

# Mechanisms of neural activity exploration and consolidation underlying neuroprosthetic skill learning

*Vivek Athalye*



Electrical Engineering and Computer Sciences  
University of California at Berkeley

Technical Report No. UCB/EECS-2017-211

<http://www2.eecs.berkeley.edu/Pubs/TechRpts/2017/EECS-2017-211.html>

December 14, 2017

Copyright © 2017, by the author(s).  
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

**Mechanisms of neural activity exploration and consolidation underlying  
neuroprosthetic skill learning**

by

Vivek Ravindra Athalye

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jose M. Carmena, Chair

Professor Kannan Ramchandran

Professor Bruno Olshausen

Fall 2017

**Mechanisms of neural activity exploration and consolidation underlying  
neuroprosthetic skill learning**

Copyright 2017  
by  
Vivek Ravindra Athalye

## Abstract

Mechanisms of neural activity exploration and consolidation underlying neuroprosthetic skill learning

by

Vivek Ravindra Athalye

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Jose M. Carmena, Chair

The Brain-Machine Interface (BMI) is an emerging technology which directly translates neural activity into control signals for effectors such as computers, prosthetics, or even muscles. Work over the last decade has shown that high performance BMIs depend on machine learning to adapt parameters for decoding neural activity, but also on the brain learning to reliably produce desired neural activity patterns. How the brain learns neuroprosthetic skill de novo is not well-understood and could inform the design of next-generation BMIs in which both the brain and machine synergistically adapt.

We view BMI learning from the brain's perspective as a reinforcement learning problem, as the brain must initially explore activity patterns, observe their consequences on the prosthetic, and finally consolidate activity patterns leading to desired outcomes. This thesis will address 3 questions about how the brain learns neuroprosthetic skill:

1. How do task-relevant neural populations coordinate during activity exploration and consolidation?
2. How can the brain select activity patterns to consolidate? Does the pairing of neural activity patterns with neural reinforcement signals drive activity consolidation?
3. Do the basal ganglia-dependent mechanisms of neural activity exploration and consolidation generalize across cortex, even to visual cortex?

First, we present the use of Factor Analysis to analyze neural coordination during BMI control by partitioning neural activity variance arising from two sources: private inputs to each neuron which drive independent, high-dimensional variance, and shared inputs which drive multiple neurons simultaneously and produce low-dimensional covariance. We found that initially, each neuron explores activity patterns independently. Over days of learning, the population's covariance increases, and a manifold emerges which aligns to the decoder. Strikingly, this low-dimensional activity drives skillful control of the decoder.

Next, we consider the role of reinforcement signals in the brain in driving neural activity consolidation. By performing experiments with a novel BMI that delivers reward through optogenetic stimulation, we found that cortical neural activity patterns which causally lead to midbrain dopaminergic neural reinforcement are consolidated. This provides evidence for a neural law of effect, following Thorndike’s behavioral law of effect stating that behaviors leading to reinforcements are repeated.

Previous work has shown that dopaminergic reinforcement signals contribute to plasticity between cortex and striatum, the input area to the subcortical basal ganglia, and that corticostriatal plasticity is necessary for BMI learning. Thus, we investigate whether the basal-ganglia dependent ability to explore and consolidate activity patterns generalizes across cortex. Indeed, we find that the brain can explore and consolidate activity patterns even in visual cortex, an area thought primarily to represent visual stimulus, and that learning requires the basal ganglia, as optogenetic inhibition of dorsomedial striatum blocks learning.

Together, these results contribute to our understanding of how the brain solves the reinforcement learning problem of learning neuroprosthetic skill, suggesting a computational role for high-dimensional private neural variance and exploration, low-dimensional shared neural variance and consolidated control, and 1) dopaminergic midbrain and 2) striatum activity and neural reinforcement.

To Shilpa. To Mom. To Baba.

# Contents

<b>Contents</b>	<b>ii</b>
<b>List of Figures</b>	<b>iv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Appendix: Clinical Motor BMI State-of-the-art . . . . .	10
<b>2 Emergence of Coordinated Neural Dynamics Underlies Neuroprosthetic Learning and Skillful Control</b>	<b>14</b>
2.1 Introduction . . . . .	14
2.2 Methods . . . . .	18
2.3 Results . . . . .	27
2.4 Discussion . . . . .	48
2.5 Appendix A . . . . .	53
<b>3 Evidence for a Neural Law of Effect</b>	<b>65</b>
3.1 Introduction . . . . .	65
3.2 Methods . . . . .	68
3.3 Results . . . . .	76
3.4 Discussion . . . . .	89
3.5 Appendix A . . . . .	91
<b>4 Volitional modulation of primary visual cortex activity requires the basal ganglia</b>	<b>108</b>
4.1 Introduction . . . . .	108
4.2 Methods . . . . .	110
4.3 Behavioral Task . . . . .	112
4.4 Results . . . . .	117
4.5 Discussion . . . . .	141
<b>5 Conclusion</b>	<b>145</b>
5.1 Thesis contributions . . . . .	145
5.2 Future work . . . . .	146

References

# List of Figures

1.1	<b>Functional priorities of quadriplegic patients suffering from spinal cord injury.</b> Figure taken from (Anderson, 2004; Tsu et al., 2015). Pie chart shows the percentage of quadriplegic patients indicating each area as their top priority for functional recovery. . . . .	1
1.2	<b>The BMI constitutes a novel limb.</b> Figure taken from (Shenoy and Carmena, 2014). Physical plant 1 represents the arm of a patient with a severed connection between the brain and spinal cord. Physical plant 2 represents the BMI. . . . .	2
1.3	<b>Closed-loop BMI System.</b> BMI systems can employ sensors which record different neural signals, from single neuron action potentials with microelectrode arrays to aggregated electrical activity with ECoG or EEG sensors. The BMI can then generate control signals for external effectors such as a computer keyboard and mouse, a robotic arm, or a user’s own muscles. Figure adapted from (Ajiboye et al., 2017; Collinger et al., 2013b; Leuthardt et al., 2006; Pandarinath et al., 2017). . . . .	4
1.4	<b>Neural action potential.</b> From (Hausser, 2000), a review of the classic work of Hodgkin and Huxley. . . . .	5
1.5	<b>Opsins used for optogenetics.</b> From (Yizhar et al., 2011) . . . . .	8
1.6	<b>ChR2 in Th-Cre Mice.</b> From (Tsai et al., 2009) . . . . .	8
1.7	<b>ChR2 stimulation of VTA.</b> Blue ticks indicate optical stimulation. From (Tsai et al., 2009) . . . . .	9
1.8	<b>Jaws inhibition.</b> Red bar indicates optical inhibition. From (Chuong et al., 2014) . . . . .	9
1.9	<b>State-of-art human BMI performance for typing.</b> From (Pandarinath et al., 2017). Caption from paper’s figure legend: Examples of text typed during three copy typing evaluations with participants T6, T5, and T7. Each example shows prompted text, followed by the characters typed within the first minute of the two-minute evaluation block. Box width surrounding each character denotes the time it took to select the character. ‘<’ character denotes selection of backspace key. . . . .	11

1.10	<b>Seven-DOF control of a prosthetic limb for reaching and grasping.</b> Experiment from (Collinger et al., 2013b), figure and caption from (Bensmaia and Miller, 2014). (A) A patient with tetraplegia had two electrode arrays implanted into the primary motor cortex. (B) The patient controlled the position and orientation of the modular prosthetic limb in order to grasp objects. . . . .	12
1.11	<b>Grasp-pour-and-stir functional movement.</b> Figure and caption from (Bouton et al., 2016). Sequential snapshots from the functional movement task showing the participant opening his hand (A), grasping the glass bottle (B), pouring its contents (dice) into a jar (C), grasping a stir stick from another jar (D), transferring the stir stick without dropping it (E), and using it to stir the dice in the jar (F). The task required the participant to evoke different cortical modulation patterns to control the opening of his hand, perform a cylindrical palmar grasp, and achieve a precision pinch grasp, while simultaneously moving his arm. . . . .	13
2.1	<b>Operant Learning BMI Paradigm Using Stable Neurons and a Fixed Decoder.</b> Monkeys use neural activity of stable neurons to control a fixed decoder over days, so improvement can only be driven by adaptive changes in neural population activity. . . . .	15
2.2	<b>Models of Learning.</b> (A) Independent Neuron Learning posits that each neuron explores and acquires independent patterns to control the BMI. (B) Constrained-Network Learning posits that the neural population learns BMI control by exploring and changing covariation patterns in a fixed shared space, e.g. because of network connectivity constraints. (C) Flexible-Network Learning posits that the neural population learns BMI control by exploring and changing both independent and covariation patterns and by using its exploration to change the shared space of covariation patterns. For intuition of how independent and covariation patterns can contribute to movement variability, see (Figure 3.11) . . . . .	16
2.3	<b>Electrophysiology.</b> Electrophysiology in rhesus macaque M1 and PMd using 64 microwire arrays. Figure courtesy of Jose M. Carmena. . . . .	19
2.4	<b>Experimental Paradigm.</b> (A) Monkeys control a 2D cursor to one of 8 instructed targets using neural activity of stable neurons over days. (B) Center-out task timeline. (C) The task uses a position decoder, which maps a neural spatiotemporal pattern into a low-dimensional control signal that produces the cursor endpoint. . . . .	20
2.5	<b>Readout Space of Wiener Filter.</b> The decoder transforming neural activity into cursor position can be decomposed into two stages of filters. The first stage consists of “spatial filters” which project neural activity into a subspace termed the decoder’s “Readout Space.” The time series of neural activity projections are termed control signals. The decoder’s second stage consists of “temporal filters” which perform linear combinations of control signals over time lags. . . . .	21
2.6	<b>Cursor control is refined.</b> Caption on next page. . . . .	28

2.6	(Previous page.) <b>(A,B)</b> Performance improved over training for Monkey P (main) and Monkey R (inset). Percent correct increased (Monkey P: $r=0.93$ , $p=5.8e-7$ , early < late $p=3.3e-10$ ; Monkey R: $r=0.96$ , $p=4.7e-4$ , early < late $p=1.5e-9$ ). Success rate in trials per minute increased (Monkey P: $r=0.96$ , $p=1.6e-8$ , early < late $p=1.4e-6$ ; Monkey R: $r=0.91$ , $p=4e-3$ , early < late $p=4.0e-6$ ). Targets (indicated by color) showed different time courses of improvement. Mean and s.e.m. error bars are shown over targets ( $n=8$ for both subjects). <b>(C, top)</b> Single trial example trajectories over three epochs, for one example target. <b>(C, bottom)</b> Workspace occupancy maps for the example target over three epochs. Intensity indicates the fraction of trials a position was occupied. <b>(D)</b> Percent of workspace entered per trial decreased over training (Monkey P: $r=-0.94$ , $p=3.2e-07$ , early > late $p=4.1e-9$ ; Monkey R: $r=-0.87$ , $p=1e-2$ , early > late $p=2.0e-4$ ). <b>(E)</b> Similarity of workspace occupancy to the final workspace map increased and stabilized with training (Monkey P: $r=0.92$ , $p=2.3e-6$ , early < late $p=5.9e-9$ ; Monkey R: $r=0.88$ , $p=2e-2$ , early < late $p=1.2e-2$ ). <b>(E, bottom inset)</b> Heat map shows the similarity between each epoch pair's occupancy maps. A heat map was calculated for each target separately; the target average is shown here. The lower left red box indicates cursor scatter during gross exploration while the upper red box indicates late training stability of refined control. <b>(F)</b> Trial-to-trial variability of workspace occupancy decreased (Monkey P: $r=-0.93$ , $p=4.6e-7$ , early > late $p=1.7e-8$ ; Monkey R: $r=-0.54$ , $p=2e-1$ , early > late $p=2.4e-2$ ). . . .	29
2.7	<b>Factor Analysis Covariance Decomposition.</b> Factor Analysis (FA) decomposes the spike count covariance matrix into the sum of a diagonal private variance matrix and a low-rank shared variance matrix. . . . .	30
2.8	<b>Variance Analyses via Factor Analysis.</b> . . . . .	31
2.8	<b>Variance Analyses via Factor Analysis.</b> <b>(A)</b> Private signals modulating two neurons' firing rates. Each scattered gray 'x' is a binned population spike count. Neuron 2 has a larger private variance than neuron 1. <b>(B)</b> Shared signals modulating two neurons' firing rates within a one-dimensional shared space. <b>(C)</b> FA describes the mixture of private and shared variance underlying population activity. <b>(D)</b> FA detects changes in the total magnitude of variance. <b>(E)</b> FA also detects changes in the shared-to-total variance ratio, quantifying how shared and private variance change in different proportions. The shared-to-total ratio plus the private-to-total ratio equals 1. <b>(F)</b> FA detects changes in how the population co-varies using the shared space alignment between Epoch A and B, which is the fraction of Epoch A's shared variance captured in Epoch B's shared space. . . .	31
2.9	<b>Private Trial-to-Trial Variability Decreases and the Shared Trial-to-Trial Variability Consolidates a Shared Space.</b> . . . . .	33

- 2.9 **Private Trial-to-Trial Variability Decreases and the Shared Trial-to-Trial Variability Consolidates a Shared Space.** (A) Trial-to-trial variability example: FA fit on 2 direct cells in early and late training for Monkey P on one example target. (B) Private trial-to-trial variability for Monkey P (main) and Monkey R (inset) decreased (Monkey P:  $r=-0.93$ ,  $p=6.7e-7$ , early > late  $p=1.1e-8$ ; Monkey R:  $r=-0.81$ ,  $p=2.7e-2$ , early > late  $p=2.2e-2$ ). (C) The shared-to-total variance ratio increased over long-term training (Monkey P:  $r=0.83$ ,  $p=1.1e-4$ , early < late  $p=8.1e-3$ ) but not significantly over shorter training (Monkey R:  $r=0.27$ ,  $p=5.5e-1$ , early < late  $p=8.2e-2$ ). (D) The private trial-to-trial variability decrease correlated with success rate improvement (Monkey P:  $r=-0.96$ ,  $p=1.9e-8$ ; Monkey R:  $r=-0.79$ ,  $p=3.5e-2$ ). (E, left) The shared space alignment map indicated consolidation. Each element is the alignment of the row epoch's shared variance with the column epoch's shared space. The upper right red block indicates that the shared space becomes relatively stable, while the preceding blue indicates the shared space is unstable in early training. (E, right) The shared space alignment between the final epoch's shared space and the preceding epochs' shared variance increased with long-term training (Monkey P:  $r=0.83$ ,  $p=2.6e-4$ , early < late  $p=5.8e-4$ ; Monkey R:  $r=-0.12$ ,  $p=8.2e-1$ , early > late  $p=4.9e-1$ ). The mean chance variance alignment is 0.07, and the 95th percentile chance alignment is 0.25 for Monkey P (see Experimental Procedures). (F) The shared space alignment with the final epoch's shared space correlated with success rate over long-term training (Monkey P:  $r=0.88$ ,  $p=3.3e-5$ ; Monkey R:  $r=0.07$ ,  $p=9e-1$ ). Changes in mean firing rate do not explain these results (Figure 3.15). These results also held for analyses over training days rather than epochs (Figure 3.12) and different window lengths (Figure 3.14). . . . . 34
- 2.10 **Decoder-timescale shared variance increases and consolidates a shared space.** . . . . . 36

- 2.10 **Decoder-timescale shared variance increases and consolidates a shared space.** (A) Decoder-timescale variance example: FA fit on 2 direct cells in early and late training for Monkey P on one example target. (B) The shared dimensionality decreased from 4 to 2 (Monkey P:  $r=-0.85$ ,  $p=5.7e-5$ , early  $>$  late  $p=1.4e-7$ ; Monkey R:  $r=-0.80$ ,  $p=3.1e-2$ , early  $>$  late  $p=9.8e-5$ ). We used a main shared dimensionality of 2 (see Experimental Methods). (C) Main shared variance increased with training (Monkey P:  $r=0.92$ ,  $p=8.1e-7$ , early  $<$  late  $p=1.9e-5$ ; Monkey R:  $r=0.84$ ,  $p=2.0e-2$ , early  $<$  late  $p=1.2e-5$ ). (D) The ratio of main shared variance to total variance increased with training (Monkey P:  $r=0.95$ ,  $p=7.1e-8$ , early  $<$  late  $p=2.2e-6$ ; Monkey R:  $r=0.76$ ,  $p=4.6e-2$ , early  $<$  late  $p=1.2e-4$ ). (E) Main shared variance correlated with success rate (Monkey P:  $r=0.98$ ,  $p=1.9e-10$ ; Monkey R:  $r=0.69$ ,  $p=8.6e-2$ ). (F, left) Shared space alignment map. The upper right red block indicates the shared space becomes stable, while the preceding blue indicates the shared space is changing in early training. (F, right) Alignment with the final epoch's shared space increased with training (Monkey P:  $r=0.95$ ,  $p=1.8e-7$ , early  $<$  late  $p=9.6e-9$ ; Monkey R:  $r=0.96$ ,  $p=2.2e-3$ , early  $<$  late  $p=2.2e-6$ ). The mean chance alignment is (Monkey P: 0.13; Monkey R: 0.20) and 95th percentile chance alignment is (Monkey P: 0.28; Monkey R: 0.40). (G) Alignment with the final epoch's shared space correlated with success rate improvement (Monkey P:  $r=0.99$ ,  $p=7.7e-11$ ; Monkey R:  $r=0.92$ ,  $p=8.3e-3$ ). Changes in mean firing rate do not explain these results (Figure 3.15). These results also held for analyses over training days rather than epochs (Figure 3.13) and different bin widths (Figure 3.14). . . . . 37
- 2.11 **Decoder-timescale private variance produces noisy yet successful movements, while shared variance generate skillful movements.** . . . . . 40

- 2.11 Decoder-timescale private variance produces noisy yet successful movements, while shared variance generate skillful movements.** (A) Both private and shared signals can mathematically produce the same desired control signal. (A, left) The desired control signal is shown over time, and its values are plotted on neuron firing rate axes. The readout value is the sum of the 2 neurons' firing rates. The dashed lines are contours which capture firing rates yielding the same sum. (A, middle) The neurons' activity is uncorrelated and produces the desired control signal. (A, right) The neurons' activity is perfectly correlated and produces the desired control signal. (B) The contribution of private and shared variance to each trial's activity can be estimated and simulated through the decoder separately to determine how they independently contribute to movement. (C) Both main shared and private variance increased best progress to target over training. (Private variance; Monkey P:  $r=0.85$ ,  $p=5.7e-5$ , early < late  $p=2.7e-6$ ; Monkey R:  $r=0.75$ ,  $p=5.3e-2$ , early < late  $p=3.0e-2$ ). (Main shared variance; Monkey P:  $r=0.94$ ,  $p=2.0e-7$ , early < late  $p=3.1e-8$ ; Monkey R:  $r=0.74$ ,  $p=5.8e-2$ , early < late  $p=9.6e-3$ ). (D) Example trials driven by main shared and private variance from the final epoch for each target. (E) In late training, successful main shared variance-driven movements were significantly faster than private variance-driven movements (main shared minus private time-to-target; Monkey P: mean = -0.80 sec, main shared < private  $p=1.8e-56$ ; Monkey R: mean = -0.66 sec, main shared < private:  $p=4.3e-11$ ). (F) In late training, successful main shared variance-driven movements were significantly more direct than private variance-driven movements (main shared minus private distance traveled; Monkey P: mean = -1.6, main shared < private  $p=3.2e-52$ ; Monkey R: mean = -4.0, main shared < private  $p=2.49e-22$ ). Interestingly, main shared variance achieved higher performance than even total activity (Figure S6). For (EF), late training trials were analyzed for which both private and main shared variance were successful (Monkey P: epochs 8 to 15, 832 trials; Monkey R: epochs 4 to 7, 385 trials). (G) Total activity-driven movements explored the workspace more than main shared variance-driven movements, especially in early training (main shared < total, Monkey P:  $p=2.5e-9$ ; Monkey R:  $p=3.2e-9$ ). . . . . 41
- 2.12 Task-relevant shared variance increases because of shared variance growth and re-alignment.** . . . . . 43

- 2.12 **Task-relevant shared variance increases because of shared variance growth and re-alignment.** (A) Shared variance can drive the decoder by producing variance in the decoder’s readout space. (B) Main shared variance in the readout space increased (Monkey P:  $r=0.90$ ,  $p=6.6e-6$ , early < late  $p=8.5e-9$ ; Monkey R:  $r=0.85$ ,  $p=1.5e-2$ , early < late  $p=8.4e-5$ ) (C) Shared variance can increase in the readout space by two mechanisms. (C, left) An increase in shared variance magnitude increases the readout space shared variance, as long as the shared space and readout space initially align. (C, right) Re-alignment of the shared space with the readout space increases the readout space shared variance. Readout space shared variance growth is the product of the magnitude growth and re-alignment. (D) Magnitude growth and re-alignment both increased for main shared variance, plotted as the ratio of each epoch’s value to the first epoch’s value. (Magnitude growth factor; Monkey P:  $r=0.93$ ,  $p=7.3e-7$ , early < late  $p=3.1e-4$ ; Monkey R:  $r=0.82$ ,  $p=2.4e-2$ , early < late  $p=1.6e-5$ ). (Re-alignment factor; Monkey P:  $r=0.84$ ,  $p=9.9e-5$ , early < late  $p=8.2e-6$ ; Monkey R:  $r=0.17$ ,  $p=7.2e-2$ , early < late  $p=6.2e-1$ ). While shared space alignment with the readout space increased, note that a significant fraction of shared variance remained outside the readout space in late learning (Figure 3.16 A). . . . . 44
- 2.13 **Shared neural trajectories are consolidated at fine-timescale which produce accurate movement.** . . . . . 46
- 2.13 **Shared neural trajectories are consolidated at fine-timescale which produce accurate movement.** (A) Shared variance can arise without a consistent neural trajectory (left) or with a consistent neural trajectory (right). (B) For one example target, every epoch’s trial-averaged main shared trajectory is plotted in the same informative 2D plane of the high-dimensional neural space (see Supplemental Experimental Procedures). (C) 6 example cells’ main shared time-course over training epochs for one target. (D) The magnitude of the main shared trajectory increased, as quantified by the squared norm (Monkey P:  $r=0.95$ ,  $p=8.8e8$ , early < late  $p=9.1e-7$ ; Monkey R:  $r=0.86$ ,  $p=1.3e-2$ , early < late  $p=1.0e-6$ ). (E) The correlation between each epoch’s main shared trajectory and the final epoch increased and stabilized (Monkey P:  $r=0.98$ ,  $p=3.0e-9$ , early < late  $p=2.2e-9$ ; Monkey R:  $r=0.93$ ,  $p=6.3e-3$ , early < late  $p=5.5e-8$ ). (F) The percent of fine-timescale main shared variance due to the trial-averaged main shared trajectory increased (Monkey P:  $r=0.96$ ,  $p=9.7e-9$ , early < late  $p=4.4e-6$ ; Monkey R:  $r=0.77$ ,  $p=4.4e-2$ , early < late  $p=7.1e-5$ ). (G, H) The trial-averaged main shared trajectory simulated through the decoder increased movement accuracy (Monkey P:  $r=0.92$ ,  $p=1.4e-6$ , early < late  $p=1.6e-8$ ; Monkey R:  $r=0.84$ ,  $p=1.8e-3$ , early < late  $p=5.9e-4$ ). . . . . 47

