Abstract

Top-down Control of Sensorimotor Behavior:
Choices, Outcomes, and Context in the Medial Frontal Cortex of the Mouse

Michael James Siniscalchi
2020

Text of Abstract (750 words)

Top-down Control of Sensorimotor Behavior: Choices, Outcomes, and Context in the Medial Frontal Cortex of the Mouse

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by Michael James Siniscalchi

Dissertation Director: Dr. Alex C. Kwan, Ph.D
Committee Chair: Dr. Daeyeol Lee, Ph.D

May 2020

Contents

List of Figures

List of Tables

Acknowledgments

To Dr. James R. Siniscalchi, Ph.D (1942-2019)

Chapter 1

Introduction

1.1 Top-down Control and Flexibility

Text of this section.

1.2 Role of the Medial Frontal Cortex

Text of this section.

1.3 Cortical Cell Types and Behavior

Chapter 2

Enhanced Population Coding for Rewarded Choices in the Medial Frontal Cortex of the Mouse

Instrumental behavior is characterized by the selection of actions based on the degree to which they lead to a desired outcome. However, we lack a detailed understanding of how rewarded actions are reinforced and preferentially implemented. In rodents, the medial frontal cortex is hypothesized to play an important role in this process, based in part on its capacity to encode chosen actions and their outcomes. We therefore asked how neural representations of choice and outcome might interact to facilitate instrumental behavior. To investigate this question, we imaged neural ensemble activity in layer 2/3 of the secondary motor region (M2) while mice engaged in a two-choice auditory discrimination task with probabilistic outcomes. Correct choices could result in one of three reward amounts (single, double or omitted reward), which allowed us to measure neural and behavioral effects of reward magnitude, as well as its categorical presence or absence. Single-unit and population decoding analyses revealed a consistent influence of outcome on choice signals in M2. Specifically, rewarded choices were more robustly encoded relative to unrewarded choices, with little dependence on the exact magnitude of reinforcement. Our results provide insight into the integration of past choices and outcomes in the rodent brain during instrumental behavior.

2.1 Introduction

Associations between past choices and their outcomes allow for efficient selection of actions likely to meet one's present goals. The mechanisms through which such associations are implemented remain unclear. However, at the behavioral level, it is well known that rewarded choices tend to be repeated at the expense of those that have yielded meager or aversive results. To gain insight into the associative mechanisms underlying goal-directed action selection, the present study focuses on the question of how populations of simultaneously recorded neurons in the cerebral

cortex represent and integrate information related to choices and their outcomes.

How does the brain selectively reinforce rewarded actions in order to bias their future implementation? Physiological studies in primates and rodents suggest that the frontal lobe plays an important role in these functions. For example, the primate prefrontal cortex is known to contain neurons that encode chosen actions and outcomes (Barraclough et al., 2004; Genovesio et al., 2006; Seo et al., 2007; Histed et al., 2009), suggesting a plausible neural substrate for their association during goal-directed behavior. Moreover, single-unit recordings have revealed that prior reward enhances the discriminability of spiking activity related to past (Donahue et al., 2013) and upcoming choices (Histed et al., 2009).

Similarly, recordings from the medial frontal cortex (MFC) of rodents have revealed neural signatures of prior choices and outcomes (Sul et al., 2010, 2011; Hyman et al., 2017). In particular, these studies have demonstrated neural representations of choice and outcome history in the most dorsal anatomical sub-region of MFC, which is referred to as secondary motor cortex (M2) in mice, and medial agranular or medial precentral cortex in rats (Sesack et al., 1989; Barthas and Kwan, 2017). Murine M2 has also been implicated in the flexible acquisition and initiation of voluntary actions (Ostlund et al., 2009; Gremel and Costa, 2013; Murakami et al., 2014; Siniscalchi et al., 2016; Barthas and Kwan, 2017; Makino et al., 2017). Based on its putative role in instrumental behavior, M2 may serve as an important interface for the mixing of choice- and reward-related signals in the rodent brain. However, the details of how reinforcement might interact with choice-related neural representations remains unclear.

One intriguing hypothesis is that a choice's outcome could affect the strength or persistence of its neural signature in M2. To explore this possibility, we trained mice on a two-choice auditory discrimination task and then introduced a probabilistic reinforcement schedule during testing. Simultaneous two-photon calcium imaging enabled the characterization of task-related neural ensemble activity in M2. Three randomly interleaved outcomes (single, double and omitted rewards) delivered following correct responses allowed us to measure the impact of reward on choice coding, as well as to distinguish effects of changes in reward magnitude from those of its absolute presence or absence.

We found that rewarding outcomes boosted the fidelity of choice signals encoded in the population activity patterns—an effect that persisted into the subsequent trial. Importantly, the reward-dependent enhancement of choice-related signals depended less on differences in reward size than on the categorical presence or absence of rewards. These results suggest one plausible cortical mechanism for the reinforcement of rewarded actions—namely, that rewarding outcomes lead to a more robust population-level read-out of recent choice history.

2.2 Materials and Methods

Animals

All procedures were performed in accordance with the regulations of the Institutional Animal Care and Use Committee at Yale University. Animals were housed on a 12/12-h light–dark cycle (lights off at 19:00) in groups of three to five per cage. Ten adult (postnatal days 111–279) male mice with a C57Bl/6J (#000664; Jackson Laboratory) genetic background were used. Although two subjects from these experiments were also used in an earlier study (Siniscalchi et al., 2016), the data and analyses in this article have not been reported elsewhere.

Behavioral Setup

Subjects were placed in a modified acrylic tube (8486K33; McMaster-Carr) and held head-fixed during the behavioral task by fastening an implanted head plate to a custom-made stainless-steel bracket (eMachineShop) couple to the tube. This setup limits gross body movements but allows for small postural adjustments and movement of the hind limbs.

Two lick spouts, mounted on a 3D-printed plastic part, were placed on either side of the mouth to allow lateralized responses and corresponding delivery of water rewards. This basic two-choice setup was modeled after an earlier study (Guo et al., 2014). Lick spouts were fabricated from 20-G hypodermic needles, which were cut and carefully filed with an abrasive wheel, and then soldered to a wire lead and connected to a battery-operated lick detection circuit (Slotnick, 2009). Output signals from the detector were digitized with a USB data acquisition device (USB-201; Measurement Computing) plugged into a desktop computer.

Water rewards were delivered through each spout by gravity-feed and actuated with a solenoid valve (EV-2-24; Clippard) controlled by TTL pulses from a second USB-201. Pulse duration for a single reward (2 μL) was calibrated for each valve by sending 100 pulses and then weighing the ejected volume (a 15–20 ms TTL pulse was typically required). On double-reward trials, 4 μL were delivered by using either a calibrated longer duration pulse or two single rewards separated by 100 ms.

Auditory stimuli were played through a pair of computer speakers (S120; Logitech) placed in front of the animal. The task structure was automated using custom scripts written for Presentation (Neurobehavioral Systems, Inc.).

During training, the behavioral apparatus was enclosed in the cabinet of an audio-visual cart (4731T74; McMaster-Carr) soundproofed with acoustic foam (5692T49; McMaster-Carr). For imaging, the setup was replicated within the enclosure of a two-photon microscope.

Two-choice Auditory Discrimination Task with Probabilistic Outcomes

Mice were trained to perform a two-choice auditory discrimination task. To motivate participation, subjects were water restricted, as follows. Six days per week, water was provided only in a single daily training session, as a reward for correct choices. On the remaining day, a water bottle was placed in the home cage for 15 min. Three phases of training were used to shape the behavior, identical to those used in our prior study (Siniscalchi et al., 2016).

During phase one, mice were habituated to head fixation and trained to lick the left or right spout for water: each lick to either spout triggered the release of 4 μL of water, with a minimum interval of 1 s between rewards.

After attaining > 100 rewards in a single session of phase one (1–2 days), subjects were advanced to phase two, in which they were required to lick for a similar number of rewards at each spout. In this case, water was only available from one 'target' spout at a time, with the target moving to the opposite side each time the mouse earned three consecutive rewards from a given target. Additionally, sound stimuli were introduced in association with rewarded licks ('hits') on a given spout. The stimuli were trains of four 500-ms-long logarithmic chirps, with starting and ending frequencies of either 5 and 15 kHz ('upsweep'), or 15 and 5 kHz ('downsweep'), respectively. Upsweeps were played immediately following a left hit, and downsweeps were played immediately following a right hit. Similar to phase one, a minimum interval of 1 s was imposed between rewards.

After attaining > 100 hits within a single session of phase two (1–2 days), subjects were advanced to phase three, in which they were trained to perform a two-choice sound discrimination task. Unlike the earlier two phases, in which the operant behavior was self-paced, phase three was structured into trials with a defined response period. Each trial began with the presentation of a 2-s-long sound cue (upsweeps or downsweeps, randomly drawn), which indicated the target spout for that trial. Upsweeps indicated 'left' and downsweeps indicated 'right'. To earn a water reward, the subject was required to lick the target spout within the final 1.5 s of the cue (the 'response window'). The first lick to either spout within the response window terminated the sound cue and triggered an immediate outcome: 2 μL of water from the target spout for a hit or playback of a 2-s-long white noise sound for an 'error', in which the wrong spout was chosen. 'Misses' were defined as trials in which the mouse failed to lick within the response window. In any case, the next trial would begin 7 s after cue offset. Thus, trial durations ranged from 7.5 to 9 s, depending on response time. Each session was terminated automatically after twenty consecutive misses.

For the current study, subjects were trained on the two-choice sound discrimination task to a high level of proficiency ($> 90\%$ hit rate) and were then tested on a task variant with probabilistic outcomes. This variant was identical to phase three of training, except that correct responses could result in one of three outcomes: 2 μL of water ('single reward') with 80% probability, no reinforcement ('omitted reward') with 10% probability, or 4 μL of water ('double reward') with 10% probability. In a subset of sessions (5/16, identified in Table 2.1), the white noise sound used in error trials

was also played at the time of an omitted reward. We analyzed the behavioral and neural data for those sessions separately but did not detect any obvious differences; therefore, we have presented the pooled results.

Exp. ID	Subject	# Cells	# Trials	Hit rate (%)	White noise in 0× trials
1	L4	35	198	99.0	Yes
2	L2	47	228	80.7	Yes
3	M6	62	321	93.2	Yes
4	M6	42	233	97.9	Yes
5	M6	55	212	95.3	Yes
6	M12	51	336	94.9	No
7	M20	46	182	91.8	No
8	M12	40	300	95.7	No
9	M13	43	313	93.0	No
10	M12	44	285	98.3	No
11	M14	38	210	99.1	No
12	M16	59	202	88.6	No
13	M16	77	269	93.7	No
14	M17	49	267	91.8	No
15	M17	33	222	92.8	No
16	M22	50	262	94.3	No

Table 2.1: Summary of Imaging Experiments Analyzed in Chapter Two

Virus Injection and Cranial Window Implantation

All subjects were treated pre-operatively with carprofen (5 mg/kg, s.c.; #024751; Butler Animal Health) and dexamethasone (3 mg/kg, s.c.; Dexaject SP, #002459; Henry Schein Animal Health). Anesthesia was then induced with 2% isoflurane in oxygen, and the animal was placed on a water-circulating heating pad (TP-700; Gaymar Stryker). Following induction, isoflurane concentration was lowered to 1.5% and the head was secured in a stereotaxic frame with ear bars (David Kopf Instruments).

The scalp was shaved with electric trimmers and cleaned with a povidone-iodine surgical scrub (Betadine; Perdue Products L.P.). A narrow portion of scalp was removed along the midline from the interaural line to a line visualized just posterior to the eyes. The incision was retracted to expose the dorsal aspect of the skull, which was then scrubbed briefly with 3% hydrogen peroxide to aid in removal of the periosteum, and washed generously with artificial cerebrospinal fluid (aCSF; in mM: 5 KCl, 5 HEPES, 135 NaCl, 1 $MgCl_2$ and 1.8 $CaCl_2$; pH 7.3).

A 3-mm-diameter circular craniotomy was made over the right hemisphere using a high-speed rotary drill (K.1070; Foredom), centered on a medial target within M2 (AP +1.5 mm, ML −0.5 mm relative to Bregma). The dura was left intact and was irrigated frequently with ACSF over the remainder of the procedure.

An adeno-associated virus encoding GCaMP6s (AAV1-Syn-GCaMP6s-WPRE-SV40; Penn Vector Core) was infused at four AP-ML coordinates through a glass

micropipette attached to a microinjection unit (Nanoject II; Drummond). The injection sites formed a 200-µm −wide square centered on the target location. Each site was injected with 46 nL of virus over 3 min, at a depth of 0.4 mm from the dura. To minimize backflow of the injected solution, the micropipette was left in place for 5 min after each infusion. A small amount of warmed agarose solution (Type III-A, High EEO agarose; 1.2% in ACSF; #A9793; Sigma Aldrich) was then applied along the perimeter of the craniotomy to fill the space between the cranial window and the surrounding tissue after implantation.

The cranial window consisted of two concentric circular glass parts, glued together with UV-activated optical adhesive (NOA 61; Norland Products, Inc.): a 3-mm-diameter, #1 thickness prefabricated glass coverslip (#64-0720-CS-3R; Warner Instruments) and a 2-mm-diameter plug cut from a #1 or 2 thickness glass coverslip. The window was carefully placed on the brain surface with the glass plug facing down, and then secured to the skull at the edge of the craniotomy using surgical tissue adhesive (Vetbond; 3M). Gentle downward pressure was applied to stabilize the implant during this procedure, using a wooden probe attached to the stereotaxic frame. A custom-made stainless-steel head plate (eMachineShop) was then bonded to the skull with dental cement (C&B Metabond; Parkell Inc.), with care taken to cover any remaining exposed skull.

Post-operative care was provided immediately and for three consecutive days following surgery. This consisted of analgesia (carprofen, 5 mg/kg, s.c.) and fluid support (preservative-free 0.9% NaCl, 0.5 mL, s.c., up to twice daily). All animals were given a one-week post-operative recovery period prior to the onset of behavioral training.

Two-Photon Calcium Imaging

The two-photon microscope (Movable Objective Microscope; Sutter Instrument) was controlled using ScanImage software (Pologruto et al., 2003). The excitation source was a Ti:Sapphire ultrafast femtosecond laser (Chameleon Ultra II, Coherent). Beam intensity was modulated using a Pockels cell (350-80-LA-02; Conoptics) and blanked with an optical shutter (LS6ZM2; Uniblitz/Vincent Associates). The beam was focused through a high-numerical aperture objective (XLUMPLFLN, 20×/0.95 NA; Olympus). Excitation wavelength was set to 920 nm, and the emission was collected behind a bandpass filter from 475 to 550 nm using a GaAsP photomultiplier tube (H7422P-40MOD; Hamamatsu). The time-averaged excitation intensity after the microscope objective was ∼100 mW.

Time-lapse image frames were acquired at 256 × 256 pixels and a nominal frame rate of 3.62 Hz, using bidirectional raster scanning. To synchronize the behavioral and imaging data, a TTL pulse was sent by Presentation 1 s prior to the start of each trial. This TTL signal was assigned as an external trigger in ScanImage to initiate a new image file. Timestamps of the TTL pulses along with timestamps of other behavioral events were written to a text file by Presentation, allowing the image files to be aligned with behavioral events.

Analysis of Behavioral Data

Timestamps of the behavioral events, including cue onsets, licks, and reinforcement onsets, were logged to a text file by Presentation. All further processing and analysis were done using custom scripts written in MATLAB (MathWorks, Inc.). The number of trials performed was defined as the number of trials in which either spout was licked within the response window. Correct rate was defined as the number of correct trials divided by the number of trials performed. Miss rate was defined as the number of misses divided by the total number of trials. The sensitivity index, or d-prime, was calculated as the difference between the inverse of the standard normal cumulative distribution for the correct rate on upsweep trials and the inverse of the standard normal cumulative distribution for the incorrect rate on downsweep trials.

Analysis of Imaging Data

Raw time-lapse image stacks corresponding to each trial were saved as multipage TIF files by ScanImage. As a first processing step, the raw stacks were merged to a single TIFF. Timestamps for the first frame of each trial, as well as the external trigger, were extracted for alignment with behavior. The merged TIF file was then processed for x–y motion correction using either the TurboReg (Thevenaz et al., 1998) or moco (Dubbs et al., 2016) plug-in for ImageJ (Schneider et al., 2012). Regions of interest (ROIs) were manually selected around cell bodies appearing in the maximal or average projection image, using a custom graphical user interface programmed in MATLAB. Pixel intensity values within each ROI were summed for each frame t to obtain $F(t)$. The baseline fluorescence, $F_0(t)$, was estimated as the 10th percentile of $F(t)$ within a 10 min moving window centered on t. $\Delta F/F(t)$ was then calculated as the fractional change in $F(t)$ relative to $F_0(t)$.

Event-Aligned Activity and Choice Selectivity

To obtain trial-averaged activity traces associated with a specific behavioral event (eg cue onset), we first aligned $\Delta F/F$ traces based on their timing relative to each instance of the event and then took the mean across traces. To estimate confidence intervals, we performed a bootstrapping procedure, as follows. For N instances of a particular event, N traces were resampled randomly with replacement and then averaged. This process was repeated 1000 times, in order to approximate the sampling distribution for the mean. Upper and lower bounds of the associated confidence interval were then estimated as percentiles of this distribution.

Choice selectivity was calculated as the difference between cue-aligned, trial-averaged traces from trials in which the ipsilateral versus contralateral spout was chosen, divided by the sum of the two traces. Therefore, choice selectivity was a function of time that could take values from -1 to 1, with negative values indicating an ipsilateral preference and positive values indicating a contralateral preference. The mean choice selectivity from 2 to 4 s after cue onset was used as a scalar estimate in comparisons between double- and omitted-reward trials.

Multiple Linear Regression

To characterize the relationship between task variables and the activity of individual neurons, we fit the following linear equation:

$$\begin{aligned}\frac{\Delta F}{F}(t) = a_0 & \\ & + a_1 C_n + a_2 C_{n-1} + a_3 C_{n-2} \\ & + a_4 R_n + a_5 R_{n-1} + a_6 R_{n-2} \\ & + a_7 C_n R_n + a_8 C_{n-1} R_{n-1} + a_9 C_{n-2} R_{n-2} \\ & + \epsilon(t),\end{aligned}$$

where $\Delta F/F(t)$ is the fractional change in fluorescence at time t relative to baseline; C_n, C_{n-1}, and C_{n-2} are the choices made on the current trial, the prior trial and the trial before last, respectively; R_n, R_{n-1}, and R_{n-2} are the outcomes for the current trial, the prior trial and the trial before last, respectively; $a_0 \ldots a_9$ are the regression coefficients; and $\epsilon(t)$ is the error term.

Choices were dummy-coded as -1 for left licks and 1 for right licks. Outcomes were dummy-coded as 0 for single-reward trials, -1 for omitted-reward trials and 1 for double-reward trials. Error trials and misses were not analysed. Regression coefficients and their p-values were estimated for each 500 ms time bin t, within an interval from -2 to 6.5 s relative to cue onset.

To summarize the pooled results from all experiments, the proportion of neurons with a p-value less than 0.01 was plotted over time for each predictor. The binomial test was used to determine whether this proportion was significantly different from chance level.

Decoding: Linear Discriminant Analysis

To determine how reliably the subject's choices were encoded in the neural ensemble activity, we constructed and tested classifiers based on linear discriminant analysis. All classifiers were constructed using the 'classify' function in MATLAB, with the 'type' parameter set to 'linear' (the default); this classification method fits a multivariate normal density function to each group, using a pooled estimate of covariance.

