for both tasks. For the visually-guided task, we combined the ipsi and contralateral trials in panel I, and used linear interpolation between 4 and 8 mW to estimate the effects at 6 mW. The data for the accumulating-towers task is from the bilateral inactivations shown in panels K, L. Performance was normalized such that 100 corresponds to control ('laser off') performance, and 0 corresponds to chance. (K) Whole-trial bilateral inactivation of 29 cortical patches during performance of the accumulating-towers task was done at three laser powers: 1 mW, 2 mW or 6 mW. Shown for each power are the resulting spatial maps of inactivation effects on behavioral performance. The size of each circle indicates the size of the effect (caption on the right), given by the difference in % correct between 'laser on' and 'laser off' trials. The color indicates sign of the effect (red: decreased, blue: increase), and the saturation is proportional to the p-value (color bar, see Methods), thresholded such that non-significant effects appear white. 13/29 bilateral locations were significant at 1 mW (n = 6 mice, 455.6 ‘laser on’ trials/location). Inactivation at 2 mW significantly affected performance in 26/29 bilateral targets (n = 7 mice, 268.5 ‘laser on’ trials/location). The 6-mW figure is reproduced from Fig. 1 for comparison. (L) Drop in performance for each of the 29 bilateral targets, for the three laser powers in panel K. Dark gray circles indicate significant effects, and light gray circles, non-significant effects. We used two-way ANOVA to quantify the effect size as a function of power and location. Power significantly affected effect size ($p = 8.6 \times 10^{-20}$), but the effect size was not significantly different across locations ($p = 0.07$). (M) We performed whole-trial laser stimulation in wild type mice not expressing ChR2 while they performed the accumulating-towers task (n = 3, 4526 ‘laser on’ trials, mean of 150.9 trials/location), targeting the same 29 bilateral locations as in the experiments with VGAT-ChR2-EYFP mice (ChR2+), with a laser power of 6 mW. (N) Comparison of change in performance of ‘laser on’ trials between ChR2– controls and ChR2+ mice. None of the 29 locations had significant effects compared to ‘laser off’ trials in ChR2– controls (light gray circles), whereas all did in ChR2+ mice (dark gray circles). When compared to ChR2– experiments, 27/29 locations in ChR2+ had significant effects (open red circles). (O) – (S) No effects of laser stimulation on a number of behavioral indicators. In each panel, the size of each circle indicates the size of the effect (caption on the left), given by the difference between laser 'off' and laser 'on' trials. The color indicates sign of the effect (red: decreased, blue: increase), and the saturation is proportional to the p-value (color bar, see Methods), thresholded such that non-significant effects appear white.

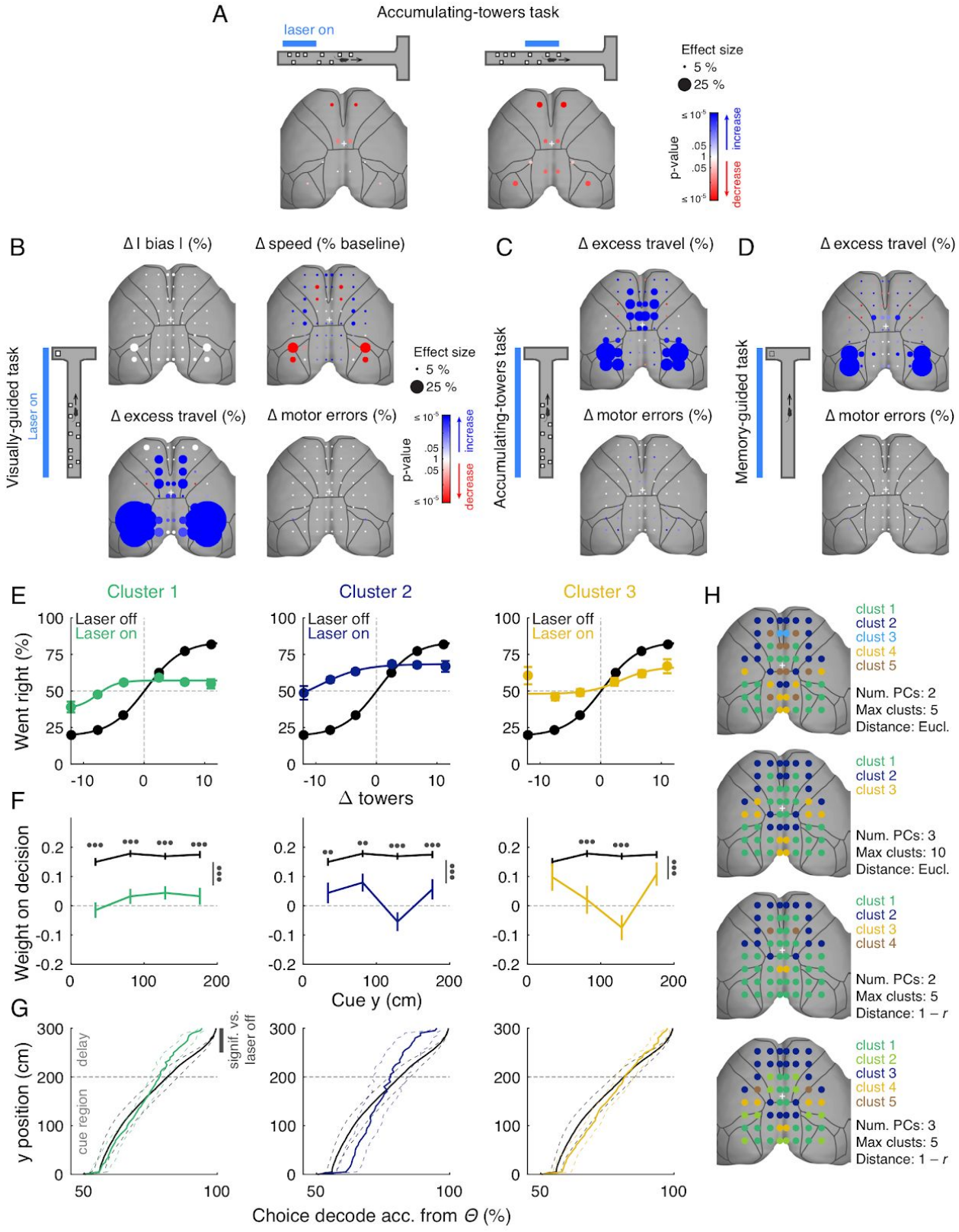
A
Accumulating-towers task
laser on
Effect size
5 %
25 %
p-value
increase
decrease
B
Δ | bias | (%)
Δ speed (% baseline)
Visually-guided task
Laser on
Δ excess travel (%)
Δ motor errors (%)
C
Δ excess travel (%)
Accumulating-towers task
Δ motor errors (%)
D
Δ excess travel (%)
Memory-guided task
Δ motor errors (%)
E
Cluster 1
Cluster 2
Cluster 3
Laser off
Laser on
Went right (%)
Δ towers
F
Weight on decision
Cue y (cm)
G
y position (cm)
delay
cue region
signif. vs. laser off
Choice decode acc. from Θ (%)
H
clust 1
clust 2
clust 3
clust 4
clust 5
Num. PCs: 2
Max clusts: 5
Distance: Eucl.
Num. PCs: 3
Max clusts: 10
Distance: Eucl.
Num. PCs: 2
Max clusts: 5
Distance: 1 – r
Num. PCs: 3
Max clusts: 5
Distance: 1 – r