- 2.14 **Manners in which private and shared variance generate movement variance depending on the readout space.** (A) Private variance is high-dimensional and therefore generates variance in any readout space. (B) Low-dimensional shared variance generates significant variance only in readout spaces with which it is aligned. (C) Illustration of two neurons with equal variance. If the two neurons exhibit shared variance as opposed to private variance, they can concentrate more variance in the readout dimensions and thus generate movement variance more efficiently. . . . . 53
- 2.15 **Trial-to-trial variability over days of training.** (A) Over days, private trial-to-trial variability (private variance divided by mean rate) decreased (Monkey P:  $r=-0.77$ ,  $p=1.7e-4$ , early > late  $p=2.2e-8$ , 8 targets consistent; Monkey R:  $r=-0.84$ ,  $p=1.8e-2$ , early > late  $p=2.4e-2$ , 6 targets consistent). The window size was 0.9 sec for Monkey P and 1.3 sec for Monkey R as in Figure 4B. (B) Private trial-to-trial variability decrease correlated with success rate (Monkey P:  $r = -0.88$ ,  $p=1.9e-6$ ; Monkey R:  $r = -0.78$ ,  $p=3.7e-2$ ). (C) The ratio of shared variance to total variance increased (no normalization by mean rate) increased (Monkey P:  $r=0.58$ ,  $p=1.2e-2$ , late > early  $p=6.5e-4$ , 7 targets consistent). (D) The shared space stabilized (Monkey P:  $r=0.55$ ,  $p=2.3e-2$ , late > early  $p=4.5e-2$ , 6 targets consistent). (E) Shared space alignment to the final day correlated with success rate (Monkey P:  $R=0.65$ ,  $p=4.7e-3$ ). . . . . 54
- 2.16 **Decoder-timescale variance over days of training.** . . . . . 55
- 2.16 **Decoder-timescale variance over days of training.** (A) The shared dimensionality started at 4 and fell to 2 in late training, thus a main shared dimensionality of 2 was used. Main shared variance increased (Monkey P:  $r = 0.79$ ,  $p=1.1e-4$ , late > early  $p=1.1e-7$ , 8 targets consistent; Monkey R:  $r=0.77$ ,  $p=4.2e-5$ , late > early  $p=5.2e-3$ , 6 targets consistent). (B) The main shared to total variance ratio increased (Monkey P:  $r=0.88$ ,  $p=1.6e-5$ , late > early  $p=3.9e-7$ , 8 targets consistent; Monkey R:  $r=0.77$ ,  $p=4.2e-5$ , late > early  $p=5.2e-3$ , 6 targets consistent). (C) Main shared variance correlated with success rate (Monkey P:  $r=0.89$ ,  $p=5.4e-7$ ; Monkey R:  $r=0.83$ ,  $p=2.1e-2$ ). (D) Main shared covariance stabilized (Monkey P:  $r=0.89$ ,  $p=1.3e-6$ , late > early  $p=2.1e-8$ , 8 targets consistent; Monkey R:  $r=0.94$ ,  $p=5.5e-3$ , late > early  $p=4.8e-5$ , 8 targets consistent). (E) The main shared space alignment to the final epoch correlated with success rate (Monkey P:  $r=0.94$ ,  $p=2.3e-8$ ; Monkey R:  $r=0.78$ ,  $p=6.7e-2$ ). . . . . 56
- 2.17 **Variance results are robust to temporal parameter choice.** . . . . . 57

- 2.17 Variance results are robust to temporal parameter choice. (A)** The 30th percentile of time to target over all targets. Minimum 30th percentile over all conditions was (Monkey P: 0.9 sec; Monkey R: 1.3 sec), which was used in Fig 4 for trial-to-trial variability. **(B)** Trial-to-trial private variability decreased for analysis window = 0.6 sec: (Monkey P:  $r=-0.91$ ,  $p=2.4e-6$ , early > late  $p=2.9e-12$ , 8 targets consistent; Monkey R:  $r=-0.45$ ,  $p=3.1e-1$ , early > late  $p=4.4e-2$ , 6 targets consistent), and for analysis window = 1 sec (Monkey P:  $r=-0.93$ ,  $p=4.3e-7$ , early > late  $p=1.2e-8$ , 8 targets consistent; Monkey R:  $r=-0.68$ ,  $p=9.2e-2$ , early > late  $p=1.4e-2$ , 6 targets consistent). **(C)** Trial-to-trial shared to total variance ratio increased for analysis window = 0.6 sec: (Monkey P:  $r=0.67$ ,  $p=6.5e-3$ , late > early  $p=1.0e-3$ , 6 targets consistent) and window = 1 sec: (Monkey P:  $r=0.60$ ,  $p=1.7e-2$ , late > early  $p=2.7e-3$ , 5 targets consistent). **(D)** The trial-to-trial shared space alignment to the final epoch increased for window = 0.6 sec (Monkey P:  $r=0.87$ ,  $p=5.4e-5$ , late > early  $p=6.7e-4$ , 7 targets consistent) and for window = 1 sec (Monkey P:  $r=0.67$ ,  $p=6.5e-3$ , late > early  $p=1.0e-3$ , 7 targets). **(E)** The decoder-timescale main shared to total variance ratio increased over a range of bin widths. Bin width = 10 ms: (Monkey P:  $r=0.72$ ,  $p=2.3e-4$ , late > early  $p=2.8e-3$ , 7 targets consistent, Monkey R:  $r=0.50$ ,  $p=2.6e-1$  N.S., late > early  $p=4.4e-2$ , 7 targets consistent), 20 ms: (Monkey P:  $r=0.89$ ,  $p=1.1e-5$ , late > early  $p=5.9e-6$ , 8 targets consistent; Monkey R: N.S.  $r=0.47$ ,  $p=2.8e-1$ , late > early  $p=0.11$ , 7 targets consistent), 50 ms: (Monkey P:  $r=0.93$ ,  $p=4.9e-7$ , late > early  $p=4.3e-5$ , 8 targets consistent; Monkey R:  $r=0.75$ ,  $p=4.9e-2$ , N.S. late > early  $p=6.6e-2$ , 8 targets consistent), 100 ms: (Monkey P:  $r=0.95$ ,  $p=7.1e-8$ , late > early  $p=2.2e-6$ , 8 targets consistent, Monkey R:  $r=0.76$ ,  $p=4.6e-2$ , late > early  $p=2.9e-2$ , 7 targets consistent), 200 ms: (Monkey P:  $r=0.92$ ,  $p=8.0e-7$ , late > early  $p=2.0e-6$ , 8 targets consistent, Monkey R:  $r=0.72$ ,  $p=6.7e-2$ , late > early  $p=2.3e-2$ , 7 targets consistent), 500 ms: (Monkey P:  $r=0.88$ ,  $p=1.6e-5$ , late > early  $p=2.2e-4$ , 8 targets consistent, Monkey R:  $r=0.58$ ,  $p=1.7e-1$  N.S., late > early  $p=5.2e-2$  N.S., 7 targets consistent). . . . . 58
- 2.18 Mean firing rate changes do not explain trial-to-trial variability and decoder-timescale variance changes.** . . . . . 59

**2.18 Mean firing rate changes do not explain trial-to-trial variability and decoder-timescale variance changes.**

**(A)** The mean firing rate increased over training (Monkey P:  $R=0.79$ ,  $p=5.1e-4$ , early < late  $p=7.1e-6$ , 8 targets consistent; Monkey R:  $R=0.49$ ,  $p=2.7e-1$ , early < late  $p=6.1e-3$ , 7 targets consistent). **(B)** We performed FA on a simulated data set of independent Poisson spike counts with window sizes used in Figure 4, the experimentally observed neurons' means, and 50 times the number of trials. Fitting FA on Poisson counts yields the best shared dimensionality as 0. In these plots, we used 1 shared dimension, matching our analysis of real neural data (Figure 4). Monkey P's simulated private trial-to-trial variability increased while Monkey R showed no trend (Monkey P,  $r=0.78$ ,  $p=5.1e-4$ , late > early  $p=2.3e-5$ ; Monkey R n.s. :  $r=0.33$ ,  $p=4.8e-1$ , late > early  $p=1.2e-1$ ). Both subjects' shared variance is close to 0 and showed no trend (Monkey P,  $r=0.29$ ,  $p=2.9e-1$ , early vs late  $p=4.9e-1$ ; Monkey R,  $r=-0.06$ ,  $p=9.1e-1$ , early vs late  $p=1.1e-1$ ). Because shared variance does not track changes in mean firing rate, we chose not to normalize shared variance by mean firing rate. **(C)** As mathematically expected, we observed that normalizing each simulated Poisson neuron's private variance by its mean yielded a value close to 1 with no significant trend over training. The value is expected to be less than 1, because some variance was absorbed as shared variance since we used a shared dimensionality of 1. **(D)** The trial-to-trial shared to total variance ratio showed no trend and is close to 0. **(E)** The trial-to-trial shared space alignment to the final epoch showed no trend (Monkey P:  $r=0.21$ ,  $p=4.7e-1$ , early vs late  $p=1.5e-1$ ; Monkey R:  $r=-0.55$ ,  $p=2.6e-1$ , early vs late  $p=6.9e-1$ ). . . . .

- 2.18 **Mean firing rate changes do not explain trial-to-trial variability and decoder-timescale variance changes.** **(F)** The experimentally observed data's decoder-timescale main shared, high-dimensional shared, and private variance, normalized by total variance. Normalized private variance decreased (Monkey P:  $r=-0.92$ ,  $p=1.4e-6$ , early > late  $p=1.3e-4$ ; Monkey R:  $r=-0.38$ ,  $p=4.1e-1$ , early > late  $p=7.6e-2$ ), and normalized high dimensional shared signal variance decreased (Monkey P:  $r=-0.80$ ,  $p=3.8e-4$ , early vs late  $p=1.2e-4$ ; Monkey R:  $r=-0.76$ ,  $p=4.8e-2$ , early > late  $p=8.7e-4$ ). Normalized main shared variance increased (Figure 5D). **(G)** Decoder-timescale shared variance changes are not explained by changes in mean firing rate. We applied FA on Poisson simulated neurons with 100 ms bins using experimentally observed average firing rates and shared dimensionality (Figure 5B). Private variance increased (Monkey P:  $r=0.76$ ,  $p=9.8e-4$ , early < late  $p=2.8e-4$ ; Monkey R:  $r=0.42$ ,  $p=3.4e-1$ , early < late  $p=2.3e-2$ ) and main shared variance was unchanged (Monkey P:  $r=0.39$ ,  $p=1.5e-1$ , early vs late  $p=1.7e-1$ ; Monkey R:  $r=0.34$ ,  $p=4.5e-1$ , early vs late  $p=4.8e-1$ ). High dimensional shared variance decreased slightly because the experimental shared dimensionality decreased (Monkey P:  $r=-0.59$ ,  $p=2.0e-2$ , early > late  $p=2.6e-4$ ; Monkey R:  $r=-0.37$ ,  $p=4.1e-1$ , early > late  $p=8.2e-3$ ). **(H)** Decoder-timescale Poisson simulated neurons' main shared variance to total variance ratio showed no trend (Monkey P:  $r=0.27$ ,  $p=3.4e-1$ , early vs late  $p=4.1e-1$ ; Monkey R:  $r=0.039$ ,  $p=9.3e-1$ , early vs late  $p=6.9e-1$ ). **(I)** Decoder-timescale Poisson simulated neurons' shared space alignment to the final epoch showed no trend (Monkey P:  $r=0.32$ ,  $p=2.6e-1$ , early vs late  $p=1.5e-1$ ; Monkey R:  $r=0.14$ ,  $p=7.9e-1$ , early vs late  $p=4.6e-1$ ). . . . . 61
- 2.19 **Decoder-timescale main shared variance and control.** . . . . . 62
- 2.19 **Decoder-timescale main shared variance and control.** **(A)** The fraction of decoder-timescale main shared variance in the readout space increased for Monkey P ( $r=0.86$ ,  $p=3.9e-5$ , late > early  $p=2.2e-4$ , 8 targets consistent) but not Monkey R ( $r=-0.28$ ,  $p=5.5e-1$ , late < early  $p=1.1e-1$ ). **(B)** Example trials driven by main shared variance and total activity from the final epoch for each target. **(C)** In late learning, successful movements driven by main shared variance are significantly faster (main shared minus total time-to-target; Monkey P: mean = -0.78 sec, main shared < total  $p=4.2e-51$ ; Monkey R: mean = -0.66 sec, main shared < total  $p=3.2e-11$ ). **(D)** In late learning, successful movements driven by main shared variance are significantly more direct (main shared minus total distance-traveled; Monkey P: mean = -2.8, main shared < total  $p=5.38e-83$ ; Monkey R: mean = -5.1, main shared < total  $p=1.8e-28$ ). For (CD), late learning trials were analyzed for which total activity and main shared variance were successful (Monkey P: epochs 8 to 15, 832 trials; Monkey R: epochs 4 to 7, 385 trials). . . 63
- 2.20 **Summary Table.** Summary of number of targets consistent with analyzed trends. . . . . 64

- 3.1 **Closed-loop BMI paradigm for pairing specific motor cortex activity patterns with phasic VTA dopaminergic activity.** Schematic of the BMI paradigm. Each mouse receives a unilateral microwire array implant in motor cortex (targeted to layer V) and a contralateral optical fiber implant in VTA. Recorded single units are arbitrarily assigned into two ensembles, and the concomitant increase of one ensemble's activity and decrease in the other ensemble's activity drives the decoder to change the auditory tone produced every 500 ms. The rare, lowest tone triggers phasic optical stimulation to the VTA, while the rare, highest tone serves as a control. . . . . 67
- 3.2 **Mouse electrode implantation in M1 and viral infection and optical fiber implantation in VTA.** Each mouse receives a unilateral microwire array implant in motor cortex (targeted to layer V) and a contralateral optical fiber implant in VTA. A coronal brain slice is shown depicting viral infection specific to the dopaminergic cells of the VTA. The immunohistochemistry labels for tyrosine hydroxylase (TH, red) and the cre-dependent fluorescent protein (YFP, yellow). 68
- 3.3 **BMI decoder calibration. (A)** Every session during the baseline period, 500 samples of 500 ms spike counts are collected from spontaneous neural activity as the mouse freely behaves in the box with no task or auditory tones. Each ensemble's firing rate modulation is defined as the sum of the member neurons' normalized spike counts (mean-centered, range-normalized) and then quantized into 4 activation states. The decoder's state is the difference between ensemble 1 and ensemble 2's activation state and is mapped into one of 7 tones. **(B)** BMI calibration on baseline period spontaneous neural activity results in a Gaussian-like distribution over tones, such that Target 1 (5 kHz) and Target 2 (19 kHz) are rare. The mean and s.e.m. baseline distribution for each session is plotted on the left, averaged over all animals. Baseline distributions show no significant change from session 1, as shown on the right. . . . . 78
- 3.4 **Target neural patterns for the BMI.** Ensemble 1 and 2 firing rate modulation before Target 1 and Target 2 hits, averaged over all recorded cells and sessions. Only Target 1 resulted in VTA self-stimulation. . . . . 79
- 3.5 **BMI trial structure.** Subjects initiate trials by entering the center tone. If subjects do not complete a trial within 60s, a timeout white noise tone plays for 1s. When subjects hit a target, the target tone cue plays for 1s, a 0.5s pause ensues, and then subjects receive Trial structure is the same for Target 1 and Target 2, except that Target 1 hit results in phasic VTA stimulation (2s train of 14 Hz pulses with 10 ms width). . . . . 79
- 3.6 **Target pattern reentrance increases during VTA optogenetic self-stimulation.** 81

### 3.6 Target pattern reentrance increases during VTA optogenetic self-stimulation.

(A) Distribution of the percent of time that each tone was occupied during baseline (gray) and BMI (cyan) blocks of Session 1 (left) and Session 4 (right) in an example mouse. (Note that no tones were actually played during the baseline block.) (B) Quantification of the behavioral changes between Session 1 and 4 for the example mouse. The Session 4 Occupancy Gain (cyan) is the Session 4 BMI distribution normalized to the Session 4 baseline distribution, then normalized to the Session 1 ratio. The 95% confidence interval of the Occupancy Gain baseline bootstrap distribution is plotted in gray. To generate the bootstrap distribution, the BMI session was simulated 10,000 times as though neural activity were drawn from that session's baseline period. (C) The Occupancy Gain is plotted over sessions 2 through 4. The 95% confidence interval of the baseline bootstrap distribution is shown in gray. The mean and s.e.m. over ChR2 animals (N=10) is shown in cyan (top) and over YFP animals (N=6) is shown in black (bottom). By Session 4, the behavioral changes were statistically different across tones for ChR2 animals but not YFP (repeated measures ANOVA, ChR2:  $F(6,48)=3.46$ ,  $p=6.4e-3$ ; YFP:  $F(6,30)=0.96$ ,  $p=0.47$ ). On Session 4, 5 kHz (Target 1) was significantly different from all tones from 8 kHz to 19 kHz (Tukey's post hoc multiple comparisons test).

82

3.7 BMI trial structure. (A) Top: the Occupancy Gain for 5 kHz (Target 1) is plotted over sessions. The 95% confidence interval of the bootstrap distribution is plotted in gray. Middle: The average of ChR2 animals (cyan) shows a significantly larger effect than expected from bootstrap from Session 2-4 (Session 2:  $p=1.2e-3$ , Session 3:  $p<1e-5$ , Session 4:  $p<1e-5$ ). Bottom: YFP animals (black) never showed an effect significantly larger than expected from bootstrap. (B) Top: the Preference Gain for 5 kHz (Target 1) versus 19 kHz (Target 2) is plotted over sessions. This is calculated as the increase factor of the probability of hitting Target 1 given that a target was hit (see methods). The 95% confidence interval of the bootstrap distribution is plotted in gray. Middle: ChR2 animals (cyan) were significantly larger than expected from bootstrap after Session 1 ( $p<1e-5$  for Sessions 2-4). Bottom: YFP animals (black) were never significantly larger than bootstrap. (C) Top: the Preference Gain for low tones (5-8 kHz, close to Target 1) versus high tones (12-19 kHz, close to Target 2) is plotted over sessions. This is calculated as the increase factor of the probability of hitting a low tone given that the center tone was not produced. Middle: ChR2 animals (cyan) were significantly larger than expected from bootstrap after Session 1 ( $p<1e-5$  for Sessions 2-4). Bottom: YFP animals (black) were never significantly larger than bootstrap. For A-C, a star indicates that the population average is significantly larger than expected from the baseline bootstrap distribution.

83

- 3.8 **Analysis of fine-timescale neural variance preceding target hit (A)** The decoder maps spike counts in 500 ms bins into quantizations of (Ensemble 1, Ensemble 2) space. Neural activity can take multiple routes to achieve Target 1. **(B)** We analyzed variance of spike counts at fine-timescale (100 ms bins) in a 3s window preceding target hit. **(C)** Factor Analysis was used to analyze the ratio of shared variance to total variance (SOT), which ranges from 0 to 1, for the full population controlling the BMI. A 2-neuron illustration shows a neural solution with SOT=0, 0.6, and 1. . . . . 85
- 3.9 **Learning correlates with increase in covariance of the neurons that produce the target pattern.** . . . . . 86
- 3.9 **Learning correlates with increase in covariance of the neurons that produce the target pattern. (A)** Correlation of change in shared variance before Target 1 hit (Neural Covariance Gain) with change in preference for Target 1 over Target 2 (Learning), over Session 2, 3 and 4. ChR2 animals (left) showed a significant correlation (ChR2 S4:  $r=0.86$ ,  $p=6.1e-3$ ; ChR2 pool S3, S4:  $r=0.71$ ,  $p=1.0e-3$ ; ChR2 pool S2, S3, S4:  $r=0.62$ ,  $p=9.8e-4$ ; ChR2 S3:  $r=0.60$ ,  $p=6.5e-2$ ; ChR2 S2:  $r=0.62$ ,  $p=1.3e-1$ ) while YFP animals (right) showed no correlation (YFP pool S2, S3, S4:  $r=-0.14$ , n.s.  $p=6.4e-1$ ; YFP S4:  $r=-0.32$ ,  $p=6.0e-1$ ; YFP S3:  $r=-0.69$ ,  $p=5.1e-1$ ; YFP S2:  $r=0.37$ ,  $p=5.4e-1$ ). **(B)** SOT of direct and indirect cells was plotted over sessions for ChR2 Learners (left, N=5), ChR2 Poor Learners (middle, N=5), and YFP subjects (right, N=5). ChR2 Learners are animals in the ChR2 group who individually showed significant Preference Gain for T1 versus T2 on both Session 3 and 4. ChR2 Poor Learners constitute the remaining animals who as a population showed significant T1 Occupancy Gain on Session 3 and 4 but did not show significant Preference Gain for T1 versus T2. For direct cells, ChR2 Learners SOT increased from early (Session 1 and 2 pooled) to late training (Session 3 and 4 pooled), while ChR2 Poor Learners and YFP did not (one-sided rank sum test; ChR2 Learners: early < late  $p=1.6e-2$ , ChR2 Poor Learners: early < late n.s.  $p=2.1e-1$ , YFP: early < late n.s.  $p=8.3e-1$ ). Raw SOT pooled over all ChR2 animals also showed significant increase from early to late training (ChR2: early < late  $p=1.7e-2$ ). For indirect cells, SOT showed no change for all groups (ChR2 Learners: early < late n.s.  $p=4.3e-1$ , ChR2 Poor learners: early < late n.s.  $p=2.7e-1$ , YFP: early < late n.s.  $p=7.1e-1$ ) Traces in the insets shows the average of each animals SOT in Session 1 and 2 (early) versus the average of Session 3 and 4 (late). . . . . 87

- 3.10 **Covariance of the neurons that produce the target pattern gradually aligns to the decoder.** (A) Analysis of shared variance alignment with the decoder’s Ensemble 1 and Ensemble 2 assignments via the angle between the shared space and the decoder’s “Ensemble 1 minus Ensemble 2” axis. (B) The angle between shared variance and the decoder axis decreased for ChR2 Learners (left), but not Poor Learners (middle) and YFP (right): (one-sided rank sum test comparing Session 1 and 2 to Session 3 and 4; ChR2 Learner: late > early  $p=2.8e-3$ ; ChR2 Poor Learner: late > early n.s.  $p=3.7e-1$ ; YFP: late > early n.s.  $p=7.5e-1$ ). Traces in the insets show the average of each animal’s angle in Session 1 and 2 (early) versus the average of Session 3 and 4 (late). . . . . 88
- 3.11 **Optical fiber placement.** Circles represent the tops of optical fibers. The implantation hemisphere was randomly assigned to each animal (left hemisphere: ChR2  $n=5$ , YFP  $n=2$ ; right hemisphere ChR2  $n=5$ , YFP  $n=4$ ). VTA, ventral tegmental area. SNc, substantia nigra pars compacta. SNr, substantia nigra pars reticulata. . . . . 91
- 3.12 **Verification of optogenetic VTA stimulation in a motor reinforcement task.** (A) ChR2 mice ( $N=2$ ) were trained on an operant lever pressing task in which active lever presses triggered optogenetic VTA stimulation while inactive lever presses resulted in no reinforcement. Performance increased over sessions. 92
- 3.13 **Number of recorded neurons and SNR waveforms.** (A) The number of recorded neurons for ChR2 and YFP groups. (B) The number of direct neurons assigned to control the brain-machine interface. In each experiment, there are 2-4 E1 and E2 cells. (C) The SNR of each recorded neurons waveform, calculated as in (Suner et al., 2005), which demonstrates well-isolated single units have SNR above 4. . . . . 93
- 3.14 **Neuroprosthetic-triggered VTA stimulation does not reinforce place preference or specific overt movements.** . . . . . 94

- 3.14 Neuroprosthetic-triggered VTA stimulation does not reinforce place preference or specific overt movements.** (A) An example mouse's head position (left) is shown at time of Target 1 hit for an early session (session 1, middle) and a late session (session 3, right). (B) The variance of position at Target 1 hit does not change for ChR2 and YFP animals (rank sum test, session 1 and 2 vs session 3 and 4). There is no difference between ChR2 and YFP animals (rank sum test, ChR2 vs YFP animals pooling sessions). (C) 1.5 second videos preceding Target 1 hit were aligned to the mouse's position and head-back orientation at Target 1 hit (see methods) and averaged over frames and trials. An example ChR2 mouse is shown over 4 sessions. (D) The mean absolute pixel change in the 1.5 second videos preceding Target 1 hit is shown for an example ChR2 mouse over 4 sessions. (E) The spread of the mean pixel change was calculated via the entropy of the mean pixel change image. It does not change from early to late training (rank sum test, session 1 and 2 vs session 3 and 4). The spread of ChR2 mean pixel change is larger than YFP (rank sum test, ChR2 > YFP,  $p=2.6e-13$ , pooling over sessions), indicating ChR2 animals do not execute a specific movement before Target 1 hit. (F) The trial-to-trial similarity of the absolute pixel change (see methods) does not change for ChR2 and YFP animals (rank sum test, session 1 and 2 vs session 3 and 4). There is no difference between ChR2 and YFP animals (rank sum test, ChR2 vs YFP animals pooling sessions), indicating ChR2 animals' movements do not become more similar from trial-to-trial. . . . . 95
- 3.15 Number of shared dimensions for FA.** The number of shared dimensions (i.e. number of factors, equivalent to the rank of the shared covariance matrix) for FA was fit using model selection by evaluating cross-validated log-likelihood. (A) The number of shared dimensions for ChR2 and YFP groups. (B) The number of shared dimensions for the ChR2 Learner and ChR2 Poor Learner groups. . . . . 96
- 3.16 Behavior of ChR2 Learners and Poor Learners.** . . . . . 97