Choices, C_n, were dummy-coded as -1 for left licks and 1 for right licks. For each 500 ms time bin t within the interval from -2 to 6.5 s relative to cue onset, the $\Delta F/F(t)$ values of all neurons were incorporated into a trial-indexed set of population activity vectors, with each vector representing the $\Delta F/F(t)$ for all neurons in the corresponding trial. A Monte Carlo cross-validation procedure was then applied across trials, as follows. First, activity vectors from a randomly drawn 80% of single-reward trials were used as the training set to construct a classifier. The classifier was then tested for accuracy using the activity vectors from the remaining 20% of single-reward trials.

Additionally, we tested the accuracy of the classifier at decoding choices in other

trial types, ie double-reward, omitted-reward, and error trials. To estimate chance-level performance for each classifier, an identical classifier was constructed and tested, with the exception that the C_n values within the training set were first shuffled randomly.

To characterize the potential advantage of simultaneous recording, we compared the accuracy of classifiers trained on actual imaging data versus 'pseudo-ensemble' data in which simultaneity had been destroyed. To build a pseudo-ensemble, the $\Delta F/F(t)$ values for each cell were randomly shuffled across all training trials with the same C_n value. Therefore, each pseudo-ensemble activity vector comprised $\Delta F/F(t)$ values drawn from many different trials, while preserving cell identity as well as choice and outcome specificities. Classifiers trained on pseudo-ensemble data versus real data were then compared for accuracy using the remaining (unshuffled) test trials.

For all of the analyses described in this section, average decoding accuracy was estimated as the mean across 30 iterations. This iterative cross-validation procedure was repeated for each time bin t.

Decoding: Random Forests

As a second approach to decoding choices from the population activity, we constructed and tested random forest classifiers (Breiman, 2001). Neural and behavioral data were treated in the same manner as for the linear discriminant analysis described above: choices, C_n, were dummy-coded as -1 and 1; time ranged in 500 ms increments from -2 to 6.5 s relative to cue onset; and single-reward trials were split into training and testing sets for 30 iterations of Monte Carlo cross-validation. For each iteration, a random forest classifier was trained using population activity vectors from a random sample of 80% of trials and then tested on the remaining 20%.

The random forest algorithm is a bootstrap aggregation ('bagging') approach consisting of many decision trees. Each decision tree takes as input a set of features (eg the activity of cells in the neural ensemble) and arrives at a predicted binary response (eg left or right choice). It does so by comparing a subset of the features to a series of corresponding threshold values ('splits') learned from the training set through a process of greedy recursive partitioning. The overall predicted response of the random forest is the majority vote of the predicted responses across all trees.

To construct each decision tree, M population activity vectors were drawn randomly with replacement from the M trials making up the training set. Each split in the tree comprised a threshold on the $\Delta F/F(t)$ value of one cell selected as the strongest predictor out of a randomly drawn subset of cells. A new subset was drawn randomly without replacement to determine each split. The number of cells in each subset was equal to the square root of N, where N is the total number of cells in the imaged ensemble.

To choose the number of trees, we tested a range of values and found that classifier performance saturated beyond 50 trees; therefore, the number of trees was set to 100. The procedure was implemented by calling the 'fitensemble' function in MATLAB with the 'Method' parameter set to 'Bag', the 'Type' parameter set to 'Classification' and the 'Learners' parameter set to 'Tree'.

Additionally, in the same manner as we did for the linear discriminant analysis, we tested classifiers constructed using single-reward trials on other trial types, determined chance-level accuracy by training classifiers on shuffled choices, and characterized the advantage of simultaneous recording by comparing with random forest classifiers constructed using pseudo-ensemble data. For all decoding analyses described in this section, average decoding accuracy was estimated as the mean accuracy across all iterations. This entire procedure was repeated for each time bin t.

Decoding: Varying the Ensemble Size

To determine how the number of neurons in an ensemble influenced decoding accuracy, we constructed and tested classifiers on neural data drawn from subsets of the imaged populations. Because the smallest ensemble imaged in our experiments had just over 30 neurons, we limited this analysis to ensemble sizes from 1 to 30 cells. We also limited the analysis to the time interval from 2 to 4 s from cue onset, the period in which our decoding analyses showed the highest decoding accuracy.

For each ensemble size, a subset of the imaged cells was drawn randomly without replacement to produce an ensemble of predetermined size, and the process was repeated for 30 draws. For each draw, a classifier was constructed using a random subsample of 80% of trials and then tested on the remaining trials. Decoding accuracy was estimated as the mean across draws. This iterative procedure was repeated for each ensemble size and for the two types of classifiers (linear discriminant and random forest).

Experimental Design and Statistical Analysis

The structure of the task was fully automated using custom scripts written for Presentation, which randomized the sound cues and outcomes presented in each trial. No further blinding was used.

Sample sizes are noted in the results and figure legends. No statistical analysis was employed to determine sample sizes; however, they were similar to those used elsewhere in the field. All behavioral and neural ensemble analyses were performed using a sample size of $N = 16$ imaging sessions from ten animals (Table 2.1). For analyses of single-unit activity (Figs. 2.4 & 2.5), the sample size was $N = 771$ cells; a subsample of $n = 226$ choice-selective cells was considered in Figure 2.5 (see Results).

A total of seven sessions were excluded from the study prior to neural activity analysis for the following reasons: poor behavioral performance (overall correct rate $< 80\%$, one session); too few trials completed and, therefore, fewer than five trials for at least one outcome type (three sessions); and residual movement artifacts after image motion correction (three sessions).

All statistical analyses were performed in MATLAB. A paired design was used for comparisons across outcome conditions, and the likelihood p of a false positive was estimated with a Wilcoxon signed-rank test. $p < 0.05$ was taken to indicate a significant difference. No corrections were made for multiple comparisons, but p-values are noted explicitly in the Results. For the multiple linear regression analysis (Fig.

2.4), the significance threshold for each predictor was set at $\alpha = 0.01$. Significant proportions were determined using a binomial test, with $\alpha = 0.01$. For neural ensemble analyses, chance-level accuracy of decoding choices from the neural activity was determined by testing classifiers constructed using shuffled choices.

2.3 Results

2.3.1 Two-Choice Auditory Discrimination Task with Probabilistic Outcomes

Water-restricted mice were trained to perform a two-choice auditory discrimination task under head restraint (Fig. 2.1A). Subjects were required to choose between two lick spouts placed on either side of the mouth, only one of which (the 'target') would be rewarded on a given trial. One of the two sound cues was presented at the start of each trial, indicating the target spout. The sound cues were trains of four 500-ms-long logarithmic chirps, with starting and ending frequencies of either 5 and 15 kHz ('upsweep') or 15 and 5 kHz ('downsweep'), respectively. Upsweeps indicated 'left' and downsweeps indicated 'right'. Licking either spout triggered an immediate outcome: delivery of a water reward if the target spout was chosen ('correct') or playback of a 2-s-long white noise sound if the wrong spout was chosen ('error'). To study the influence of reward on behavior and neural activity, correct responses were reinforced probabilistically with one of three water amounts: 2 μL (single reward) with 80% probability, 4 μL (double reward) with 10% probability, or 0 μL (omitted reward) with 10% probability. In a subset of sessions, the white noise sound used in error trials was also played at the time of an omitted reward (Table 2.1). Each session was terminated automatically after 20 trials without a response ('misses').

The use of probabilistic outcomes allowed us to systematically investigate neural and behavioral effects of reward during sensorimotor decision making. In particular, the inclusion of omitted-reward trials permitted the effects of reward absence to be explicitly characterized during correct trials, thus eliminating differences in decision accuracy as a potential confound. Additionally, the influence of reward size (single reward vs. double reward) could be compared to that of its categorical presence or absence (single reward or double reward vs. omitted reward).

We obtained 16 imaging sessions from 10 mice while they performed this task (range: 1–3 sessions per mouse; Table 2.1). Figure 2.1 shows the behavioral performance from one such session. Subjects made 259 ± 11 choices per session, with an overall correct rate of $94 \pm 1\%$ ($mean \pm SEM$; Fig. 2.1C) and a sensitivity index (d') of 3.4 ± 0.2. Within each session, they encountered an average of 195 ± 9 single-reward trials, 24 ± 1 double-reward trials and 25 ± 2 omitted-reward trials, and made 16 ± 3 errors. Correct rates were similar for upsweep ($94 \pm 2\%$) and downsweep ($94 \pm 1\%$) trials. Thus, subjects maintained a high level of accuracy following the introduction of variable outcomes.

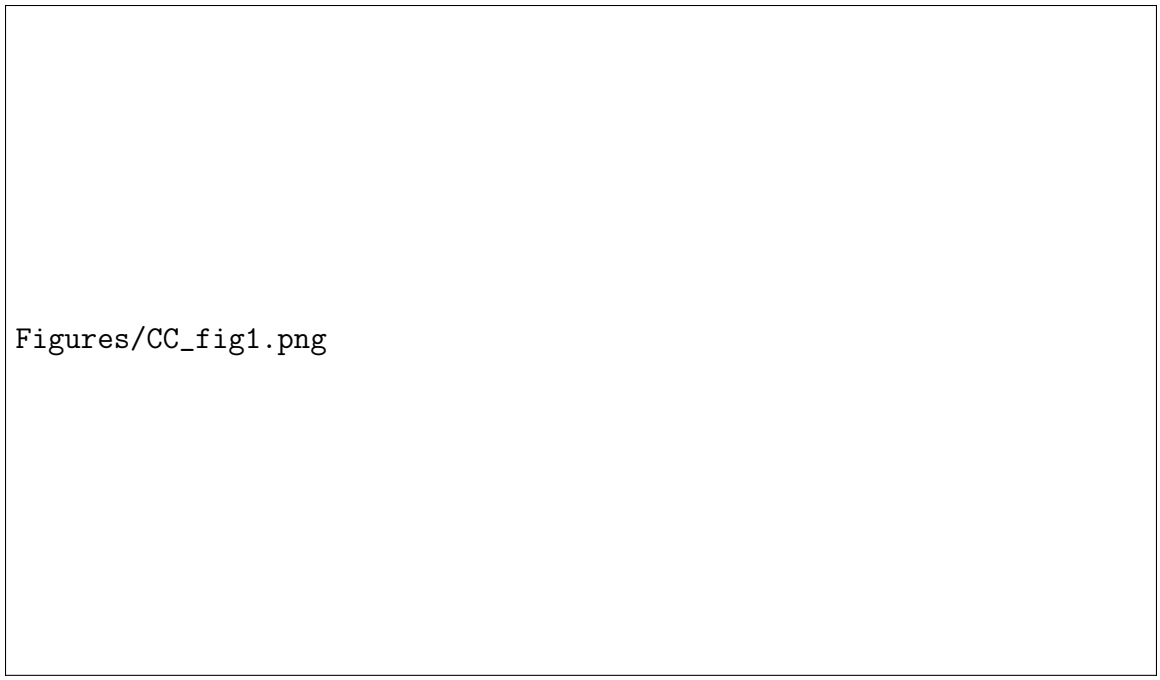


Figure 2.1: Two-choice auditory discrimination task with probabilistic outcomes. On each trial, mice were required to lick the target spout (left or right) indicated by a sound cue (upsweep or downsweep, respectively). Correct responses were rewarded probabilistically with one of three water amounts. (A) Flow diagram of the trial structure on correct trials. Each trial began with one of the two sound cues. The first lick to the target spout within 0.5–2 s following cue onset (Response window) immediately triggered one of three outcomes (Reinforcement): single reward (1×), double reward (2×) or omitted reward (0×), with probabilities 80, 10 and 10%, respectively. The next trial would begin 7 s after cue offset. (B) Behavioral performance from an example session (Experiment 1 in Table 2.1). Occurrences of each sound cue (top), choice (middle) and outcome (bottom) are displayed in raster form according to trial number. Errors occurred when the non-target spout was chosen for the first lick within the response window. Misses were defined by the failure to respond within the response window. (C) Summary of behavioral performance. Gray triangles, individual sessions. Black cross-hairs, $mean \pm SEM$. (D) Number of occurrences of each outcome per session. For all figures, $N = 16$ sessions from 10 mice unless otherwise noted.

2.3.2 Outcome-Dependent Behavioral Adjustments on Multiple Timescales

Subjects were well-trained in auditory discrimination before probabilistic outcomes were introduced, and the optimal stimulus–response relationships remained the same. Therefore, behavioral adjustments based on the new distribution of outcomes would not increase the overall amount of reward obtained. This raises an important question: were subjects aware of the different outcomes? Two lines of evidence indicate that they were. First, the number of licks to the target spout during the period following reward delivery increased in a graded manner with the volume of water reward given (Fig. 2.2A). On average, 13.9 ± 0.8 licks were registered at the target spout in single-reward trials ($mean \pm SEM$, N = 16 sessions). Relative to single-reward trials, the mean number of licks per trial rose by $21 \pm 4\%$ in double-reward trials (vs. single-reward trials: $p = 0.001$, Wilcoxon signed-rank test, $N = 16$ sessions) and decreased by $48 \pm 3\%$ in omitted-reward trials (vs. single-reward trials: $p = 4 \times 10^{-4}$, Wilcoxon signed-rank test, $N = 16$ sessions). Because the different outcome types were interleaved randomly across correct trials, these results indicate that the mice could detect changes in reward volume and adjusted their consummatory licking accordingly.

Second, behavioral performance varied as a function of the prior trial's outcome. The most notable effect was on the number of misses (Fig. 2.2B). The likelihood of a miss was $7 \pm 1\%$ for trials following a single reward ($mean \pm SEM$, $N = 16$ sessions). It was significantly lower at $4 \pm 1\%$ for trials following a double reward ($p = 0.04$, Wilcoxon signed-rank test, $N = 16$ sessions) but significantly higher at $11 \pm 2\%$ if reward was omitted in the previous trial (vs. single-reward trials: $p = 0.005$, Wilcoxon signed-rank test, $N = 16$ sessions). However, when subjects did respond, the correct rate was consistently high at 95 ± 1, 95 ± 2 and $95 \pm 1\%$ for trials following a single, double and omitted rewards, respectively (Fig. 2.2C; $mean \pm SEM$, $N = 16$ sessions). Taken together, the observed adjustments in licking and the effect on miss rate in the subsequent trial provide clear evidence that the mice monitored trial outcomes. Notably, the magnitude of reinforcement affected subjects' willingness to respond, but did not influence the accuracy of decisions.

2.3.3 Persistent Decline in Performance Following Errors

Errors were uncommon in these experiments. However, inference to the causes of residual inaccuracy could help to illuminate internal processes underlying decision making. Our task design permitted us to examine and rule out two potential sources of error. First, errors could have resulted mainly from stochastic fluctuations in perceptual performance. In this case, the likelihood of a correct response should not depend significantly on recent performance history. Thus, similar levels of performance would be expected in trials following errors and correct responses. However, relative to correct responses, errors were associated with deficits in both decision accuracy and willingness to respond in the subsequent trial. Correct rate dropped to $86 \pm 3\%$ in trials following an error (Fig. 2.2C; vs. single-reward trials: $p = 0.01$; vs.

Figure 2.2: Subjects adjusted their behavior based on trial outcome. (A) Mean lick density across sessions as a function of time for the left and right spouts, averaged separately from trials in which the sound cue was an upsweep (top row) or downsweep (bottom row), and outcome was single reward (solid), double reward (dotted) or omitted reward (dashed). Black error bar, 95% confidence interval for time of outcome. (B) Fraction of trials missed immediately following each outcome. Gray triangles, individual sessions. Black crosshairs, $mean \pm SEM$. Wilcoxon signed-rank test: $^{*}P < 0.05$, $^{**}P < 0.005$, n.s., not significant. (C) Fraction of trials with a correct response immediately following each outcome.

omitted-reward trials: $p = 0.01$; Wilcoxon signed-rank test, $N = 16$ sessions), and miss rate increased to $23 \pm 5\%$ (Fig. 2.2B; vs. single-reward trials: $p = 0.004$; vs. omitted-reward trials: $p = 0.04$, Wilcoxon signed-rank test, $N = 16$ sessions). We also asked whether the likelihood of an error could be explained by outcome-dependent processes, such as exploration following the absence of an expected reward. This hypothesis is not supported by the observation that error rates following correct trials were similar regardless of the prior trial's outcome (Fig. 2.2C). Collectively, our results suggest that the drop in task performance associated with errors tends to persist and cannot be solely accounted for by the trial outcome.

2.3.4 Single-unit Activity Related to Choices and Outcomes in Mouse M2

To characterize neural activity in M2 while mice engaged in the task, we simultaneously imaged the brain at cellular resolution using a two-photon microscope (Denk et al., 1990) (Fig. 2.3A). An adeno-associated virus encoding GCaMP6s (AAV1.Syn.GCaMP6s.WPRE.SV40) was injected into layer 2/3 of M2, and a chronic glass window was implanted for optical imaging (Fig. 2.3B). GCaMP6s is a genetically encoded calcium indicator that exhibits an approximately 25% rise in fluorescence intensity per action potential in cortical pyramidal neurons (Chen et al., 2013). In 16 imaging sessions, we recorded fluorescence transients from an average of 48 ± 3 neurons in layer 2/3 of M2 ($mean \pm SEM$, range: 33–77 cells; Fig. 2.3C). All imaging was done in the right hemisphere; therefore, left and right licks were always contralateral and ipsilateral to the imaged neurons, respectively.

Many M2 neurons exhibited changes in fluorescence following the sound cue, indicating task-driven neural activity. Figure 2.3D shows a neuron with preferential activity on trials in which the left spout was chosen. The activity of the same neuron was also monotonically modulated by reward size: activity was highest for double rewards, moderate for single rewards and lowest for omitted rewards. The influence of choice and outcome on this neuron was approximately additive. However, other neurons had more complex responses. Some neurons showed a choice preference on single- and double-reward trials that was not observed on omitted-reward trials, suggesting a choice–outcome interaction (Fig. 2.3E). Other neurons only exhibited significant choice preference during omitted-reward trials (Fig. 2.3F). These results highlight the diversity of choice- and outcome-related signals in M2 at the level of individual neurons.