Figure S3. Distinct but overlapping behavioral effects of inactivating different regions, related to Figures 1 and 2.

(A) Widespread effects of inactivating a subset of five regions during the first half (n = 10 – 12 mice/location, average of 1078.6 'laser on' trials/location) or second half (n = 6 – 12 mice/location, average of 987.0 'laser on' trials/location) of the cue region in the accumulating-towers task. The size of each circle indicates the size of the effect (caption on the right), given by the difference in % correct between laser 'off' and laser 'on' trials. The color indicates the sign of the effect (red: decreased, blue: increase), and the saturation is proportional to the p-value (color bar, see Methods), thresholded such that non-significant effects appear white. (B) Effects of inactivation of each of 29 bilateral regions on absolute magnitude of response side bias, running speed, excess travel (trajectory longer than 110% of nominal maze length), and the rate of motor errors (trials with early turns, large view angles, erroneous arm entries, very low speed or high excess travel, see STAR Methods for details) during the visually-guided task. The size of each circle indicates the size of the effect (caption on the right), given by the difference between laser 'off' and laser 'on' trials. The color indicates sign of the effect (red: decreased, blue: increase), and the saturation is proportional to the p-value (color bar, see Methods), thresholded such that non-significant effects appear white. (C) Same as B, for excess travel and rate of motor errors during the accumulating-towers task. (D) Same as C, for the memory-guided task. (E) – (G) Combined effects of inactivation of all cortical regions within each of the three clusters (Fig. 1K – J), for the accumulating-towers task. (E) Psychometric functions for 'laser on' and 'laser off' trials for the accumulating-towers task. Error bars: binomial confidence intervals, lines: best-fitting functions. (F) Logistic regression analysis of the amount of evidence per spatial bin on the mice's decision (as in Fig. S1B), fitted separately for 'laser on' and 'laser off' trials. Error bars: ± SD over bootstrapping iterations (n = 10,000). Closed circles indicate significance (two: $p < 0.01$, three: $p < 0.001$). Vertical bars with associated circles indicate significant changes in average evidence weights across spatial bins. (G) Decoding accuracy of final choice based on the view angle at each binned y position in the maze, performed separately for 'laser on' and 'laser off' trials. Larger numbers indicate larger separation in the view angle trajectories between left- and right-choice trials. Thick lines: average, thin dashed lines: ± S.D. over bootstrapping iterations (n = 10,000). Vertical bar: positions for which decoding accuracy was significantly different between 'laser on' and 'laser off' trials. (H) Cluster identities in the accumulating-towers task are fairly stable for different algorithm parameters (captions on the right). Clusters have been numbered and color coded to reflect proximity.