- 3.16 **Behavior of ChR2 Learners and Poor Learners.** The ChR2 group (n=10) was split into two groups based on their degree of learning: Learners (n=5) and Poor Learners (n=5). Each individual of the Learners group showed significant Preference Gain for T1 versus T2 for both Session 3 and 4. Poor Learners as a group showed significant T1 Occupancy Gain for both Session 3 and 4, but they also showed significant T2 Occupancy Gain, and thus did not show significant Preference Gain for T1 versus T2. Learners and Non-Learners' performance pooling over Session 3 and 4 was significantly different (non-parametric rank sum test: T1 Occupancy Gain  $p=1.7e-2$ ; T1 Preference Gain  $p=2.6e-4$ ; Low vs High Tones Preference Gain  $p=4.1e-3$ ). (A) Occupancy Gain for Learners over sessions 2 through 4. The 95% confidence interval of the baseline bootstrap distribution is shown in gray. (B) Occupancy Gain for Poor Learners over sessions 2 through 4. (C) The Occupancy Gain for Target 1 (5 kHz). (D) Preference Gain for Target 1 (5kHz) versus Target 2 (19kHz). (E) Preference Gain for low tones (5-8kHz, close to Target 1) versus high tones (12-19 kHz, close to Target 2). . . . . 98
- 3.17 **Covariance gain over learning.** (A) The covariance gain increased for ChR2 Learners, but not for ChR2 Poor Learners and YFP. This covariance gain correlated with learning (Figure 3D). . . . . 99
- 3.18 **Significance of FA results comparing experimental data to shuffled distributions.** A shuffled neural activity data set was generated by shuffling the time labels of each neuron's binned spike count. This maintains each neuron's marginal spike count distribution, but breaks the population's joint distribution. FA was performed on 1000 shuffles of each data set to generate the FA shuffle distributions, and 95% confidence intervals are plotted in gray. (A) In ChR2 animals, SOT starts roughly at chance and increases. YFP animals are above chance over training. (B) In ChR2 animals and ChR2 Learners, the angle between the shared space and the decoder axis decreases below chance. In ChR2 Non-Learners and YFP animals, the angle remains roughly at chance. Colored errorbars are s.e.m. over animals. The gray shuffle distribution is the 95% confidence interval. 100
- 3.19 **FA fit on the full population of recorded motor cortex neurons, including direct and indirect neurons.** . . . . . 101
- 3.19 **FA fit on the full population of recorded motor cortex neurons, including direct and indirect neurons.** (A) For these analyses, FA was fit on the full population of recorded motor cortex neurons, including direct and indirect neurons. (B) After decomposing the spike count covariance matrix into private and shared variance, we investigated how much shared and private variance the direct neurons have and how much the indirect neurons have. The direct neurons' shared variance includes variance shared with indirect neurons. (C) Trends are consistent with (Figure 3.9B), in which the direct cells' shared variance is plotted from fitting FA only on direct neurons. (D) Covariance gain trends using FA fit on the fully-recorded neural population. (E) Shared dimensionality yielded by FA model fitting for Learners, Poor Learners, and YFP. . . . . 102

3.20	<b>Significance of FA fit on all neurons (direct plus indirect) compared to shuffled distributions</b> For these analyses, FA was fit on the full population of recorded motor cortex neurons, including direct and indirect neurons. <b>(A)</b> SOT for direct cells. The SOT is above chance for all data, other than Chr2 Non-Learners on Session 2. Only Chr2 Learners show an SOT with increasing trend over training. <b>(B)</b> SOT for indirect cells. . . . .	103
3.21	<b>Individual neuron firing properties over training do not correlate with learning.</b> . . . . .	104
3.21	<b>Individual neuron firing properties over training do not correlate with learning.</b> Each analysis in this figure other than waveform SNR was performed using neural activity in the 3-second window preceding Target 1 hit binned in 100 ms bins. The s.e.m. is over animals within a session. Changes over training were analyzed with a Wilcoxon rank sum test between early and late sessions. <b>(A)</b> The SNR of neural recordings do not change over learning. <b>(B)</b> The average firing rate of direct neurons does not change over training. <b>(C)</b> The average fine-timescale modulation of direct neurons does not change over training. This was calculated as the norm of each z-scored PSTH and then averaged over each animal's neurons. <b>(D)</b> Fine-timescale firing rate variance does not change over training. <b>(E)</b> The fine-timescale firing rate Fano Factor (the variance divided by the mean using the 3-second window preceding Target 1 hit over 100 ms bins) does not change over training. . . . .	105
3.22	<b>No effect of laser stimulation on average M1 firing rates.</b> . . . . .	106
3.22	<b>No effect of laser stimulation on average M1 firing rates.</b> In this figure, red dotted lines indicate laser pulse times. Gray lines indicate individual neuron PSTHs. Neurons recorded on separate sessions were considered separate. <b>(A)</b> In Chr2 animals, laser onset does not change the firing rate of direct or indirect cells calculated in 100 ms bins. There were n=89 Ensemble 1 neurons, n=89 Ensemble 2 neurons, and n=550 indirect neurons, pooling over animals and sessions. <b>(B)</b> In Chr2 animals, laser pulses do not trigger phasic firing responses. PSTHs calculated in 1 ms bins. <b>(C)</b> In YFP animals, there is no average firing rate change after laser onset. There were n=48 Ensemble 1 neurons, n=48 Ensemble 2 neurons, and n=400 indirect neurons, pooling over animals and sessions. <b>(D)</b> In YFP animals, there are no phasic firing responses locked to laser pulses. . . .	107
4.1	<b>Rat V1 histology and recordings.</b> <b>(A)</b> Left: schematic showing the location of rat V1 electrode arrays, verified by histology for 6 of 10 animals. Right image shows an example section with electrode tracks. <b>(B)</b> Sample V1 waveforms from an example session. Units outlined in green were chosen for Ensemble 1; units outlined in blue were chosen for Ensemble 2; and units outlined in black were recorded but had no direct relationship to cursor control (Indirect units). . . .	110

- 4.2 **Rat DMS histology and number of recordings.** (A) Left: schematic showing the location of rat DMS electrode arrays, verified by histology for 6 of 10 animals. Right image shows an example section with electrode tracks. (B) Total number of well-isolated DMS units for 8 animals recorded across several training sessions. Solid line shows the mean and s.e.m.; grey lines show data for individual animals. . . . . 111
- 4.3 **Decoder calibration.** (A) Plot shows a histogram of binned E1-E2 values (i.e. total spike counts for all ensemble 1 neurons minus total spike counts for all ensemble 2 neurons for each 100 ms bin) during a 15-minute baseline period. A mixture of gaussians (dashed lines) are fitted to generate the PDF (solid black line). Based on a simulation of the task using recorded baseline neural activity, thresholds are calculated for a low target (unrewarded in this plot), a high target (rewarded), and a midpoint (Baseline). (B) The line fitted to the T2, Baseline, and T1 values that is used to compute a feedback frequency based on the binned E1-E2 value. . . . . 113
- 4.4 **Jaws inhibition of DMS.** (A) Jaws-mediated inhibition of DMS neurons. Inset shows a coronal section stained for neuronal cell bodies (blue) and Jaws expression (red). Zoomed inset shows a magnification with labeling apparent on single neurons (white arrows). Histogram plot shows suppression of spike activity in one example DMS neuron during Jaws-mediated inhibition (red bar). (B) Mean firing rates of all recorded DMS units during Jaws-induced inhibition with red light (red bar); N=5. Firing rates were normalized to baseline firing rate values. Shaded areas indicated s.e.m. (C) Fraction of inhibition for recorded DMS units during Jaws-induced inhibition, where a value of 1 indicates complete inhibition of all spike activity. Grey circles show individual units. . . . . 115
- 4.5 **V1 BMI paradigm.** Activity of well-isolated V1 units (top left) were used to generate auditory tones using a differential transform (top right). Animals were rewarded for producing a target tone (red). A second tone (black) at the opposite end of the frequency range terminated the trial but was not rewarded. . . . . 117
- 4.6 **Average Ensemble 1 and 2 modulations preceding target hit.** (A) Average z-scored firing rates of V1 neurons arbitrarily assigned to Ensemble 1 (green), Ensemble 2 (blue), or unassigned (indirect; black), time-locked to rewarded targets. Shaded areas show s.e.m. (B) Same as (A), but time-locked to the unrewarded target. Increased activity in Ensemble 1 units moved the tone frequency in the opposite direction as increased activity in Ensemble 2 units. . . . . 118

- 4.7 **Task performance over training.** **(A)** Timecourse of learning across training days. Bold line shows the mean and s.e.m. across 9 rats; grey lines show learning curves for animals individually. One animal only completed 4 sessions; data for this animal has been excluded from this plot. Dashed lines bound the range of chance performance levels. Orange highlighted region shows data range classified as the early learning phase (first 3 days) for all animals; red region shows range for late learning phase (last 3 days). **(B)** Comparison of performance between early learning phase and late learning phase. (N=9 animals, early mean = 39.6% rewarded; late mean = 67.2%; P=0.00162 (paired t-test)) . . . . . 119
- 4.8 **Time to target hit.** Latencies to rewarded vs unrewarded targets for early training (first three days of training) compared to late training (last three days of training). Mean rewarded targets, early =21.4 sec; mean rewarded late = 18.0 sec; P=.0156 (paired t-test). Mean unrewarded targets, early=20.7 sec; unrewarded late=18.1; P=0.068 (paired t-test). . . . . 120
- 4.9 **Neural modulations during passive tone playback.** **(A)** Ensemble 1 (green) and Ensemble 2 (blue) modulations during passive playback of tones decoupled from ongoing neural activity, time-locked to the rewarded target tone. **(B)** Modulation depth of Ensemble 1 and Ensemble 2 during online control compared to tone playback. Mean modulation depth online control = 4.34; mean tone playback = 1.74; P=0.00035 (paired t-test). Black bars show mean and s.e.m. . . . 121
- 4.10 **Within-session learning increases over training sessions.** . . . . . 123

- 4.10 **Within-session learning increases over training sessions.** (A) Comparison of within-session improvements during the early learning period (first 3 days) of training relative to the late learning period (last 3 days) for each animal (N = 9 animals), expressed as the ratio of rewarded to unrewarded targets. Data to compute the ratio for first trials (shaded in cyan) was averaged over the first 10 mins of each session, while data for the last trials (shaded in blue) was averaged over the last 10 mins. During the early learning period, the mean rewarded/unrewarded ratio for first trials was 1.12, last trials session = 3.848; P = 0.275 (paired t-test). For the late period, mean ratio early in session = 1.195, mean late in session = 11.667; P = 0.00130 (paired t-test). (B) Mean distribution of cursor values for all animals for the first trials (first 10 mins) compared to the last trials (last 10 mins) of sessions during the late period. Cyan bars show the initial distribution, based on baseline activity, used to set the task parameters, while blue bars show the distribution at the end of the training session for the last trials. Dashed lines show the thresholds corresponding to the rewarded and unrewarded targets. (C) Timecourse of mean within-session task learning during the late period of training for all animals. Shaded areas represent s.e.m. Chance rates for rewarded and unrewarded targets were set at approximately 30% at the start of each training session. Comparisons in performance were made during over first 10 mins of each session (first trials, cyan bar) and over the last 10 minutes (last trials, blue bar) of each session. N = 10 animals, mean of 3 sessions per animal. (D) Quantification of rewarded and unrewarded target hits for the first trials compared to the last trials; data same as c. N = 10 animals, mean of the late period (last 3 sessions) for each animal. Paired t-test between first trials and last trials for rewarded targets: P = 0.00055; mean early = 36.9%, mean late = 59.1%. Paired t-test between first trials and last trials for unrewarded targets: P = 0.00145; mean early = 33.0%, mean late = 20.6%. Red and blue indicate the rewarded and unrewarded targets, respectively (\*\* indicates P < 0.001). Black lines show mean and s.e.m. . . . . . 124
- 4.11 **Animals do not learn without auditory feedback.** (A) Timecourse of performance when animals were not given auditory feedback about cursor position during training. (B) Comparison of rewarded and unrewarded target hits during sessions without auditory feedback. Data is the mean over the last 10 minutes of each session. N = 7 animals (mean of 1.5 sessions per animal; range 1-2 sessions). Mean rewarded targets = 38.1%; mean unrewarded = 41.7%; P = 0.738 (unpaired t-test). Horizontal black line shows mean and vertical bars show standard error. 125

- 4.12 **V1 modulation is sensitive to task contingency.** (A) Percentage of rewarded trials for contingency degradation sessions. Bars show means, lines show individual animals. N=8 animals, mean of 2 sessions per animal. P: pre-degradation, mean=64.5%. CD: peri-degradation (reward decoupled from cursor), mean = 43.1%. R; reinstatement of reward, mean = 61.4%. Paired t-test between pre- (P) and peri- (CD) degradation:  $P = 0.0038$ . Paired t-test between CD and reinstatement (R):  $P = 0.0283$ . Paired t-test between pre-degradation (P) and reinstatement (R):  $P = 0.289$ . (B) Quantification of performance in contingency reversal sessions; N = 7 animals; average of 1.57 sessions per animal (range 1-2). P: pre-reversal (mean = 85.8%); Rev: perireversal (mean = 47.8%); R: recovery of performance with the decoder still reversed (mean = 81.2%). Paired t-test between P and Rev:  $P = 0.0435$ ; paired t-test between Rev and R:  $P = 0.447$ ; paired t-test between P and R:  $P = 0.689$ . . . . . 127
- 4.13 **V1 modulation is sensitive to changes in visual context.** . . . . . 129
- 4.13 **V1 modulation is sensitive to changes in visual context.** (A) Timecourse showing the mean percentage of rewarded trials within all sessions under lighted conditions (yellow) compared to dark conditions (black). Shaded areas represent s.e.m. (B) Quantification of data in A, using the first trials and last trials in a session. Train light: N = 4 animals; average of 7.5 sessions per animal (range = 5 to 10 sessions). Paired t-test between first trials and last trials:  $P = 0.0113$ ; mean first trials = 37.8%; mean last trials = 74.0%. Train dark: N = 6 animals; 8 sessions per animal. Paired t-test between first trials and last trials:  $P = 0.00167$ ; mean early = 0.396; mean late = 0.635. Unpaired t-test between last trials for light sessions and last trials for dark sessions:  $P = 0.205$  (\* indicates  $P < .05$ ; black crosses show mean and s.e.m.). (C) Timecourse showing the mean percentage of rewarded trials when animals learned a decoder under dark conditions, and were switched to a lighted condition mid-session (“train dark, test light”). (D) Mean percentage of rewarded trials when animals were trained in dark, and then tested in the light (same data as c). N = 5 animals, mean of 1.8 sessions per animal (range 1-2). Data taken from last 15 mins of dark training and first 15 mins of light testing. Mean train dark: 87.4%, mean test light 42.6%;  $P = 0.0309$  (paired t-test). Error bars show s.e.m.; horizontal lines show mean (\* indicates  $P < 0.05$ ). (E) Mean percentage of rewarded trials when animals were trained in the light and tested in the dark. N = 4 animals, mean of 1.5 sessions per animal (range 1-2). Mean train light: 77.9%. Mean test dark: 35.4%.  $P = 0.043$  (paired t-test). . . . . 130
- 4.14 **V1 firing rates are sensitive to visual context.** Mean spike rates for all V1 units when the light conditions were changed midsession. N = 6 rats. Paired t-test between light and dark conditions:  $P = 0.0314$ . . . . . 130
- 4.15 **Changes in neural dynamics over learning.** . . . . . 132

- 4.15 **Changes in neural dynamics over learning.** **(A)** Mean pairwise correlations between units within Ensemble 1 (green) or within Ensemble 2 (blue) during training. Correlations were calculated using 1 ms bins. **(B)** Change in normalized correlation coefficients (cc) from the first trials to the last trials within sessions. CC calculated between units within Ensemble 1 (mean change = 4.418;  $P = 0.0417$ ; paired t-test), within Ensemble 2 (mean change = 0.942;  $P = 0.775$ ; paired t-test), or between units in ensembles 1 and 2 (mean change = 1.361;  $P = 0.491$ ; paired t-test).  $N = 10$  rats; mean of 4.4 sessions per rat (range 3-6). **(C)** Example data from one session showing the relationship between changes in correlation of Ensemble 1 units (green) and task performance (percent of correct trials, black). **(D)** Mean correlation between Ensemble 1 units during a 2 second window prior to target hits or timeouts. Mean cc prior to rewarded targets = 0.0285; mean cc prior to unrewarded targets = 0.0241 ( $P = 0.00167$ ; paired t-test). Mean cc prior to timeouts: 0.0240. Comparison between rewarded targets and timeouts:  $P = 0.0362$ ; comparison between unrewarded targets and timeouts:  $P = 0.549$ . **(E)** Mean spectrogram of V1 LFP time-locked to rewarded target hits. **(F)** Spike-field coherence between direct units and V1 LFP for late learning periods during task performance (red) compared to nonengaged passive behavior (blue). Shaded area represents s.e.m. **(G)** Spike-field coherograms showing the evolution of coherence over time for LFPs in V1 and spikes from direct (combined Ensemble 1 and Ensemble 2) units in during early training periods (days 1-3; top plot), late training periods (last 3 days, middle plot), and indirect units (no direct relationship to cursor control) for late periods (bottom plot) time-locked to rewarded targets. **(H)** Coherence between direct units and V1-LFP in the 10-25 Hz band for early training compared to late training. Solid lines show the mean for 10 animals and shaded areas represent s.e.m. . . . . . 133
- 4.16 **Comparing E1 and E2 neural dynamics over learning.** **(A)** Comparison between spike-field coherence between E1 or E2 units and V1 LFP in late training. **(B)** Spike-field coherence between E1 or E2 units time-locked to rewarded targets and isolated between 10 and 25 Hz for early training periods. **(C)** Coherence as in (B) for late training periods. . . . . . 134

- 4.17 **Striatal activity becomes task-related with learning.** (A) Raster plot of an example DMS unit time-locked to V1 ICMS. (B) histogram of evoked spike counts from unit shown in (A). (C) Mean z-scored firing rate of units recorded in the dorsal medial striatum (DMS) time-locked to rewarded targets for early compared to late learning. (D) Modulation depth of DMS units in a 1-second window surrounding rewarded target hits. Mean for early periods = 2.872; mean for late periods = 3.691;  $P = 0.00015$ . (E) Proportion of variance of direct unit spikes in V1 explained by DMS unit spikes in a 500 ms window prior to rewarded target hits, computed using a linear regression analysis on individual training days. Solid black lines show mean and s.e.m.; grey lines show data for individual animals. Mean of first two days = 0.0788; mean of last 2 days = 0.276;  $P = 0.0196$  (paired ttest). (F) Mean spectrogram of DMS LFP time-locked to rewarded target hits for all animals during the early period (sessions 1-3), left), compared to the late period (last 3 sessions, right). (G) Field-field coherograms between V1 LFP and DMS LFP for early (left) compared to late training (right) sessions time-locked to rewarded targets. (H) Field-field coherence between V1 and DMS in the 0-5 Hz band in for early and late training, time locked to rewarded targets. Solid lines show mean for all animals; shaded areas represent s.e.m. . . . 136
- 4.18 **Striatal inhibition impairs learning.** (A) Schematic showing the replication of rat experiments in Jaws-injected mice. Left: mice were injected bilaterally with rAAV8/CamKII-Jaws-KGC-GFP-ER2. Middle: After Jaws expression stabilized, mice were implanted unilaterally with an electrode array in V1 L5 and bilaterally in DMS. Right: animals were trained on the same task as rats (see Figure 4.5) with the addition of optogenetic inhibition. Bottom: timecourse of experiments. LED 50% indicates that Jaws was activated via red LED light in 50% of trials. (B) Success rates for LED-on VS LED-off trials in Jaws-injected animals across training sessions. (C) Percentage of rewarded trials for Jaws compared to control GFP mice over the course of several days of training. Red bar indicates sessions where the LED was active on 50% of trials for both groups (LED 50). (D) Quantification of performance across days for Jaws-injected animals (red;  $N = 4$ ) and GFP injected animals (green,  $N = 4$ ) Each time period is a mean over 3 sessions. Mean GFP, first 3 sessions = 39.7%, Mean GFP, last 3 sessions = 69.7%;  $P = 0.00072$  (paired t-test). Mean Jaws, first 3 sessions = 29.3%, mean Jaws, last 3 sessions LED 50 = 33.3%;  $P = 0.183$ . Mean jaws, last 3 sessions LED off = 62.1%. Paired t-test between Jaws last 3 sessions LED 50 and Jaws LED off:  $P = 0.00131$ . (EF) Mean spectrogram of V1 LFP in Jaws animals time-locked to rewarded targets for LED-50 sessions (E) and LED-off sessions (F). LED-50 sessions include both trials with and without Jaws activation. . . . . 138

- 4.19 **Striatal inhibition impairs within-session learning but does not impair execution.** **(A)** Mean gamma power (25-60 Hz, solid lines) in V1 time-locked to rewarded targets during LED-50 vs LED-off sessions. Shaded areas show s.e.m. **(B)** Mean performance of trained Jaws-injected animals after several days of training. Data is plotted in the order that the training sessions occurred. Black bars indicate sessions without LED activation. Striped red bars indicate sessions where the red LED was active on 50% of trials. N = 4 animals. Mean LED off, first session = 63.3%; mean LED 50 % (second session) = 38.9%; mean LED off, last session = 59.6%. Paired t-test between LED off, first session and LED 50%: P = 0.0122. Paired t-test between LED 50% and LED off, second last session: P = 0.0259. Paired t-test between LED off, first and last sessions: P = 0.502. . . . 140
- 4.20 **Indirect neuron activity decreases modulation depth but predicts target choice.** **(AB)** Absolute z-scored firing rates for E1, E2 and indirect units, time-locked to rewarded (A) and unrewarded (B) target hits. **(C)** Mean modulation depth of direct (pooled E1 and E2) units and indirect units in early (sessions 1-3) and late (sessions 8-10) training. Data taken during a 3-second window prior to rewarded target hits. Mean direct early = 3.988, mean direct late = 3.942, P = 0.743 (unpaired t-test). Mean indirect early = 3.135, mean indirect late = 2.718, P = 0.0218 (unpaired t-test). **(D)** Total number of indirect units recorded across training sessions. Black lines show mean and s.e.m.; grey lines show values for individual animals. **(E)** Mean percent of single indirect units with significant predictability for target choice over all training days. Significance determined by permutation test with P < 0.05. Mean = 10.782%. Black lines show mean and s.e.m. for all animals; dots show mean for each animal individually. **(F)** Prediction of target choice using population activity. Direct (E1 and E2) units or indirect units were used to predict target choice using a logistic regression analysis. Spike data was taken in a 500 ms window prior to target hits. Grey bars show chance accuracy rates calculated by permutation test. Mean direct = 0.964 ; mean indirect = 0.729. . . . . 144

## Acknowledgments

I want to start by acknowledging my great honor to write acknowledgements. I have always wanted to do something worthy of thanking people for; it appears my time has come.

Jose and Rui: I came to both of you not knowing what I was doing. How do I begin to thank you for giving me a chance at something I wanted so badly? You gave me all the freedom, time, and resources to thoroughly lose myself, and you gave me the mentorship, enthusiasm, butt kicks, and love to help me find my way. Thank you for the lesson: focus on the goal, and the way will reveal itself. You both mean so much to me.

Jose, your spark and your vision never fail to inspire me. You have shown me the beauty and potential of scholarship built on excitement and passion. Of course, beautiful potential is nothing without the relentless drive to finish. May I ever improve at your ability to ‘just doooo it.’

Rui, I cherish your pure joy, beginner’s mind, and expansive curiosity. For many things, thank you. You taught me: slow down, we’re in a rush. It has been my honor to travel in a rush with you and Jose, may we travel far.

Kannan, Bruno, Michel, and Miki: thank you for advising. I admire you greatly as teachers, researchers, and people.

Carmena and Costa Lab Mates: I love you dearly. You have been my family during my PhD. Collaborators Preeya, Ryan, Aaron, and Fernando: many of my best memories are from working and hanging out with you. Preeya: nir nir nir psh. Ryan: ah, drinks after surgery training completion. Aaron: from Pikachu to SM\*, we went a long way. Fernando: naco na pedra, my first dinner in Lisbon.

Thanks: Kelvin, Sid, and Suraj: For your mentorship and friendship, you are invaluable. Ana Vaz, Mariana, Gabi, Cristina, Teresa: For rescue at many random times. Fatuel, Joaquim, and Andreas: For deep command, wisdom, and comfort. Alex: For profundity. Sevinc: For dancing under pressure and isoflurane. Nuno, Ivo, Rodrigo: For laughter.

Open Lab of Champalimaud Center for the Unknown, Jimmy, Julia, Mert, Tomas, Ines, Sam, Sofia, Diogo, Susana Lima, Asma, Joao, Ivan, Geoff, Ved, Adair, Ritesh: you know what it is.