2.3.5 Persistence of Choice- and Outcome-Related Signals

To more systematically characterize choice- and outcome-related signals in M2 neurons, we used multiple linear regression analysis. For each cell, we fit a linear equation (see Materials and Methods) to estimate the dependence of its fluorescence intensity on the following predictors: choice, outcome, and the interaction of choice and outcome; for the current trial, last trial, and the trial before last (Fig. 2.4A). The analysis revealed choice-dependent activity in a substantial fraction of M2 neurons (210/771

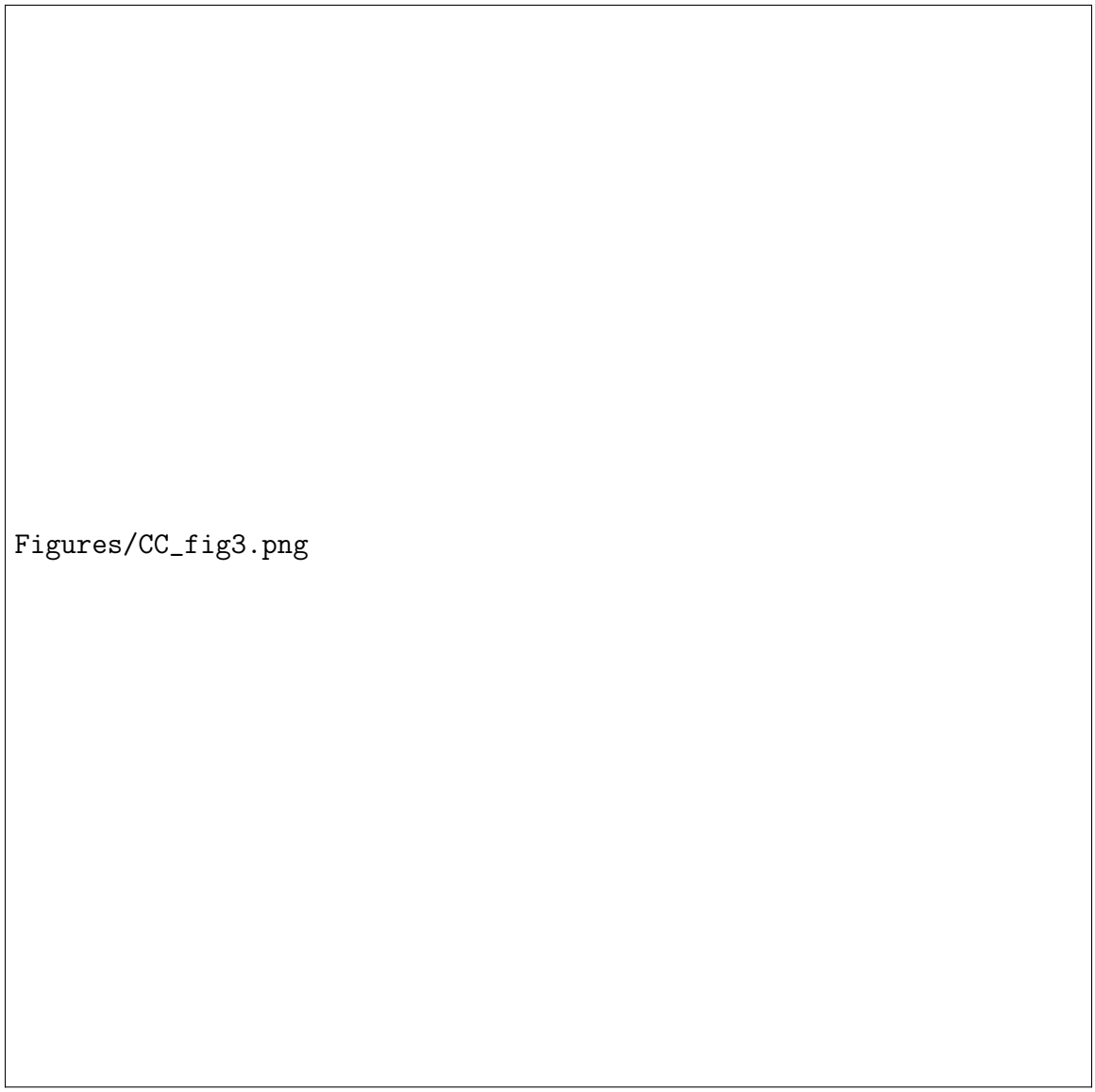
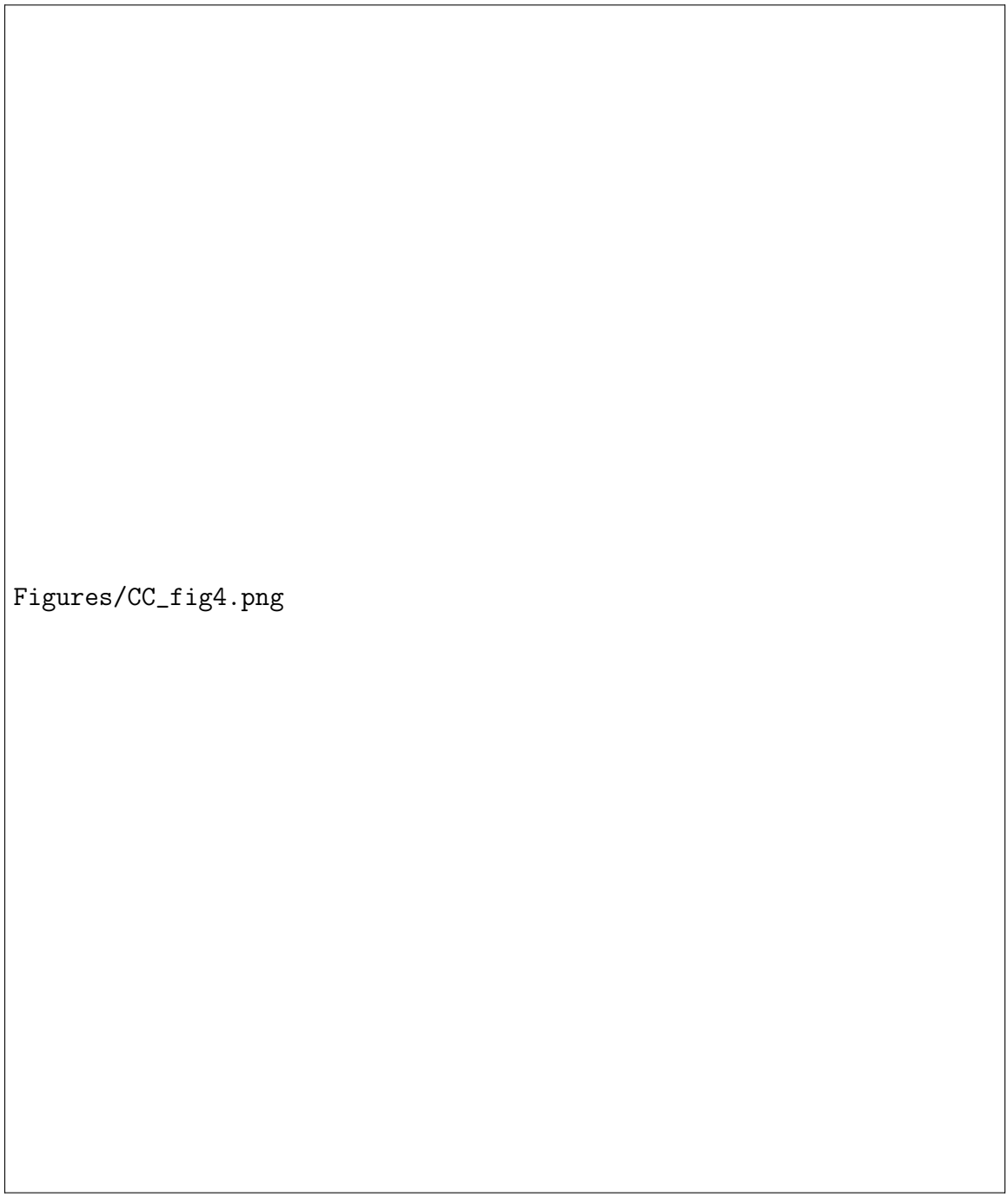


Figure 2.3: Two-photon calcium imaging of choice- and outcome-related signals in secondary motor cortex (M2). (A) Schematic representation of experimental setup for behavior with simultaneous two-photon imaging. (B) Schematic representation of preparation for in vivo two-photon imaging of M2. PrL, prelimbic cortex; Cg1, cingulate area 1; M1, primary motor cortex. (C) An example field-of-view in layer 2/3 of M2 containing GCaMP6s-expressing neurons. The image is a mean projection of the full time-lapse image stack from Experiment 13 in Table 1. (D) Mean fluorescence traces from an example cell, aligned to the sound cue and averaged across different subsets of trials. In the leftmost three panels, traces from left (red) and right (blue) trials are overlaid for each trial outcome. The rightmost two panels display the same data, with traces from single- (solid), double- (dotted) and omitted-reward trials (dashed) overlaid for each chosen action. Shading, 90% confidence interval. (E–F) Same as D for two additional cells.

cells, 27%; $p < 0.01$ for the corresponding regression coefficient in at least five consecutive time-bins, binomial test; Fig. 2.4B). Dependence on outcome (42 cells, 5%; Fig. 2.4C) or a choice–outcome interaction (25 cells, 3%; Fig. 2.4D) was evident in comparatively fewer cells. Notably, significant fractions of M2 neurons continued to show choice- and outcome-dependent signals well into the next trial, persisting even after the next choice was made (black bars denoting $p < 0.01$, binomial test, middle panel, Fig. 2.4B–D).

The observation that M2 neurons can represent chosen actions across more than one trial led us to ask whether current and prior choices are represented at the single-unit level by the same or different populations. To address this question, we quantified the number of neurons with a significant regression coefficient only for the current choice ($p < 0.01$ for a_1, the regression coefficient for the current choice estimated 2 s after cue onset, and $p \geq 0.01$ for a_2, the regression coefficient for the prior choice estimated at the time of cue onset), only for the prior choice ($p \geq 0.01$ for a_1 and $p < 0.01$ for a_2) and for both the current and prior choices ($p < 0.01$ for both a_1 and a_2). We found that most choice-dependent M2 neurons were sensitive exclusively to the current (147/771 cells, 19%) or prior (66/771 cells, 9%) choice (Fig. 2.4E). Within the small proportion of neurons sensitive to both current and prior choices (36/771 cells, 5%), the corresponding regression coefficients were correlated and generally did not change signs (Fig. 2.4F). Therefore, the choice preferences of these cells were maintained across time-points in consecutive trials. Taken together, these results indicate that very few of the choice-selective neurons in M2 represent both the current and prior choice during the period immediately following cue onset.

Figure 2.4 *(following page)*: Sustained representations of choices and their outcomes in M2. (A) A schematic representation of the multiple linear regression model that was fit to the fluorescence of each neuron in each 500 ms time bin. (B) The proportion of cells with significant choice-dependent activity as quantified by the regression model, plotted as a function of time. The regression model accounted for the influence of choices made on the current trial (left), the last trial (middle) and the trial before last (right), as well as the additional predictors shown in C–D. Significance of each predictor was tested at $\alpha = 0.01$. Black bars, bins in which the proportion of cells with significant regression coefficients was above chance level ($P < 0.01$, binomial test). Gray shading, the significance threshold for the binomial test. Black error bar, 95% CI for time of outcome. $N = 771$ cells from 16 sessions from 10 mice. (C) Same as B for trial outcome. (D) Same as B for the interaction of choice and outcome. (E) The proportion of neurons with a significant regression coefficient for the choice made in the current trial only, in the prior trial only, and in both the current and prior trials. (F) Scatter plot of the neurons with significant regression coefficients for both the current and prior choice. The coefficient for the current choice, a_1, is plotted against the coefficient for the prior choice, a_2. Right inset, the same plot expanded to show the five data points outside the range of the main axis.


Figures/CC_fig4.png

2.3.6 Effect of Outcome on Choice Selectivity of Individual M2 Neurons

To further characterize choice representations in M2, we focused on the 226 'choice-selective' neurons whose activity was found in the multiple linear regression analysis to be significantly modulated by choice, or the interaction of choice and outcome, or both. For each of these neurons, fluorescence traces were first averaged across subsets of trials according to whether the contralateral (left) or ipsilateral (right) spout was chosen (trial-averaged traces for each cell during single reward, left trials are shown in Fig. 2.5A). A time-varying choice selectivity index was then calculated as the normalized difference between the two mean traces. During the period following cue onset in single-reward trials, the majority of choice-selective neurons preferred the contralateral choice (134/226, 59%; Fig. 2.5B). Similar to the degree of temporal variation observed across neurons in their trial-averaged activity traces, peak choice selectivity was also temporally distributed across neurons relative to the sound cue (Fig. 2.5A,B).

Unexpected outcomes signify inaccuracies within a subject's internal representation of the environment, including the values assigned to specific actions. Such information can be crucial to instrumental behavior. Thus, we asked whether choice representations in M2 were modified when low-probability outcomes occurred. Specifically, we examined double- and omitted-reward trials, which occurred a similar number of times per session. The choice selectivity of individual neurons clearly varied by trial type, as evidenced by the degree of scatter from the unity line in Figure 2.5C. Nevertheless, across the choice-selective neurons, the choice selectivity values for double- and omitted-reward trials were significantly correlated (Pearson's $R = 0.31$, $p = 5 \times 10^{-6}$, $N = 226$ cells). Notably, choice selectivity magnitudes were reduced on omitted reward as compared to double-reward trials ($p = 5 \times 10^{-5}$, Wilcoxon signed-rank test; Fig. 2.5D). Extending this analysis to include all imaged neurons yielded similarly correlated choice selectivity values ($R = 0.20$, $p = 2 \times 10^{-8}$, $N = 771$ cells) and a similar reduction in choice selectivity magnitudes for omitted reward relative to double-reward trials ($p = 0.002$, Wilcoxon signed-rank test). These results indicate that trial outcomes can substantially influence choice representations in M2. In particular, the absence of an expected reward in omitted-reward trials was associated with weaker representations of chosen actions at the level of single neurons.

2.3.7 Reward Omission Weakens Ensemble-level Choice Representations

Frontal cortical neurons often exhibit complex patterns of selectivity for task variables, and mounting evidence suggests that the associated neural representations may be best understood by focusing on the activity of large populations of neurons, rather than single units. Therefore, relative to the single-unit analyses presented above, the measure of accuracy at decoding choices from ensemble activity could provide a more reliable estimate of the choice information available to downstream brain areas.

To test the effect of trial outcomes on population-level choice representations in

Figure 2.5: Choice representations in M2 are modified by trial outcome. (A) Heat map of trial-averaged fluorescence as a function of time for all choice-selective neurons during single-reward, left trials. Cells are sorted by the center-of-mass of their trial-averaged fluorescence traces. $n = 226$ cells with significant encoding of choice or an interaction of choice and outcome as determined by multiple linear regression (see Methods). (B) Heat map of choice selectivity for the neurons in A as a function of time during single-reward trials. Choice selectivity was calculated as the normalized difference between mean fluorescence traces from left and right trials. Red and blue shadings indicate preference for left and right choices, respectively. Cells are sorted first by mean choice preference and then by the center-of-mass of their choice selectivity traces. (C) Scatter plot of the neurons in A, plotting the choice selectivity of each cell in omitted-reward trials against double-reward trials. CC, Pearson correlation coefficient. (D) Empirical cumulative distribution of choice selectivity magnitudes for double-reward (solid) and omitted-reward (dotted) trials.

M2, we trained linear classifiers on the ensemble activity recorded during single-reward trials and then compared their accuracy for decoding choices across trials with different outcomes (see Materials and Methods). Baseline decoding accuracy was estimated with a Monte Carlo cross-validation procedure, using ensemble activity patterns from single-reward trials as training and testing sets. For each of 30 iterations, a classifier was constructed from a random sample of 80% of these trials and then tested on the remaining fraction. Decoding accuracy for the other outcome types was estimated by testing the same classifiers on the ensemble activity associated with the full set of double-reward, omitted-reward and error trials.

The choices made in correct trials could be decoded from the population activity with above chance-level accuracy at every time-point in the current trial following the animal's response (Fig. 2.6A–C, left; black bar denotes $p < 0.05$, Wilcoxon signed-rank test, vs. shuffle). An ensemble representation of the chosen action could also be detected for at least 1 s into the subsequent trial (Fig. 2.6A–C, right). For example, in single-reward trials, a maximal decoding accuracy of $78 \pm 2\%$ was reached at 1.75 s after cue onset ($mean \pm SEM$, $N = 16$ sessions; Fig. 2.6A, left). At the time of the next cue onset, accuracy remained above chance, at $63 \pm 2\%$ ($mean \pm SEM$, $N = 16$ sessions; Fig. 2.6A, right). By contrast, in error trials, the decoding accuracy increased gradually from the time of cue onset (Spearman's $\rho = 0.87$, $p < 10^{-4}$) and only rose to significance late in the trial (Fig. 2.6D, left). It reached a maximum of $66 \pm 4\%$ at the time of the next sound cue ($mean \pm SEM$, $N = 16$ sessions; Fig. 2.6D, right).

For explicit comparisons of decoding accuracy across outcome conditions, we computed the mean accuracy over all time-points in the current and subsequent trial (Fig. 2.6E). Relative to single-reward trials, mean decoding accuracy dropped significantly in omitted-reward ($p = 0.004$, Wilcoxon signed-rank test, $N = 16$ sessions) and error trials ($p = 0.002$, Wilcoxon signed-rank test, $N = 16$ sessions), with no detectable difference in double-reward trials ($p = 0.8$, Wilcoxon signed-rank test, $N = 16$ ses-

Figure 2.6 *(following page)*: The accuracy of decoding chosen actions from the neural ensemble activity was diminished during omitted-reward and error trials. Choices were decoded using classifiers based on linear discriminant analysis, and accuracy was estimated with Monte Carlo cross-validation (repeated random subsampling). (A) The accuracy of decoding choices made on single-reward trials (left), or trials in which the previous outcome was single reward (right), plotted as a function of time. Data are presented as $mean \pm SEM$. Chance-level accuracy (black dashed line) was determined by testing classifiers constructed using shuffled choices. Black horizontal bars, bins significantly different from chance ($P < 0.05$, Wilcoxon signed-rank test). Black error bar, 95% confidence interval for time of outcome. (B-D) Same as A for double-reward, omitted-reward, and error trials. Results from single-reward trials are overlaid for visual comparison (gray triangles). Lower gray bars, bins with a significant difference in decoding accuracy relative to single-reward trials. (E) Mean decoding accuracy across all time-points shown in A–D for each trial outcome. Gray triangles, individual sessions. Black crosshairs, $mean \pm SEM$. Wilcoxon signed-rank test: $^{**}P < 0.01$; n.s., not significant. (F) Same as E using random forest classifiers.

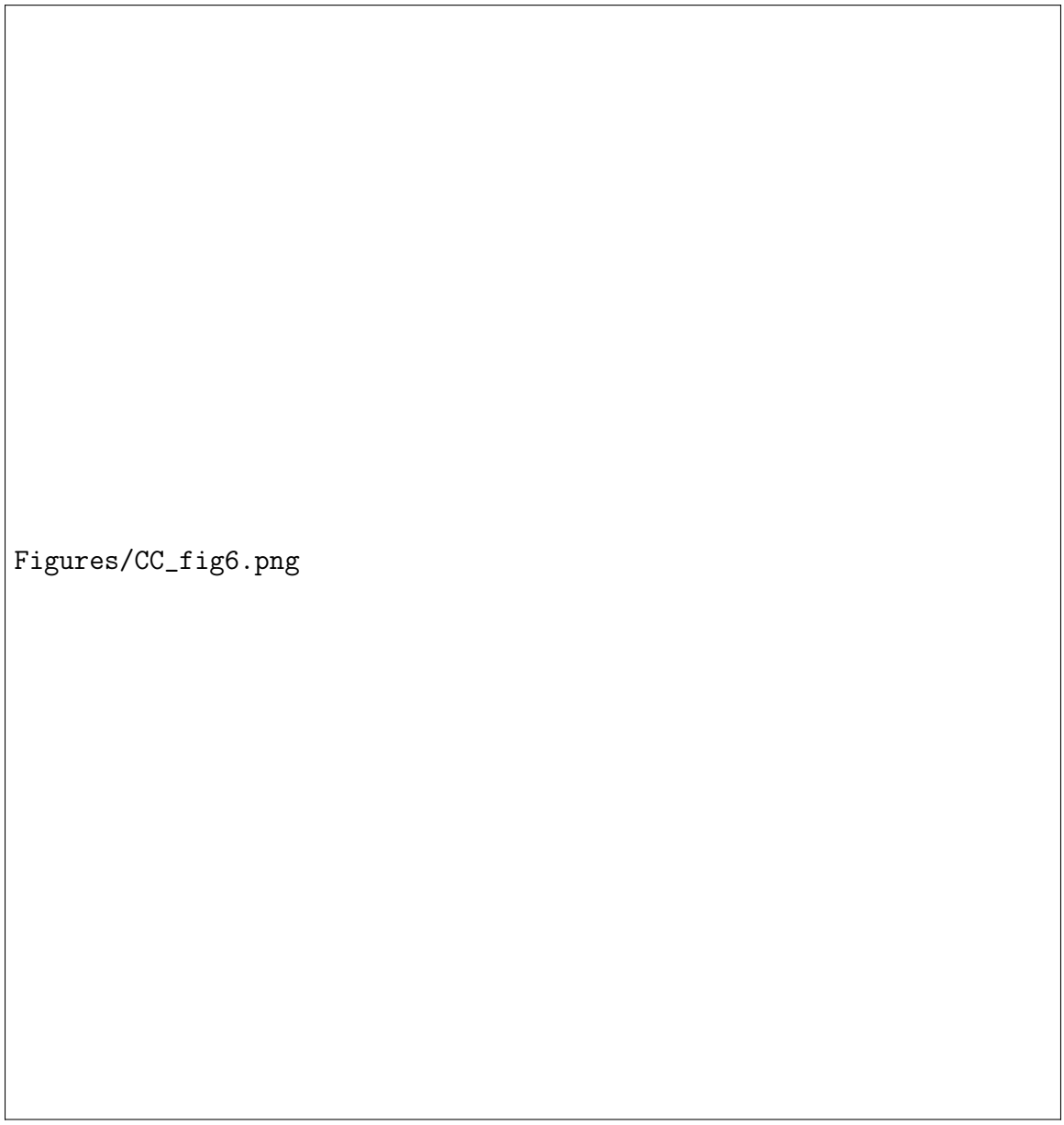


Figure 2.6 Decoding accuracy diminished during omitted-reward and error trials.

sions).