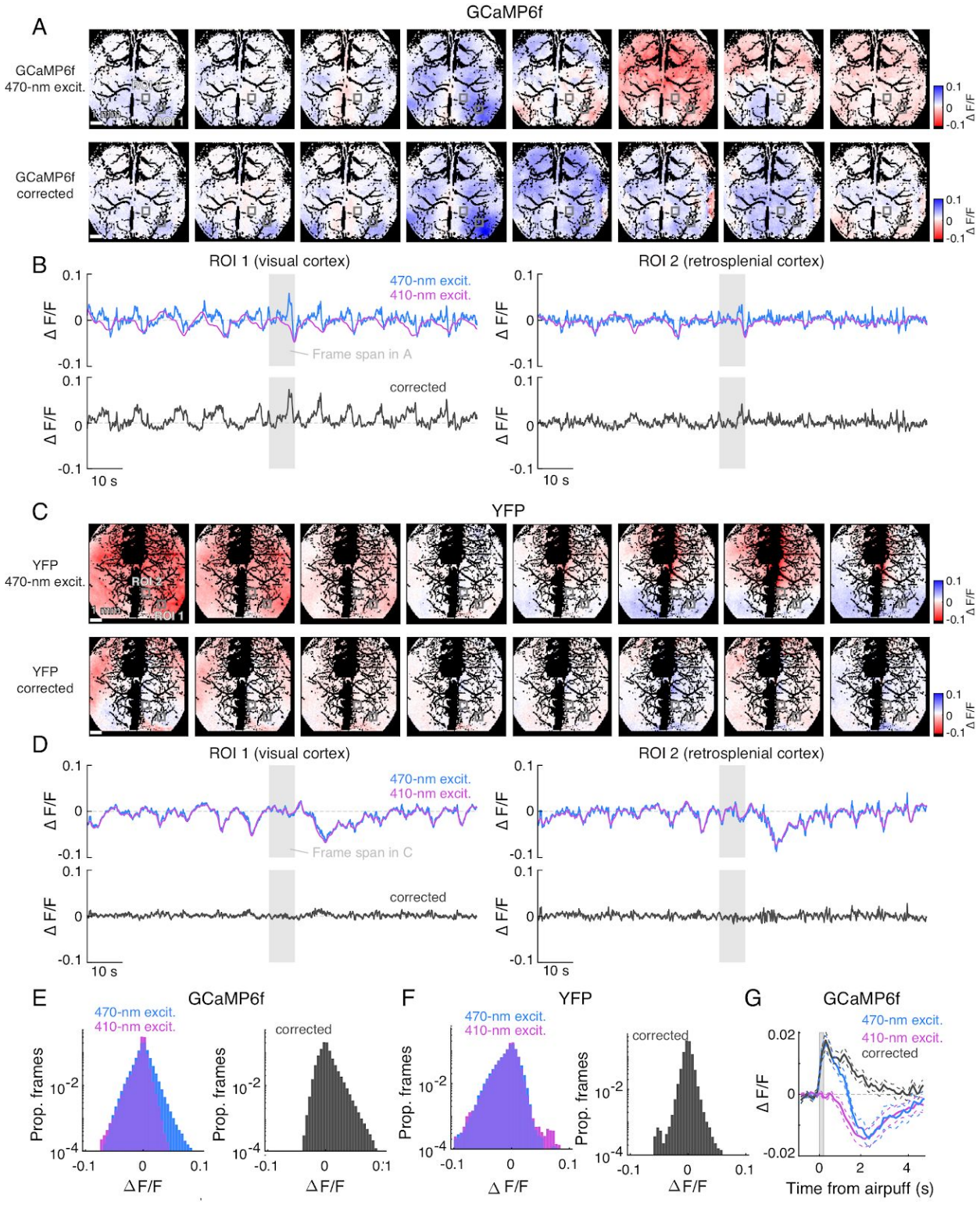
A
GCaMP6f
GCaMP6f
470-nm excit.
GCaMP6f
corrected
ROI 2
ROI 1
1 mm
0.1
0
-0.1
ΔF/F
B
ROI 1 (visual cortex)
ROI 2 (retrosplenial cortex)
470-nm excit.
410-nm excit.
Frame span in A
corrected
ΔF/F
10 s
C
YFP
YFP
470-nm excit.
YFP
corrected
D
ROI 1 (visual cortex)
ROI 2 (retrosplenial cortex)
470-nm excit.
410-nm excit.
Frame span in C
corrected
10 s
E
GCaMP6f
470-nm excit.
410-nm excit.
corrected
Prop. frames
ΔF/F
F
YFP
470-nm excit.
410-nm excit.
corrected
Prop. frames
ΔF/F
G
GCaMP6f
470-nm excit.
410-nm excit.
corrected
ΔF/F
Time from airpuff (s)

Figure S4. Isosbestic correction of hemodynamic artifacts in widefield Ca^{2+} imaging, related to Figure 3 and STAR Methods.

(A) Top: example imaging frames (downsampled to 1 Hz) showing pixel-wise ΔF/F for 470-nm excitation for a mouse expressing GCaMP6f. Bottom: same frames as top, after applying the isosbestic correction (see STAR Methods for details). (B) Top: example ΔF/F traces shown separately for frames with violet (410 nm) and blue (470 nm) excitation. Bottom: result of applying the isosbestic correction to the blue trace on top. Notice the correction of negative transients coming from hemodynamic contamination. Shown on the left is an example of correction on a visual cortical ROI (~600 x 600 μm, ROI 1 in panel A) and on the right a retrosplenial cortical ROI (ROI 2 in panel A), illustrating similar correction near large midline blood vessels. (C) – (D) Same as A and B, for a mouse expressing YFP. (E) Left: distribution of ΔF/F values across all experiments (n = 25 sessions from 6 Emx1-Ai93 mice) for 470-nm and 410-nm excitation GCaMP6f fluorescence for a visual cortical ROI. Right: distribution of corrected ΔF/F values across all experiments. (F) Same as E, for YFP fluorescence. (G) Trial-averaged response to 30 200-ms long air puffs to the right whisker pad, obtained by averaging left-hemisphere pixels with significant changes in fluorescence (see STAR Methods for details). Triggered averages are shown separately for violet and blue-excitation fluorescence, as well as the corrected trace, further illustrating the hemodynamic contamination and the outcome of the correction algorithm.