Shilpa, Mom, and Baba: Finally... there was no path without you, and now we made it. You are my source of deepest support, and my seers into the cosmos.

To PhD students reading this: I extend my love to you. Your day is coming to write acknowledgements :)

# Chapter 1

## Introduction

Millions of people suffer from neurological injuries and diseases which disable movement including spinal cord injury, amputation, stroke, and amyotrophic lateral sclerosis (ALS). People suffering from paralysis strongly desire therapies which restore arm and hand movement (Figure 1.1) and thus provide a sense of independence (Anderson, 2004; Brandman et al., 2017; Tsu et al., 2015). The Brain-Machine Interface (BMI) is an emerging technology promising to restore movement and communication by directly translating neural activity into control signals for effectors such as computers, prosthetics, or even muscles, thus bypassing broken neural pathways. The state-of-the-art BMI can already achieve impressive performance, for example permitting human typing at up to 40 correct characters per minute (Pandarinath et al., 2017).

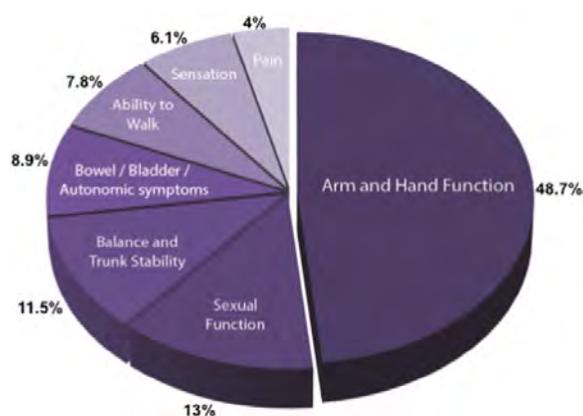


Figure 1.1: **Functional priorities of quadriplegic patients suffering from spinal cord injury.** Figure taken from (Anderson, 2004; Tsu et al., 2015). Pie chart shows the percentage of quadriplegic patients indicating each area as their top priority for functional recovery.

However, to enable widespread clinical use, critical challenges remain regarding the tech-

nology's performance, longevity, and robustness, such as the significant hardware challenge of interfacing with the brain to stably record from many neurons for an entire lifetime. This is not the subject of this thesis. This thesis is concerned with what would happen even with the ideal physical brain-machine link: a conceptual challenge remains in the closed-loop interaction between the brain and the machine. Early efforts to read information out of the brain attempted to model how the brain encodes its intentions for natural movement. However, when the brain uses a BMI to achieve its goals in closed-loop and receives continuous feedback about neuroprosthetic movement, it confronts the underlying truth: the neuroprosthetic is not the natural limb. The neuroprosthetic is a novel limb (Figure 1.3), and to control it, the brain learns a novel skill which can result in dramatic reorganization of neural activity (Carmena et al., 2003; Ganguly and Carmena, 2009; Ganguly et al., 2011). Work over the last decade has shown that high performance BMIs depend on machine learning to adapt parameters for decoding neural activity as the brain uses the BMI (Dangi et al., 2014; Gilja et al., 2012; Shanechi et al., 2016; Taylor et al., 2002), but also on the brain learning to reliably produce desired neural activity patterns (Carmena et al., 2003; Ganguly and Carmena, 2009; Orsborn et al., 2014). The neuroscience of neuroprosthetic skill learning is not well-understood and could inform the design of next-generation BMIs in which both the brain and machine synergistically adapt (Shenoy and Carmena, 2014).

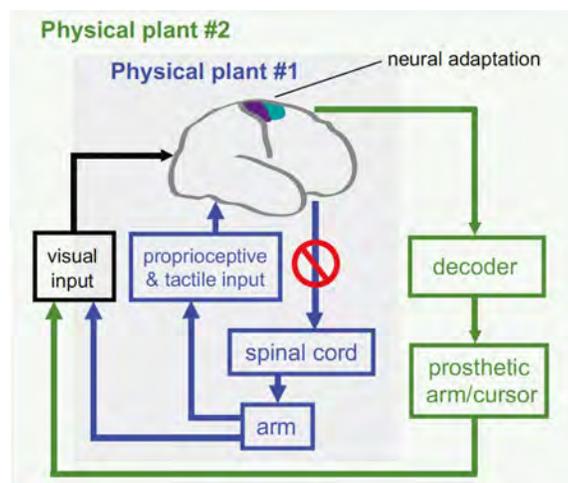


Figure 1.2: **The BMI constitutes a novel limb.** Figure taken from (Shenoy and Carmena, 2014). Physical plant 1 represents the arm of a patient with a severed connection between the brain and spinal cord. Physical plant 2 represents the BMI.

In this thesis, we view neuroprosthetic skill learning from the brain's perspective as a reinforcement learning problem, as the brain must initially explore activity patterns, observe their consequences on the prosthetic, and finally consolidate activity patterns leading to desired outcomes. Notably, the process of neural activity exploration and consolidation may also underlie how the brain learns natural motor skills. Initially, overt movements

and underlying neural activity patterns exhibit variability from attempt to attempt, that decreases over learning as movements (Cohen and Sternad, 2009; Shmuelof et al., 2012; Skinner, 1938; Thorndike, 1898) and neural activity (Cao et al., 2015; Costa et al., 2004; Peters et al., 2014) are re-entered which lead to success. Thus, we will draw connections between neuroprosthetic skill learning and the brains natural mechanisms of motor learning.

This thesis will address 3 questions about how the brain learns neuroprosthetic skill. First, how do the neurons driving neuroprosthetic output coordinate during activity exploration and consolidation? We investigate the degree to which each neuron serves as an independent explorer and controller versus the entire population exploring and controlling in coordination. The neural signals uncovered in this analysis could inform future adaptive BMI systems.

Next, having studied the structure of neural variability over learning, we ask: how can the brain select activity patterns to consolidate? Thorndikes law of effect states that motor behavior leading to reinforcement tend to be repeated more often (Thorndike, 1898). Seeking evidence for a neural law of effect, we performed experiments to determine if pairing neural activity patterns with neural reinforcement signals drives activity consolidation.

Finally, we ask: does the observed mechanism of neural activity exploration and consolidation generalize across cortex? Previous work shows the basal ganglia are involved in motor reinforcement learning (Barnes et al., 2005; Hikosaka et al., 1999; Yin et al., 2009) and motor cortex BMI learning (Koralek et al., 2012), but the basal ganglia receives input from most of cortex (Hintiryan et al., 2016; Kemp and Powell, 1970; McGeorge and Faull, 1989; Webster, 1965) and recurrently feeds back through the thalamus (Redgrave et al., 2011). We performed experiments to test if animals could learn to consolidate activity patterns in primary visual cortex to receive reward, and if this learning depended on the basal ganglia. Together, these results contribute to our understanding of how the brain solves the reinforcement learning problem of exploring and consolidating activity patterns underlying neuroprosthetic skill.

## **Closed-Loop Motor Brain-Machine Interface Systems**

The motor BMI system consists of 4 main components. First, BMI sensors measure the activity of neurons in the brain. Second, a mathematical algorithm called the decoder translates brain activity into intentions and ultimately control signals for an effector. Third, the effector (e.g. computers, prosthetics, or muscles) performs the users' intended action. Fourth, the feedback system provides the user with information about the effectors state in real-time (e.g. visual feedback of a computer screen). Critically, the motor BMI is a closed-loop feedback system, such that a users neural activity produces actions which the user can observe and/or feel and then make corrections or adjust future actions.

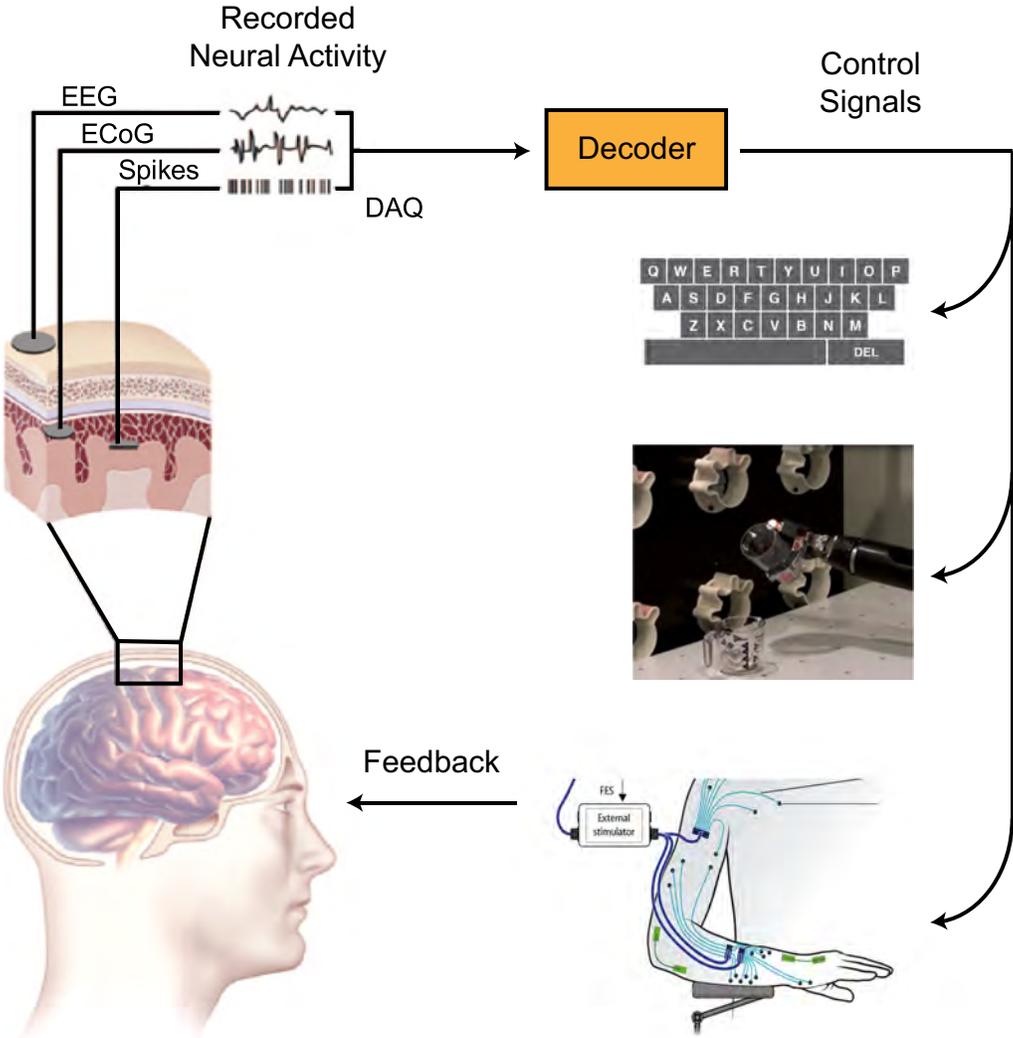


Figure 1.3: **Closed-loop BMI System.** BMI systems can employ sensors which record different neural signals, from single neuron action potentials with microelectrode arrays to aggregated electrical activity with ECoG or EEG sensors. The BMI can then generate control signals for external effectors such as a computer keyboard and mouse, a robotic arm, or a user’s own muscles. Figure adapted from (Ajiboye et al., 2017; Collinger et al., 2013b; Leuthardt et al., 2006; Pandarinath et al., 2017).

BMI performance depends on the signals measured from the brain. Neurons communicate by firing stereotyped electrical impulses known as action potentials, one of the great discoveries of neuroscience (Husser, 2000) (Figure 1.4). In this work, we focus on intra-cortical BMIs which measure the activity of hundreds of individual neurons by utilizing microelectrode arrays implanted into the brain's cortex. While other classes of BMI sensors measure electrical activity aggregated over many neurons, like ECoG electrodes placed on the cortical surface or overlying dura, or EEG electrodes placed on the scalp, the highest performance BMIs to date leverage invasive microelectrode arrays' superior neural and temporal resolution.

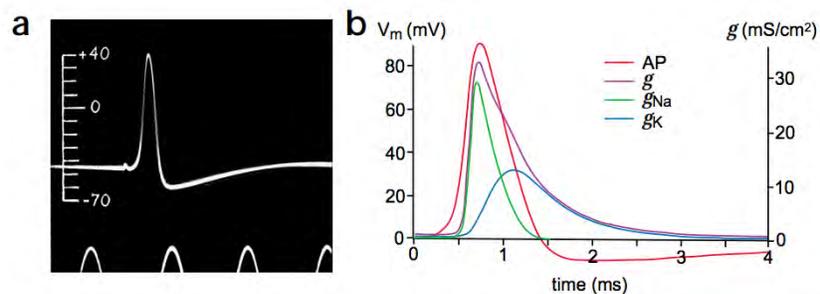


Fig. 1. The action potential: from experiment to theory. (a) The first intracellular recording of an action potential, from squid axon. Time calibration, 2 ms. Modified from ref. 2. (b) Separation of ionic conductances underlying the action potential (AP) in the H-H model. Modified from ref. 6.

Figure 1.4: **Neural action potential.** From (Hausser, 2000), a review of the classic work of Hodgkin and Huxley.

## Foundations of BMIs in the Neuroscience of Motor Control and Learning

Two scientific findings provided basic principles underlying BMI technology development. First, information about natural motor control is encoded in the activity of populations of neurons in motor cortex and can be decoded to reconstruct physical movement (Chapin et al., 1999; Georgopoulos et al., 1986; Humphrey et al., 1970; Wessberg et al., 2000). Researchers wondered if these very signals could be used to control a closed-loop BMI. Thus, so-called ‘biomimetic’ decoders were engineered to take advantage of the natural neural code for movement. Second, in the early 1970s, foundational and prescient experiments were conducted in which single neurons in the motor cortex of behaving monkeys controlled an audiovisual feedback signal (Fetz, 1969; Fetz and Baker, 1973; Fetz and Finocchio, 1971). The monkeys learned to increase single neurons’ firing rates to reach a target value and receive food reward. These studies were the first to demonstrate that animals can learn to directly control neural activity to achieve behavioral outcomes and formed the basis of later experiments showing that animals can improve BMI performance with experience.

## Closed-loop BMI Control Benefits from Machine and Brain Learning

One early prominent view was that good BMI design depended mostly on building models of how neural activity encodes movements (Georgopoulos et al., 1986; Nessler et al., 2011). With improved encoding models, decoders could be built to use neural activity to better reconstruct a subject’s movement with his native body. This “offline prediction”, so-called because the decoder’s output is not being used to control a device online, was thought to be an important predictor of online performance.

After early proofs of principle (Chapin et al., 1999; Serruya et al., 2002), studies asking how 2-dof and 3-dof BMI systems could work revealed new, basic neuroscientific questions. These studies observed that the neural encoding of movement changed when subjects began controlling the BMI in closed loop (i.e. producing movements and correcting errors in real-time), and further, that subjects improved BMI control significantly with practice (Carmena et al., 2003; Musallam et al., 2004; Taylor et al., 2002). How does the brain perform closed-loop control? This question contrasted with early questions about models to better reconstruct native movements offline and gave rise to new algorithms for adapting decoders to the neural activity patterns used in closed-loop BMI control rather than offline calibration (Dangi et al., 2013, 2014; Gilja et al., 2012; Mahmoudi and Sanchez, 2011; Orsborn et al., 2012, 2014; Shanechi et al., 2016; Taylor et al., 2002).

How does the brain learn to control a BMI, how does learning change neural encoding of BMI control, and how does adapting the decoder interact with the brain learning? These questions depend on the basic neuroscience of how the brain learns skills. Initial experiments revealed the brain’s capacity for profound plasticity during BMI training. (Ganguly and Carmena, 2009) paired stable neurons with a fixed decoder over weeks of training, casting the decoder as a new, fixed limb the brain would learn to control. Each neuron changed its movement encoding until the brain converged on a stable neural solution to BMI control which was readily recalled day after day. If experimenters provided the subjects a new decoder every day fit from offline reconstruction of native movements, this made BMI performance less stable.

Thus, for 2- and 3-dof BMIs, sophisticated encoding and decoding models were not the critical challenge, as relatively simple linear decoders such as the Kalman Filter (Wu et al., 2006) yielded performance rivaling the native arm (Gilja et al., 2012). Designing high-performance BMIs requires synergizing the machine learning algorithms decoding the brain as well as the neurobiological learning which adapts the user to the BMI (Shenoy and Carmena, 2014).

## BMIs as a Tool to Study the Brain

Early efforts to test BMIs found that basic neuroscience questions were important for BMI system design. An exciting realization then followed: BMIs themselves serve as unique tools to study the brain. In particular, a BMI constitutes a simplified sensorimotor loop in the brain, in which the experimental paradigm specifies the neurons which drive behavior and

the causal transformation between neural activity and behavior (the decoder) (Orsborn and Carmena, 2013). This is in stark contrast to the natural motor system, in which there are parallel neural pathways for motor control and vastly more motor-related neurons than can be measured. By permitting experimenters to observe all neurons directly producing behavior and to specify the neural-movement transformation, BMIs have enabled study of 1) the neural basis of closed-loop motor control and 2) the neural basis of skill learning.

BMI experiments dissecting the neural basis of closed-loop motor control include studies of neural dynamics during movement (Kao et al., 2015), the comparison of neural activity patterns in neurons which causally drive movement versus surrounding neurons (Ganguly et al., 2011), the brain’s internal models for movement generation (Golub et al., 2015), how feedback is incorporated into online control (Suminski et al., 2010), and how movement-error information is isolated from motor output (Stavisky et al., 2017). In addition, BMI experiments have enabled researchers to ask how subjects adapt to new contexts and perturbations (Chase et al., 2012; Hwang et al., 2013; Jarosiewicz et al., 2008; Legenstein et al., 2010; Sadtler et al., 2014).

The focus of this thesis is to use the BMI to study how the brain learns skills *de novo* (Costa, 2011; Ganguly and Carmena, 2009, 2010; Hliot et al., 2010). BMI learning shows striking parallels to natural motor learning, with the additional benefit of providing a reduced functional circuit for action to study. Animals initially show increased movement and neural variability which decreases during both motor learning (Cohen and Sternad, 2009; Costa et al., 2004; Peters et al., 2014; Shmuelof et al., 2012) and BMI learning (Arduin et al., 2013; Zacksenhouse et al., 2007), resulting in consolidated, task-relevant neural ensembles and activity patterns (Cao et al., 2015; Clancy et al., 2014; Ganguly and Carmena, 2009; Koralek et al., 2012; Peters et al., 2014). The exploration and consolidation of activity patterns for both BMI and motor learning depends on shared mechanisms. Activity and plasticity in the basal ganglia supports motor learning (Barnes et al., 2005; Hikosaka et al., 1999; Yin et al., 2009) and BMI learning (Koralek et al., 2013, 2012). Finally, sleep is critical for motor learning (Ramanathan et al., 2015) and BMI learning (Gulati et al., 2014, 2017). Thus, general learning mechanisms may underlie motor and BMI skill learning.

## Optogenetics for Studying BMI Learning

In this thesis, after first using the monkey BMI paradigm to study the structure of neural variability over learning, we will then use the rodent BMI paradigm to ask questions of how neural circuits interact to enable BMI learning. In particular, the rodent paradigm enables the use of optogenetics: a tool combining optics and genetics permitting experimenters to either excite or inhibit specific neurons by illuminating them. The technique works by genetically modifying neurons, typically by injecting a virus, to express light-sensitive ion channels (opsins) (Yizhar et al., 2011) (Figure 1.5).

When the ion channels are photo-stimulated, photocurrents across the neuron membrane are induced which can excite or inhibit the cell’s spiking. Critically, expression of the opsins can be targeted to specific cell types by leveraging the *cre-lox* system. Transgenic mice

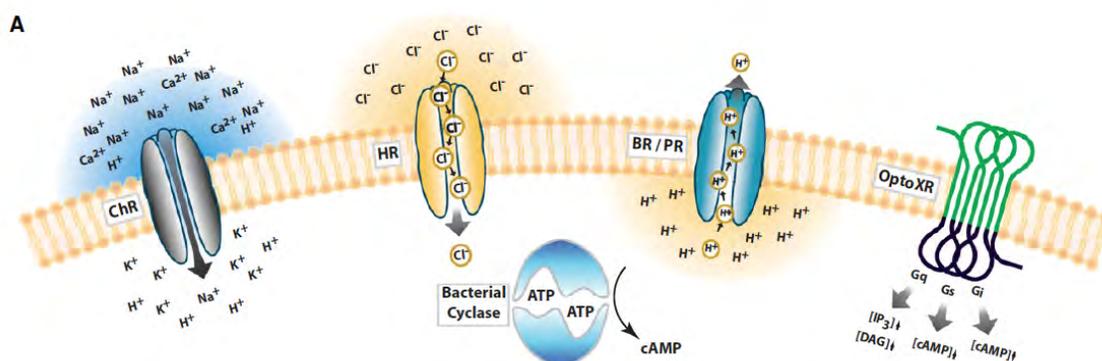


Figure 1.5: **Opsins used for optogenetics.** From (Yizhar et al., 2011)

can be developed to express Cre-Recombinase, an enzyme which catalyzes recombination between two lox-P sites, in specific neurons in the brain. Then, by injecting a viral vector containing the optogenetic actuator gene between two lox-P sites, only the neurons expressing Cre-Recombinase will express the desired opsin.

In Chapter 3, we deliver neural reinforcement signals to the brain by exciting dopaminergic neurons in the Ventral Tegmental Area (VTA) (Tsai et al., 2009). We used ChR2, a channel which when opened via photostimulation permits the flow of cations down the electrochemical gradient into the neuron, resulting in depolarization. ChR2 was virally-delivered to Th-Cre mice, enabling DA neuron-specific expression of ChR2-EYFP (Figure 1.6) and optical stimulation (Figure 1.7).

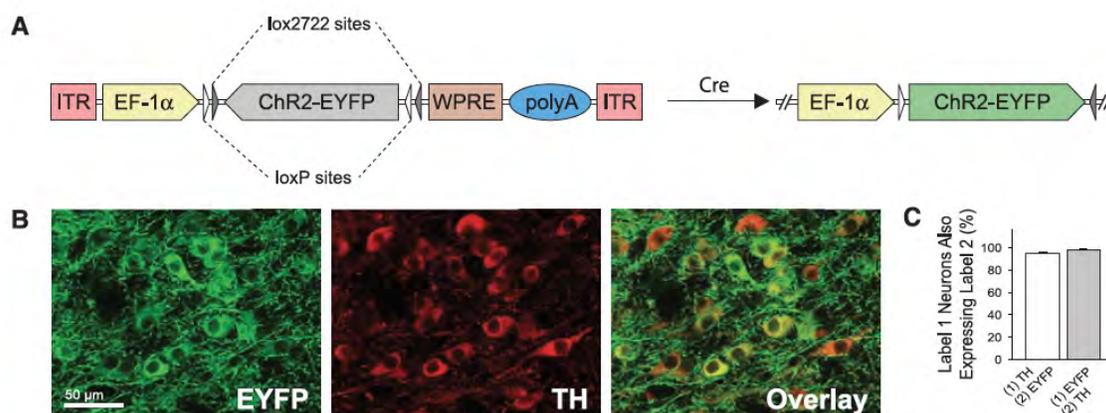


Figure 1.6: **ChR2 in Th-Cre Mice.** From (Tsai et al., 2009)

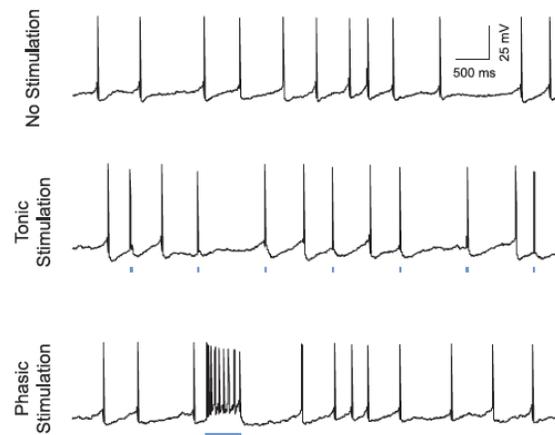


Figure 1.7: **ChR2 stimulation of VTA.** Blue ticks indicate optical stimulation. From (Tsai et al., 2009)

In Chapter 4, we test the role of neural activity in dorsomedial striatum during BMI learning by inhibiting dorsomedial striatum activity. We used *Jaws*, a red-shifted light-driven chloride pump which hyperpolarizes the cell (Chuong et al., 2014), virally delivered to dorsomedial striatum.