It seems unlikely that the choice information present in M2 ensemble activity would be read out by the brain in exactly the same manner as a linear classifier. Therefore, we sought an alternative approach to determine whether the results would generalize to other methods of decoding. Specifically, we turned to random forest classification, a bootstrap aggregation method based on decision trees. Random forest classifiers operate on a fundamentally different principle than linear classifiers (see Materials and Methods), but overall, they yielded very similar results. In particular, comparisons of mean decoding accuracy again revealed marked differences between single-reward trials and omitted-reward or error trials ($p = 0.006$ and 0.003, respectively, Wilcoxon signed-rank test, $N = 16$ sessions), but no difference between single- and double-reward trials ($p = 0.2$, Wilcoxon signed-rank test, $N = 16$ sessions; Fig. 2.6F). Thus, the results of two distinct classification approaches support the same conclusion: that choice information was encoded with higher fidelity during trials in which choices were rewarded. More specifically, increases in reward magnitude (ie double reward) had little impact on decoding accuracy, whereas reward absence in omitted-reward and error trials significantly diminished the accuracy with which chosen actions could be decoded from neural ensemble activity patterns in M2.

2.3.8 Simultaneous Recording Improves Decoding Accuracy

Our measurements of neural activity came from two-photon calcium imaging, which enabled the simultaneous acquisition of fluorescence transients from ensembles of at least 30 neurons. This allowed us to address an important, unresolved methodological question: does simultaneous recording confer a decoding advantage, relative to recording from each cell individually and then combining the data post hoc *in silico*?

As a basis for comparison, we generated 'pseudo-ensemble' data for each session, in which the activity traces from each neuron were shuffled across single-reward trials where the same action was chosen. Thus, the correspondence between individual neural activity traces and left or right choices was preserved. However, at the population level, the ensemble activity patterns no longer reflected simultaneously recorded activity. In particular, the shuffle should preserve co-fluctuations in neural activity related to the sound cue, choice and outcome of a given trial, while disrupting the residual correlations.

To determine the extent to which correlations in neural activity were disrupted by the shuffling procedure, we first determined the correlation between each pair of neurons across trials, using the mean cellular fluorescence over the time interval from 2 to 4 s after cue onset (Fig. 2.7A). As expected, pairwise correlations were reduced within the pseudo-ensembles (Fig. 2.7B). Across all sessions, the mean Pearson correlation coefficient decreased from 0.173 ± 0.020 in simultaneously recorded ensembles to 0.022 ± 0.007 in pseudo-ensembles ($p = 4 \times 10^{-4}$, Wilcoxon signed-rank test; Fig. 2.7C, left). The mean magnitude of correlation also decreased, from 0.192 ± 0.017 in simultaneously recorded ensembles to 0.054 ± 0.005 in pseudo-ensembles ($p = 4 \times 10^{-4}$, Wilcoxon signed-rank test; Fig. 2.7C, right).

Figure 2.7D shows the accuracy of decoding choices from actual and pseudo-

ensemble data as a function of time during single-reward trials. To directly compare the two conditions, we again focused on the time interval from 2 to 4 s after cue onset, when decoding accuracy was highest. Over this interval, linear classifiers constructed with either simultaneous or pseudo-ensemble activity from 30 cells could decode choices with high accuracy, at 75 ± 2 and $72 \pm 2\%$, respectively ($mean \pm SEM$, $N = 16$ sessions; black and red bars denoting $p < 0.01$ vs. shuffle, Wilcoxon signed-rank test, Fig. 2.7E). In a head-to-head comparison, decoders constructed from simultaneous activity outperformed pseudo-ensemble decoders by $4.0 \pm 0.8\%$ ($mean \pm SEM$, $N = 16$ sessions). Random forest classifiers also exhibited a significant simultaneity effect, although it was slightly smaller than for linear classifiers. Decoding accuracy for ensemble and pseudo-ensemble activity was $76 \pm 2\%$ and $73 \pm 2\%$, respectively, with a

Figure 2.7 *(following page)*: Simultaneous recording imparted a decoding advantage that increased as a function of ensemble size. The accuracy of decoding choices from the activity of simultaneously imaged ensembles of neurons was compared to that of pseudo-ensembles in which simultaneity was disrupted by shuffling the activity traces from each neuron across trials in which the same choice was made. Only correct trials resulting in a single reward were used for this analysis. Classification accuracy was tested using Monte Carlo cross-validation (repeated random subsampling). Chance-level accuracy was determined by testing classifiers constructed using shuffled choices. (A) Pearson correlation matrix for all cells from one example session, under three conditions: with simultaneity preserved ('ensemble'), after shuffling across trials with the same chosen action ('pseudo-ensemble'), and after shuffling across trials irrespective of chosen action ('full scramble'). Correlations were estimated from cellular fluorescence averaged over the interval from 2 to 4 s following cue onset. (B) Histogram of the Pearson correlation coefficients estimated for all pairs of cells imaged in all experiments, using simultaneous ensembles (top) and pseudo-ensembles (bottom). (C) Mean pairwise correlation (left), and pairwise correlation magnitude (right) across all sessions. Gray lines, means from individual sessions. Black crosshairs, $grandmean \pm SEM$. (D) Performance of decoders based on linear discriminant analysis, plotted as a function of time for single-reward trials (left), or trials in which the previous outcome was a single reward (right). Accuracy of ensemble classifiers (black circles) is overlaid with that of pseudo-ensemble classifiers (red triangles) for visual comparison. Black dashed line, chance-level accuracy. (E) Decoder performance as a function of the number of cells used to decode the chosen action. Performance of ensemble (black circles) and pseudo-ensemble (red triangles) classifiers was estimated as the mean classification accuracy over the interval from 2 to 4 s following cue onset in single-reward trials. The number of cells was varied from 1 to 30 by drawing cells randomly from the full ensemble or pseudo-ensemble without replacement. Black dashed line, chance-level accuracy. Horizontal bars, bins in which ensemble (black, upper bar) or pseudo-ensemble (red, lower bar) classifiers performed significantly better than chance ($P < 0.05$, Wilcoxon signed-rank test). (F) Marginal percentage point change in decoding accuracy, plotted as a function of ensemble size for ensembles (black) and pseudo-ensembles (red). (G) Difference in accuracy of the ensemble and pseudo-ensemble decoders shown in E–F, plotted as a function ensemble size. Black horizontal bars, bins in which the accuracy of ensemble and pseudo-ensemble classifiers differed significantly ($P < 0.05$, Wilcoxon signed-rank test). (H–K) Same as D–G for random forest classifiers. Data in D–K are presented as $mean \pm SEM$.

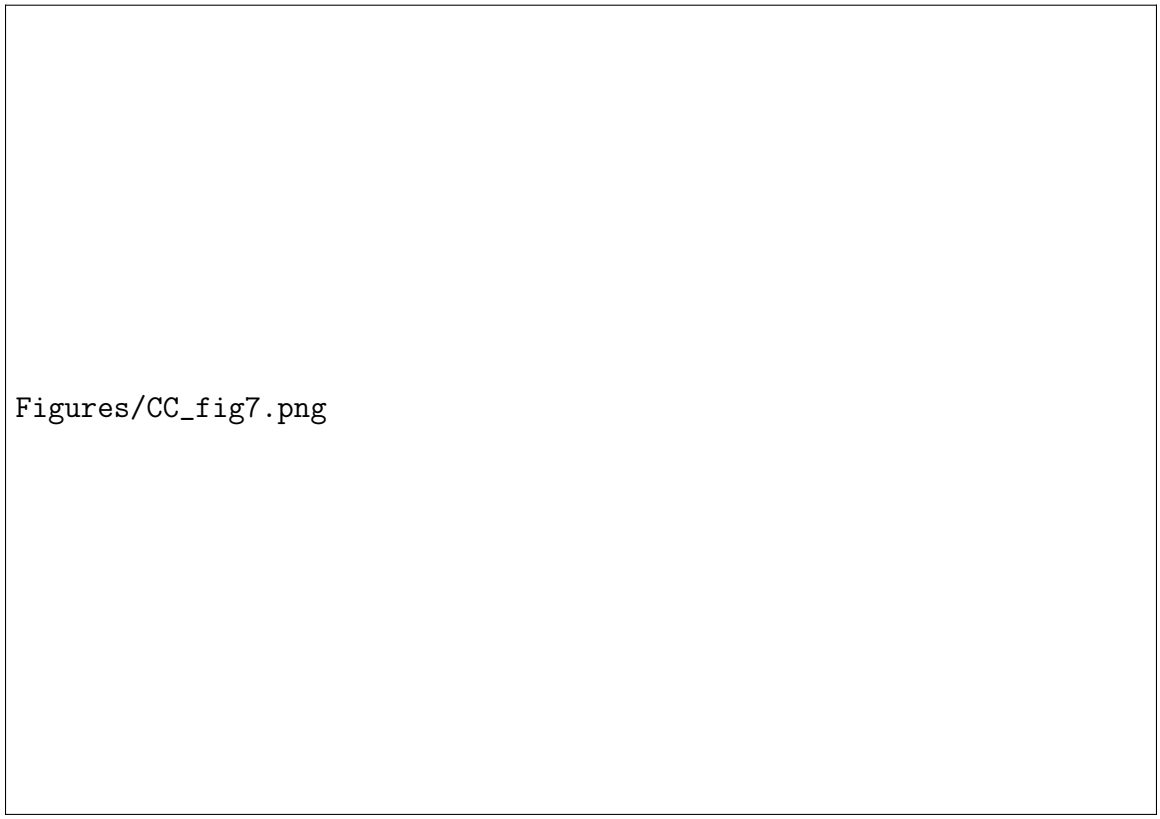


Figure 2.7 Decoding advantage of simultaneous recording increased with ensemble size.

mean difference of $3.3 \pm 0.5\%$ in a head-to-head comparison using an ensemble size of 30 cells ($mean \pm SEM$; Fig. 2.7H).

Additionally, we assessed the impact of ensemble size on decoding accuracy by training and testing classifiers using random samples of 1 to 30 neurons. Classifiers constructed from either ensemble or pseudo-ensemble data could decode choices with an accuracy exceeding chance for every ensemble size tested (Fig. 2.7E, black and red bars bar denoting $p < 0.01$ for ensembles and pseudo-ensembles, respectively, Wilcoxon signed-rank test). In both cases, mean decoding accuracy increased as function of ensemble size (ensembles: Spearman's $\rho = 0.997$, $p = 0$; pseudo-ensembles: $\rho = 0.996$, $p = 0$). However, the marginal change in decoding accuracy dropped rapidly (Fig. 2.7F). For ensembles, it decreased from 3.6 ± 0.7 to $1.3 \pm 0.4\%$ after the second and tenth cell was added, respectively ($mean \pm SEM; p = 0.02$, Wilcoxon signed-rank test). For pseudo-ensembles, it decreased from 3.1 ± 0.7 to $0.6 \pm 0.4\%$ ($mean \pm SEM; p = 0.02$, Wilcoxon signed-rank test).

Interestingly, the decoding advantage associated with simultaneous recording was also related to ensemble size. The difference in mean decoding accuracy was significant for all ensemble sizes larger than nine cells (Fig. 2.7G, black bar denoting $p < 0.01$, Wilcoxon signed-rank test) and increased as a function of the number of cells up to the largest ensemble size examined (Spearman's $\rho = 0.97, p = 0$). An identical analysis using random forests yielded similar results (Fig. 2.7I–K).

Taken together, our analyses reveal that choices can be decoded more accurately from simultaneously recorded population activity, relative to pseudo-ensembles in which the correlations in neural activity associated with simultaneity have been disrupted. This difference increased with the number of neurons in an ensemble, across the range of ensemble sizes tested. Moreover, the marginal decoding accuracy decreased rapidly regardless of whether ensembles or pseudo-ensembles were used—a result consistent with high levels of redundancy in the population code for chosen actions in M2.

2.4 Discussion

How does the outcome of a chosen action influence how it is represented in the brain? In this study, we used a two-choice discrimination task with probabilistic outcomes to investigate this question in the M2 region of the murine MFC. The results help to illuminate how information related to choices and their outcomes are integrated within the frontal lobe. M2 neurons were found to robustly encode rewarded choices; however, choice-related signals diminished when a rewarding outcome was omitted. Furthermore, an increase in the magnitude of reinforcement had far less impact on choice representations than did its categorical presence or absence. The preferential encoding of rewarded choices in M2 provides a plausible mechanism that may underlie its established role in the learning and implementation of reinforced actions during instrumental behavior.

2.4.1 Cortical Representation of Prior Choices in Rodents

Optimal performance in the discrimination task required the subject to choose strictly based on auditory cues. In principle, information about past actions could be discarded or ignored. Therefore, in these well-trained animals it was somewhat surprising to observe robust and persistent choice representations, both at the level of single neurons (Figs. 2.4 and 2.5) and ensembles (Figs 2.6 and 2.7). However, similar task-irrelevant information coding has been reported elsewhere—in the monkey prefrontal cortex (Genovesio et al., 2014), as well as in the posterior parietal cortex of rodents (Morcos and Harvey, 2016). At the behavioral level, response biases based on choice and outcome history have been observed in human subjects during perceptual tasks even after extensive training (Fründ et al., 2014; Abrahamyan et al., 2016). Our findings and previous results therefore suggest that under some circumstances, higher-order cortical areas continue to monitor past choices and outcomes, even if task performance does not strictly require such information.

Previous studies have reported that neurons in the rodent MFC encode past choices (Sul et al., 2011; Siniscalchi et al., 2016) and their outcomes (Kargo et al., 2007; Sul et al., 2011; Yuan et al., 2014). However, sustained choice-related signals are not unique to this brain region. They have also been found in other nodes of the frontal-striatal network including the dorsomedial striatum (Kim et al., 2013) and orbitofrontal cortex (Sul et al., 2010), as well as the posterior parietal cortex (Hwang et al., 2017). This is not to say that all cortical regions exhibit persistent signals associated with chosen actions. For example, our previous study detected only very brief choice signals in the mouse primary visual cortex (Siniscalchi et al., 2016). Similarly, choice signals like those found in dorsomedial striatum lasted only transiently in dorsolateral striatum (Kim et al., 2013). This point is notable because M2 and other medial frontal areas send dense projections to dorsomedial striatum, while afferents in dorsolateral striatum come mostly from primary motor cortex (Reep and Corwin, 1999).

The extended time course of the choice signals we observed could serve as an eligibility trace that keeps recently performed actions available for learning. In the current study, the representation of choice-related information in M2 neurons persisted into the middle of the next trial (Fig. 2.6A). It is worth noting that our previous study detected significant choice-related signals over an even longer duration—up to two trials after the corresponding action was chosen (Siniscalchi et al., 2016). Interestingly, the prior study employed a rule-switching task in which subjects were required to monitor choices and their outcomes. This raises an intriguing possibility: that the temporal scale of choice history signals may depend on the task demands (Bernacchia et al., 2011; Donahue et al., 2013). Namely, the optimal learning rate depends on the volatility of the environment (Behrens et al., 2007; Farashahi et al., 2017). If the persistence of choice representations is indeed a flexible parameter, then it could allow the system to adapt to changes in volatility by serving as a point of adjustment for the temporal integration of choice information.

2.4.2 Enhanced Population Coding for Rewarded Choices

Our results reveal that neural representations of chosen actions in mouse M2 are outcome-dependent. This finding agrees in principle with a previous study demonstrating that rewarded choices are more reliably encoded relative to unrewarded choices in the primate supplementary eye field and dorsolateral prefrontal cortex during a matching-pennies task (Donahue et al., 2013). Interestingly, the effect was not evident in recordings from the same set of neurons during a visual search task in which a visuospatial cue instructed the correct response at the beginning of each trial. The critical difference between these two tasks may be the presence of an instructive cue—a feature which the visual search task shares with the task used in our current study. Notably, we did find a robust reward-associated enhancement of choice coding under these circumstances. One possible explanation is that, in contrast to visuospatial instruction, the cues used in our auditory discrimination task only came to be associated with the correct choice through learning. Therefore, the reward-associated enhancement we observed may reflect an action-monitoring process associated with the maintenance of arbitrary sensorimotor associations—a process which may be unnecessary during the less demanding visual search task. Another possible explanation regards our use of ensemble decoding, which may be more sensitive than decoding from single units and can thus be expected to detect smaller differences between outcome conditions.

In general, the neural correlates of choice and outcome history have been studied using binary outcome conditions in which a reward is either provided or withheld on a given trial. We sought to extend the results of these prior studies by comparing effects of multiple reward magnitudes. Additionally, we were able to dissociate effects of performance and outcome by measuring the impacts of unexpected reward omissions and windfalls in mice trained to a very high level of proficiency (¿90

At the behavioral level, reward size affected consummatory licking as well as the likelihood of a response to the next cue, but failed to influence the accuracy of responses (Fig. 2.2A–C). This suggests that infrequent changes in reward magnitude impacted motivation without significantly influencing choices. Moreover, the neural ensemble representations of chosen actions did not differ between single- and double-reward trials. Instead, the greatest contrast was found between rewarded and unrewarded choices. Our results therefore suggest that the influence of outcome on choice representations in M2 is driven less by the magnitude of reward than by its explicit presence or absence. However, because actions were reinforced immediately in our task design, it remains possible that floor or ceiling effects could have limited the outcome sensitivity of the choice-selective population.

What physiological mechanisms underlie the outcome dependence of choice signals observed in M2? One intriguing possibility concerns the role of neuromodulation, which may directly reconfigure the local network dynamics, or act on inputs to M2. In particular, dopaminergic (Schultz et al., 1997) and cholinergic (Hangya et al., 2015) neurons are known to carry signals related to reward. Furthermore, reward-dependent activation of dopaminergic projections to nearby primary motor cortex have been implicated in motor skill learning (Hosp et al., 2011; Leemburg et al.,

2018). It is therefore interesting to speculate on whether similar mechanisms might contribute to associative learning (Takehara-Nishiuchi and McNaughton, 2008) and more specifically, to the auditory-motor associations necessary for performance of the task presented here. In any case, the impact of neuromodulators on motor cortical choice signaling will comprise an exciting topic for future research.

2.4.3 Persistent Neural and Behavioral Effects Associated with Errors

The actions chosen on error trials were decoded least accurately from the corresponding ensemble activity (Fig. 2.6E,F). This result is consistent with an earlier study that revealed disrupted MFC ensemble representations for choices and their outcomes during periods when rats committed multiple errors in a radial arm maze (Lapish et al., 2008; Hyman et al., 2012b). The discretized trial structure of our auditory discrimination task allowed us to build upon this prior result by measuring the reliability of ensemble representations into the next trial. Furthermore, the inclusion of omitted-reward trials allowed direct comparisons of ensemble representations associated with correct and incorrect choices, independent of the associated outcomes.