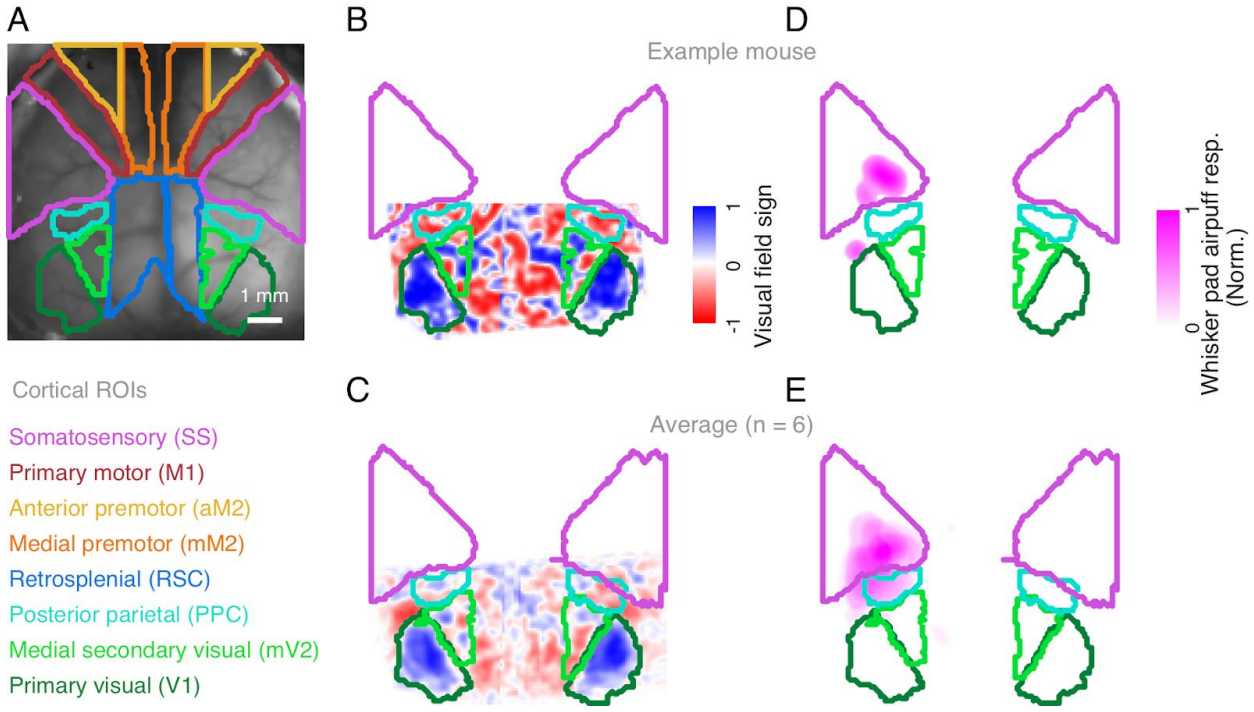

Figure S5. Anatomically-defined imaging regions of interest correspond well to measured sensory maps, related to Figure 3 and STAR Methods.

(A) Outlines of 8 different areas for each hemisphere were obtained from the Allen Brain Mouse Reference Atlas (ccv3) and fit to each mouse's brain by performing rigid registration between binary images containing V1 pixels (from the atlas vs. obtained with visual mapping). Cortical regions are color coded according to the naming conventions underneath. See STAR Methods for exact correspondence between Allen Brain nomenclature and the names used here. (B) Comparison between visual field sign maps (see Methods) and outlines of 4 ROIs from atlas fits for the same animal as in A. (C) Average visual field signs (n = 6) and ROI outlines. Notice, for example, the good correspondence between V1 (dark green) and the large blue patch of negative field sign, which corresponds to the primary visual cortex. (D) Overlay between a heat map of normalized responses to air puffs delivered to the right whisker pad (shown only for significant pixels) and 4 ROIs for the same animal as in panels A and B. In this case responses are mostly restricted to the somatosensory cortex ROI. (E) Average responses to air puffs overlaid with the ROIs. Notice that on average responses are not restricted to somatosensory cortex, being also observed in the posterior parietal cortex ROI.