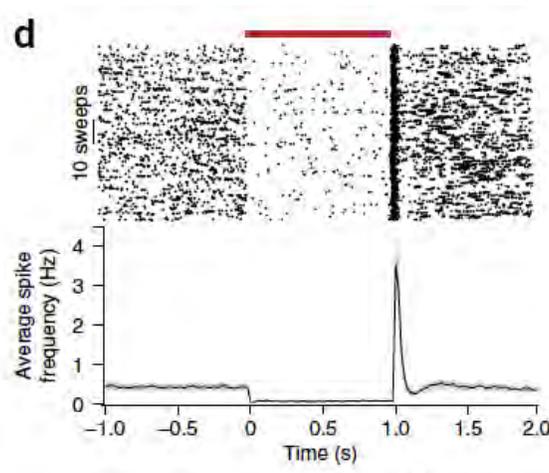


Figure 1.8: **Jaws inhibition.** Red bar indicates optical inhibition. From (Chuong et al., 2014)

Our work constitutes some of the first to use optogenetics to study neural circuits underlying BMI learning.

## Chapter Previews

In Chapter 2, we use Factor Analysis to study the coordination of neurons driving neuro-prosthetic output as they explore and consolidate neural activity patterns.

In Chapter 3, we ask if neural reinforcement signals drive consolidation of antecedent neural activity patterns. We introduce a new BMI in which reinforcement signals are delivered through optogenetic stimulation of the VTA.

In Chapter 4, we study the generalization of basal ganglia-dependent neural reinforcement learning to primary visual cortex (V1). We inhibit striatum activity and uncover its relevance to BMI learning.

## 1.1 Appendix: Clinical Motor BMI State-of-the-art

Broadly, BMIs can 1) read information out of the brain, and 2) write information into the brain or modulate its state. While this technology may sound as though it belongs to the future, BMIs which electrically stimulate the brain are prevalent in the clinic (Tsu et al., 2015). Cochlear implants, invented in the 1960s, restore hearing by translating sounds into patterned stimulation of the cochlear nerve (Clark et al., 2013; Kral and Sharma, 2012; Merzenich et al., 1980). Deep brain stimulation (DBS) is a highly successful implantable stimulation device for treating Parkinsons disease and essential tremor that has been in use since 1997 (Awan et al., 2009; Perlmutter and Mink, 2006). More recently, in 2013 the NeuroPace RNS stimulator was FDA-approved to treat medically refractory epilepsy (Morrell, 2011).

Here we overview the state of the art in motor BMI translation to the clinic. While the highest performance BMIs leverage invasive microelectrode arrays, ultimately, the question of suitability of an implanted versus external BMI system depends on the individual patient's comparison of potential benefit and risk (Hochberg and Anderson, 2012; Hochberg and Cochrane, 2013).

Recent studies of humans using intra-cortical motor BMIs have achieved important performance milestones for the control of computers. In 2017, three human subjects with ALS achieved state-of-art BMI control of computer cursor movement and typing as part of the BrainGate2 pilot clinical trial (<https://www.clinicaltrials.gov/ct2/show/NCT00912041>) (Pandarinath et al., 2017). In a copy-typing task in which subjects typed pre-determined phrases, the three subjects were able to achieve 32, 39, and 14 correct characters per minute (ccpm), or 6.3, 7.8, and 2.7 words per minute (wpm) (Figure 1.9).

Clearly, significant room remains for BMI technology improvement to match communication rates for able-bodied subjects using smart phones (12-19 wpm), touch typing (40-60 wpm), and speaking (90-170 wpm). However, a survey of people with ALS found that 59% of respondents would be satisfied with a BMI that achieved 10-14 ccpm (2-2.9 wpm), and 72% would be satisfied with 15-19 ccpm (3-3.8 wpm), placing this BMIs performance within clinically-viable range (Huggins et al., 2015). Another survey indicated that over 50% of

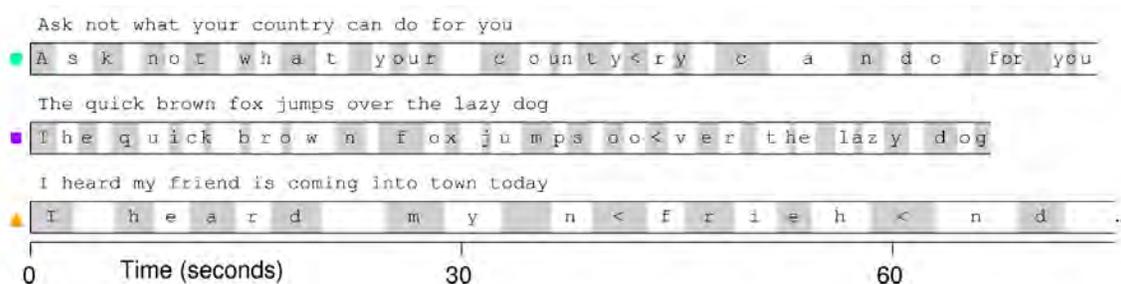


Figure 1.9: **State-of-art human BMI performance for typing.** From (Pandarinath et al., 2017). Caption from paper’s figure legend: Examples of text typed during three copy typing evaluations with participants T6, T5, and T7. Each example shows prompted text, followed by the characters typed within the first minute of the two-minute evaluation block. Box width surrounding each character denotes the time it took to select the character. ‘<’ character denotes selection of backspace key.

people with spinal cord injury would definitely or very likely undergo an implant surgery for a high-performance BMI (Collinger et al., 2013b). Thus, people with paralyses express clear willingness to undergo a surgical procedure if it can provide significant functional improvement.

In addition to restoring communication, intra-cortical motor BMIs have made progress in the control of prosthetics and even users’ own muscles (Bensmaia and Miller, 2014). In 2013, an individual with tetraplegia resulting from spinocerebellar degeneration gradually acquired control of a 7 degree-of-freedom (dof) robot to perform reach, grasp, and placement of diverse objects (Collinger et al., 2013a) (Figure 1.10). In 2016-17, people with paralysis due to cervical spinal cord injury used intra-cortical BMIs to activate forearm muscles with electrical stimulation to restore cortical control of isolated finger movement, reaching, and grasping (Ajiboye et al., 2017; Bouton et al., 2016) (Figure 1.11).

The current state-of-art BMIs are testament to the progress of BMI technology in human trials over the past 15 years (Aflalo et al., 2015; Gilja et al., 2015; Hochberg et al., 2006, 2012; Jarosiewicz et al., 2015; Kim et al., 2008).

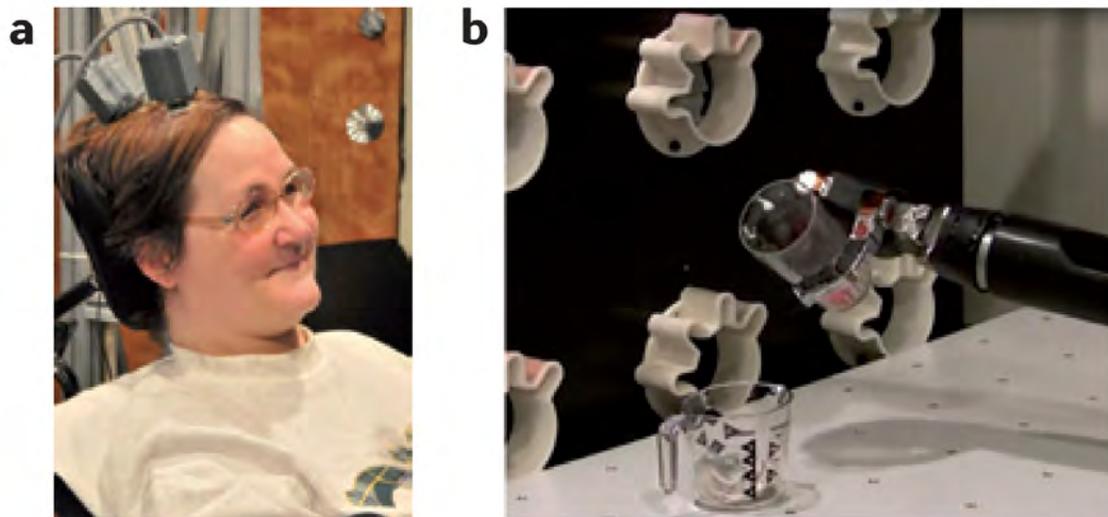


Figure 1.10: **Seven-DOF control of a prosthetic limb for reaching and grasping.** Experiment from (Collinger et al., 2013b), figure and caption from (Bensmaia and Miller, 2014). **(A)** A patient with tetraplegia had two electrode arrays implanted into the primary motor cortex. **(B)** The patient controlled the position and orientation of the modular prosthetic limb in order to grasp objects.

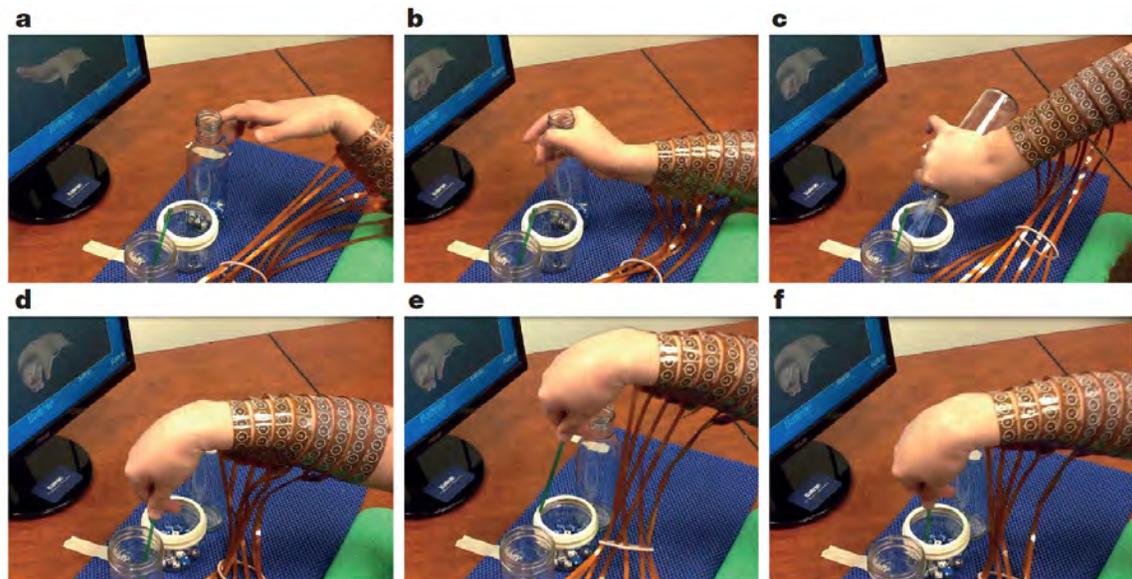


Figure 1.11: **Grasp-pour-and-stir functional movement.** Figure and caption from (Bouton et al., 2016). Sequential snapshots from the functional movement task showing the participant opening his hand (**A**), grasping the glass bottle (**B**), pouring its contents (dice) into a jar (**C**), grasping a stir stick from another jar (**D**), transferring the stir stick without dropping it (**E**), and using it to stir the dice in the jar (**F**). The task required the participant to evoke different cortical modulation patterns to control the opening of his hand, perform a cylindrical palmar grasp, and achieve a precision pinch grasp, while simultaneously moving his arm.

## Chapter 2

# Emergence of Coordinated Neural Dynamics Underlies Neuroprosthetic Learning and Skillful Control

During motor learning, movements and underlying neural activity initially exhibit large trial-to-trial variability which decreases over training, resulting in consolidated motor and neural patterns. However, it is unclear how task-relevant neural populations coordinate to explore and consolidate activity patterns underlying behavioral improvement. Exploration and consolidation could happen for each neuron independently, across the population jointly, or both. We disambiguated among these possibilities by investigating how subjects learned *de novo* to control a brain-machine interface using a fixed motor cortex population. We decomposed population activity into the sum of private and shared signals, which produce uncorrelated and correlated neural variance respectively, and examined how these signals' evolution causally shapes behavior. We found initially large trial-to-trial movement and private neural variability reduce over learning. Concomitantly, task-relevant shared variance increases, consolidating a manifold containing consistent neural trajectories that generate refined control. These results suggest that motor cortex acquires skillful control by leveraging both independent and coordinated variance to explore and consolidate neural patterns.

### 2.1 Introduction

When we begin learning a novel skill, we have no experience about how to move to achieve a particular goal. Early in training, movements exhibit variability from attempt to attempt, allowing the motor system to explore actions and select them based on consequences (Sutton and Barto, 1998; Tumer and Brainard, 2007; Wu et al., 2014). Gradually, movement variability decreases as the motor system consolidates particular movements that lead to success (Cohen and Sternad, 2009; Shmuelof et al., 2012). This suggests that a task-relevant neural population explores neural activity patterns to generate novel motor commands, and se-

lects and consolidates those that achieve the desired goal. Indeed, neurophysiological motor learning studies have found that movement-related neural activity in various species follows a similar trend, exhibiting high variability in early training which reduces as particular ensembles and activity patterns are consolidated in late training (Arduin et al., 2013; Barnes et al., 2005; Cao et al., 2015; Costa et al., 2004; Kao et al., 2005; Mandelblat-Cerf et al., 2009; Peters et al., 2014; Santos et al., 2015; Zacksenhouse et al., 2007). However, it is unclear how exploration of particular patterns in a task-relevant neural population is accomplished, and also how goal-achieving patterns are refined and consolidated to improve performance.

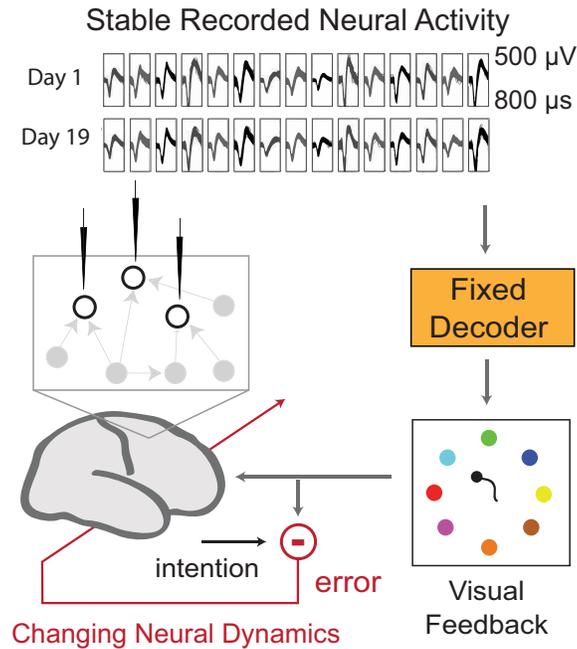


Figure 2.1: **Operant Learning BMI Paradigm Using Stable Neurons and a Fixed Decoder.** Monkeys use neural activity of stable neurons to control a fixed decoder over days, so improvement can only be driven by adaptive changes in neural population activity.

The motor system is distributed and redundant, with parallel pathways and many more neurons than muscles, making it difficult to investigate the causal relationship between observed neural activity patterns and motor output (Briggman and Kristan, 2008). We therefore took advantage of a paradigm in which we could identify the output neurons that control behavior as well as the explicit transformation between output neuron activity and behavior. We used an operant learning brain-machine interface (BMI) in which stable recordings from ensembles of primary motor cortex (M1) neurons in macaque monkeys are input to a fixed mathematical transform (decoder) outputting prosthetic two-dimensional cursor movements. The BMI provided a closed-loop feedback system operating within the natural motor system (Figure 4.5). Because we defined the decoder to take input from observed cells (direct cells),

this neuroprosthetic skill learning paradigm (Ganguly and Carmena, 2009, 2010) is uniquely advantageous to investigate how a task-relevant neural population coordinates to acquire skillful control.

Given that the decoder is novel, subjects must initially explore neural population activity space. This is a non-trivial search because population activity space is high-dimensional, with each neuron’s activity represented as one dimension. Over training, subjects must use the behavioral consequences of explored activity patterns to select and refine goal-achieving patterns. This consolidation process is not well-understood, especially as there are many neural activity patterns which can elicit the same behavioral output in systems mapping many neurons to fewer outputs, such as a BMI or our natural motor system. How might this learning happen? We consider three possibilities.

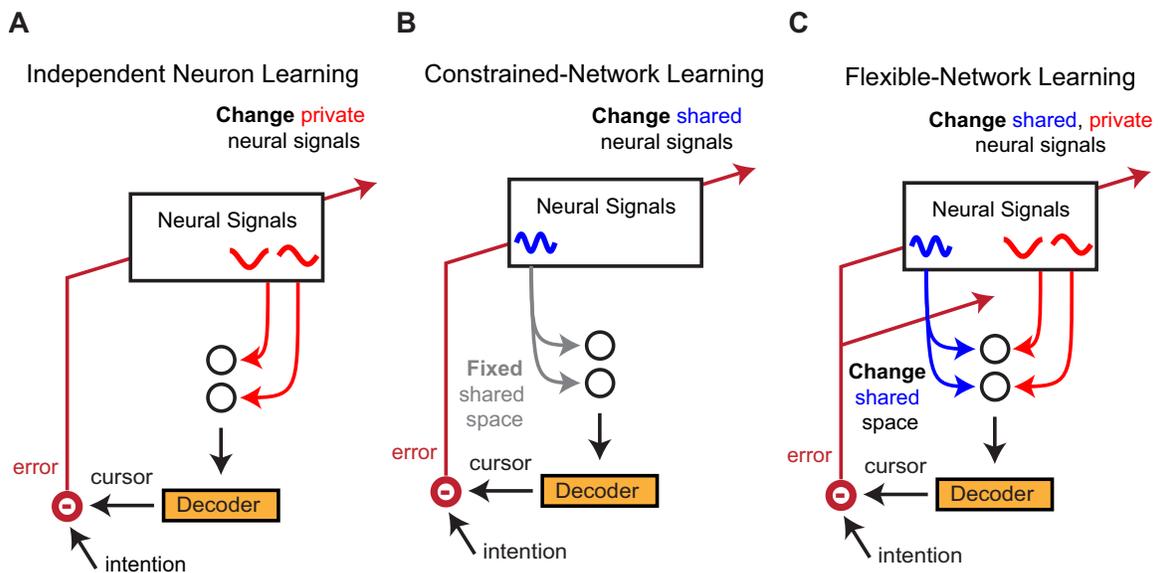


Figure 2.2: **Models of Learning.** (A) Independent Neuron Learning posits that each neuron explores and acquires independent patterns to control the BMI. (B) Constrained-Network Learning posits that the neural population learns BMI control by exploring and changing covariation patterns in a fixed shared space, e.g. because of network connectivity constraints. (C) Flexible-Network Learning posits that the neural population learns BMI control by exploring and changing both independent and covariation patterns and by using its exploration to change the shared space of covariation patterns. For intuition of how independent and covariation patterns can contribute to movement variability, see (Figure 3.11)

In “independent neuron learning,” each neuron explores independently, as has been proposed in computational models (Héliot et al., 2010; Legenstein et al., 2010), and each neuron tunes its independent patterns over learning (Figure 2.2A). Because independent neuron learning explores every dimension of population activity space, this learning can acquire

any firing rate configuration and thus control any decoder. However, this learning cannot explain how subjects rapidly acquire (within one day) decoders which leverage neural covariance (Sadtler et al., 2014).

In contrast, “constrained-network learning” posits that the neural population possesses network connectivity constraints and thus preferentially produces covariation patterns within a fixed, lower-dimensional subspace of the full activity space. Indeed, both BMI and motor control studies have found that motor cortical neurons exhibit task-relevant, low-dimensional covariance during execution of well-learned behavior (Churchland et al., 2012; Kao et al., 2015; Sadtler et al., 2014). Under constrained-network learning, the neural population explores and changes covariation patterns within the fixed subspace (Sadtler et al., 2014) (Figure 2.2B). One limitation of the constrained exploration is that this learning can only acquire decoders reading out activity within the populations existing covariance subspace.

Finally, we consider “flexible-network learning,” which both permits the learning of many decoders and can explain the emergence of new covariance. In this model, the neural population explores and changes both independent and covariation patterns. Critically, independent exploration permits subjects to initially explore a higher dimensional space of activity patterns. Upon discovering patterns with rewarding consequences, subjects may shape their network to reliably reproduce rewarding patterns, constraining patterns to lie within a low-dimensional subspace. This mechanism thus enables consolidation of covariance in a different subspace from initial covariance (Figure 2.2C).

To determine which of these three possibilities best explains how the task-relevant neural population coordinates to acquire neuroprosthetic skill *de novo*, we sought to disentangle how independent and coordinating sources of neural variance changed and contributed to BMI control over the course of learning. We used Factor Analysis (FA) (Everitt, 1984) to decompose population activity into the sum of private and shared signals, which estimate the effects of both private inputs to each cell and coordinating inputs which drive multiple cells simultaneously (Churchland et al., 2010; Yu et al., 2009). Using FA, we arbitrated between our three learning models in addressing how a task-relevant neural population explores, consolidates, and optimizes activity to acquire a neuroprosthetic skill *de novo*.

First, we asked how private and shared sources influence the exploration of different activity patterns from trial-to-trial. To isolate trial-to-trial variability from within-trial variance relevant for control, we focused on the variability of the population firing rate at coarse-timescale (time window  $\sim 1$ s) for a fixed cursor state (the center of the workspace) with a fixed movement intention (the target). Second, we asked how private and shared sources contribute to within-trial control over the course of learning. To assess control-relevant neural activity, we focused on spike counts at the decoder-timescale of 100 ms. This decoder-timescale neural variance can be interpreted as signal strength for BMI control. Finally, we asked how the spatial and temporal structure of late-training neural activity supported skillful control. In particular, we asked whether neural activity preferentially occupied the control-relevant dimensions of the decoder, and whether a particular temporal sequence was consolidated which generated skillful control. We focused on fine-timescale trajectories at 50 ms resolution, faster even than the decoder-timescale.

Exploration and consolidation via independent neuron learning predicts that private trial-to-trial variability decreases and private decoder-timescale variance is used for control. Constrained-network learning predicts the opposite; shared trial-to-trial variability decreases and shared decoder-timescale variance is used for control. Combining these changes under flexible-network learning, we would expect that in early learning, private trial-to-trial variability contributes to activity exploration, and private decoder-timescale variance produces variable movements, while we would expect that in late learning, shared decoder-timescale variance consolidates and contributes to skillful control. Indeed, our analyses revealed that private and shared variance changed differentially to sculpt neural spatiotemporal patterns and improve movements, providing insight into how motor cortex coordinates to explore, consolidate, and optimize population activity underlying novel skills.

## 2.2 Methods

### Experimental model and subject details

Two adult male rhesus monkeys (*Macaca mulatta*) (Monkey P, age: 7 years 1 month, weight: 15 kg; Monkey R, age 6 years, weight: 10.7 kg) were chronically implanted in the brain with arrays of 64 Teflon-coated tungsten microelectrodes (35  $\mu\text{m}$  in diameter, 500  $\mu\text{m}$  pitch; Innovative Neurophysiology, Durham NC) in an  $8 \times 8$  array configuration (Ganguly and Carmena, 2009). Monkey P was implanted in the left hemisphere in the arm area of both primary motor cortex (M1) and dorsal premotor cortex (PMd), and in the right hemisphere in the arm area of M1, with a total of 192 microwires across three implants. Monkey R was implanted bilaterally in the arm area of M1 and PMd (256 microwires across four implants). Only activity from M1 was used in these BMI experiments (Monkey P: right M1; Monkey R: left M1). Array implants were targeted for pyramidal tract neurons in layer 5. Localization of target areas was performed using stereotactic coordinates from a neuroanatomical atlas of the rhesus brain (Paxinos et al., 2000).

All procedures were conducted in compliance with the National Institutes of Health *Guide for the Care and Use of Laboratory Animals* and were approved by the University of California at Berkeley Institutional Animal Care and Use Committee.

### Electrophysiology

Neural activity was recorded using the MAP system (Plexon) (Figure 2.3). Stable units were selected based on waveform shape, amplitude, relationship to other units on the same channel, inter-spike interval distribution, and the presence of an absolute refractory period. Only units from primary motor cortex were used which had a clearly identified waveform with signal-to-noise ratio of at least 4:1. Activity was sorted prior to recording sessions using an online spike-sorting application (Sort Client; Plexon). Stability of waveforms was confirmed by analyzing the stability of PCA projections over days (Wavetracker; Plexon).