Another previous study demonstrated a tight relationship between sustained error signals in MFC and behavioral performance in the next trial, measured as post-error slowing during a timing task (Narayanan et al., 2013). Similarly, our analysis revealed not only an error-related decrement in the fidelity of neural choice representations, but also a behavioral performance decrement following error trials that could not be explained by reward omission alone. These results may provide some insight into the sources of error for these well-trained subjects. Specifically, the prolonged time course of the neural and behavioral effects associated with errors suggests that they may have arisen in part due to factors that spanned multiple trials—such as periods of hypo- or hyper-arousal. In any case, these results together with prior studies indicate that errors are often associated with persistent internal states that can impact subsequent behavioral performance.

2.4.4 Simultaneous Recording Confers a Modest Decoding Advantage in M2

Prior theoretical work has demonstrated that correlated variability in neural populations can either degrade or enhance population coding, depending on the interaction between signal and noise correlations (Averbeck and Lee, 2003; Averbeck et al., 2006). The analysis shown in Fig. 2.7 revealed that chosen actions could be decoded more accurately from simultaneously recorded ensembles, relative to pseudo-ensembles in which the correlations in neural activity associated with simultaneity (ie noise correlations) had been disrupted. In particular, the observed effect of simultaneity seems to have resulted from the preservation of largely positive correlations in trial-to-trial neural variability unrelated to the chosen action (Fig. 2.7B,C). Possible sources of noise correlations in our recordings could include unobserved behavior such as whisk-

ing, features of the network architecture, or changes in internal state associated with motivation or arousal.

The effect of simultaneity increased with the number of neurons in an ensemble, across the range of ensemble sizes tested (Fig. 2.7G). Notably, an earlier study in the primate supplementary motor area found no statistically significant effect of correlated spike-count variability on the encoding of movements by ensembles of three to eight neurons (Averbeck and Lee, 2006). Our analyses only revealed a consistent simultaneity effect for ensembles larger than nine neurons, which may highlight the utility of large-scale recordings for addressing this question. However, it should be emphasized that even for ensembles of 30 cells, the comparative advantage for ensembles was modest (4%), and choices could still be decoded from pseudo-ensembles with above chance-level accuracy at every ensemble size. Furthermore, estimated correlations between neurons tend to strengthen at longer timescales (Averbeck and Lee, 2003). Hence, the wider time-bins used in our study (500 vs. 66 ms), as well as the slower dynamics associated with calcium imaging could explain why our analyses were more sensitive to correlated variability. We also found that marginal decoding accuracy decreased rapidly as cells were added to the population, for both ensembles and pseudo-ensembles (Fig. 2.7F,J). This result suggests a high level of redundancy in M2 population codes, similar to previous results found in the rat primary motor cortex during a simple reaction time task (Narayanan et al., 2005).

2.4.5 Insights into the Role of M2 in Goal-Directed Behavior

The choice selectivity magnitudes of individual neurons (Fig. 2.5D) and the accuracy of decoding choices from ensemble activity (Fig. 2.6E,F) both decreased from double- to omitted-rewarded trials, and then further decreased in error trials. How do the observed physiological changes ultimately impact behavior? Causal perturbations aimed at addressing this question will require a more detailed understanding of how genetically (Kvitsiani et al., 2013; Pinto and Dan, 2015; Kamigaki and Dan, 2017) or anatomically identified subtypes of frontal cortical neurons (Li et al., 2015; Chen et al., 2017; Otis et al., 2017) contribute to the choice signals observed in our experiments.

Goal-directed behavior requires the capacity to adjust the current policy for action selection according to the impact of past choices on the likelihood of a desired outcome. Our results demonstrate that sustained neural representations of chosen actions in mouse M2 are sensitive to their resultant outcomes, such that rewarded choices are more robustly encoded. In turn, the preferential encoding of rewarded choices could allow the frontal cortex to bias the influence of recent, positively reinforced actions on future decisions. This proposed mechanism would help to explain effects of lesioning (Passingham et al., 1988; Gremel and Costa, 2013) and inactivation (Siniscalchi et al., 2016; Makino et al., 2017) that have implicated M2 more broadly in the learning and implementation of voluntary behavior. In summary, our results contribute to a growing body of evidence supporting a role for MFC, and M2 more specifically, in the flexible execution of goal-directed actions.

2.5 Acknowledgments

The contents of this chapter were modified from a previously published research article:

Siniscalchi MJ, Wang H, Kwan AC. Enhanced population coding for rewarded choices in the medial frontal cortex of the mouse. Cerebral Cortex. 2019 Jan 7;29(10):4090-106.

Author Contributions

M.J.S. and A.C.K. designed the experiments. M.J.S. conducted the experiments. M.J.S., H.W. and A.C.K. analyzed the data. M.J.S. and A.C.K. wrote the article.

Funding

National Institute of Mental Health (grant R01MH112750 and R21MH118596 to A.C.K.); National Institute on Aging (grant P50AG047270 to A.C.K.); SFARI Explorer Award (A.C.K.); National Institutes of Health (training grant T32NS041228 to M.J.S.); National Science Foundation Graduate Research Fellowship (DGE-1122492 to M.J.S.); China Scholarship Council-Yale World Scholars Fellowship (H.W.).

Chapter 3

Neural Ensemble Dynamics during Flexible Sensorimotor Behavior

The ability to shift between repetitive and goal-directed actions is a hallmark of cognitive control. Previous studies have reported that adaptive shifts in behavior are accompanied by changes in neural activity in frontal cortex. However, neural and behavioral adaptations can occur at multiple time scales, and their relationship remains poorly defined. Here we developed an adaptive sensorimotor decision-making task for head-fixed mice, requiring them to shift flexibly between multiple auditory–motor mappings. Two-photon calcium imaging of secondary motor cortex (M2) revealed different ensemble activity states for each mapping. When adapting to a conditional mapping, transitions in ensemble activity were abrupt and occurred before the recovery of behavioral performance. By contrast, gradual and delayed transitions accompanied shifts toward repetitive responding. These results demonstrate distinct ensemble signatures associated with the initiation and termination of sensory-guided behavior and suggest that M2 leads the engagement of goal-directed response strategies that require sensorimotor associations.

3.1 Introduction

Operant behaviors are structured around stimuli, actions and outcomes. Successful execution of a task requires selecting actions that are consistent with the contingencies between these task variables. Control of action selection in the brain should be both stable and flexible. On the one hand, stability allows a subject to sustain high performance to maximize reward. On the other hand, flexibility is essential for quickly adjusting behavior when a change in contingencies occurs. Striking the delicate balance between stability and flexibility is therefore a key requirement of adaptive decision-making. Moreover, a lack of balance between these opposing aspects of cognitive control is a hallmark of psychiatric disorders (Griffiths et al., 2014).

How do we know when to be stable or flexible in a changing environment? In tasks without explicit contextual cues, subjects may adjust their response strategy through

reward feedback. Prior studies have observed task-dependent differences in neuronal firing rates and selectivity in multiple frontal cortical regions (Asaad et al., 2000; Rich and Shapiro, 2009; Rodgers and DeWeese, 2014). During periods of behavioral adjustment, evolution of cortical activity was found to be gradual and late, occurring on time courses that generally match or lag the improvement in task performance (Mitz et al., 1991; Chen and Wise, 1995; Pasupathy and Miller, 2005; Antzoulatos and Miller, 2011). However, neurons in the frontal cortex exhibit substantial cell-to-cell variability in such time courses (Mitz et al., 1991). Population activity may therefore be more useful for capturing circuit dynamics (Mante et al., 2013; Stokes et al., 2013; Wills et al., 2005). Using ensemble recordings, two studies examined reward-guided adaptations, and they found the corresponding changes in network activity to be surprisingly abrupt (Durstewitz et al., 2010; Karlsson et al., 2012). Determining the functional significance of these findings, however, will require quantitative comparisons of ensemble activity transitions that differ in their dynamics. Transitions that are relatively gradual versus abrupt, or that differ in onset with respect to behavioral changes, could reflect distinct underlying mechanisms for cognitive control.

To study adaptive sensorimotor decision-making in mice, we designed a head-fixed task that required animals to shift many times between three sets of stimulus–response contingencies. This task is a variant of arbitrary sensorimotor mapping, a classic model in which subjects are required to follow conditional rules (Bunge et al., 2005; White and Wise, 1999), such as "for stimulus A, perform one action; for stimulus B, perform another action." Once learned, the stimulus–response contingencies can then be switched, requiring the learning of novel mappings or retrieval of familiar associations. Associations are made by linking nonspatial stimuli or conditions to actions, and are therefore termed "arbitrary" (Wise and Murray, 2000). A number of brain regions are involved in arbitrary sensorimotor mapping, including the frontal lobe, striatum, hippocampus and thalamus (Wise and Murray, 2000). Within the frontal lobe, the dorsal premotor cortex has been implicated in the selection of motor programs based on antecedent conditions, as evidenced by the results of lesion studies (Petrides, 1985; Halsband and Passingham, 1985; Nixon et al., 2004), electrophysiology (Mitz et al., 1991; Chen and Wise, 1995), functional imaging (Toni et al., 2001; Boettiger and D'Esposito, 2005) and transcranial stimulation (Rushworth et al., 2002) in humans and nonhuman primates.

Secondary motor cortex (M2) has been described as a potential rodent homolog of primate higher-order motor areas (Murray et al., 2000; Preuss, 1995). Its location, adjacent to the medial prefrontal and primary motor regions, suggests that it may function as a cognitive–motor interface. A long line of research has linked the premotor cortex and neighboring regions to the generation of volitional movements (Nachev et al., 2008; Schall et al., 2002; Isoda and Hikosaka, 2007). Recent studies in rodents have also focused on the role of M2 in driving movements. Random-ratio lever-pressing was shown to become insensitive to reward devaluation in M2-lesioned mice, suggesting a role for M2 in goal-directed actions (Gremel and Costa, 2013). Neural activity is modulated before movement, reflecting involvement in action preparation and initiation (Sul et al., 2011; Erlich et al., 2011; Murakami et al., 2014). Moreover, M2 neurons encode not only current action, but also prior choice and outcome, indicating

a broader role in decision-making (Sul et al., 2011). One early study showed that rats with lesions of medial frontal cortex—a broader region that includes M2—had deficits in a visual conditional motor task (Passingham et al., 1988).

To elucidate the relationship between frontal ensemble activity and adaptive behavior, we used two-photon calcium imaging to record from M2 neurons in behaving mice. We found distinct population activity patterns associated with each of the three sets of stimulus–response contingencies. Moreover, following a contingency switch, transitions between ensemble patterns occurred earlier and were more abrupt when animals were required to abort repetitive actions and use a conditional rule. In fact, this change in ensemble activity state could be detected after only a few error trials, preceding the more gradual recovery of behavioral performance. Our results uncover distinct neural transitions associated with different phases of voluntary behavior and identify a leading role for M2 in engaging actions that require the use of sensorimotor associations.

3.2 Materials and Methods

Animals

We used adult male mice with C57BL/6J genetic background. Mice were housed in groups of 3–5, in 12-h/12-h light-dark cycle (lights off at 19:00), and most experiments were performed in late afternoons and evenings (16:00–midnight). At the start of experiments, mice were P51–117. No statistical tests were used to predetermine sample sizes, but sample sizes for this study are similar to those generally employed in the field. All experimental procedures were approved by the Institutional Animal Care and Use Committee, Yale University.

Surgery

Mice underwent two surgeries. For each surgery, the mouse was anesthetized with 2% isoflurane in oxygen during induction, which was then lowered to 1–1.5% for the remainder of the surgery. The mouse was placed over a water-circulating heating pad (TP-700, Gaymar Stryker) in a stereotaxic frame (David Kopf Instruments). Before the operation, the mouse was injected with carprofen (5 mg/kg, s.c.; #024751, Butler Animal Health) and dexamethasone (3 mg/kg, s.c.; Dexaject SP, #002459, Henry Schein Animal Health). The mouse was injected with carprofen immediately after surgery (5 mg/kg, s.c.) and each day for the following 3 d (5 mg/kg, s.c.).

For the first surgery, an incision was made to expose the skull. Based on stereotaxic coordinates, the center location of the mouse secondary motor cortex (M2; AP 1.5 mm and ML −0.5 mm relative to bregma) was marked over the right hemisphere. In other experiments, we targeted the anterior-lateral motor cortex (ALM; AP 2.5 mm, ML −1.5 mm) or the primary visual cortex (V1; AP −3.8 mm, ML 2.0 mm) in the right hemisphere. A stainless steel head plate (eMachineshop) was cemented in place with clear C&B Metabond (Parkell, Inc.); care was taken to cover the entire exposed

area of skull. Mice were given at least 1 week to recover before behavioral training (see below).

Once mice reached a performance criterion of $> 90\%$ correct rate on three consecutive days and was ready for imaging experiments, a second surgery was performed under anesthesia. Using a dental drill, a 3-mm-diameter craniotomy was made at the target location, which had been marked previously and remained visible through the Metabond. Dura was left intact and was irrigated with artificial cerebrospinal fluid (aCSF, in mM: 5 KCl, 5 HEPES, 135 NaCl, 1 MgCl2, 1.8 CaCl2; pH 7.3).

Using a glass micropipette attached to a microinjection system (Nanoject II, Drummond), 32–46 nL of AAV1-Syn-GCaMP6s-WPRE-SV40 (5×10^{13} titer; UPenn Vector Core) was injected at a depth of 400 μm below dura at each of four locations: the vertices of a square 200 μm wide, centered on the target coordinates. The glass micropipette was left in place for 5 min after injection to reduce backflow. A drop of warmed agarose solution (1.2% in ACSF, Type III-A, High EEO, A9793, Sigma-Aldrich) was then applied to the cortical surface.

A two-layer glass window was fabricated by first etching out a 2-mm-diameter circle from #0 thickness glass cover slip, then bonding with UV-activated polymer (61, Norland Optical Adhesive) to a #1 thickness, 3-mm-diameter round glass cover slip (64-0720 CS-3R, Warner Instruments). This glass window was then placed against the cortical surface. While applying light pressure, cyanoacrylate glue was added to the rim to attach the glass to the skull and Metabond. Mice were again given at least 1 week to recover before resuming behavioral training. Imaging experiments began when behavioral performance criterion was reached.

Eight of 11 mice went through this procedure involving two surgeries. For the remaining three mice, the head plate implant, viral injection and window implant procedures were performed in the same surgery before behavioral training.

Behavioral Setup

For head-fixed mouse behavior, we used a training apparatus with two lick ports, thus enabling two alternative choices (Guo et al., 2014). Two metal screws were used to affix the head plate of the mouse onto a stainless steel mount. The mouse was then restrained inside an acrylic tube, which restricted gross body movements but allowed postural adjustments.

The lick ports were fabricated from stainless steel 20-gauge needles, which were positioned at 90° and 270° with respect to the mouse's head orientation, and held in place by a 3D-printed plastic part mounted on a micromanipulator for fine positional adjustment. Water was delivered at the ports by gravity feed and the liquid volume was controlled by pneumatic valves (EV-2-24, Clippard) calibrated with an intravenous dripper to deliver ~ 2 μL per pulse. A battery-operated touch detector circuit signaled when the mouse's tongue contacted a lick port.

Auditory stimuli were played through computer speakers placed directly in front of the animal. The intensity of the auditory stimuli was calibrated to ~ 85 dB peak amplitude.

Water delivery, lick detection and sound presentation were connected to a desktop computer via a data acquisition board (USB-201, Measurement Computing). Presentation software (Neurobehavioral Systems) controlled the entire behavioral system.

Behavioral training was performed inside the closed compartment of an audio-visual cart that was dark and soundproofed with acoustic foam (5692T49, McMaster-Carr). An infrared webcam was used to monitor the animal while in the rig. For imaging, mice were tested using a replica of the behavioral training setup under a two-photon microscope.

Adaptive Decision-Making Task

To motivate participation in the task, water consumption was restricted to behavioral sessions. Mice were trained for 1 session per day, 6 d per week. On the non-training day, water was provided *ad libitum* in the home cage for 15 min.

The mice were trained through four phases to shape their behavior. Phase one ($\sim$ 2 d): mice were habituated to head fixation in the behavior box and trained to lick either one of the two ports for water reward. Mice were advanced to the next phase when they made $>$ 100 responses in a session.

Phase two ($\sim$ 2 d): mice were trained to sample both ports. Here mice were required to lick the left port to obtain water rewards three times, followed by the right port for the next three rewards, and so on. Mice were advanced to the next phase when they made $>$ 100 correct responses in a session.

Phase three ($>$ 15 d): animals underwent training for two-choice auditory discrimination. One of two auditory cues was presented to begin each trial—a 2-s-long train of 0.5-s-long logarithmic frequency modulated sweeps from 5 to 15 kHz ('upsweep') or from 15 to 5 kHz ('downsweep'). These stimuli were interleaved randomly from trial to trial. At 0.5 s following the onset of the auditory cue, a response window opened, lasting for a duration of 1.5 s. The first lick within this response window was registered as its response for the trial. All other licks were logged but had no consequences. Once a response was recorded, playback of the auditory cue was terminated.

A correct response, i.e., a left lick for upsweep or a right lick for downsweep, resulted in immediate delivery of 2 µL of water from the corresponding port. The next trial would begin 7 s following response. Incorrect responses resulted in 2 s of white noise presentation, with the next trial beginning 5 s later. Thus, each trial had a total duration within a range from 7.5 to 9 s. Animals were allowed to perform trials until satiated (20 consecutive misses), typically after $\sim$ 60 min.

Training continued daily until a correct rate of $>$ 90% was attained for 3 consecutive days. For imaging experiments, mice were then trained under the two-photon microscope (with laser turned off) for habituation to the recording setup. All mice were able to discriminate at $>$ 90% correct rate after 1–3 d of re-training.

Finally, mice were tested on the adaptive decision-making task. The task always began with a sound block (S) indistinguishable from the two-choice auditory discrimination task. However, once the mouse reached a performance criterion of $>$ 85% correct for the last 20 trials, the stimulus–response-outcome contingencies changed from being sound- to action-guided. In action-guided trials, task structure

was identical to sound-guided trials. However, the correct response became fixed to one response direction, for example, always left, regardless of the stimulus identity. No cue signaled the change in contingencies.

When the mouse reached performance criterion again, another block switch occurred. A sound block was always followed by an action block and vice versa. The second block was randomly chosen for each experiment to be action-left (AL) or action-right (AR). However, once the first action block was chosen, the block sequence became fixed for the remainder of the session. Therefore, the sequence of blocks could be one of two possibilities: (S-AL-S-AR-S- . . .) or (S-AR-S-AL-S . . .).

Each session was terminated after 20 consecutive misses (trials with no response). Mice typically performed the adaptive decision-making task for 60–90 min. Following each adaptive decision-making test, mice resumed daily two-choice auditory discrimination until the next recording session, up to a maximum of 7 adaptive decision-making tests.

Two-Photon Calcium Imaging

The two-photon microscope (Movable Objective Microscope, Sutter Instrument) was controlled using ScanImage software51. The excitation source was a Ti:Sapphire femtosecond laser (Chameleon Ultra II, Coherent). Excitation intensity was controlled by a Pockels cell (350-80-LA-02, Conoptics) and focused onto the sample with a 20×, N.A. 0.95 water immersion objective (Olympus). The time-averaged excitation laser intensity was 90–100 mW after the objective.

To image fluorescence transients from GCaMP6s-expressing neurons, excitation wavelength was set at 920 nm, and emission was collected from 475–550 nm with a GaAsP photomultiplier tube. Time-lapse images were acquired at a resolution of 256 × 256 pixels and a frame rate of 3.62 Hz using bidirectional scanning. To synchronize behavior with imaging, a TTL pulse was sent at the beginning of each trial from the data acquisition board of the behavioral system to the imaging system to act as an external trigger for initiating image acquisition.

Neural Inactivation

Mice were implanted with a head plate. The locations of M2 were marked on both hemispheres (AP 1.5 mm, ML −0.5 mm), and then covered with a thin layer of clear Metabond. Mice were then trained as described above, in preparation for the adaptive decision-making test.