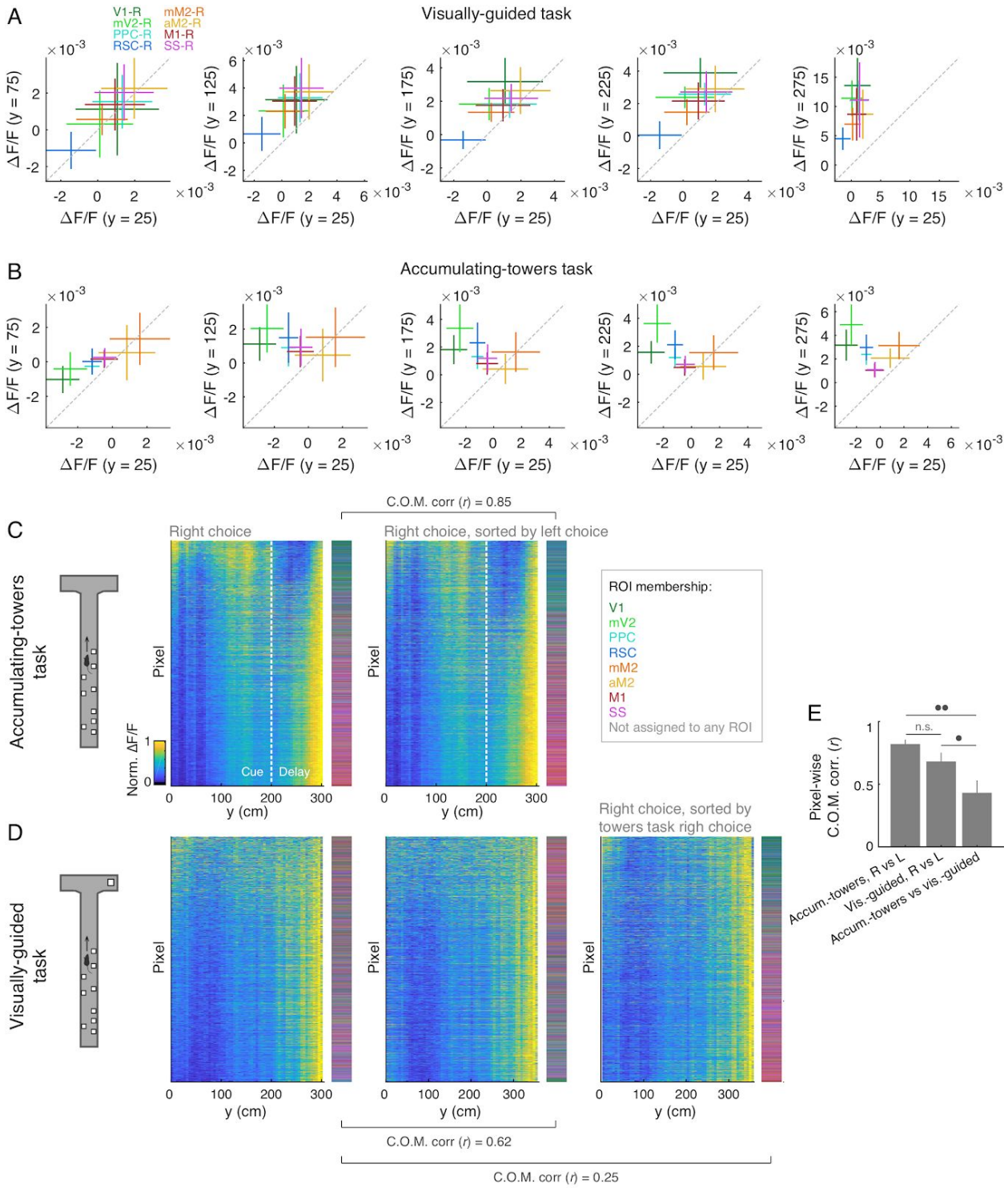
A
Visually-guided task
V1-R
mV2-R
PPC-R
RSC-R
mM2-R
aM2-R
M1-R
SS-R
ΔF/F (y = 75)
ΔF/F (y = 125)
ΔF/F (y = 175)
ΔF/F (y = 225)
ΔF/F (y = 275)
ΔF/F (y = 25)
B
Accumulating-towers task
C
Accumulating-towers task
Right choice
Right choice, sorted by left choice
C.O.M. corr (r) = 0.85
Pixel
Norm. ΔF/F
Cue
Delay
y (cm)
ROI membership:
V1
mV2
PPC
RSC
mM2
aM2
M1
SS
Not assigned to any ROI
D
Visually-guided task
Right choice, sorted by towers task righ choice
C.O.M. corr (r) = 0.62
C.O.M. corr (r) = 0.25
E
Pixel-wise C.O.M. corr. (r)
n.s.
Accum.-towers, R vs L
Vis.-guided, R vs L
Accum.-towers vs vis.-guided

Figure S6. Global activation in the accumulating-towers and the visually-guided tasks, but with different time courses, related to Figure 3.

(A) Comparison of the average ΔF/F during the visually-guided task between a single reference maze y position (25 cm, x axis), and 5 equally spaced later positions (y axis). Shown are averages for right-hemisphere ROIs (labels on top left) calculated across all correct trials. Bars indicate ± SEM across mice (n = 6). (B) Same as A, for the accumulating-towers task. (C) Left: Normalized average ΔF/F for correct right-choice trials in the accumulating-towers task from an example session. Rows correspond to pixels (n = 10,823), sorted in ascending order by C.O.M. Color code to the right indicates ROI membership (caption on far right). Note higher activity in posterior than frontal cortical pixels during the accumulation period (y < 200 cm). Right: same pixels, sorted according to C.O.M. calculated for correct left trials. Correlation coefficient *r* is between pixel-wise C.O.M.s calculated separately for correct right and left trials. (D) Left: normalized ΔF/F for the same pixels as in A, shown for right-choice trials in the visually-guided task. Pixels are sorted by right-choice trial C.O.M. Note that the activity of most pixels ramps up. Middle: Same pixels, sorted according to left-trial C.O.M. Left: sorting according to right-choice trials in the accumulating-towers task (C), illustrating the different dynamics. (E) Average correlation between the vectors given by the C.O.M. of each pixel (~10,000/session), for right- vs. left-choice trials in the towers and visually-guided tasks, and between each task overall (calculated separately for right and left trials). Error bars, ± SEM (n = 6 mice). Using full activity instead of C.O.M. yielded very similar *r* values (not shown).