## Chronic Multi-Site Multi-Electrode Recordings

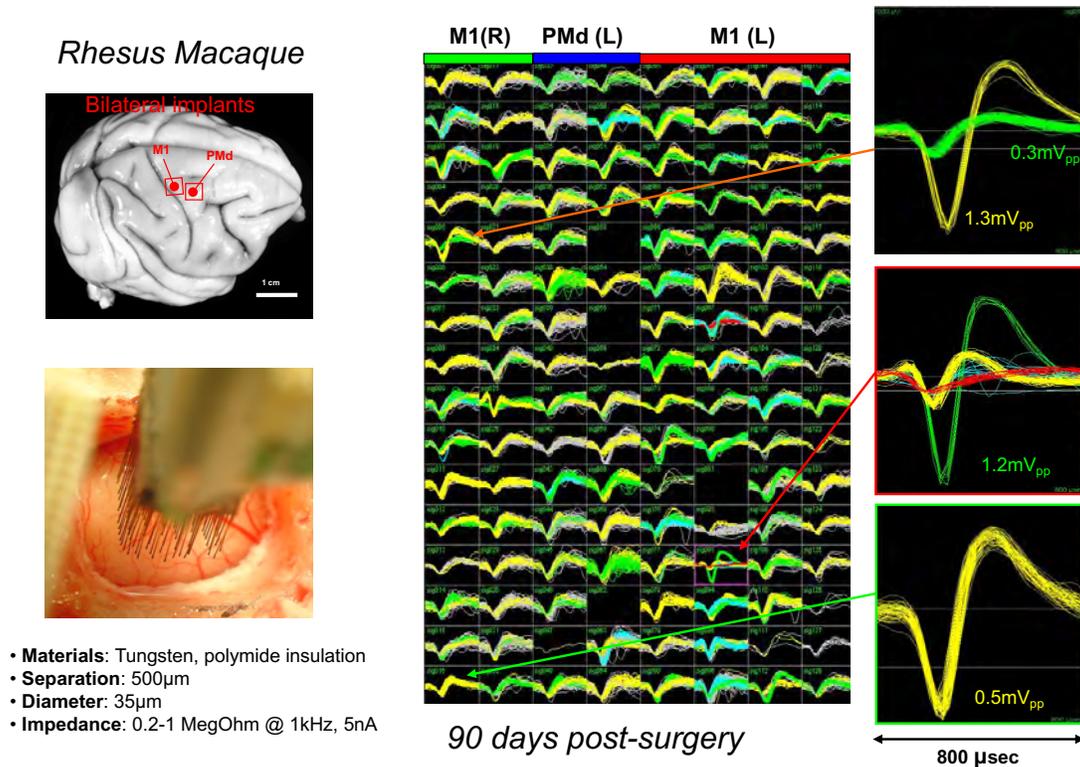


Figure 2.3: **Electrophysiology.** Electrophysiology in rhesus macaque M1 and PMd using 64 microwire arrays. Figure courtesy of Jose M. Carmena.

## BMI Task

Subjects performed a center-out reaching task to eight targets (Figure 2.4A). The cursor was continuously controlled by neural activity. Subjects self-initiated trials by moving to the center target. Peripheral targets were presented in pseudorandom order. A successful trial required a short hold at the center, moving to the peripheral target within 15 seconds, and a brief hold at the target (Figure 2.4B). Successful trials resulted in a liquid reward; failed trials were repeated. During BMI control, both arms were removed from the workspace, lightly restrained, and thus unable to perform reaches. During selected sessions, video and surface electromyogram (EMG) recordings from proximal muscle groups were performed. Neither animal moved their upper extremity during brain control. Before starting the BMI learning experiments, subjects were overtrained on the task performed with arm movements using a Kinarm (BKIN Technologies) exoskeleton which restricted shoulder and elbow to move in the horizontal plane.

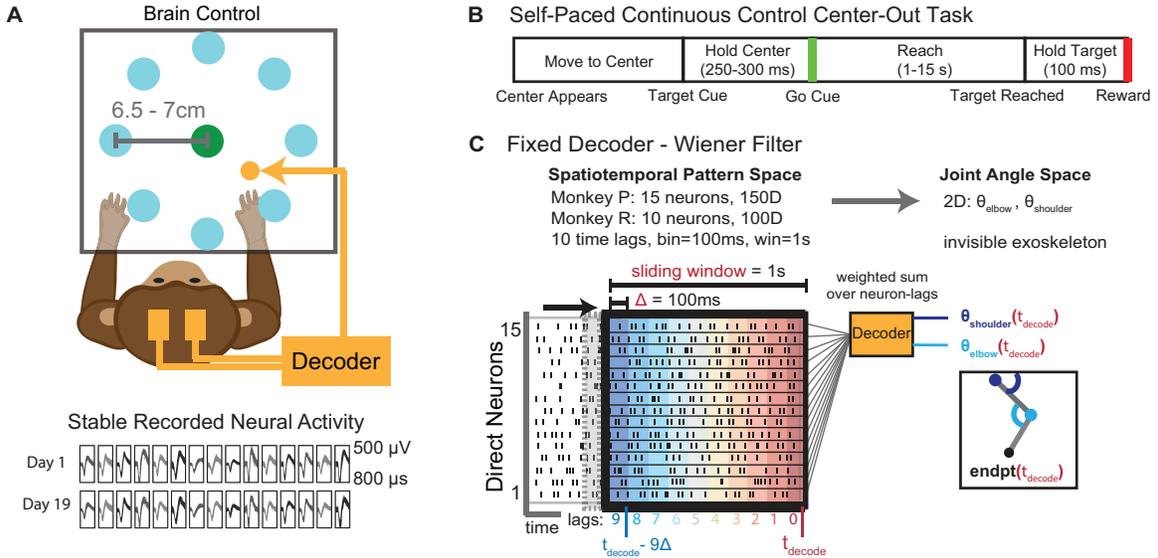


Figure 2.4: **Experimental Paradigm.** (A) Monkeys control a 2D cursor to one of 8 instructed targets using neural activity of stable neurons over days. (B) Center-out task timeline. (C) The task uses a position decoder, which maps a neural spatiotemporal pattern into a low-dimensional control signal that produces the cursor endpoint.

## BMI Decoder

### Decoding Algorithm

In this study, the decoder used neural activity to control the joint angles of a virtual, invisible two-link arm (Figure 2.4C). A Wiener Filter linearly mapped the binned spike counts of  $N$  neurons  $x(t) \in \mathbf{R}^N$  in the last one second into virtual shoulder and elbow joint angles  $y(t) \in \mathbf{R}^2$  via:

$$y(t) = \mathbf{b} + \sum_{k=0}^M \mathbf{a}(k\Delta)x(t - k\Delta) \quad (2.1)$$

The fixed decoder parameters  $\mathbf{a}(k\Delta) \in \mathbf{R}^{2 \times N}$  for  $k = 0, \dots, M$  and  $\mathbf{b} \in \mathbf{R}^2$  used in this study were fit via least-squares linear regression to predict the subjects' actual shoulder and elbow joint angles in a calibration manual control session (Ganguly & Carmena, 2009). In this experiment, num lags = 10 =  $M + 1$ , and the spike count bin size was  $\Delta = 100$  ms. The joint kinematics forward model mapped shoulder and elbow angles  $y(t)$  to endpoint position, the only feedback signal the subjects received. The joint positions  $y(t)$  and endpoint feedback signal were updated every  $\Delta = 100$  ms.

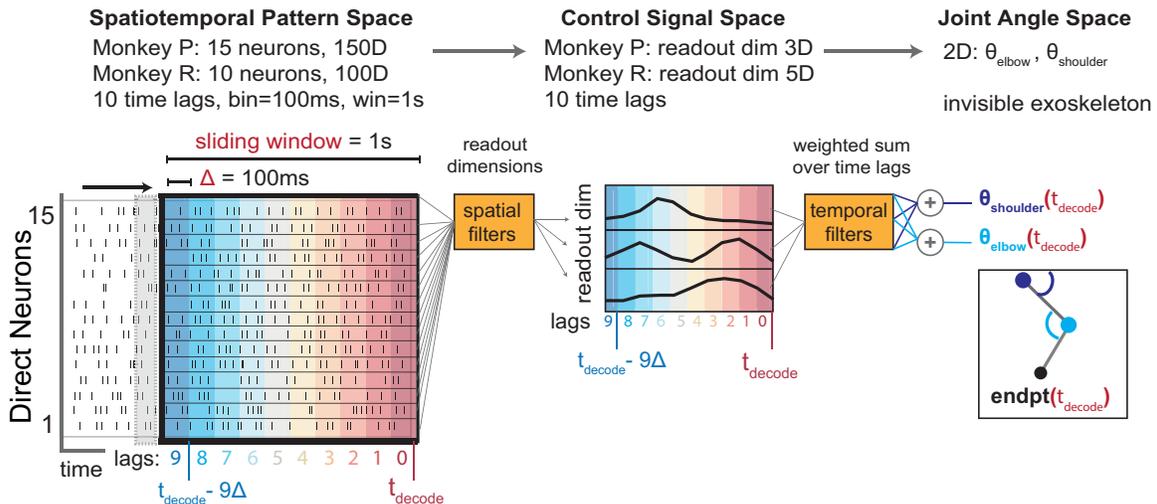


Figure 2.5: **Readout Space of Wiener Filter.** The decoder transforming neural activity into cursor position can be decomposed into two stages of filters. The first stage consists of “spatial filters” which project neural activity into a subspace termed the decoder’s “Readout Space.” The time series of neural activity projections are termed control signals. The decoder’s second stage consists of “temporal filters” which perform linear combinations of control signals over time lags.

### Decoder Readout Space

We analyzed the decoder’s weights and determined that applying neural activity through the decoder is equivalent to first projecting neural activity into a subspace (the decoder readout space) and then applying it through the decoder (Figure 2.5). The decoder produced 2-dimensional output by summing num lags = 10 linear combinations of neural activity, so we analyzed how many neural dimensions the decoder primarily leveraged to produce control. At maximum, it could be 10, e.g. if each lag used a different individual neuron’s activity. We found the readout space dimensionality by evaluating the expression:

$$\mathbf{a}(k\Delta) \approx \mathbf{a}(k\Delta)\mathbf{P} \quad (2.2)$$

for  $k = 0, \dots, M$ , where  $\mathbf{a}(k\Delta)$  are the decoder parameters,  $\mathbf{P} \in \mathbf{R}^{N \times N}$  is a projection matrix of rank  $L$ , and num lags = 10 =  $M + 1$ . (Note that for  $L = N$  we have perfect equality.)  $\mathbf{P}$  was found by performing PCA on the decoder parameters to compress neural dimensionality (as opposed to temporal dimensionality).  $L = 3$  was sufficient for  $\mathbf{a}(k\Delta)\mathbf{P}$  to capture  $\sim 97\%$  of  $\mathbf{a}(k\Delta)$  parameter variance for Monkey P, and  $L = 5$  captured  $\sim 99\%$  of parameter variance for Monkey R.

The “decoder readout space” is the column space of  $\mathbf{P}$ , as we can reproduce BMI cursor movements by first projecting neural activity into this space. Thus, we note that for our

decoder, the readout space dimensionality is not the same as the cursor dimensionality.

## Quantifying Learning Over Training Epochs

We analyzed the behavioral and neural learning process for each target separately over “training epoch” blocks of a constant number of trials. We chose the “training epoch” rather than the more standard “training day” because we wished to eliminate the effect of variable trial numbers per training day on analysis results. Note that because subjects performed a different number of trials to each target overall, the number of trials in a training epoch differs for each target. For Monkey P (Monkey R), the smallest epoch size is 23 trials (31 trials) and the largest is 36 trials (56 trials).

### Behavioral Metrics

We quantified task performance using percent correct (percent of initiated trials completed successfully) and success rate (successful trials per minute). Success rate combines accuracy and speed and is given by  $\frac{\text{num successful trials}}{\sum_{\text{all trials}} \text{time to initiate trial} + \text{time to complete trial}}$ .

### Cursor Variability Analyses

To analyze cursor variability, we analyzed the cursor trial-to-trial Spatial Occupancy Map (SOM) by discretizing the workspace (each bin was 0.34% of the workspace) and computing the fraction of trials the cursor would enter each position bin. This was computed for each target and training epoch as:  $\text{SOM}(\text{position}_{ij}) = \frac{\text{num trials cursor enters position}_{ij}}{\text{num trials}}$ . We performed three analyses. First, we assessed control refinement via the fraction of workspace

entered per trial:  $\frac{1}{\text{num trials}} \sum_{\text{trial}_i=1}^{\text{num trials}} \frac{\text{num bins entered}(\text{trial}_i)}{\text{num bins}}$ . Second, we assessed the stabil-

ity of the learned cursor control strategy via the correlation between the final epoch SOM and previous epochs:  $\text{corr}(\text{SOM}_{\text{epoch}_i}, \text{SOM}_{\text{epoch}_j})$ , for  $\text{epoch}_i = 1, \dots, \text{epoch}_{\text{final}-1}$ , where

$\text{corr}(a, b) = \frac{\sum_{i=1}^N a_i b_i}{\|a\| \|b\|}$  is the standard vector correlation for  $a, b \in \mathbf{R}^N$ . Third and finally, we assessed the cursor’s trial-to-trial movement variability by considering each trial as a binary vector of 0’s and 1’s indicating the cursor’s spatial occupancy. We computed the covariance matrix using this parameterization of each trial and calculated the variability via the trace. We note that a decrease in trial-to-trial variability is not a mathematical consequence of reduction in workspace explored, as it is possible for subjects to enter the same large fraction of the workspace consistently as well as to enter different small fractions of the workspace from trial-to-trial. The trends found by these analyses are not sensitive to the discretization resolution; the same results were achieved with coarser resolution, e.g. bins 4 times larger (data not shown).

## Factor Analysis

### Model

We used FA to model the joint distribution of  $N$  neurons' spike counts  $x \in \mathbf{R}^N$  as the sum of 1) a mean rate  $\mu \in \mathbf{R}^N$ , 2) private signals with diagonal covariance  $\Psi \in \mathbf{R}^{N \times N}$ , and 3) shared signals due to a low-dimensional latent variable  $z \in \mathbf{R}^k$ ,  $k < N$ :

$$\begin{aligned} z &\sim N(0, I) \\ x|z &\sim N(\mu + Uz, \Psi) \\ x &\sim N(\mu, UU^T + \Psi) \end{aligned} \tag{2.3}$$

$U \in \mathbf{R}^{N \times k}$  are the weights (or factors) capturing how  $z$  drives  $x$ . We make explicit FA's decomposition of population activity  $x$  into shared signals  $x^{\text{shared}} = Uz \in \mathbf{R}^N$  and private signals  $x^{\text{private}} \in \mathbf{R}^N$ :

$$\begin{aligned} x^{\text{private}} &\sim N(0, \Sigma^{\text{private}}) \\ x^{\text{shared}} &\sim N(0, \Sigma^{\text{shared}}) \\ x^{\text{private}} &\perp x^{\text{shared}} \\ x &= \mu + x^{\text{shared}} + x^{\text{private}} \end{aligned} \tag{2.4}$$

where  $\Sigma^{\text{private}} = \Psi$  and is diagonal,  $\Sigma^{\text{shared}} = UU^T$  and is low rank, and  $\Sigma^{\text{total}} = \Sigma^{\text{shared}} + \Sigma^{\text{private}}$ .

For each target separately, we fit the model's parameters by maximizing the log-likelihood of the data with the EM algorithm for a chosen shared dimensionality  $k < N$ . The best-fitting shared dimensionality  $k^*$  is estimated by using cross-validated log-likelihood to determine which  $k$  best describes held-out data (Dempster et al., 1977). To ensure we didn't choose too high of shared dimensionality, we analyzed the eigenvalues of the shared variance found using the dimensionality which maximized cross-validated log-likelihood and evaluated how many dimensions were needed to capture 90% of shared variance. We used this estimate of shared dimensionality throughout the work.

### Comparison to Pairwise Correlation and PCA

To quantify correlated population activity, FA has advantages over more familiar measures such as average pairwise correlation and Principal Components Analysis (PCA). First, pairwise correlation addresses how well one cell's activity can predict another's, while the cell's shared-to-total variance ratio indicates how well that cell's activity can be predicted by the rest of the population together. Second, average pairwise correlation does not tell us the data's dimensionality. While PCA provides information about dimensionality, FA provides additional features. FA is a probabilistic model which posits total covariance as the sum of shared and private variance, permitting analysis of the two sources separately, while PCA analyzes dimensionality of total covariance.

### Shared and Private Variance and the Shared-to-total Variance Ratio

Neuron  $i$ 's shared variance is  $\Sigma_{ii}^{\text{shared}}$ , private variance is  $\Sigma_{ii}^{\text{private}}$ , and total variance is  $\Sigma_{ii}^{\text{total}} = \Sigma_{ii}^{\text{shared}} + \Sigma_{ii}^{\text{private}}$ . The population-average variance is  $\frac{1}{N}\text{trace}(\Sigma) = \frac{1}{N} \sum_{i=1}^N \Sigma_{ii}$ . The population's shared variance to total variance ratio is  $\frac{\text{trace}(\Sigma^{\text{shared}})}{\text{trace}(\Sigma^{\text{total}})}$ .

## Data Analysis using FA

### Extracting Main Shared Variance

After fitting FA, we ordered shared signals by their variance by diagonalizing  $\Sigma^{\text{shared}}$  via the singular value decomposition (SVD) (Yu et al., 2009). Throughout the work we compared the decoder-timescale ‘‘main shared’’ variance. i.e. variance within the top  $p$  orthogonalized factors which provide the best rank- $p$  approximation of  $\Sigma^{\text{shared}}$ . For both subjects,  $p = 2$ , as that was the shared dimensionality in late learning.

### Shared Space Alignment

We used the ‘‘shared space alignment’’ to measure the similarity between the shared variance (or main shared variance) of epoch A and epoch B. The shared space alignment is the fraction of epoch A shared variance captured in epoch B's shared space and thus ranges from 0 to 1. For some geometric intuition, in the one-dimensional case (i.e.,  $\text{rank}(\Sigma^{\text{A, shared}}) = \text{rank}(\Sigma^{\text{B, shared}}) = 1$ ), the space alignment is equivalent to  $\cos(\theta)$ , where  $\theta$  is the angle between epoch A and epoch B's one-dimensional shared space. We note that the shared space alignment is asymmetric when shared dimensionality is greater than 1, such that alignment of A with B need not be equal to the alignment of B with A.

Let  $\Sigma^{\text{A, shared}}$  be epoch A's shared variance, and epoch B's shared space is  $\text{col}(U^{\text{B}})$ , the column space of epoch B's factor matrix  $U^{\text{B}}$ . We computed the shared space alignment in three steps:

1) Compute  $P_{U^{\text{B}}} \in \mathbf{R}^{N \times N}$ , the projection matrix into  $\text{col}(U^{\text{B}})$ . This is  $P_{U^{\text{B}}} = VV^T$ , where  $V$  is an orthonormal basis for  $\text{col}(U^{\text{B}})$ , e.g. calculated via the singular value decomposition  $\text{SVD}(U^{\text{B}}U^{\text{B}T}) = VSV^T$ .

2) Project  $\Sigma^{\text{A, shared}}$  onto  $\text{col}(U^{\text{B}})$ , which is  $P_{U^{\text{B}}}\Sigma^{\text{A, shared}}P_{U^{\text{B}}}^T$ .

3) Compute the alignment via  $\frac{\text{trace}(P_{U^{\text{B}}}\Sigma^{\text{A, shared}}P_{U^{\text{B}}}^T)}{\text{trace}(\Sigma^{\text{A, shared}})}$

This computation is used in Figure 2.9 E, 2.10 F, and 2.12 D. In Figure 2.12 D, main shared variance is projected into the readout space.

We computed the chance alignment by generating 100,000 random pairs of 1) a projection subspace of high dimensional neural space and 2) identity covariance in a random subspace of high dimensional neural space, and calculating the mean and 95th percentile of alignment. For Figure 2.9 E, the projection subspace and covariance are both 1-dimensional in a 15-dimensional ambient space (yielding mean chance alignment of 0.07 and 95th percentile chance alignment of 0.25). For Figure 2.10 F, the projection subspace and covariance are

both 2-dimensional in a 15-dimensional ambient space for Monkey P (yielding mean chance alignment of 0.13 and 95th percentile chance alignment of 0.28) and 10-dimensional ambient space for Monkey R (yielding mean chance alignment of 0.20 and 95th percentile chance alignment of 0.40).

### Trial-to-trial Variability of Shared and Private Signals

We analyzed trial-to-trial variability by applying FA to a data matrix in which each column is one trial’s population spike counts in a window ( $\sim 1$  second) locked to Go Cue. For each monkey, we chose the window length to be the longest amount of time such that at least 70% of trials for every target and training epoch were longer (Monkey P 0.9 seconds; Monkey R 1.3 seconds) (Figure 3.14 A). We removed neurons with an average rate below 1.5 Hz because we had too few trials of non-zero activity to accurately estimate their variance. Over targets and epochs, a median of 11 out of 15 neurons were used for Monkey P and 9 out of 10 neurons were used for Monkey R. For all targets and epochs, the best-fitting shared dimensionality was 1, thus shared space alignment over training was between shared spaces of equal dimensionality (Figure 2.9 E).

We considered two components to spike count variability: fixed firing rate spiking variability (such as in a Poisson process) and underlying firing rate variability (Churchland et al. 2010). We analyzed whether neural variability changes differed from changes expected with Poisson spiking variability by 1) normalizing each neuron’s private variance by its mean rate and 2) performing FA on Poisson-simulated neurons (Figure 3.15 B-E). The ratio of shared variance to total variance was calculated without any normalization of private variance (Figure 2.9 C).

### Decoder-timescale Private and Shared Variance

We analyze decoder-timescale population variance for each target and epoch by binning trials in 100 ms bins and applying FA. Neurons with firing rate below 1.5 Hz were excluded, as in the trial-to-trial variability analysis.

### Sample-by-sample Contribution of Private and Shared Variance

Given an observed spike count at time  $t$  :  $x(t)$ , we used FA to estimate the shared and private variance contribution:

$$\begin{aligned}
 \hat{x}^{\text{shared}} &= E[x^{\text{shared}}(t) | x(t)] = E[Uz(t) | x(t)] \\
 &= UU^T(UU^T + \Psi)^{-1}(x(t) - \mu) \\
 \hat{x}^{\text{private}} &= x(t) - \hat{x}^{\text{shared}}(t) - \mu
 \end{aligned}
 \tag{2.5}$$

The main shared variance contribution is the linear projection of  $\hat{x}^{\text{shared}}(t)$  into the main shared subspace (the top eigenvectors of  $\Sigma^{\text{shared}}$ ).

### Trial-to-trial Decoder Simulations

To test the how decoder-timescale main shared variance and private variance independently contribute to control, we separately supplied each trial’s main shared trajectory  $\hat{x}^{\text{shared}}(t) + \mu$  and private trajectory  $\hat{x}^{\text{private}}(t) + \mu$  as input to the decoder for the length of the observed trial. We quantified simulated performance using best progress to target, time to target hit (seconds), and normalized distance traveled to target hit. Best progress to target was computed for each trial as  $\frac{\max(\text{closest distance to target} - \text{target radius}, 0)}{\text{distance from center to target}} \in [0, 1]$ , such that a value of 1 indicates the target was reached and 0 indicates the cursor got no closer than the starting center point. Normalized distance traveled to achieve the target was normalized by the distance from center to target, such that a value of 1 indicates the cursor moved perfectly straight to the target. To compare late learning shared and private control, we used trials for which both were successful (Figure 2.11 E,F).

### Relationship of Shared Variance to the Readout Space

We quantified shared variance in the decoder’s readout space via  $\text{trace}(P_R \Sigma^{\text{shared}} P_R^T)$ , where  $P_R$  is the projection matrix onto the readout space (Figure 2.12 B). It is the product of the shared variance magnitude:  $\text{trace}(\Sigma^{\text{shared}})$  and shared space alignment with the readout space:  $\frac{\text{trace}(P_R \Sigma^{\text{shared}} P_R^T)}{\text{trace}(\Sigma^{\text{shared}})}$ , i.e. the fraction of shared variance captured in the readout space (Figure 2.12 D, 3.16 A). For each target and training epoch, we calculated the main shared variance magnitude and readout space alignment and normalized by training epoch 1 values to yield a “magnitude growth factor” and an “alignment growth factor” (Figure 2.12 D).

### Shared Neural Trajectories at Fine-Timescale

To quantify shared neural trajectories, we zoomed in bin resolution to fine-timescale with 50 ms bins, performed FA on each target and epoch separately, and calculated

$$E[x^{\text{main shared}}(t) \mid x(t)]$$

on each spike count time sample. We then trial averaged each trial’s main shared trajectory (Figure 2.13 BC). We quantified the magnitude of the main shared trajectory by taking the squared Frobenius norm (Figure 2.13 D) and the stability of the trajectory by taking the correlation between each epoch’s main shared trajectory and the final main shared trajectory (Figure 2.13 E). The fraction of main shared variance explained by the average main shared trajectory was calculated as follows. We took the difference between each trial’s main shared trajectory and the trial-averaged main shared trajectory within a 2 second window and calculated how much main shared variance remained in these residuals. The main shared variance due to the trial-averaged main shared trajectory was  $1 - \frac{\text{residual main shared variance}}{\text{main shared variance}}$  (Figure 2.13 F). Finally, we supplied the trial-averaged main shared trajectory as input to the decoder and quantified the progress to target (Figure 2.13 H).