On the first day of testing, craniotomies were performed at the marked locations. Using a glass micropipette attached to a microinjection system (Nanoject II, Drummond), aCSF, with or without muscimol (5 mM, 46 nL per hemisphere; cat. #195336, MP Biomedical), was injected at a depth of 400 μm into M2 of both hemispheres. Behavioral testing began 1–3 h following injection.

The same mice were tested after saline and muscimol treatments on consecutive days in a counter-balanced design, with no blinding. The mice were randomized to receive either saline or muscimol first in an alternating manner depending on

the order in which they reached the behavioral performance criterion. Twelve mice were allocated for this experiment; however, one was excluded due to equipment malfunction during testing.

Histology

Following experiments, mice were transcardially perfused with chilled formaldehyde solution (4% in phosphate-buffered saline). The brains were sectioned with a vibratome and imaged with an inverted wide-field fluorescence microscope.

Analysis of Behavioral Data

Timestamps of stimulus presentation, licks and water delivery were logged in a text file by Presentation software (Neurobehavioral Systems, Inc.). Scripts were written in MATLAB to parse the log files.

'Perseverative errors' were defined as incorrect responses that would have been correct according to the contingencies of the last block of trials. For example, during an action-left block, the stimulus–response pairings of upsweep–left lick and downsweep–left lick would be 'correct'. Downsweep–right lick would be a perseverative error, because this stimulus–response pairing would have been correct in the preceding sound-guided block. The remaining possible stimulus–response pairing, upsweep–right lick, would be classified as an 'other error'.

The number of trials performed included all correct and error trials, but excluded 'miss' trials where the mouse failed to lick within the response window. Miss trials typically occurred near the end of the session when the mouse was satiated.

The number of trials to criterion was defined as the number of trials performed in a certain trial block before reaching a performance criterion of 85% correct for the last 20 trials. Therefore, the minimum value of this quantity is 20.

Mean trials to criterion for each session was calculated excluding the first sound block, because contingency switches had not yet begun. Mean blocks per 100 trials, mean perseverative errors per block and mean other errors per block were calculated excluding the last block (i.e., trials after the last block switch). We often compared conditions before and after block switches, which were defined as the 20 trials before or following a block switch.

The first lick time was defined as the time of the first lick after sound onset for each trial, even if this occurred before the start of the response window. The first lick time is thus a sum of the reaction time and movement time. For this measurement, we excluded trials in which the mouse licked within 0.5 s before cue onset, in which case the first lick may represent the continuation of a spontaneous lick bout rather than a reaction to the stimulus.

Analysis of Imaging Data

Time-lapse fluorescence images were first corrected for x–y motion using the TurboReg plug-in (Thevenaz et al., 1998) for ImageJ (Schneider et al., 2012).

We wrote a graphical user interface in MATLAB to select cell bodies as regions of interest (ROIs). Values of pixels within an ROI were averaged within each frame to yield the cellular fluorescence measurement $F_C(t)$. For each cell, we estimated the neuropil signal by approximating the ROI area as a circle to estimate a radius r (Peron et al., 2015), then creating an annulus-shaped neuropil area with inner and outer radii of $2r$ and $3r$. This neuropil area excluded pixels if they were part of the ROI of another cell body. Values of pixels within the annulus-shaped neuropil area were averaged to generate $F_N(t)$. To subtract the neuropil signal, we calculated $F(t) = F_C(t) - \alpha F_N(t)$, where α is a correction factor ranging from 0.2 to 0.6. The value of α was calibrated for each experiment to avoid overcorrection, by making sure that $F(t) > 0$ for each cell.

For each ROI, the fractional change in fluorescence, $\Delta F/F(t)$, was calculated as

$$\frac{\Delta F}{F}(t) = \frac{F(t) - F_0(t)}{F_0(t)}$$

where $F_0(t)$ is the baseline fluorescence as a function of time.

To estimate baseline, we first obtained $F_{image}(t)$, the mean pixel intensity for the entire 256×256 pixel field of view as a function of time. $F_0(t)$ was then calculated as

$$F_0(t) = F^* \times \frac{F_{0,image}(t)}{F^*_{0,image}}$$

where $F_{0,image}(t)$ is the 10th percentile of $F_{image}(t)$ within a sliding window of 10 min duration. F^* and $F^*_{0,image}$ are the 10th percentile of $F(t)$ and $F_{0,image}(t)$ within the first 10 min of the session, respectively. We verified that $F_{0,image}(t)/F^*_{0,image}$ does not vary with specific choices or rule blocks, and thus serves the purpose of compensating for slow, full-field signal drifts due to non-physiological sources.

We repeated the ensemble analyses with two other methods for calculating baseline: firstly, estimating $F_0(t)$ using the 10th percentile of $F(t)$, on a per-cell basis, with a moving window of 10 min duration; and secondly, estimating $F_0(t)$ using the 10th percentile of $F(t)$ from the entire session, i.e., without a moving window. These different ways to estimate baseline led to qualitatively similar results for all the ensemble analyses.

Analysis of Task-Related Activity and Choice Encoding

To calculate trial-averaged fluorescence transients, we created time bins that were 0.5 s wide and assigned each $\Delta F/F(t)$ value at a particular time t to the corresponding time bin relative to the animal's response. The binned $\Delta F/F(t)$ values were averaged to obtain trial-averaged $\Delta F/F(t)$. To estimate the uncertainty of the trial-averaged $\Delta F/F(t)$, a bootstrap analysis was performed by drawing fluorescence transients per trial, with replacement, up to the same number used to construct the trial average. The median and 95% confidence intervals of trial-averaged $\Delta F/F(t)$ were estimated from 1,000 iterations of this bootstrap analysis.

To quantify choice encoding, we performed multiple linear regression analysis on

the $\Delta F/F(t)$ of each cell using the following equation:

$$\frac{\Delta F}{F}(t) = a_0 + a_1 C_n + a_2 C_{n-1} + a_3 C_n C_{n-1} + a_4 C_{n-2} + \epsilon(t),$$

where C_n is the choice made in the current trial, C_{n-1} is the choice from the prior trial, C_{n-2} is the choice from two trials ago, $\epsilon(t)$ is the error term, and $a_0 \ldots a_n$ are the regression coefficients. Left choices were coded as 1 and right as -1.

A cell was deemed to encode one of the choice parameters or their interaction if $P < 0.01$ for the corresponding regression coefficient. To avoid confounds from rule and reward signals, we analyzed only sound-guided trials in which the outcomes of the current trial and prior trial were both reward. We did not analyze action trials because parameters such as C_n and C_{n-1} were highly correlated by virtue of the task structure—thus breaking a primary assumption underlying this type of analysis.

Analysis of Neural Ensemble Trajectories

For state-space analyses, we used demixed principal component analysis (Machens et al., 2010) (dPCA). To prepare the imaging data for dPCA, we aligned $\Delta F/F$ traces from the first six seconds of each trial following the response. This alignment produced an array with dimensions ($cells \times time \times trials$). We then averaged across four trial types: $C_n = 1$ for pre-switch sound trials; $C_n = -1$ for pre-switch sound trials; $C_n = 1$ for pre-switch action trials; and $C_n = -1$ for pre-switch action trials—in all cases using only rewarded trials. This trial-averaged array ($cells \times time \times 4$) was input into the dPCA algorithm to demix time- and task-dependent variances and obtain principal components (PCs).

To calculate neuronal trajectories, single-trial or trial-averaged $\Delta F/F$ traces were projected onto the first three PCs. To characterize similarities between trajectories across blocks, we calculated the trajectory for each block using the trial-averaged fluorescence across the 20 trials prior to the rule switch.

The similarity between a pair of trajectories was quantified by calculating the mean Euclidean distance between trajectories at matching time points in state space. For pooled comparisons across experiments, the Euclidean distances were normalized by the dispersion population vectors from the corresponding experiment, calculated as the root mean square of the distances between all population vectors and the centroid of the vectors.

To quantify how neural ensemble trajectories evolved on a trial-to-trial basis, we used the Mahalanobis distance, which is a measure of distance between one point and another collection of points. We defined the origin and destination as the population activity from the 20 trials preceding the current rule switch and the next rule switch, respectively. We were interested in the relative separation between the origin, an individual trial that occurred in between, and the destination. Therefore, for each time point within a trial, we calculated Mahalanobis distances, $d_{origin}(t)$ and $d_{dest}(t)$, from the corresponding population activity vector (one three-dimensional value) to those of the origin and destination, respectively (20 three-dimensional values each, estimated as the median across time for each trial). For each individual trial n,

$d_{origin}(n)$ and $d_{dest}(n)$ were then estimated as the median distances for the ~ 30 time points within a trial.

The location of an individual trial relative to the origin and destination was estimated as the ratio of Mahalanobis distances, $d_{origin}(n)/(d_{origin}(n) + d_{dest}(n))$. To summarize the relationship between this distance ratio and the number of trials from the rule switch, we then fit a logistic function,

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}} + L_{min}$$

where x_0 is the midpoint trial, k is the steepness, L is the range, and L_{min} is the minimum value. The parameter L_{min} was not fitted, but rather was estimated for each transition by calculating the mean Mahalanobis distance ratio using the five trials prior to each rule switch.

We fit every neural ensemble transition using this method, but excluded those in which the midpoint trial $x_0 < -5$ or $x_0 > 200$, indicating a poor fit. Based on this criterion, we excluded none (0/33) of the action-to-sound shifts and 8% (3/38) of the sound-to-action shifts in our analysis of M2 neural ensembles. For analysis of the ALM data set, we excluded 8% (2/26) of the action-to-sound shifts and 3% (1/32) of the sound-to-action shifts.

When comparing behavioral and neural transitions, we defined 'behavioral transition trial' as the number of trials to criterion (85% correct for 20 consecutive trials) minus 20, i.e. the first in a series of 20 trials preceding the rule switch. The 'neural transition trial' was defined as the trial when the first term of the logistic fit reached a value of 75% L. That is, the trial x that satisfies this equation:

$$0.75L = \frac{L}{1 + e^{-k(x-x_0)}}$$

This definition was arbitrary because it was unknown how much the population activity pattern must resemble the final pre-switch ensemble state in order to qualify as a transition. Therefore, in another analysis we first fit each neural transition with the logistic function and identified the behavioral trial corresponding to each 5% L step of neural transition from 10 to 90% L. We then calculated the mean hit and error rates at the corresponding behavioral trials. The analysis provided a second description of the relationship between behavioral performance and neural transition without explicitly defining a transition trial.

Neural Ensemble Decoding

To determine how accurately trial type could be predicted from the ensemble activity, we gathered imaging frames that occurred between 0 to 6 s from time of response out of the frame-by-frame imaging data (i.e., $\Delta F/F(t)$). We then projected these $\Delta F/F(t)$ values onto the PCs determined from dPCA to obtain population activity vectors. This procedure reduced the dimensionality of our data from $(frames \times cells)$ to $(frames \times 3)$.

Population activity vectors in this analysis were sampled from the last 20 trials

of completed sound, action-left, or action-right blocks, and restricted to rewarded trials. Other trial types were not considered for the decoding analysis. We trained classifiers on a randomly chosen fraction (80%) of the population activity vectors. Classifiers were based on linear discriminant analysis, using Mahalanobis distances with stratified covariance estimates (the 'classify' function in MATLAB with 'Mahalanobis' option). We then tested the performance of each classifier on the remaining 20% of the population activity vectors, comparing classification results with actual trial types. This five-fold cross-validation process was repeated 1,000 times to obtain a median estimate of classifier accuracy.

To examine decoding accuracy as a function of time, we constructed a new set of population activity vectors from the population activity within each 280 ms time-bin following the first response in each trial. This duration is the inverse of frame rate, which was 3.6 Hz. We then measured decoding accuracy separately for each time-bin. To decode from single-cell activity, $\Delta F/F(t)$ of each cell was used instead of population activity vectors to train or test the classifier.

Statistics

Statistical tests were performed in MATLAB and are indicated in the main text or figure legends. Unless otherwise noted, a Wilcoxon signed-rank test was used for all paired comparisons. For two-sample, unpaired comparisons, a Wilcoxon rank-sum test was used. Paired t-tests were used for bin-wise analysis of lick rates. For quantification of choice signals as a function of time, multiple linear regression was first performed as detailed above; a binomial test was then applied to the proportion of cells significantly encoding choice within each time-bin. For ensemble decoding analyses, mean classification accuracy was tested against chance level using a one-sample t-test. For t-tests, the sampling distribution of the mean was assumed to be normal, but this was not formally tested. All t-tests were two-tailed.

3.3 Results

3.3.1 An Adaptive Decision-Making Task for Head-Fixed Mice

We trained head-fixed mice to perform a task requiring flexible sensorimotor mapping. In each trial, fluid-restricted mice were presented with one of two randomized auditory stimuli, either logarithmic frequency-modulated sweeps from 5 to 15 kHz ('upsweep') or from 15 to 5 kHz ('downsweep'), and had to respond with a lick to the left or right port (FIGREF 1a). A correct response was rewarded with 2 μL of water, while an incorrect response resulted in white noise. Trials were organized into blocks (Fig. 1b), each with a distinct set of stimulus–response contingencies: 'sound-guided' (upsweep-left; downsweep-right), 'action-left' (upsweep-left; downsweep-left), and 'action-right' (upsweep-right; downsweep-right). When performance reached a criterion of 85% correct over 20 trials, a new block began with different contingencies. Sound and action blocks alternated, and no contextual cue was given to signal the

block transition. Therefore, performance beyond the first block required flexible response selection and outcome monitoring. Mice were prepared for this task by initial training to an expert level on two-choice auditory discrimination; i.e., $\sim$ 30 d on a task with only sound-guided trials. Here, we present data from mice with fewer than 6 sessions of experience in the adaptive decision-making task.

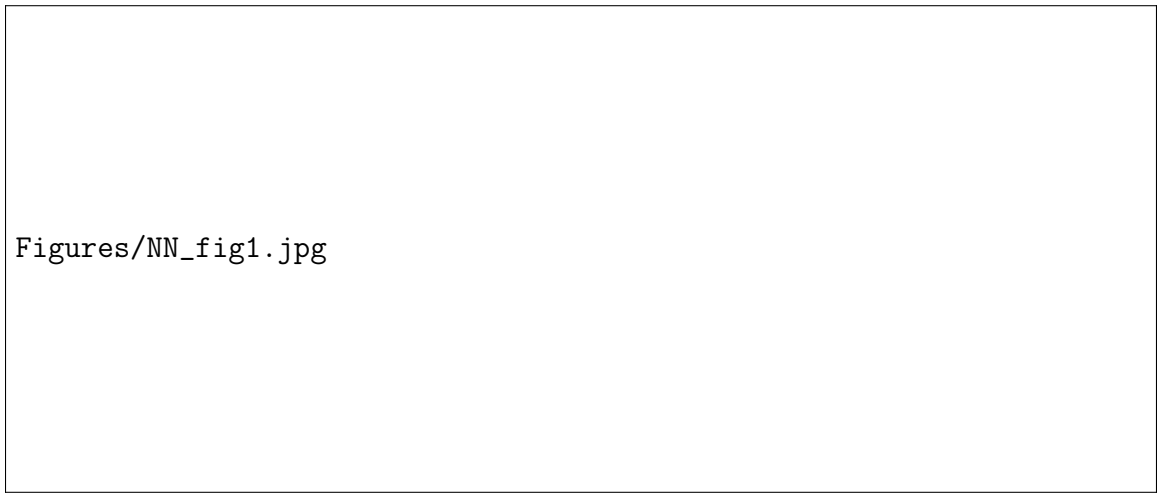


Figure 3.1: Behavioral performance of head-fixed mice in an adaptive sensorimotor decision-making task. (A) Schematic of experiment. Each trial begins with an auditory cue. A response window starts 0.5 s after cue onset, during which the first lick is recorded as the response for that trial. Water reward is delivered contingent on a correct response. ITI, intertrial interval. (B) Schematic of stimulus–response mappings and block design. (C) Behavioral performance surrounding a block switch from action to sound (left) or sound to action (right). Filled circles, hit rate. Open circles, perseverative error rate. Dotted line, other error rate. $N = 33$ action-to-sound switches and 38 sound-to-action switches. (D) Performance from one example behavioral session. Trial outcomes: correct (filled circles), perseverative error (open circles), other error (open triangles), or miss (cross). Vertical line, rule switch. (E) Left and right lick rates for upsweep or downsweep sound cues during all correct sound-guided (black), action-left (red) and action-right (blue) trials. For each choice direction, lick rates during action trials were compared to sound trials in 0.1 s bins; black bars, significant differences ($p < 0.01$, paired t-test). All data presented as $mean \pm SEM$. $N = 9$ sessions from 5 mice.

As expected, a switch in contingencies was associated with an immediate drop in correct response rate (FIGREF 1c,d). Most incorrect responses were perseverative errors, indicating a failure to update response strategy for $\sim$ 20 trials after the switch. We obtained concurrent calcium imaging and behavioral data during 9 sessions from 5 mice (REF Table 1). On average, these mice performed 418 ± 49 trials per session, including 296 ± 38 rewarded trials and 9 ± 1 block switches ($mean \pm SEM$; range: 6–19 switches; REF Supplementary Fig. 1a).

To quantify motor output, we calculated the mean lick rates and the time of first lick for different trial types. Overall, licks were tightly locked to the time of auditory cue during correct trials (REF Fig. 1e). For congruent trials (in which stimulus–

response contingencies match), lick rates were indistinguishable across sound and action blocks. For incongruent trials (for example, left action for upsweep during sound block versus downsweep during action-left block), there was a noticeable difference in mean lick rates and an increased latency to first lick (REF Supplementary Fig. 2). Nevertheless, the major determinant for the shape of the lick distribution was response direction: i.e., whether the animal chose left or right (REF Fig. 1e).

Additionally, we used video tracking to monitor whisker and hindpaw positions, and found that their movements also depended mostly on response direction (REF Supplementary Fig. 3). Therefore, although licks were the means for making operant responses in this head-fixed setup, mice performed more complex motor programs to indicate their choices.

3.3.2 Silencing M2 Selectively Impairs Shift to Sound-Guided Actions

To determine whether frontal cortical activity is necessary for adaptive decision-making in our task, we used the $GABA_A$ receptor agonist muscimol to inactivate M2 bilaterally. Muscimol (5 mM, 46 nL per hemisphere) or saline vehicle was injected $\sim$ 1 h before behavioral testing ($n = 11$ mice; REF Fig. 2a and Supplementary Table 1). We injected low-molecular-weight fluorescein to estimate the extent of the affected region, which included M2 and part of cingulate cortex (Cg1), but not other neighboring regions (REF Supplementary Fig. 4).

Compared to controls, muscimol-injected mice performed fewer trials (Fig. 2b; saline: 608 ± 42, muscimol: 476 ± 31, $mean \pm SEM$; $p = 0.007$, $W = 62$, Wilcoxon signed-rank test), although there was no difference in the number of switches per 100 trials (saline: 2.7 ± 0.2, muscimol: 2.7 ± 0.1, $mean \pm SEM$; $p = 0.96$, $W = 34$). Inactivation had no effect on the timing or rates of lick motor output (REF Supplementary Fig. 5).

Notably, separate analyses of sound and action blocks revealed selective impairments in the animals' ability to engage sound-guided actions, evidenced by a marked (55%) increase in the number of perseverative errors per block (Fig. 2c; saline: 5.7 ± 1.1, muscimol: 8.9 ± 1.9, $mean \pm SEM$; $p = 0.042$, $W = 10$, Wilcoxon signed-rank test), and a greater number of trials to reach criterion (saline: 38 ± 4, muscimol: 48 ± 6, $mean \pm SEM$; $p = 0.042$, $W = 10$). Nevertheless, muscimol-injected mice eventually reached the criterion of $\geq 85\%$ correct, indicating that the transition to a high level of performance was slowed, but not blocked, by M2 inactivation.