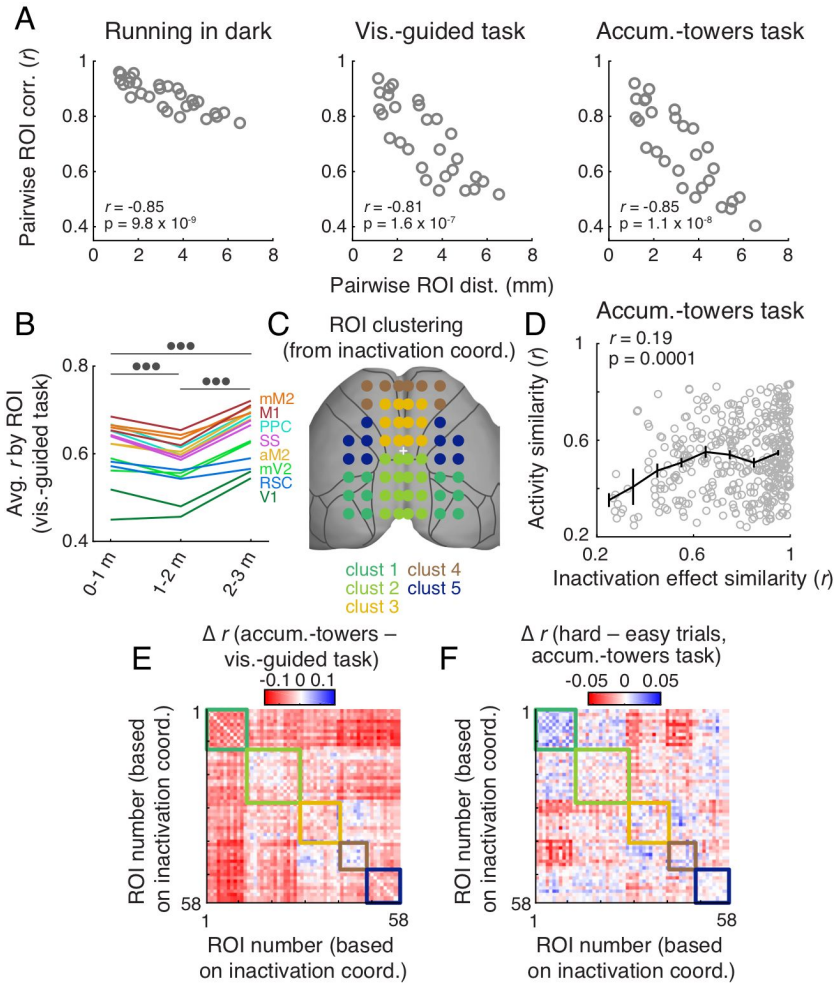

Figure S7. Spatial structure of activity correlations, related to Figure 4.
(A) Correlations between ROIs similarly depend on distance for all behavioral contexts. In each panel, each data point corresponds to an ROI pair within the same hemisphere, averaged across left and right, and across mice (n = 6). Distance between ROIs was calculated using the centroids, separately for each animal. (B) Average correlation between each ROI and all the other 15 (color-coded lines, caption), as a function of maze region in the visually-guided task. Circles: $p < 0.001$. (C) ROIs were redefined to match whole-trial-inactivation coordinates, and clustered according to their pairwise correlations during the towers task. For direct comparisons with inactivations, we enforced that homotopic locations belong to the same cluster by averaging intra-hemispheric correlations. Each of the 29 bilateral cortical regions is color-coded according to the cluster they belong to (captions). Cluster numbers and colors are arranged according to clustering proximity. (D) Scatter plot showing the correlation between the similarity in activity and inactivation effects between the 29 pairs of ROIs. Activity similarity is given by correlations as defined in Fig. 4. Inactivation similarity is given by the correlation between effects on five behavioral indicators as defined in Figs. 2 and S3. Black line, average y-axis values according to binned x-axis, shown for illustration purposes only. Error bars, ± SEM across ROI pars. (E) Difference in correlation matrices between the accumulating-towers and the visually-guided tasks, for the inactivation-grid-defined ROIs. ROIs are sorted according to their cluster membership (Panel C). (F) Same as D, for the difference in correlation matrices between hard and easy trials in the accumulating-towers task.

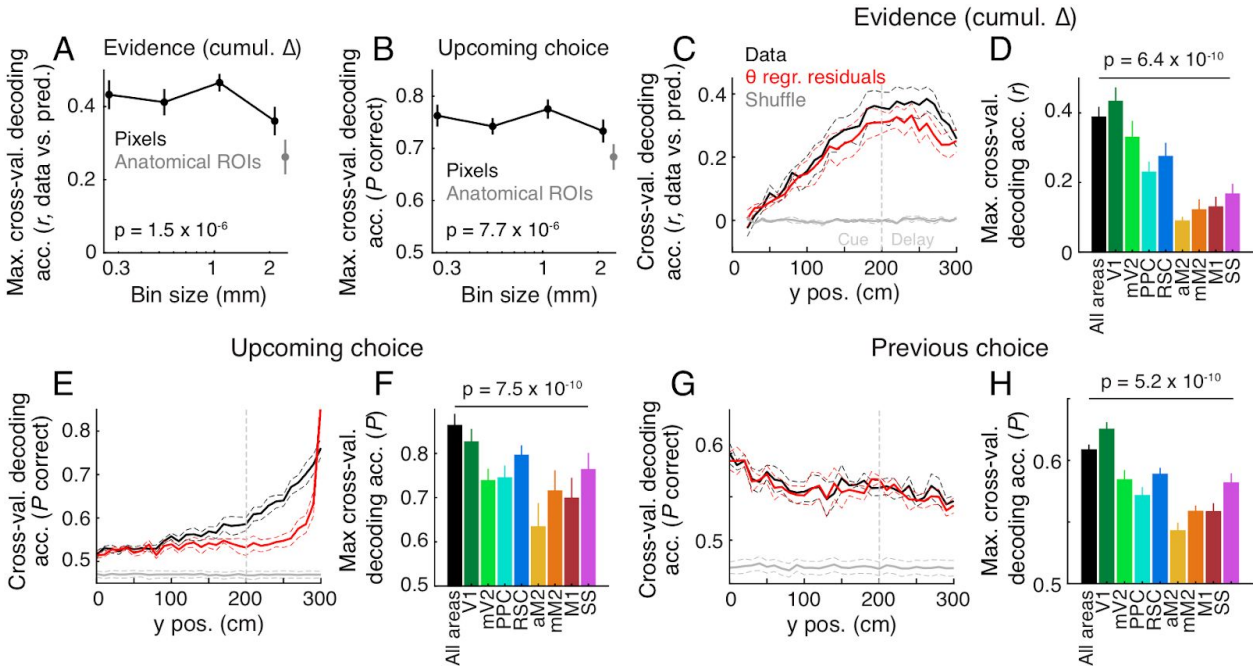

Figure S8. Decoding analysis controls, related to Figure 6.