## Quantification and Statistical Analyses

All analyses were performed for each target separately, and each plotted trend shows the mean and s.e.m. over targets ( $n=8$ ) for each training epoch, as indicated in the Results and figure legends. Trends were analyzed for significance with linear regression. Additionally, for both subjects we grouped the first 3 epochs into an early block and the final 3 epochs into a late block, resulting in  $n=24$  in each group, from 3 epochs of 8 targets. For each analysis, we then performed a non-parametric one-sided Wilcoxon rank sum test of the appropriate comparison between early and late median. P values  $< 0.05$  were considered statistically significant. All statistical analyses were performed with custom scripts in Matlab.

## 2.3 Results

### Neural control refinement increases movement success rate, directness, and consistency

We analyzed data from operant learning BMI experiments (Ganguly and Carmena, 2009) in which two rhesus macaques chronically implanted with microelectrode arrays in M1 learned to perform a two-dimensional, self-paced, continuous control center-out BMI reaching task (Figure 2.4A, B). Each trial, subjects drove the cursor under neural control to an instructed target (randomly selected from eight possible targets) for juice reward and controlled the cursor back to center to initiate the next trial. The decoder was optimized to predict upper limb movement using input from a fixed subset of recorded neurons (“direct cells”) chosen because of their stability over days (see Experimental Procedures) (Ganguly and Carmena, 2009). Because both the direct cells and the decoder parameters were held fixed for the duration of training (Monkey P: 19 days; Monkey R: 7 days), performance improvement could only be driven by adaptive changes in population activity (Figure 4.5).

The decoder transformed a neural activity sequence into the cursor’s x-y position (Figure 2.5). The direct cells’ activity in the last one second was binned into 10 time lags of 100 ms bins, and linear combinations of the neuron-lag activity produced the cursor’s position. Because the number of neurons exceeded the two cursor dimensions (Monkey P: 15 neurons; Monkey R: 10 neurons), (infinitely) many activity patterns could produce the same behavior, as is the case in our motor system which possesses more neurons than muscles. Changes in neural activity did not translate into changes in cursor position unless they occurred within the decoder’s control-relevant neural dimensions, the “readout space”, defined by the decoder weights.

Under these conditions, subjects gained proficient control, improving both accuracy and success rate for all targets (Figure 2.6 A,B). Changes over learning were analyzed over “training epochs,” where each epoch contains a constant number of trials. This division of trials ensured that analysis results were not influenced by unequal numbers of trials performed on each day. Observed trends were consistent with analyses performed across days

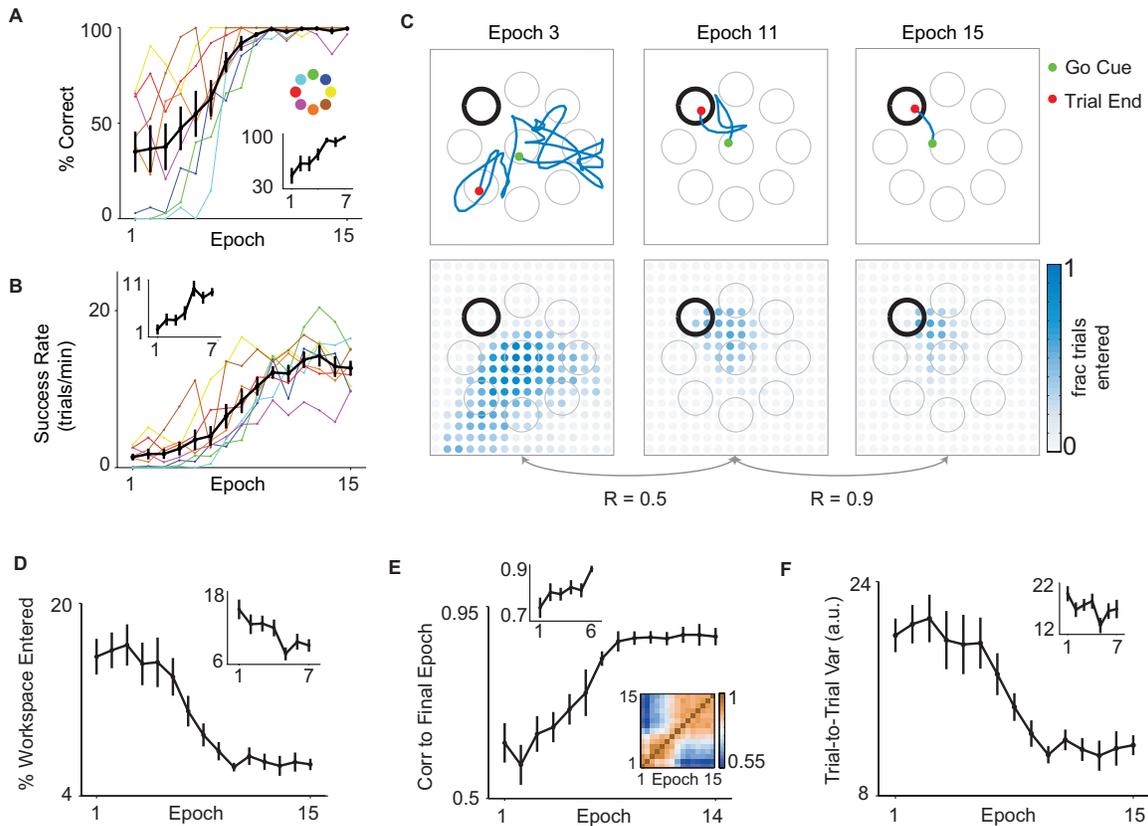


Figure 2.6: **Cursor control is refined.** Caption on next page.

(Figure 3.12, 3.13). Note that analyses of behavioral and neural changes were performed for each target separately. All trends are plotted with error bars showing the standard error of the mean (s.e.m.) over targets.

We asked how control was refined on a trial-to-trial basis by analyzing the positions the cursor entered in each trial. We computed the cursor's trial-to-trial workspace occupancy for each individual target, i.e. the probability that the cursor entered a given position within a trial (Figure 2.6 C). We found that the fraction of workspace entered per trial decreased as the cursor movements became increasingly direct (Figure 2.6 D), and that the subjects' workspace occupancy became consistent over late-training epochs (Figure 2.6 E) (see Experimental Procedures). These two results show that the cursor's trial-to-trial exploration of the workspace became refined and stable with training, extending previous reports on the straightening and stabilization of trial-averaged trajectories (Ganguly and Carmena, 2009). Finally, we found a decrease in the trial-to-trial variability of the entire set of positions entered within a trial (Figure 2.6 F) (see Experimental Procedures). These results show that neural control was refined and consolidated over long-term BMI training, increasing movement success rate, directness, and consistency.

Figure 2.6: (Previous page.) **(A,B)** Performance improved over training for Monkey P (main) and Monkey R (inset). Percent correct increased (Monkey P:  $r=0.93$ ,  $p=5.8e-7$ , early < late  $p=3.3e-10$ ; Monkey R:  $r=0.96$ ,  $p=4.7e-4$ , early < late  $p=1.5e-9$ ). Success rate in trials per minute increased (Monkey P:  $r=0.96$ ,  $p=1.6e-8$ , early < late  $p=1.4e-6$ ; Monkey R:  $r=0.91$ ,  $p=4e-3$ , early < late  $p=4.0e-6$ ). Targets (indicated by color) showed different time courses of improvement. Mean and s.e.m. error bars are shown over targets ( $n=8$  for both subjects). **(C, top)** Single trial example trajectories over three epochs, for one example target. **(C, bottom)** Workspace occupancy maps for the example target over three epochs. Intensity indicates the fraction of trials a position was occupied. **(D)** Percent of workspace entered per trial decreased over training (Monkey P:  $r=-0.94$ ,  $p=3.2e-07$ , early > late  $p=4.1e-9$ ; Monkey R:  $r=-0.87$ ,  $p=1e-2$ , early > late  $p=2.0e-4$ ). **(E)** Similarity of workspace occupancy to the final workspace map increased and stabilized with training (Monkey P:  $r=0.92$ ,  $p=2.3e-6$ , early < late  $p=5.9e-9$ ; Monkey R:  $r=0.88$ ,  $p=2e-2$ , early < late  $p=1.2e-2$ ). **(E, bottom inset)** Heat map shows the similarity between each epoch pair's occupancy maps. A heat map was calculated for each target separately; the target average is shown here. The lower left red box indicates cursor scatter during gross exploration while the upper red box indicates late training stability of refined control. **(F)** Trial-to-trial variability of workspace occupancy decreased (Monkey P:  $r=-0.93$ ,  $p=4.6e-7$ , early > late  $p=1.7e-8$ ; Monkey R:  $r=-0.54$ ,  $p=2e-1$ , early > late  $p=2.4e-2$ ).

## Modeling neural variance changes over learning with Factor Analysis

Having quantified control refinement over training, we sought to model how the direct cells coordinated to explore and change neural activity patterns underlying control. Under what conditions would independent patterns (described by private variance, Figure 2.2 A) or covariation patterns (described by shared variance, Figure 2.2 B) be preferred for generating movement variability? When a subject has no idea which neural readout dimensions the decoder uses, private variance is beneficial because this high-dimensional exploration would generate variability in any readout space (Figure 3.11 A). Low-dimensional shared variance might be misaligned with the readout space and thus produce little movement variability (Figure 3.11 B). On the other hand, when the subject has experience with the readout space, low-dimensional shared variance aligned with the readout space is more efficient in generating movements because it concentrates more variance in fewer neural dimensions (Figure 3.11 C) and it leads to faster learning (Sadtlter et al., 2014). Thus, flexible-network learning combines these properties, allowing the population to expansively explore using private variance and then consolidate control-specific shared variance. Importantly, exploring the behavioral consequences of activity outside the initial shared space could guide the shared space to rotate or change dimensionality (Figure 2.2 C).

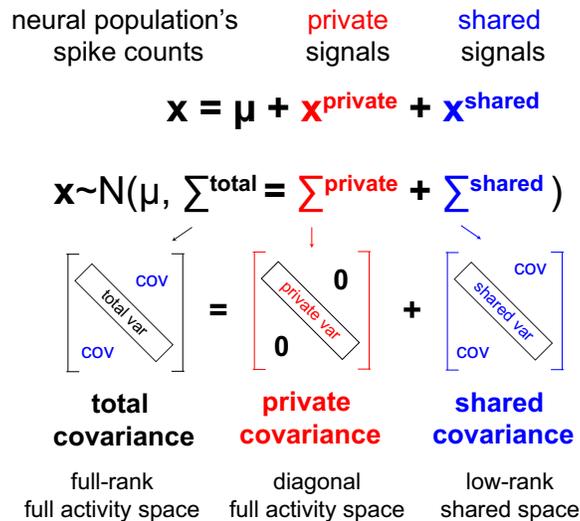


Figure 2.7: **Factor Analysis Covariance Decomposition.** Factor Analysis (FA) decomposes the spike count covariance matrix into the sum of a diagonal private variance matrix and a low-rank shared variance matrix.

Seeking evidence for these neural activity changes, we used FA (Figure 2.7) to decompose population firing rate activity into the sum of 1) a mean rate, 2) private signals which drive each neuron without correlation (Figure 2.8 A), and 3) low-dimensional shared signals which

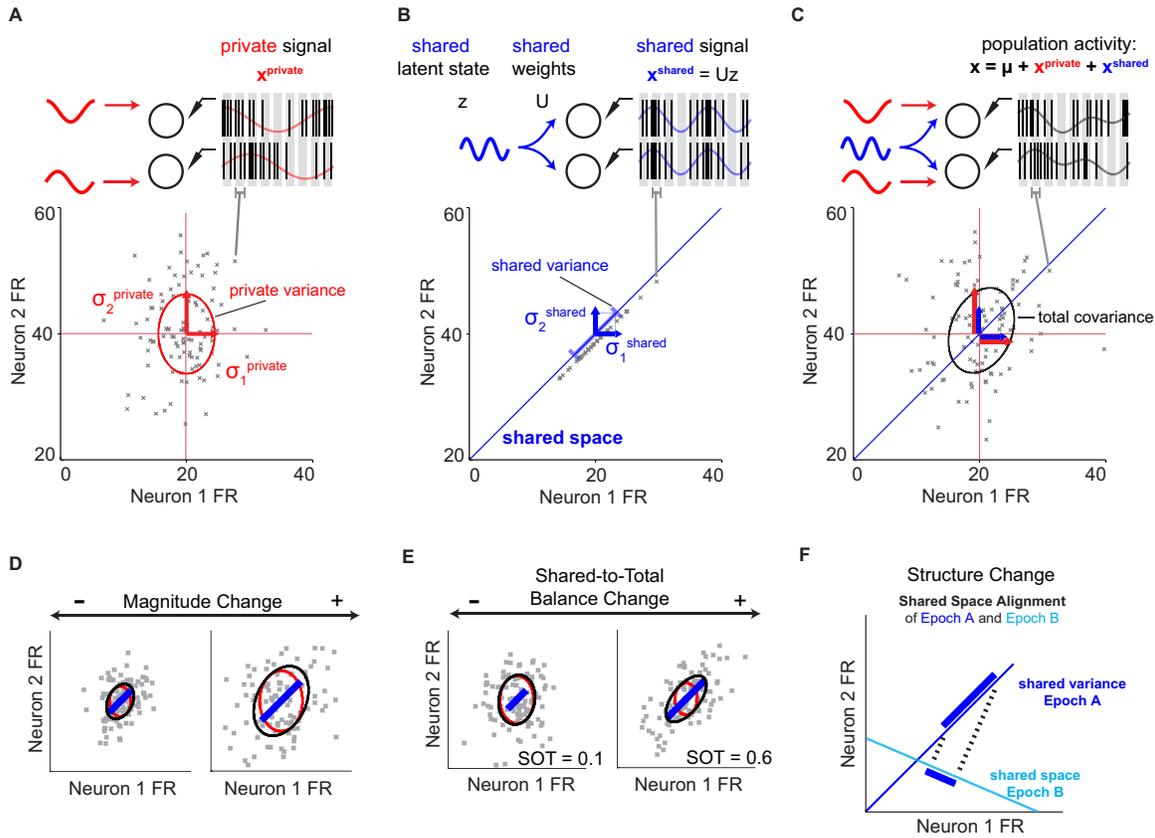


Figure 2.8: Variance Analyses via Factor Analysis.

Figure 2.8: Variance Analyses via Factor Analysis. (A) Private signals modulating two neurons’ firing rates. Each scattered gray ‘x’ is a binned population spike count. Neuron 2 has a larger private variance than neuron 1. (B) Shared signals modulating two neurons’ firing rates within a one-dimensional shared space. (C) FA describes the mixture of private and shared variance underlying population activity. (D) FA detects changes in the total magnitude of variance. (E) FA also detects changes in the shared-to-total variance ratio, quantifying how shared and private variance change in different proportions. The shared-to-total ratio plus the private-to-total ratio equals 1. (F) FA detects changes in how the population co-varies using the shared space alignment between Epoch A and B, which is the fraction of Epoch A’s shared variance captured in Epoch B’s shared space.

drive multiple neurons simultaneously, produce correlations, and constrain activity to the shared space (Figure 2.8 B). Thus, FA models the population’s total covariance matrix as the

sum of the diagonal covariance matrix due to uncorrelated private signals and the low-rank covariance matrix due to shared signals, permitting variance only within the shared space (Figure 2.7, 2.8 C). Our use of private variance and shared variance specifically refers to their respective covariance matrices. We emphasize that FA is fit for each target separately on the full simultaneously-recorded direct cell population, not just neuron-pairs as illustrated for geometric intuition (Figure 2.8). Model-fitting assessed the shared space dimensionality (number of shared signals) needed to best describe the firing rates (see Experimental Procedures).

We used FA to analyze how the neural population’s private and shared variance change over learning (Figure 2.7). First, the magnitude of total variance can change (Figure 2.8D). Second, the balance of shared and private variance can change, a measure of coordination which we quantified with the shared-to-total variance ratio (Figure 2.8 E). Third, the shared space containing shared variance can change, which we quantified by calculating the alignment between two epochs’ shared variances (Figure 2.8 F, Experimental Procedures). The shared space alignment, ranging between 0 and 1, asks how well Epoch B’s shared signals can approximate Epoch A’s shared signals. If Epoch A and B have perfectly aligned shared spaces, then the alignment is 1, and if the shared spaces are orthogonal, the alignment is 0. Thus, if the shared space stays fixed over learning as predicted by constrained-network learning, we expect alignment values to be high throughout training. In contrast, if the shared space consolidates with learning as predicted by flexible-network learning, we expect high alignment values only in late-learning. We used these three analyses to study how private and shared sources of variance contribute to the exploration and consolidation of activity patterns.

## Over learning, private trial-to-trial variability decreases while the shared space changes and stabilizes

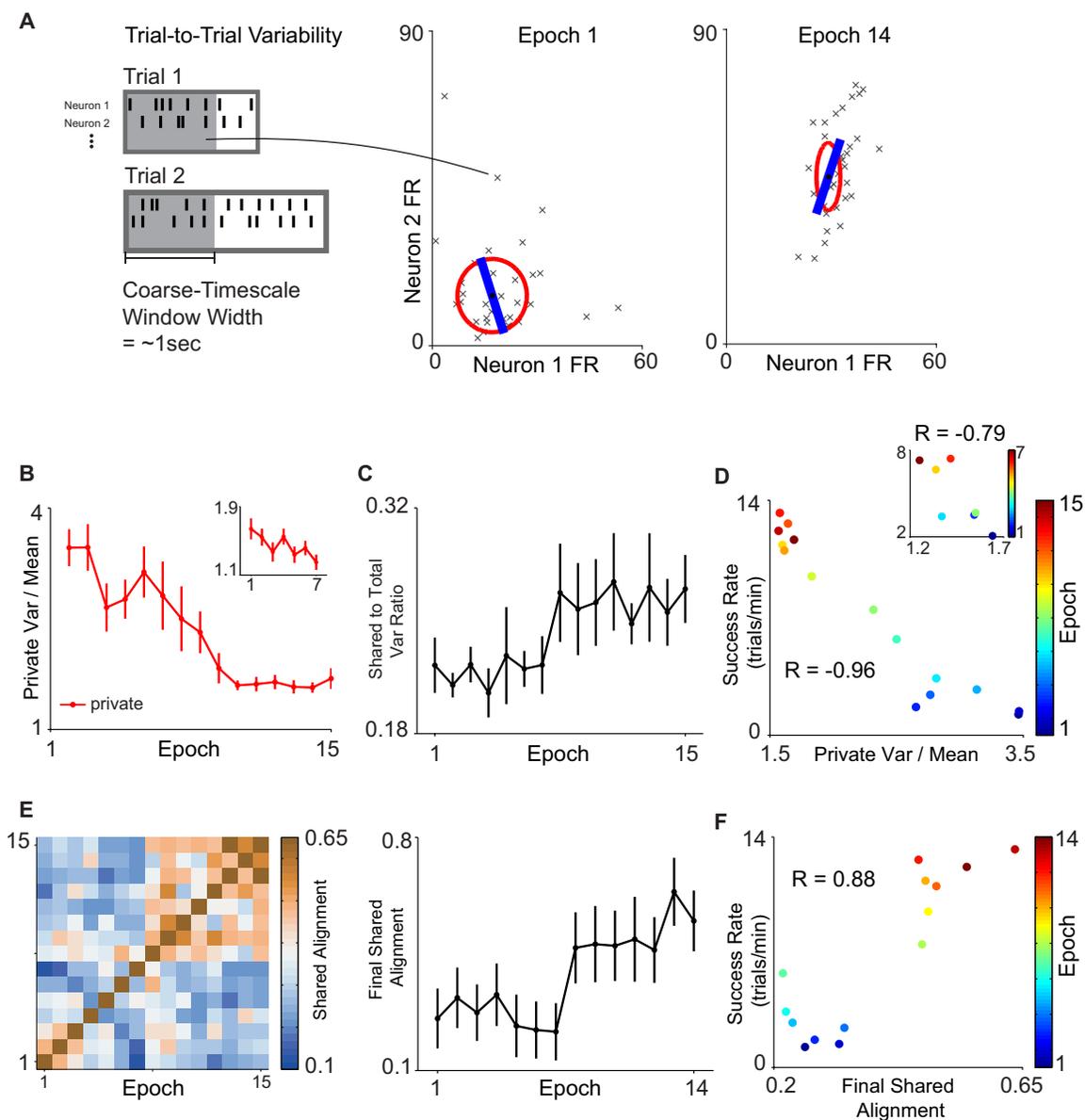


Figure 2.9: **Private Trial-to-Trial Variability Decreases and the Shared Trial-to-Trial Variability Consolidates a Shared Space.**

While previous studies have found trial-to-trial neural variability decreases for a particular intended action over motor learning, we used FA to ask how private and shared sources

Figure 2.9: **Private Trial-to-Trial Variability Decreases and the Shared Trial-to-Trial Variability Consolidates a Shared Space.** **(A)** Trial-to-trial variability example: FA fit on 2 direct cells in early and late training for Monkey P on one example target. **(B)** Private trial-to-trial variability for Monkey P (main) and Monkey R (inset) decreased (Monkey P:  $r=-0.93$ ,  $p=6.7e-7$ , early  $>$  late  $p=1.1e-8$ ; Monkey R:  $r=-0.81$ ,  $p=2.7e-2$ , early  $>$  late  $p=2.2e-2$ ). **(C)** The shared-to-total variance ratio increased over long-term training (Monkey P:  $r=0.83$ ,  $p=1.1e-4$ , early  $<$  late  $p=8.1e-3$ ) but not significantly over shorter training (Monkey R:  $r=0.27$ ,  $p=5.5e-1$ , early  $<$  late  $p=8.2e-2$ ). **(D)** The private trial-to-trial variability decrease correlated with success rate improvement (Monkey P:  $r=-0.96$ ,  $p=1.9e-8$ ; Monkey R:  $r=-0.79$ ,  $p=3.5e-2$ ). **(E, left)** The shared space alignment map indicated consolidation. Each element is the alignment of the row epoch’s shared variance with the column epoch’s shared space. The upper right red block indicates that the shared space becomes relatively stable, while the preceding blue indicates the shared space is unstable in early training. **(E, right)** The shared space alignment between the final epoch’s shared space and the preceding epochs’ shared variance increased with long-term training (Monkey P:  $r=0.83$ ,  $p=2.6e-4$ , early  $<$  late  $p=5.8e-4$ ; Monkey R:  $r=-0.12$ ,  $p=8.2e-1$ , early  $>$  late  $p=4.9e-1$ ). The mean chance variance alignment is 0.07, and the 95th percentile chance alignment is 0.25 for Monkey P (see Experimental Procedures). **(F)** The shared space alignment with the final epoch’s shared space correlated with success rate over long-term training (Monkey P:  $r=0.88$ ,  $p=3.3e-5$ ; Monkey R:  $r=0.07$ ,  $p=9e-1$ ). Changes in mean firing rate do not explain these results (Figure 3.15). These results also held for analyses over training days rather than epochs (Figure 3.12) and different window lengths (Figure 3.14).

influenced trial-to-trial variability. We analyzed firing rate variability at the start of trials for each target separately. To avoid capturing within-trial variance used for control, we extracted one spike count vector from each trial in a large  $\sim 1$  second window locked to Go Cue (Figure 2.9 A, left). Subjects’ trial times determined the window size (Monkey P: 0.9s; Monkey R: 1.3s, see Experimental Procedures for details), but the results were insensitive to window choice (Figure 3.14 B-D). We assessed private trial-to-trial variability beyond that expected for a Poisson process by normalizing each neuron’s private variance by its mean rate (see Experimental Procedures). Shared trial-to-trial variability was calculated with each neuron’s un-normalized shared variance because in Poisson simulations, shared variance did not scale with the experimentally observed firing rates (Figure 3.15 B). This analysis enabled us to disambiguate between independent-neuron learning’s private trial-to-trial variability decrease, constrained-network learning’s shared trial-to-trial variability decrease, and flexible-network learning’s private trial-to-trial variability decrease and shared space consolidation. FA applied to a neuron-pair in early and late training is shown for illustration (Figure 2.9 A), showing a decrease in private trial-to-trial variability, an increase in shared-to-total variance ratio, and rotation of the shared space. At the direct cell population

level, FA found only 1 shared dimension was needed for each target and epoch. Consistent with the example, we found initially large private trial-to-trial variability which decreased prominently over training (Figure 2.9 B, 3.12 A). In contrast, shared trial-to-trial variability showed no significant trend, resulting in an increase in the shared-to-total variance ratio (un-normalized by firing rate) over long-term training (Figure 2.9 C, 3.12 C). These results suggest that neural activity space exploration in early learning was primarily accomplished with private trial-to-trial variability. The decrease in private trial-to-trial variability was correlated with success rate over training (Figure 2.9 D, 3.12 B), indicating the subjects produced desired activity as private trial-to-trial variability reduced. While the magnitude of shared trial-to-trial variability showed no change, its subspace consolidated over training. We assessed the shared space alignment between epoch pairs (Figure 2.9 E left, 3.12 D left) and found the alignment with the final epoch's shared space increased with training (Figure 2.9 E right, 3.12 D right), correlating with the subject's success rate (Figure 2.9 F, 3.12 E). Notably, the emergent shared space possessed a significantly different alignment from the initial shared space. Altogether, these results support flexible-network learning, showing that initial trial-to-trial variability is private to each neuron, and as private trial-to-trial variability decreased, the shared space evolved and consolidated.