Silencing had the opposite effect on shifts into action blocks, during which the mice required fewer trials to reach criterion, although this effect fell short of statistical significance (saline: 43 ± 4, muscimol: 32 ± 2, $mean \pm SEM$; $p = 0.054$, $W = 55$).

These results indicate a causal role for M2 in the flexible control of action selection. Additionally, the opposing effects of silencing are useful for understanding how the mouse performs the outlined task. One solution to the task would be to forget and relearn the relevant stimulus–response associations after each contingency change, as in a reversal task33. This approach predicts symmetric changes in behavior following

perturbations. An alternative approach would be to rely on these associations for sound-guided trials and then ignore them during action blocks to favor repeated selection of the same response. In this case, the mouse would perform the task by shifting the balance between conditional and unconditional means of responding. The asymmetric deficits observed in our experiments are consistent with the second approach and implicate M2 in the breaking of repetitive actions and biasing choices based on learned associations.

3.3.3 Imaging Task-Related Activity at Cellular Resolution in M2

To characterize neural activity, we injected adeno-associated viruses encoding GCaMP6s into layer 2/3 of M2 (AAV1-Syn-GCaMP6s-WPRE-SV40; Fig. 3a). GCaMP6s is a genetically encoded calcium indicator that exhibits a $\sim 25\%$ rise in fluorescence intensity per action potential in cortical pyramidal neurons34. While mice performed the adaptive decision-making task, we used two-photon microscopy to record from 62 ± 6 cells per field of view ($mean \pm SEM$; range, 26–83 cells; $N = 9$ sessions from 5 mice; Fig. 3b).

Figure 3c shows four example M2 neurons with fluorescence transients ($\Delta F/F$) concurrent with responses during sound-guided trials. To examine how the use of conditional rules affects the activity of individual neurons, we averaged $\Delta F/F$ across correct trials for the congruent upsweep–left and downsweep–right conditions separately for sound and action blocks. Neural responses were diverse, even for neurons within the same field of view (Fig. 3d). During sound-guided trials, neurons could exhibit higher $\Delta F/F$ for specific associations—i.e., upsweep–left (cell 2) or downsweep–right (cells 1 and 3)—or have no preference (cell 4). The use of conditional rules modulated $\Delta F/F$ in some neurons (cells 1, 2 and 3) and in other cases had no effect (cell 4).

3.3.4 Neural Transition is More Rapid during Shift to Sound Rule

The observed heterogeneity of neural responses opened the question of whether single-neuron activity in M2 reflects the components of an ensemble representation for specific task variables. If so, then population-level analyses might more effectively capture the content of such representations. Toward this end, we calculated population activity vectors from $\Delta F/F$ and used demixed principal component analysis (dPCA)35,36 to project the vectors in a reduced representational space (see Methods). Plotting these vectors over time generates trajectories describing the time-dependent evolution of ensemble activity during behavior.

To determine how the ensemble activity evolved around block switches on a trial-by-trial basis, we calculated the Mahalanobis distances between population activity vectors of each trial and those of the 20 trials before the last or next block switch. Following a contingency switch, we found that the ensemble activity migrated away

from the previous representational subspace toward a new subspace associated with the new rule (REF Fig. 4a).

Comparisons of the transition dynamics following a switch into conditional versus unconditional rules uncovered marked differences. Out of 33 action-to-sound and 38 sound-to-action transitions, 33 and 35 switches, respectively, could be fit with a logistic function to compare the onset and rate of shifts in population activity patterns (Fig. 4b,c). State transitions associated with the shift to sound-guided responses occurred after only several trials, much earlier than with shifts into repeated actions (REF Fig. 4d,g; sound: midpoint trial, $x_0 = 4.0$, action: $x_0 = 10.4$, median; $p = 0.007$, $z = 2.70$, Wilcoxon rank-sum test). Furthermore, breaking from repetitive to sound-guided responding involved transitions that were more abrupt (sound: steepness, $k = 1.02$, action: $k = 0.35$, median; $p = 0.03$, $z = -2.17$, Wilcoxon rank-sum test). These differences in neural dynamics were not due to behavioral differences, because in this set of experiments trials to criterion were similar for the two rule types (sound: 39, action: 38, median; $p = 0.9$, $z = 0.09$, Wilcoxon rank-sum test; REF Supplementary Fig. 1a).

Overall, these results suggest that ensemble activity patterns in M2 shift earlier and more steeply when animals are required to abort repetitive actions and engage conditional associations to perform sound-guided behavior.

To what extent must population activity resemble the final ensemble state in order to improve behavior? To address this question, we performed two analyses to compare the timing of neural and behavioral transitions. In the first analysis, we defined 'transition trials' for behavior (trials to criterion minus 20, the sliding window for assessing criterion) and neural ensemble activity (Mahalanobis distance ratio equaling 75% L based on logistic fit; see Methods). Block-by-block paired comparisons of neural and behavioral transition trials showed that ensemble activity in M2 shifted before the recovery of behavioral performance when adapting to conditional rules (Fig. 4e and Supplementary Fig. 6; $p = 0.003$, $z = -2.96$; Wilcoxon signed-rank test). By contrast, neural and behavioral changes occurred at around the same time for shifts to unconditional responding (Fig. 4h; $p = 0.19$, $z = 1.32$; Wilcoxon signed-rank test).

We should note, however, that the definitions used for transition trials were arbitrary. Therefore, we performed a second, less biased analysis in which we determined the mean performance at the behavioral trial corresponding to a series of different neural transition locations. Compared with shifts to action trials (Fig. 4i), transitions to sound-guided trials were associated with hit and error rates that diverged later (Fig. 4f), indicating that behavioral improvement occurred later along the time course of neural transitions.

Taken together, results of these two analyses suggest that when shifting to sound-guided actions, neural ensemble transitions in M2 are nearly complete before behavioral performance improvement can be detected.

3.3.5 Distinct Activity Patterns Accompany Rule Implementations

Our results indicated that rule shifts are associated with distinct transitions in network activity. This leads naturally to the question of what ensemble dynamics accompany successful rule implementation. We examined trajectories associated with correct responses in the 20 trials before the switch, when response strategies had stabilized ($\geq 85\%$ correct by task design).

Figure 5a shows the trajectories of a 56-cell ensemble for left and right responses during sound-guided trials. The trajectories were initially indistinguishable and then diverged sharply after the animal made a response. Expanding this analysis to include action blocks revealed population activity patterns that occupied additional, distinct subspaces within the same representational space (Fig. 5b).

To quantify rule representations present in the population code, we asked how accurately block type could be predicted from individual population activity vectors. For each session, we constructed a classifier based on linear discriminant analysis (see Methods). Testing each classifier with five-fold cross-validation revealed that in all cases trial type could be decoded well above chance (Fig. 5c; sound: $78 \pm 3\%$, action-left: $86 \pm 4\%$, action-right: $82 \pm 3\%$; versus chance level of 33%, $p = 1 \times 10^{-6}$, 1×10^{-6}, 5×10^{-7}; $t(8) = 13.2$, 12.8, 14.3; one-sample t-test; $N = 9$ sessions).

Repetition of this analysis using a moving window yielded high decoding accuracy at all times during a trial (Fig. 5d), consistent with a global shift in engagement of the network rather than a simple change in the processing of cue, action or outcome related signals.

Next we asked whether accuracy of the ensemble classifier could have been driven by a few highly rule-selective cells. When classifiers were trained on $\Delta F/F$ of individual cells, we found that 27% of the cells could be used to decode block types at rates above chance; however, accuracy fell along a continuum and at levels below the accuracy of the ensemble (Fig. 5e).

To ensure that the differences in trajectories and decoding accuracies were not due to simple sensory or motor parameters, we computed trajectories with matched stimulus, prior choice, current choice and outcome conditions. Analyses of these congruent trials, which differed only by rule, yielded similar results (Fig. 5f–i and Supplementary Fig. 7a–d).

Taken together, these results indicate that the behavioral implementation of specific conditional and unconditional rules is associated with distinct network activity patterns in M2, such that population activity from any time during behavior can be used to decode task contingencies with high accuracy.

3.3.6 Activity Toggles between Rule-Related Patterns

When animals solve trials with the same contingencies a second time, do M2 ensembles revisit similar activity patterns or does population activity migrate to a previously uncharted region of state space? Our task was well suited to address this question because blocks of the same trial type were presented multiple times within the same

behavioral session.

Figure 6a shows an example set of neural circuit trajectories for the first 12 trial blocks within one behavioral session, in which trajectories could be clearly grouped by block type and not by their temporal order.

To quantify the representational similarity of ensemble dynamics on a block-by-block basis, we calculated the mean Euclidean distances between all possible pairwise comparisons of trajectories within an experiment (see Methods). We found that neural circuit trajectories from blocks of the same type had a relatively small distance of separation and were similarly compact (Fig. 6b and Supplementary Fig. 7e,f; for sound (S), action-left (AL) and action-right (AR): S–S versus AL–AL: $p = 0.6$, $W = 3$; S–S versus AR–AR, $p = 0.5$, $W = 13$; Wilcoxon signed-rank test). By contrast, trajectories from different block types were represented by markedly different ensemble activity (S–S versus S–AL, $p = 0.004$, $W = 0$; S–S versus S–AR, $p = 0.004$, $W = 0$; S–S versus AL–AR, $p = 0.004$, $W = 0$; corrected $\alpha = 0.01$, Wilcoxon signed-rank test with Bonferroni correction).

These results indicate that, during adaptive decision-making, M2 toggles between distinct functional configurations as the animal repeatedly engages corresponding changes in task demands.

3.3.7 Comparison of Task-Related Neural Dynamics in M2, ALM and V1

Next, we sought to determine whether the observed neural dynamics are specific to M2 or may also be found in other brain regions. For this purpose, we imaged neural ensembles in layer 2/3 of anterior lateral motor cortex (ALM; 65 ± 6 cells per field of view, $mean \pm SEM$; $N = 8$ sessions from 4 mice; Supplementary Fig. 1b) and primary visual cortex (V1; 57 ± 7 cells per field of view; $N = 4$ sessions from 2 mice; Supplementary Fig. 1c) to compare with the data from M2 (62 ± 6 cells per field of view; $N = 9$ sessions from 5 mice). ALM has been implicated in motor planning and execution37,38; however, it is ~ 1.5 mm distant from M2, and the relationship between the two frontal cortical regions is not understood. V1 was chosen as a control region because the task was performed in the dark and involved no visual stimulus.

Multiple linear regression analysis showed that M2 neurons robustly encoded not only the choice of the current trial, but also the choices of the two prior trials (Fig. 7a). A higher proportion of cells in ALM encoded the current choice; however, the signals decayed faster, resulting in weaker encoding of prior choices (Fig. 7b). Activity of M2 and ALM neurons could prefer either the ipsilateral or contralateral direction (Supplementary Fig. 8a,b), consistent with prior studies30,38. Unexpectedly, choice signals were also observed in V1 (Fig. 7c). Choice selectivity in V1 was relatively weak, and $\Delta F/F$ was almost always higher when animals made an ipsilateral choice (REF Supplementary Fig. 8c). Because choice signals in V1 were transient and animals performed the task in the dark, we conjecture that the selectivity might relate to corollary discharge.

To investigate ensemble activity, we employed the same dPCA and linear classifier

analyses used for M2. We found that rule type could be decoded with high accuracy using ensemble activity from ALM (matched sound, action-left trials: $78 \pm 4\%$, $t(7) = 6.94$; $p = 2 \times 10^{-4}$; matched sound, action-right trials: $78 \pm 3\%$, $t(7) = 8.68$; $p = 5 \times 10^{-5}$; versus chance level of 50%, one-sample t-test; REF Fig. 7d), but at a much worse rate for V1 (matched sound, action-left trials: $58 \pm 5\%$, $t(3) = 1.88$; $p = 0.2$; matched sound, action-right trials: $67 \pm 4\%$, $t(3) = 3.76$; $p = 0.03$). Therefore, both ALM and M2 exhibited task-specific ensemble activity patterns.

However, unlike what we found in M2, characterization of ensemble transitions in ALM did not reveal significant differences between switches to sound versus action blocks (sound: $x_0 = 8.2$, action: $x_0 = 10.2$, median; $z = 1.62$, $p = 0.11$; sound: $k = 0.37$, action: $k = 0.56$, median; $z = 0.89$, $p = 0.4$; Wilcoxon rank-sum test; REF Fig. 7e). There were also no detectable timing differences between neural and behavioral transitions in ALM (sound: $p = 0.5$, $z = 0.61$; action: $p = 0.13$, $z = -1.50$; Wilcoxon signed-rank test; neural transition defined as 75% L).

Taken together, these data indicate regionally specific ensemble dynamics associated with adaptive behavior.

3.4 Discussion

The results of our study support two novel insights regarding the function of higher-order motor cortex in adaptive choice behavior. Firstly, fast and slow ensemble transitions are neural signatures for distinct phases of voluntary behavior. A comparison between transitions was possible because our task design allowed for multiple shifts between multiple contingencies within a single behavioral session. Secondly, the relative timing of neural and behavioral shifts, as well as the specific deficits following inactivation, highlighted a leading role for this region in the engagement of sensory cue-guided actions (Fig. 3.2).

This conclusion contrasts with previous studies of homologous or nearby prefrontal cortical regions, in which neural changes closely match or lag the time course of behavioral adaptation (Mitz et al., 1991; Pasupathy and Miller, 2005; Durstewitz et al., 2010). One explanation may be that prior studies have focused on the learning of novel sensorimotor mappings or new rules, whereas our task required animals to repeatedly disengage and re-engage the learned associations needed in sound-guided trials.

Although this task shares important features with other assays for flexibility, there are also crucial differences. In contrast to paradigms that use a contextual cue to instruct rapid executive control on a trial-by-trial basis (Mante et al., 2013; Stokes et al., 2013; Duan et al., 2015), animals in our experiment adapted on a time scale of tens of trials (Fig. 1c). This relatively slow rate of adaptation resembles that of learning during arbitrary visuomotor mapping, where the animal's basis for action selection is updated gradually based on reward feedback (Pasupathy and Miller, 2005; Asaad et al., 1998).

Our task also differs from other strategy- or set-shifting tasks for rodents (Durstewitz et al., 2010; Darrah et al., 2008), in the sense that we used nonspatial stimuli that

Figures/NN_figS9.jpg

Figure 3.2: Summary of the task, behavioral performance, and neural dynamics. Recovery of behavioral performance following a rule switch (middle) was approximated as an exponential function for this schematic. Dashed lines reflect the median number of trials to a behavioral transition (sound: 19.0, action: 17.5) or neural transition (bottom; sound: 5.1, action: 14.6) relative to the rule switch. The plot of M2 ensemble activity state (bottom) was based on the midpoint trial (sound: 4.0, action: 10.4), steepness (sound: 1.02, action: 0.35), and range (sound: 0.36, action: 0.38) parameters estimated from logistic fits of the neural transition data (see **??**).

do not conform to classical definitions of exemplars or sets. Instead, the paradigm we used consisted of blocks of trials that required the subject to shift between conditional and non-conditional approaches to action selection.

Analysis of the types of errors made during training suggests that mice perform two-choice auditory discrimination in part by suppressing a prepotent tendency to repeat a rewarded choice. Action trials could thus be considered a natural strategy to the animal, whereas sound-guided trials require weeks of training to achieve high performance. One caveat for our task is that animals are likely to have different degrees of learned and intrinsic familiarity for sound versus action trials. In principle, mice may solve the task by ignoring sensory information completely during action blocks. However, the temporally structured lick rates during action blocks (Fig. 1e) strongly suggest use of the stimulus for gating lick responses.

We found that bilateral inactivation of M2 selectively impaired the shift into sound-guided actions. This observation is highly consistent with results of dorsal premotor lesions in primates, which disrupt both the learning of novel visuomotor associations and the engagement of previously learned mappings (Petrides, 1985; Halsband and Passingham, 1985; Nixon et al., 2004). Adaptation to action blocks was facilitated by M2 inactivation. This effect could result from a tendency to repeat the prior choice (Sul et al., 2011): if M2 normally biases animals toward sensory-cue-guided actions, then inactivation may remove an important brake on the unconditional strategy. In our experiments, M2 inactivation slowed but did not preclude the eventual transition to high performance on sound-guided trials. This suggests that, at least for trained mice, two-choice auditory discrimination alone does not require M2 and may be subserved by other circuits (Znamenskiy and Zador, 2013). Furthermore, we take the opposing effects of inactivation on shifts to sound-guided versus repeated actions as evidence that mice perform the task by balancing the use of conditional and unconditional responses.

A key finding of this study concerns how specific parameters of ensemble activity transitions may relate to behavior. We found that ensemble transitions were more abrupt when animals needed to retrieve and begin using conditional associations. These fast transitions may be related to those observed in medial prefrontal cortex, which have been interpreted as neural correlates of insight (Durstewitz et al., 2010) or as the abandonment of an inadequate internal model at the onset of exploration (Karlsson et al., 2012).

On what quantitative basis should transitions be classified as abrupt or gradual? We compared the steepness of these transitions directly to the slower transitions that accompanied adaptation to action blocks. Moreover, ensemble transitions occurred after only a few errors, whereas behavioral improvements took tens of sound-guided trials (Fig. 4e,f). The difference in neural and behavioral timing suggests that M2 neural activity had mostly adjusted while the animal was still systematically responding in an unconditional manner. M2 may facilitate the engagement of sound-guided behavior by biasing the use of sensory information, suppressing repetitive actions, or both. By contrast, prior studies show that when an animal must acquire novel arbitrary associations, changes in cortical activity track behavioral improvements (Mitz et al., 1991; Brasted and Wise, 2004) and lag the more rapid remapping in the stria-

tum (Pasupathy and Miller, 2005). A major difference between these studies of fast learning and our study is that the auditory–motor associations were already well learned in our task.

We found that multiple rules were each associated with a distinct subset of population activity patterns. Such task-dependent changes in neural activity are reported in multiple frontal cortical regions across species (Asaad et al., 2000; Rich and Shapiro, 2009; Rodgers and DeWeese, 2014; Durstewitz et al., 2010; Wallis et al., 2001). By asking the animal to shift repeatedly during a single session, we found that the network could return to a previously employed functional configuration to meet similar behavioral demands. This back-and-forth toggling of ensemble activity is reminiscent of the ensemble remapping observed in CA1 of the hippocampus during repeated exposure to spatial contexts (Wills et al., 2005; Leutgeb et al., 2005). One study reported that changes in environmental context also cause network activity shifts in the rodent medial prefrontal cortex. However, the ensemble code was not identical upon re-exposure, potentially owing to a systematic drift over time (Hyman et al., 2012a). The divergent findings of repeatable versus drifting network states in the rodent frontal cortex could reflect regional differences or differences in how frontal areas encode cognitive versus environmental variables.

Several lines of evidence support the idea that the neural dynamics in M2 reflect changes in internal processes (for example, representation of task contingencies or motor planning and preparation) rather than differences in overt physical movements. Firstly, three different ensemble analyses with matched, congruent trial conditions indicated distinct neural dynamics in sound and action blocks (Fig. 5f–i and Supplementary Fig. 7), despite a lack of observable difference in motor output for the same sets of trials (Fig. 1e and Supplementary Fig. 2). Secondly, neural signals related to motor execution should be strongest at the time of response. Instead, we found that the rule-specific separation of population activity patterns was substantially above chance at all times across a trial (Fig. 5d). Perhaps the strongest evidence is that muscimol inactivation of M2 had no detectable effect on motor output (Supplementary Fig. 5), while clearly affecting behavioral flexibility.

What is the purpose of functional reconfiguration during adaptive decision-making? Ensemble activity patterns within multiple network subspaces reflect the diversity of neural representations in M2. Recent studies indicate functional roles for long-range projections from rodent M2 to sensory cortices (Schneider et al., 2014; Manita et al., 2015) and dorsal striatum (Rothwell et al., 2015). Appropriate shifts in neural representations could allow M2 to exert differential top-down control in a task-dependent manner. Further study regarding the downstream impacts of frontal network transitions may yield insights into neuropsychiatric disorders in which cognitive flexibility is impaired. Plausibly, the cognitive rigidity characteristic of disorders such as schizophrenia could result from an inability of frontal cortical networks to shift or maintain stable ensemble states.