(A) Maximal cross-validated decoding accuracy for cumulative evidence (Δ towers) as a function of the pixel size over which activity was averaged (black), as well as for anatomically-defined ROIs (gray). The position of the ROI data point on the x axis corresponds to the average ROI size (square root of area). Circles: average across mice (n = 6), error bars: ± SEM. P-value is from one-way ANOVA. (B) Same as A, for choice. (C) Cross-validated performance of linear decoders of cumulative evidence (Δ towers) across maze y positions. For each position, a decoder was trained to predict evidence using the activity of all spatially binned pixels (~280 x 280 μm). Average across mice (n = 6). Dotted lines: ± SEM. Red: we first regressed the activity of each ROI against view angle (θ), and used the residuals of the regression as inputs to the decoder; black: same decoder as in Fig. 6 for comparison (with no view angle regression). (D) Maximal cross-validated decoding accuracy (i.e. taken from y position with highest value) for decoders using pixels from the whole cortex (black) or only from homotopic ROI pairs, with view angle correction. Error bars: ± SEM across mice (n = 6). (E) – (F) Same as C – D, for upcoming choice. Note that the view angle control analysis is a very stringent test, since it may remove real choice-dependent activity for not explicitly accounting for the choice vs. view angle correlation. Nevertheless, the results from this analysis supported most of our conclusions, but yielded decreased choice decoding accuracy in the late cue region and delay, consistent with behavioral analysis suggesting view angle and choice are confounded in this period (Pinto et al., 2018). (G) – (H) Same as C – D, for previous choice.

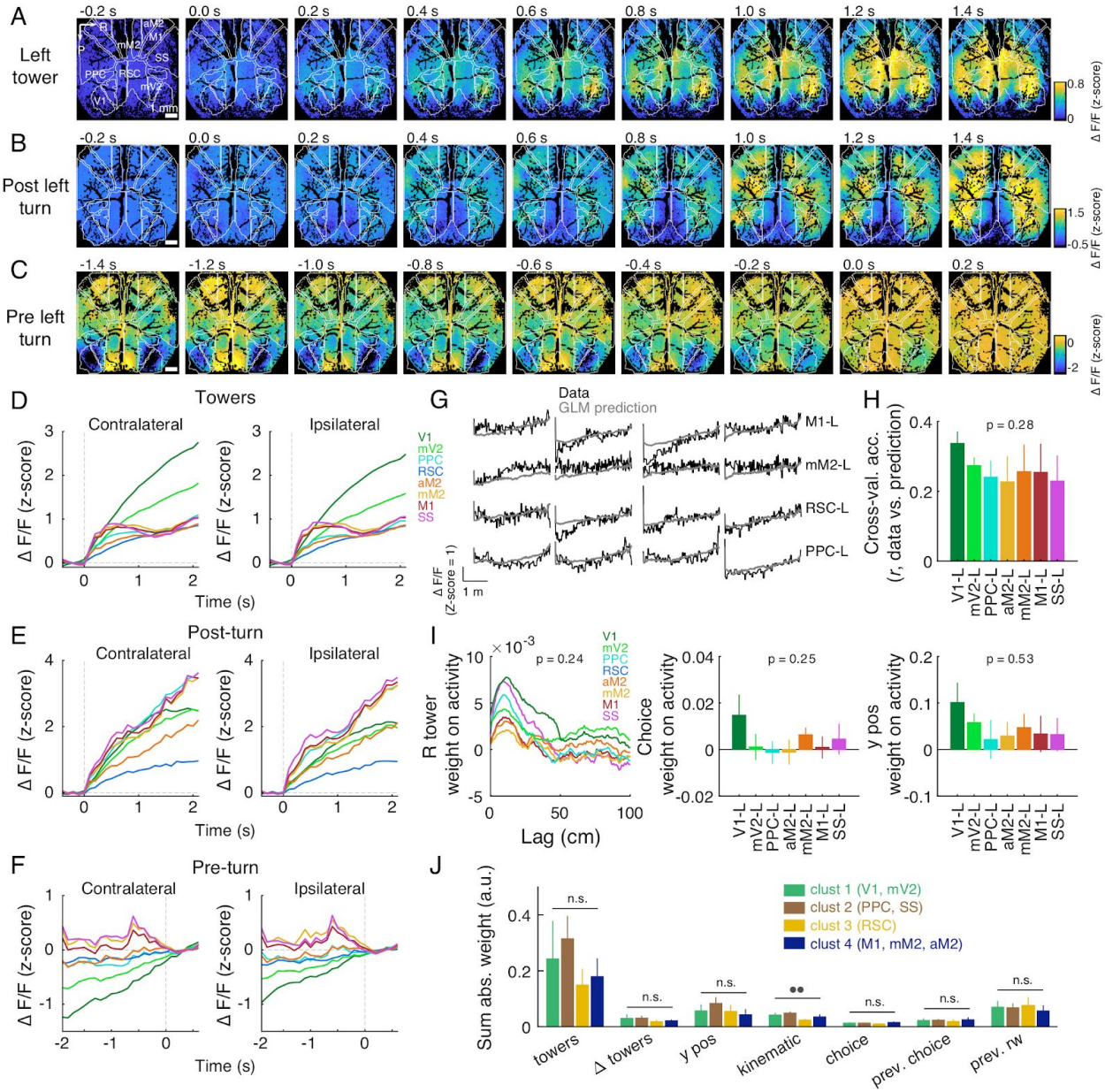

Figure S9. Widespread accumulating-towers task-related activity throughout the dorsal cortex, related to Figure 6.

(A) Representative example of average pixel-wise activity triggered by the onset of left towers during correct trials. Frames have been downsampled to 5 Hz for display. The activity of each pixel is z-scored to a 500-ms baseline period preceding the onset of each tower. (B) Same as A, following left turns (see STAR Methods for definition of turns). (C) Same as A, for activity preceding the turns. In this case activity was z-scored to a 500-ms period following the turn. (D) Average tower-triggered response for each anatomical ROI. Lines are averages across mice (n = 6). Error bars are omitted for clarity. Left: contralateral towers. Right: ipsilateral towers. (E) – (F) Same as D, for post- and pre-turn activity, respectively. (G) We fitted a GLM to the activity of each ROI with task events as predictors. Four single-trial examples of spatially binned activity for four ROIs (black), along with cross-validated GLM predictions (gray). (H) Average cross-validated prediction accuracy (Pearson's correlation between data

and prediction) for all left-hemisphere ROIs. Error bars, ± SEM across mice (n = 6). (I) Example average GLM weights for three different predictors, shown for all right-hemisphere ROIs (left: response to right towers as a function of spatial lag, middle: no-lag weight of upcoming choice, right: no-lag weight of y position in the maze). Error bars, ± SEM. Errors on tower predictors are suppressed for clarity. (J) Average absolute value of grouped GLM weights for the four ROI clusters, in Fig. 4. Tower and Δ tower weights are summed across lags; kinematic weights include speed, view angle θ and $d\theta/dy$, summed across spatial lags. The other predictors are zero-lag only. Error bars, ± SEM across mice. N.s.: not significant ($p > 0.05$), circles: $p < 0.01$ (one-way ANOVA with repeated measures).

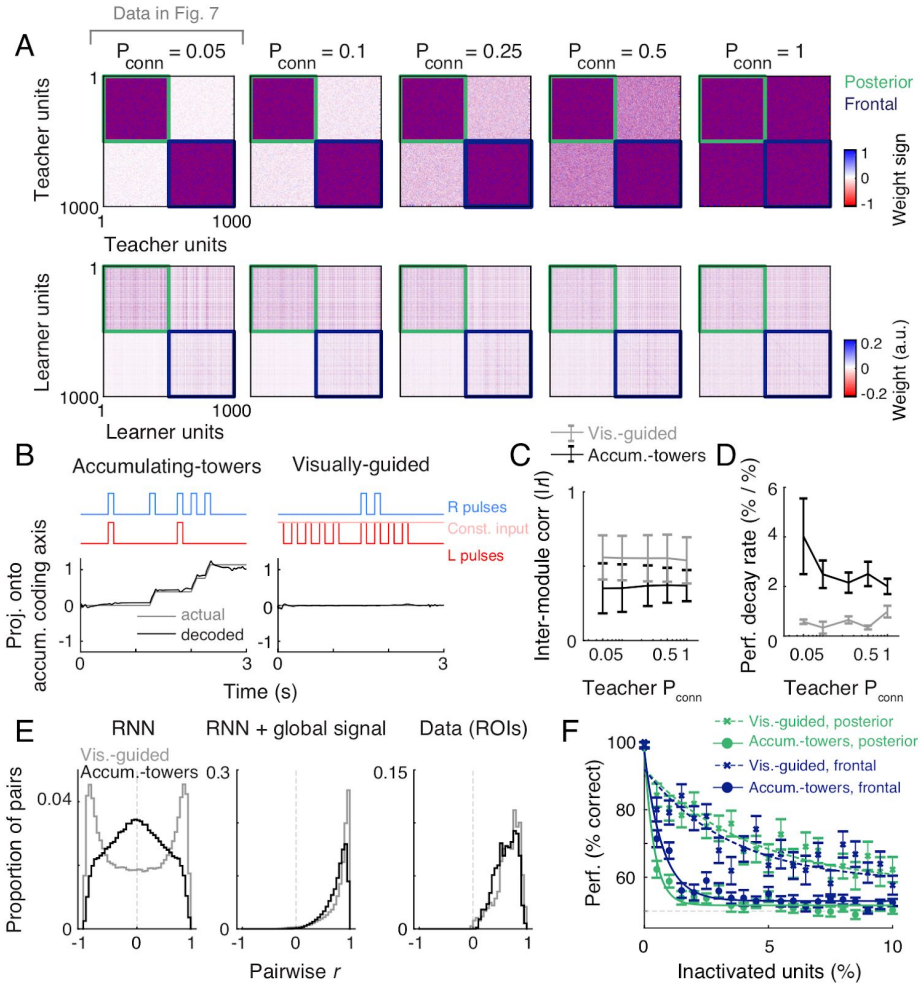

Figure S10. RNN result details, related to Figure 7.

(A) We tested whether different parameter values affected our conclusions. For instance, we varied the levels of inter-module connectivity (P_{conn}) in the teacher network. Top: Matrices showing the sign of pairwise connection weights for varying levels of inter-module P_{conn}, highlighting decreasing sparseness with increased P_{conn}. Bottom: resulting weight matrices in the learner networks, which did not have the sparseness constraints. (B) Projection of network activity onto the accumulated evidence coding axis, given by the optimal linear readout of the cumulative sum of Δ pulses. Shown is one example trial for each task, illustrating that the cumulative sum of pulses is not encoded in the network during the guided task. (C) Population average of inter-module correlations during either task as a function of teacher P_{conn}. Results are qualitatively similar across a range of values. Error bars, SD over units (N = 1000). (D) Same as C, for the average rate in performance drop as a function of increasing number of silenced network units (randomly from either module). Rate is the exponential decay parameter derived from fits to average data, bootstrapped. Error bars, SD over 100 bootstrapping iterations. (E) Distribution of pairwise activity correlations during either task. Left: RNN units; middle: RNN units with an added global signal (equal to the quadratic go signal with maximal magnitude of 0.75; other magnitudes did not change our conclusions); right: ROI pairs from the actual dataset. (F) Effects of silencing units from specific modules during one of the tasks. As for random inactivations, RNN is less sensitive to perturbations when performing the visually-guided task, but there are no differences between modules ($p_{module} = 0.33$, $p_{task} = 2.7 \times 10^{-114}$, 3-way ANOVA with factors task, module, percentage of silenced units; n = 50 runs for each data point).