3.5 Supplementary Figures

3.6 Acknowledgments

Chapter 4

Representation of Choices, Outcomes, and Context in Four Cell-Types of the Medial Frontal Cortex

Test of Abstract

Bibliography

Arman Abrahamyan, Laura Luz Silva, Steven C Dakin, Matteo Carandini, and Justin L Gardner. Adaptable history biases in human perceptual decisions. *Proceedings of the National Academy of Sciences*, 113(25):E3548–E3557, 2016.

Evan G Antzoulatos and Earl K Miller. Differences between neural activity in prefrontal cortex and striatum during learning of novel abstract categories. *Neuron*, 71(2):243–249, 2011.

Wael F Asaad, Gregor Rainer, and Earl K Miller. Neural activity in the primate prefrontal cortex during associative learning. *Neuron*, 21(6):1399–1407, 1998.

Wael F Asaad, Gregor Rainer, and Earl K Miller. Task-specific neural activity in the primate prefrontal cortex. *Journal of Neurophysiology*, 84(1):451–459, 2000.

Bruno B Averbeck and Daeyeol Lee. Neural noise and movement-related codes in the macaque supplementary motor area. *Journal of Neuroscience*, 23(20):7630–7641, 2003.

Bruno B Averbeck and Daeyeol Lee. Effects of noise correlations on information encoding and decoding. *Journal of neurophysiology*, 95(6):3633–3644, 2006.

Bruno B Averbeck, Peter E Latham, and Alexandre Pouget. Neural correlations, population coding and computation. *Nature reviews neuroscience*, 7(5):358, 2006.

Dominic J Barraclough, Michelle L Conroy, and Daeyeol Lee. Prefrontal cortex and decision making in a mixed-strategy game. *Nature neuroscience*, 7(4):404, 2004.

Florent Barthas and Alex C Kwan. Secondary motor cortex: where 'sensory'meets 'motor'in the rodent frontal cortex. *Trends in neurosciences*, 40(3):181–193, 2017.

Timothy EJ Behrens, Mark W Woolrich, Mark E Walton, and Matthew FS Rushworth. Learning the value of information in an uncertain world. *Nature neuroscience*, 10(9):1214, 2007.

Alberto Bernacchia, Hyojung Seo, Daeyeol Lee, and Xiao-Jing Wang. A reservoir of time constants for memory traces in cortical neurons. *Nature neuroscience*, 14(3): 366, 2011.

Charlotte A Boettiger and Mark D'Esposito. Frontal networks for learning and executing arbitrary stimulus-response associations. *Journal of Neuroscience*, 25(10): 2723–2732, 2005.

Peter J Brasted and Steven P Wise. Comparison of learning-related neuronal activity in the dorsal premotor cortex and striatum. *European Journal of Neuroscience*, 19 (3):721–740, 2004.

Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

Silvia A Bunge, Jonathan D Wallis, Amanda Parker, Marcel Brass, Eveline A Crone, Eiji Hoshi, and Katsuyuki Sakai. Neural circuitry underlying rule use in humans and nonhuman primates. *Journal of Neuroscience*, 25(45):10347–10350, 2005.

LONGTANG L Chen and STEVEN P Wise. Neuronal activity in the supplementary eye field during acquisition of conditional oculomotor associations. *Journal of Neurophysiology*, 73(3):1101–1121, 1995.

Tsai-Wen Chen, Trevor J Wardill, Yi Sun, Stefan R Pulver, Sabine L Renninger, Amy Baohan, Eric R Schreiter, Rex A Kerr, Michael B Orger, Vivek Jayaraman, et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, 499 (7458):295, 2013.

Tsai-Wen Chen, Nuo Li, Kayvon Daie, and Karel Svoboda. A map of anticipatory activity in mouse motor cortex. *Neuron*, 94(4):866–879, 2017.

Justin M Darrah, Mark R Stefani, and Bita Moghaddam. Interaction of n-methyl-d-aspartate and group 5 metabotropic glutamate receptors on behavioral flexibility using a novel operant set-shift paradigm. *Behavioural pharmacology*, 19(3):225, 2008.

Winfried Denk, James H Strickler, and Watt W Webb. Two-photon laser scanning fluorescence microscopy. *Science*, 248(4951):73–76, 1990.

Christopher H Donahue, Hyojung Seo, and Daeyeol Lee. Cortical signals for rewarded actions and strategic exploration. *Neuron*, 80(1):223–234, 2013.

Chunyu A Duan, Jeffrey C Erlich, and Carlos D Brody. Requirement of prefrontal and midbrain regions for rapid executive control of behavior in the rat. *Neuron*, 86 (6):1491–1503, 2015.

Alexander Dubbs, James Guevara, and Rafael Yuste. moco: Fast motion correction for calcium imaging. *Frontiers in neuroinformatics*, 10:6, 2016.

Daniel Durstewitz, Nicole M Vittoz, Stan B Floresco, and Jeremy K Seamans. Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron*, 66(3):438–448, 2010.

Jeffrey C Erlich, Max Bialek, and Carlos D Brody. A cortical substrate for memory-guided orienting in the rat. *Neuron*, 72(2):330–343, 2011.

Shiva Farashahi, Christopher H Donahue, Peyman Khorsand, Hyojung Seo, Daeyeol Lee, and Alireza Soltani. Metaplasticity as a neural substrate for adaptive learning and choice under uncertainty. *Neuron*, 94(2):401–414, 2017.

Ingo Fründ, Felix A Wichmann, and Jakob H Macke. Quantifying the effect of inter-trial dependence on perceptual decisions. *Journal of vision*, 14(7):9–9, 2014.

Aldo Genovesio, Peter J Brasted, and Steven P Wise. Representation of future and previous spatial goals by separate neural populations in prefrontal cortex. *Journal of Neuroscience*, 26(27):7305–7316, 2006.

Aldo Genovesio, Satoshi Tsujimoto, Giulia Navarra, Rossella Falcone, and Steven P Wise. Autonomous encoding of irrelevant goals and outcomes by prefrontal cortex neurons. *Journal of Neuroscience*, 34(5):1970–1978, 2014.

Christina Gremel and Rui Costa. Premotor cortex is critical for goal-directed actions. *Frontiers in computational neuroscience*, 7:110, 2013.

Kristi R Griffiths, Richard W Morris, and Bernard W Balleine. Translational studies of goal-directed action as a framework for classifying deficits across psychiatric disorders. *Frontiers in systems neuroscience*, 8:101, 2014.

Zengcai V Guo, Nuo Li, Daniel Huber, Eran Ophir, Diego Gutnisky, Jonathan T Ting, Guoping Feng, and Karel Svoboda. Flow of cortical activity underlying a tactile decision in mice. *Neuron*, 81(1):179–194, 2014.

U Halsband and RE Passingham. Premotor cortex and the conditions for movement in monkeys (macaca fascicularis). *Behavioural brain research*, 18(3):269–277, 1985.

Balázs Hangya, Sachin P Ranade, Maja Lorenc, and Adam Kepecs. Central cholinergic neurons are rapidly recruited by reinforcement feedback. *Cell*, 162(5):1155–1168, 2015.

Mark H Histed, Anitha Pasupathy, and Earl K Miller. Learning substrates in the primate prefrontal cortex and striatum: sustained activity related to successful actions. *Neuron*, 63(2):244–253, 2009.

Jonas A Hosp, Ana Pekanovic, Mengia S Rioult-Pedotti, and Andreas R Luft. Dopaminergic projections from midbrain to primary motor cortex mediate motor skill learning. *Journal of Neuroscience*, 31(7):2481–2487, 2011.

Eun Jung Hwang, Jeffrey E Dahlen, Madan Mukundan, and Takaki Komiyama. History-based action selection bias in posterior parietal cortex. *Nature communications*, 8(1):1242, 2017.

James M Hyman, Liya Ma, Emili Balaguer-Ballester, Daniel Durstewitz, and Jeremy K Seamans. Contextual encoding by ensembles of medial prefrontal cortex neurons. *Proceedings of the National Academy of Sciences*, 109(13):5086–5091, 2012a.

James M Hyman, Jennifer Whitman, Eldon Emberly, Todd S Woodward, and Jeremy K Seamans. Action and outcome activity state patterns in the anterior cingulate cortex. *Cerebral Cortex*, 23(6):1257–1268, 2012b.

James Michael Hyman, Clay Brian Holroyd, and Jeremy Keith Seamans. A novel neural prediction error found in anterior cingulate cortex ensembles. *Neuron*, 95 (2):447–456, 2017.

Masaki Isoda and Okihide Hikosaka. Switching from automatic to controlled action by monkey medial frontal cortex. *Nature neuroscience*, 10(2):240–248, 2007.

Tsukasa Kamigaki and Yang Dan. Delay activity of specific prefrontal interneuron subtypes modulates memory-guided behavior. *Nature neuroscience*, 20(6):854, 2017.

William J Kargo, Botond Szatmary, and Douglas A Nitz. Adaptation of prefrontal cortical firing patterns and their fidelity to changes in action–reward contingencies. *Journal of Neuroscience*, 27(13):3548–3559, 2007.

Mattias P Karlsson, Dougal GR Tervo, and Alla Y Karpova. Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. *Science*, 338 (6103):135–139, 2012.

Hoseok Kim, Daeyeol Lee, and Min Whan Jung. Signals for previous goal choice persist in the dorsomedial, but not dorsolateral striatum of rats. *Journal of Neuroscience*, 33(1):52–63, 2013.

D Kvitsiani, S Ranade, B Hangya, H Taniguchi, JZ Huang, and A Kepecs. Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature*, 498(7454):363, 2013.

Christopher C Lapish, Daniel Durstewitz, L Judson Chandler, and Jeremy K Seamans. Successful choice behavior is associated with distinct and coherent network states in anterior cingulate cortex. *Proceedings of the National Academy of Sciences*, 105(33):11963–11968, 2008.

Susan Leemburg, Tara Canonica, and Andreas Luft. Motor skill learning and reward consumption differentially affect vta activation. *Scientific reports*, 8(1):687, 2018.

Stefan Leutgeb, Jill K Leutgeb, Carol A Barnes, Edvard I Moser, Bruce L McNaughton, and May-Britt Moser. Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science*, 309(5734):619–623, 2005.

Nuo Li, Tsai-Wen Chen, Zengcai V Guo, Charles R Gerfen, and Karel Svoboda. A motor cortex circuit for motor planning and movement. *Nature*, 519(7541):51, 2015.

Christian K Machens, Ranulfo Romo, and Carlos D Brody. Functional, but not anatomical, separation of "what" and "when" in prefrontal cortex. *Journal of Neuroscience*, 30(1):350–360, 2010.

Hiroshi Makino, Chi Ren, Haixin Liu, An Na Kim, Neehar Kondapaneni, Xin Liu, Duygu Kuzum, and Takaki Komiyama. Transformation of cortex-wide emergent properties during motor learning. *Neuron*, 94(4):880–890, 2017.

Satoshi Manita, Takayuki Suzuki, Chihiro Homma, Takashi Matsumoto, Maya Odagawa, Kazuyuki Yamada, Keisuke Ota, Chie Matsubara, Ayumu Inutsuka, Masaaki Sato, et al. A top-down cortical circuit for accurate sensory perception. *Neuron*, 86(5):1304–1316, 2015.

Valerio Mante, David Sussillo, Krishna V Shenoy, and William T Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. *nature*, 503 (7474):78–84, 2013.

Andrew R Mitz, Moshe Godschalk, and Steven P Wise. Learning-dependent neuronal activity in the premotor cortex: activity during the acquisition of conditional motor associations. *Journal of Neuroscience*, 11(6):1855–1872, 1991.

Ari S Morcos and Christopher D Harvey. History-dependent variability in population dynamics during evidence accumulation in cortex. *Nature neuroscience*, 19(12): 1672, 2016.

Masayoshi Murakami, M Inês Vicente, Gil M Costa, and Zachary F Mainen. Neural antecedents of self-initiated actions in secondary motor cortex. *Nature neuroscience*, 17(11):1574, 2014.

Elisabeth A Murray, Timothy J Bussey, and Steven P Wise. Role of prefrontal cortex in a network for arbitrary visuomotor mapping. In *Executive Control and the Frontal Lobe: Current Issues*, pages 114–129. Springer, Berlin, Heidelberg, 2000.

Parashkev Nachev, Christopher Kennard, and Masud Husain. Functional role of the supplementary and pre-supplementary motor areas. *Nature Reviews Neuroscience*, 9(11):856–869, 2008.

Nandakumar S Narayanan, Eyal Y Kimchi, and Mark Laubach. Redundancy and synergy of neuronal ensembles in motor cortex. *Journal of Neuroscience*, 25(17): 4207–4216, 2005.

Nandakumar S Narayanan, James F Cavanagh, Michael J Frank, and Mark Laubach. Common medial frontal mechanisms of adaptive control in humans and rodents. *Nature neuroscience*, 16(12):1888, 2013.

Philip D Nixon, Kathryn R McDonald, Patricia M Gough, Iona H Alexander, and Richard E Passingham. Cortico-basal ganglia pathways are essential for the recall of well–established visuomotor associations. *European Journal of Neuroscience*, 20 (11):3165–3178, 2004.

Sean B Ostlund, Neil E Winterbauer, and Bernard W Balleine. Evidence of action sequence chunking in goal-directed instrumental conditioning and its dependence on the dorsomedial prefrontal cortex. *Journal of Neuroscience*, 29(25):8280–8287, 2009.

James M Otis, Vijay MK Namboodiri, Ana M Matan, Elisa S Voets, Emily P Mohorn, Oksana Kosyk, Jenna A McHenry, J Elliott Robinson, Shanna L Resendez, Mark A Rossi, et al. Prefrontal cortex output circuits guide reward seeking through divergent cue encoding. *Nature*, 543(7643):103, 2017.

RE Passingham, C Myers, N Rawlins, V Lightfoot, and S Fearn. Premotor cortex in the rat. *Behavioral neuroscience*, 102(1):101, 1988.

Anitha Pasupathy and Earl K Miller. Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature*, 433(7028):873–876, 2005.

Simon P Peron, Jeremy Freeman, Vijay Iyer, Caiying Guo, and Karel Svoboda. A cellular resolution map of barrel cortex activity during tactile behavior. *Neuron*, 86(3):783–799, 2015.

Michael Petrides. Deficits on conditional associative-learning tasks after frontal-and temporal-lobe lesions in man. *Neuropsychologia*, 23(5):601–614, 1985.

Lucas Pinto and Yang Dan. Cell-type-specific activity in prefrontal cortex during goal-directed behavior. *Neuron*, 87(2):437–450, 2015.

Thomas A Pologruto, Bernardo L Sabatini, and Karel Svoboda. Scanimage: flexible software for operating laser scanning microscopes. *Biomedical engineering online*, 2(1):13, 2003.

Todd M Preuss. Do rats have prefrontal cortex? the rose-woolsey-akert program reconsidered. *Journal of cognitive neuroscience*, 7(1):1–24, 1995.

RL Reep and JV Corwin. Topographic organization of the striatal and thalamic connections of rat medial agranular cortex. *Brain research*, 841(1-2):43–52, 1999.

Erin L Rich and Matthew Shapiro. Rat prefrontal cortical neurons selectively code strategy switches. *Journal of Neuroscience*, 29(22):7208–7219, 2009.

Chris C Rodgers and Michael R DeWeese. Neural correlates of task switching in prefrontal cortex and primary auditory cortex in a novel stimulus selection task for rodents. *Neuron*, 82(5):1157–1170, 2014.

Patrick E Rothwell, Scott J Hayton, Gordon L Sun, Marc V Fuccillo, Byung Kook Lim, and Robert C Malenka. Input-and output-specific regulation of serial order performance by corticostriatal circuits. *Neuron*, 88(2):345–356, 2015.

MFS Rushworth, KA Hadland, T Paus, and PK Sipila. Role of the human medial frontal cortex in task switching: a combined fmri and tms study. *Journal of neurophysiology*, 87(5):2577–2592, 2002.

Jeffrey D Schall, Veit Stuphorn, and Joshua W Brown. Monitoring and control of action by the frontal lobes. *Neuron*, 36(2):309–322, 2002.

Caroline A Schneider, Wayne S Rasband, and Kevin W Eliceiri. Nih image to imagej: 25 years of image analysis. *Nature methods*, 9(7):671, 2012.

David M Schneider, Anders Nelson, and Richard Mooney. A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature*, 513(7517):189–194, 2014.

Wolfram Schultz, Peter Dayan, and P Read Montague. A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599, 1997.

Hyojung Seo, Dominic J Barraclough, and Daeyeol Lee. Dynamic signals related to choices and outcomes in the dorsolateral prefrontal cortex. *Cerebral Cortex*, 17 (suppl_1):i110–i117, 2007.

Susan R Sesack, Ariel Y Deutch, Robert H Roth, and Benjamin S Bunney. Topographical organization of the efferent projections of the medial prefrontal cortex in the rat: an anterograde tract-tracing study with phaseolus vulgaris leucoagglutinin. *Journal of Comparative Neurology*, 290(2):213–242, 1989.

Michael J Siniscalchi, Victoria Phoumthipphavong, Farhan Ali, Marc Lozano, and Alex C Kwan. Fast and slow transitions in frontal ensemble activity during flexible sensorimotor behavior. *Nature neuroscience*, 19(9):1234, 2016.

Burton Slotnick. A simple 2-transistor touch or lick detector circuit. *Journal of the experimental analysis of behavior*, 91(2):253, 2009.

Mark G Stokes, Makoto Kusunoki, Natasha Sigala, Hamed Nili, David Gaffan, and John Duncan. Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, 78(2):364–375, 2013.

Jung Hoon Sul, Hoseok Kim, Namjung Huh, Daeyeol Lee, and Min Whan Jung. Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. *Neuron*, 66(3):449–460, 2010.

Jung Hoon Sul, Suhyun Jo, Daeyeol Lee, and Min Whan Jung. Role of rodent secondary motor cortex in value-based action selection. *Nature neuroscience*, 14 (9):1202, 2011.

Kaori Takehara-Nishiuchi and Bruce L McNaughton. Spontaneous changes of neocortical code for associative memory during consolidation. *Science*, 322(5903):960–963, 2008.

Philippe Thevenaz, Urs E Ruttimann, and Michael Unser. A pyramid approach to subpixel registration based on intensity. *IEEE transactions on image processing*, 7 (1):27–41, 1998.

Ivan Toni, Narender Ramnani, Oliver Josephs, John Ashburner, and Richard E Passingham. Learning arbitrary visuomotor associations: temporal dynamic of brain activity. *Neuroimage*, 14(5):1048–1057, 2001.

Jonathan D Wallis, Kathleen C Anderson, and Earl K Miller. Single neurons in prefrontal cortex encode abstract rules. *Nature*, 411(6840):953–956, 2001.

Ilsun M White and Steven P Wise. Rule-dependent neuronal activity in the prefrontal cortex. *Experimental brain research*, 126(3):315–335, 1999.

Tom J Wills, Colin Lever, Francesca Cacucci, Neil Burgess, and John O'Keefe. Attractor dynamics in the hippocampal representation of the local environment. *Science*, 308(5723):873–876, 2005.

Steven P Wise and Elisabeth A Murray. Arbitrary associations between antecedents and actions. *Trends in neurosciences*, 23(6):271–276, 2000.

Yuan Yuan, Hongwei Mao, and Jennie Si. Cortical neural responses to previous trial outcome during learning of a directional choice task. *Journal of neurophysiology*, 113(7):1963–1976, 2014.

Petr Znamenskiy and Anthony M Zador. Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination. *Nature*, 497(7450):482–485, 2013.