## Decoder-timescale shared variance strengthens and consolidates to coordinate population activity

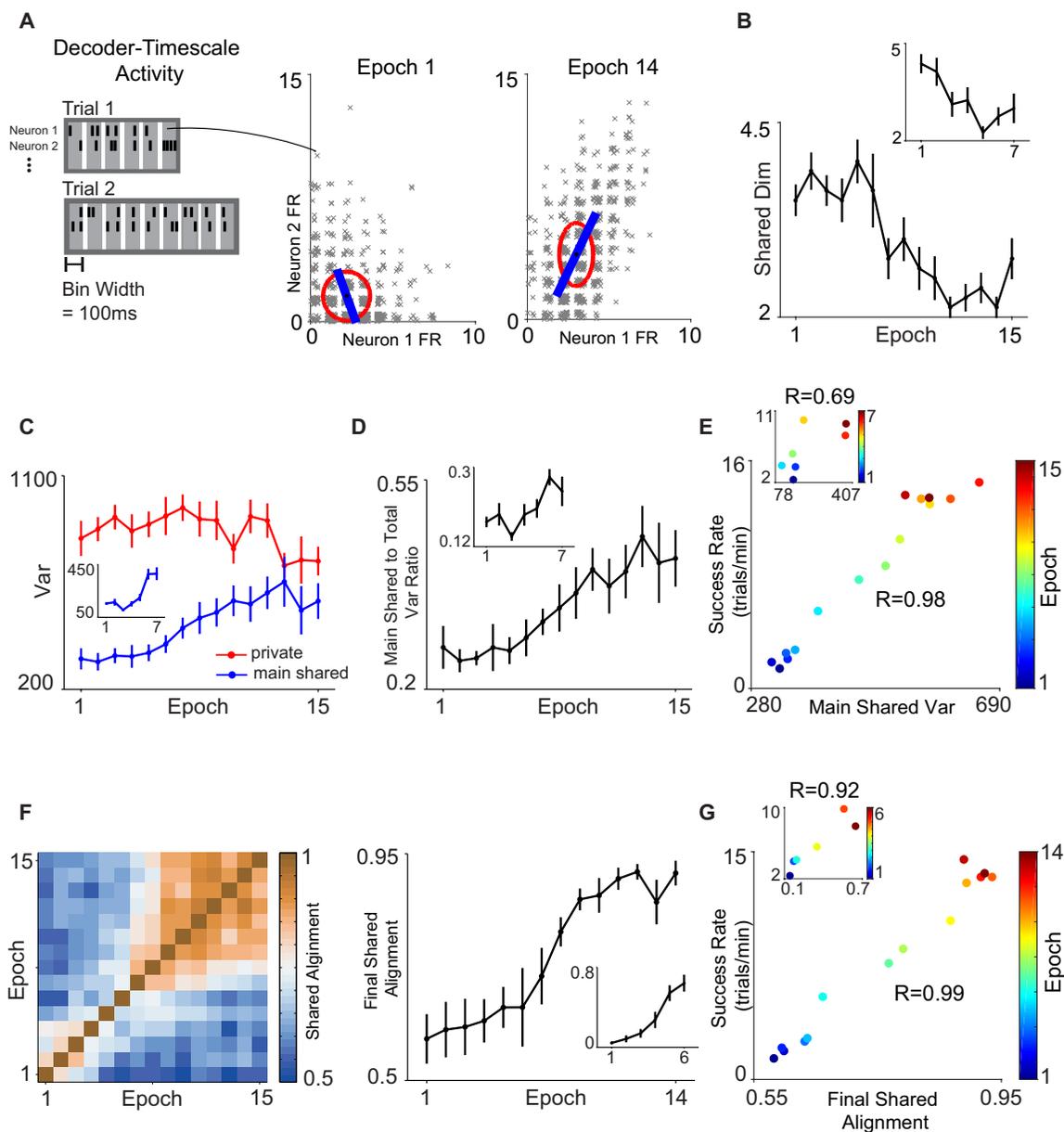


Figure 2.10: Decoder-timescale shared variance increases and consolidates a shared space.

Having analyzed coarse-timescale trial-to-trial variability, we next zoomed in temporal resolution and asked how private and shared sources contributed to control-relevant activ-

Figure 2.10: **Decoder-timescale shared variance increases and consolidates a shared space.** **(A)** Decoder-timescale variance example: FA fit on 2 direct cells in early and late training for Monkey P on one example target. **(B)** The shared dimensionality decreased from 4 to 2 (Monkey P:  $r=-0.85$ ,  $p=5.7e-5$ , early > late  $p=1.4e-7$ ; Monkey R:  $r=-0.80$ ,  $p=3.1e-2$ , early > late  $p=9.8e-5$ ). We used a main shared dimensionality of 2 (see Experimental Methods). **(C)** Main shared variance increased with training (Monkey P:  $r=0.92$ ,  $p=8.1e-7$ , early < late  $p=1.9e-5$ ; Monkey R:  $r=0.84$ ,  $p=2.0e-2$ , early < late  $p=1.2e-5$ ). **(D)** The ratio of main shared variance to total variance increased with training (Monkey P:  $r=0.95$ ,  $p=7.1e-8$ , early < late  $p=2.2e-6$ ; Monkey R:  $r=0.76$ ,  $p=4.6e-2$ , early < late  $p=1.2e-4$ ). **(E)** Main shared variance correlated with success rate (Monkey P:  $r=0.98$ ,  $p=1.9e-10$ ; Monkey R:  $r=0.69$ ,  $p=8.6e-2$ ). **(F, left)** Shared space alignment map. The upper right red block indicates the shared space becomes stable, while the preceding blue indicates the shared space is changing in early training. **(F, right)** Alignment with the final epoch’s shared space increased with training (Monkey P:  $r=0.95$ ,  $p=1.8e-7$ , early < late  $p=9.6e-9$ ; Monkey R:  $r=0.96$ ,  $p=2.2e-3$ , early < late  $p=2.2e-6$ ). The mean chance alignment is (Monkey P: 0.13; Monkey R: 0.20) and 95th percentile chance alignment is (Monkey P: 0.28; Monkey R: 0.40). **(G)** Alignment with the final epoch’s shared space correlated with success rate improvement (Monkey P:  $r=0.99$ ,  $p=7.7e-11$ ; Monkey R:  $r=0.92$ ,  $p=8.3e-3$ ). Changes in mean firing rate do not explain these results (Figure 3.15). These results also held for analyses over training days rather than epochs (Figure 3.13) and different bin widths (Figure 3.14).

ity at the decoder-timescale of 100 ms bins (decoder bin size) for each target separately. Decoder-timescale neural variance can be interpreted as signal strength for BMI control, in contrast to our previous analysis on trial-to-trial variability. We thus use FA to study if the independent-neuron, constrained-network, or flexible network model best describes how learning changes decoder-timescale population activity driving control. We first observed population activity became lower dimensional, as the shared dimensionality describing the 100 ms spike counts within all trials to a single target decreased from 4 to 2 in late training (Figure 2.10 B). Because the brain found a two-dimensional neural solution in late training, we compared shared variance over training in the two neural dimensions which captured the most shared variance, which we call the “main shared variance” (see Experimental Procedures). Note that high-dimensional shared variance did not constitute a significant fraction of total variance (Figure 3.15 F). FA applied to a neuron-pair in early and late training is shown for illustration, in which shared variance increased and rotated (Figure 2.10 A). In concordance, in the entire direct cell population, we found main shared variance increased prominently while private variance decreased mildly over long-term training (Figure 2.10 C, 3.13 A), increasing the main shared-to-total variance ratio by a factor of 2 (Figure 2.10 D, 3.13 B). These results are not specific to our bin width selection, as they held over a range of bin sizes (Figure 3.14 E). Further, main shared variance correlated with success rate

(Figure 2.10 E, 3.13 C), suggesting that it was used for control. We next tested whether a shared space was consolidated over training for decoder-timescale variance. By analyzing the main shared space alignment, we found a period of relative instability in early learning followed by a period of stability in late learning (Figure 2.10 F, left). Notably, the stable main shared space of late training was significantly different from the main shared spaces explored in early training (Figure 2.10 F, right). Further, the alignment of main shared variance to the final epoch correlated with success rate (Figure 2.10 G), suggesting main shared variance consolidation supported learning. These results support the flexible-network learning mechanism, suggesting that learning proceeds by increasing shared variance within an emergent shared space which coordinates activity used for BMI control.

## **Decoder-timescale private variance contributes noisy yet successful control, while consolidated shared variance supports skillful control**

Decoder-timescale private and shared variance changed differently over training, so we leveraged the decoder to ask how they contributed to cursor control. Mathematically, both sources are feasible strategies for generating a desired control signal (Figure 2.11 A, left). Private variance represents the limit of high-dimensional activity, such that the population exhibits no correlations, and still can possess the temporal structure to generate desired control signals in the readout space (Figure 2.11 A, middle). Shared variance captures low-dimensional covariation patterns and also can generate desired control signals in the readout space (Figure 2.11 A, right). For each target and epoch, we fit an FA model on decoder-timescale spike counts and used it to calculate the contribution of main shared and private variance to each individual trial of activity (see Experimental Procedures). We then used the main shared and private variance contributions as separate inputs to the decoder and compared the resulting control signals (Figure 2.11 B). Best trials produced by each source are shown, illustrating that both private and main shared variance could produce good control.

Under independent neuron learning, the population uses private variance for BMI control, while for constrained-network learning, the population uses shared variance. Because flexible-network learning permits both sources to contribute to control, we might expect private variance to dominate exploratory control in the beginning, while shared variance might contribute refined control in the end. Interestingly, both private and main shared variance improved how close they drove the cursor to the target (Figure 2.11 C). However, when we compared late training trials for which both sources achieved the target (Monkey P: 832 trials; Monkey R: 385 trials; Figure 2.11 D), we observed that main shared movements were faster (average time difference Monkey P: -0.80 sec; Monkey R: -0.66 sec) and more direct (average distance difference Monkey P: -1.6, Monkey R: -4.0, in units of distance from center to target) (Figure 2.11 EF). Thus, in late training, main shared variance contributed more skillful control. Indeed, offline analysis suggests main shared variance can even produce more skillful control than the total activity (Figure 3.16 B-D).

Main shared variance was not always better than private variance, however. In early training, private variance achieved positions closer-to-target (Figure 2.11 C), and total activity-driven movements explored more of the workspace than main shared movements (Figure 2.11 G). This suggests that initial main shared variance did not vary within the readout space, and that private variance served an exploratory role in early training by contributing behavioral variability. In support of flexible-network learning, private and main shared variance provided different contributions to performance improvement, with main shared variance contributing more skillful control in late learning.

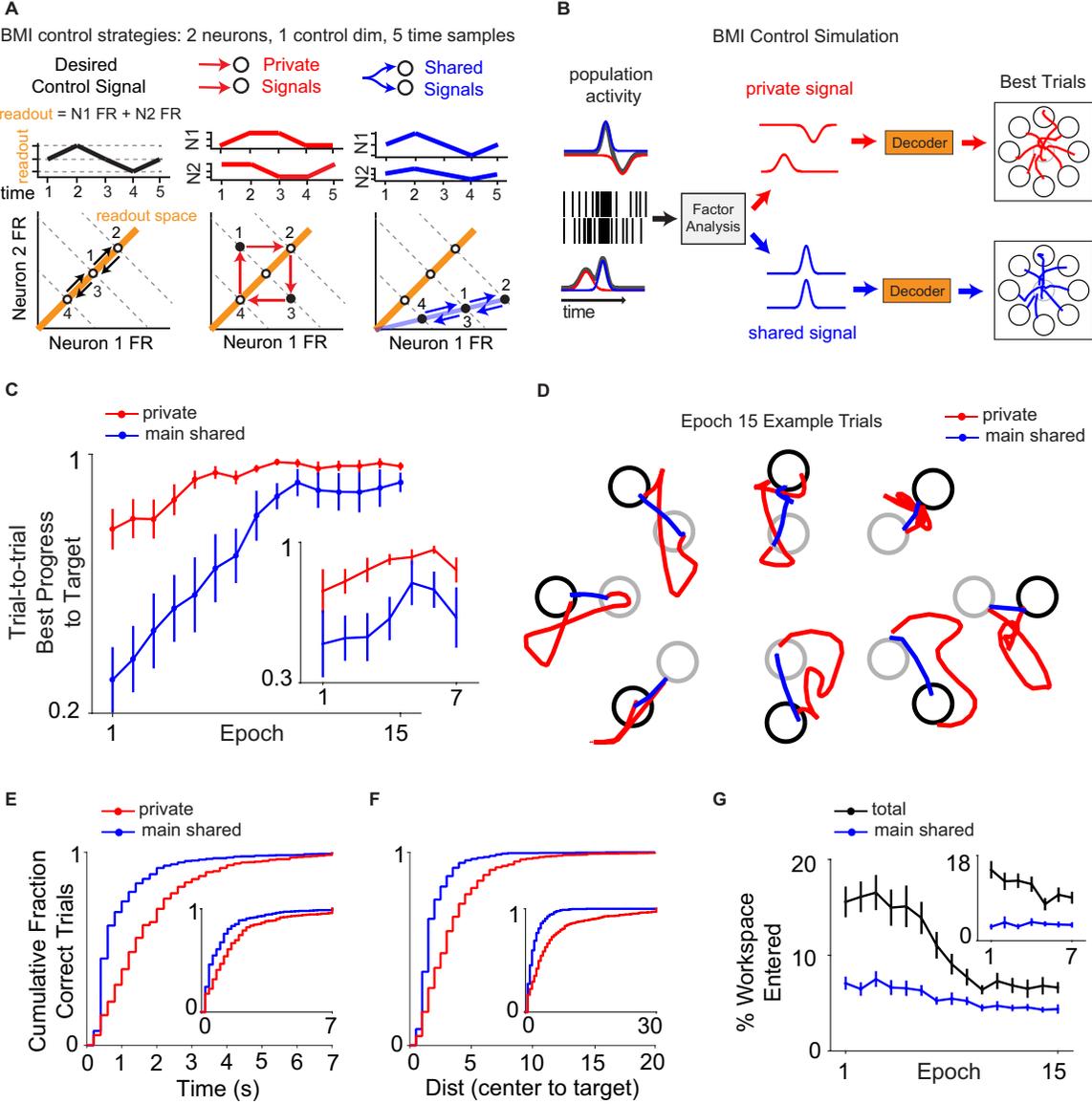


Figure 2.11: Decoder-timescale private variance produces noisy yet successful movements, while shared variance generate skillful movements.

Figure 2.11: **Decoder-timescale private variance produces noisy yet successful movements, while shared variance generate skillful movements.** (A) Both private and shared signals can mathematically produce the same desired control signal. (A, left) The desired control signal is shown over time, and its values are plotted on neuron firing rate axes. The readout value is the sum of the 2 neurons' firing rates. The dashed lines are contours which capture firing rates yielding the same sum. (A, middle) The neurons' activity is uncorrelated and produces the desired control signal. (A, right) The neurons' activity is perfectly correlated and produces the desired control signal. (B) The contribution of private and shared variance to each trial's activity can be estimated and simulated through the decoder separately to determine how they independently contribute to movement. (C) Both main shared and private variance increased best progress to target over training. (Private variance; Monkey P:  $r=0.85$ ,  $p=5.7e-5$ , early < late  $p=2.7e-6$ ; Monkey R:  $r=0.75$ ,  $p=5.3e-2$ , early < late  $p=3.0e-2$ ). (Main shared variance; Monkey P:  $r=0.94$ ,  $p=2.0e-7$ , early < late  $p=3.1e-8$ ; Monkey R:  $r=0.74$ ,  $p=5.8e-2$ , early < late  $p=9.6e-3$ ). (D) Example trials driven by main shared and private variance from the final epoch for each target. (E) In late training, successful main shared variance-driven movements were significantly faster than private variance-driven movements (main shared minus private time-to-target; Monkey P: mean = -0.80 sec, main shared < private  $p=1.8e-56$ ; Monkey R: mean = -0.66 sec, main shared < private:  $p=4.3e-11$ ). (F) In late training, successful main shared variance-driven movements were significantly more direct than private variance-driven movements (main shared minus private distance traveled; Monkey P: mean = -1.6, main shared < private  $p=3.2e-52$ ; Monkey R: mean = -4.0, main shared < private  $p=2.49e-22$ ). Interestingly, main shared variance achieved higher performance than even total activity (Figure S6). For (EF), late training trials were analyzed for which both private and main shared variance were successful (Monkey P: epochs 8 to 15, 832 trials; Monkey R: epochs 4 to 7, 385 trials). (G) Total activity-driven movements explored the workspace more than main shared variance-driven movements, especially in early training (main shared < total, Monkey P:  $p=2.5e-9$ ; Monkey R:  $p=3.2e-9$ ).

## **Control-relevant shared variance increases because of shared variance growth and re-alignment**

How does decoder-timescale shared variance support skillful control (Figure 2.11 D)? We asked how main shared variance occupied the decoder's readout space to produce control signals (Figure 2.12 A), finding that main shared variance in the readout space increased over training (Figure 2.12 B). Two independent mechanisms exist which increase readout space shared variance. First, shared variance itself can increase (magnitude growth, Figure 2.12 C left), as we already found (Figure 2.11 A). Second, the shared space can rotate to align with the readout space (re-alignment), thus making shared variance more efficient in producing control signals (Figure 2.12 C, right). We quantified the relative contributions of these two mechanisms by computing the factor by which main shared variance grew and the factor by which main shared space alignment with the readout space grew. The product of these two yields the factor by which readout space main shared variance grew. Interestingly, both mechanisms significantly increased over training (Figure 2.12 D). This provides one explanation for the changes in the main shared space; it can align with the readout space, making main shared variance more effective in producing control signals. These results provide a glimpse of how motor cortex can strengthen and optimize population coordination for neuroprosthetic control.

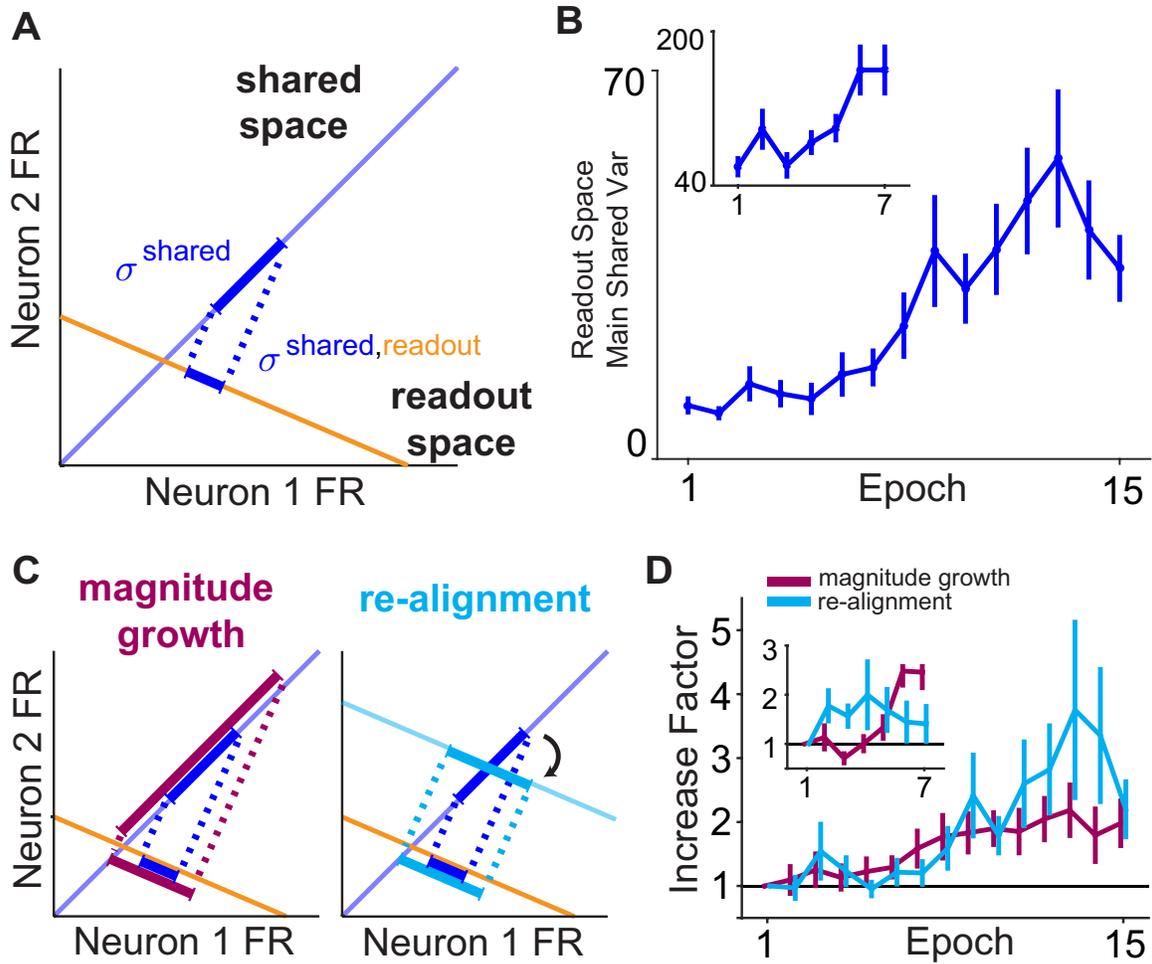


Figure 2.12: Task-relevant shared variance increases because of shared variance growth and re-alignment.

Figure 2.12: **Task-relevant shared variance increases because of shared variance growth and re-alignment.** (A) Shared variance can drive the decoder by producing variance in the decoder’s readout space. (B) Main shared variance in the readout space increased (Monkey P:  $r=0.90$ ,  $p=6.6e-6$ , early < late  $p=8.5e-9$ ; Monkey R:  $r=0.85$ ,  $p=1.5e-2$ , early < late  $p=8.4e-5$ ) (C) Shared variance can increase in the readout space by two mechanisms. (C, left) An increase in shared variance magnitude increases the readout space shared variance, as long as the shared space and readout space initially align. (C, right) Re-alignment of the shared space with the readout space increases the readout space shared variance. Readout space shared variance growth is the product of the magnitude growth and re-alignment. (D) Magnitude growth and re-alignment both increased for main shared variance, plotted as the ratio of each epoch’s value to the first epoch’s value. (Magnitude growth factor; Monkey P:  $r=0.93$ ,  $p=7.3e-7$ , early < late  $p=3.1e-4$ ; Monkey R:  $r=0.82$ ,  $p=2.4e-2$ , early < late  $p=1.6e-5$ ). (Re-alignment factor; Monkey P:  $r=0.84$ ,  $p=9.9e-5$ , early < late  $p=8.2e-6$ ; Monkey R:  $r=0.17$ ,  $p=7.2e-2$ , early < late  $p=6.2e-1$ ). While shared space alignment with the readout space increased, note that a significant fraction of shared variance remained outside the readout space in late learning (Figure 3.16 A).

## **Fine-timescale shared neural trajectories are consolidated which produce accurate movement**

Having analyzed how the consolidated main shared variance relates to the readout space, we asked whether the covariation patterns consolidated a temporal sequence which supported skillful control, since shared variance can arise with or without consistent neural trajectories (Figure 2.13 A). To investigate temporal structure, we used finer bins (50 ms) than the decoder-timescale (100 ms), and for each individual target and epoch, we fit FA and calculated the main shared variance contribution to each trial's activity and trial-averaged to form the main shared trajectory (see Experimental Procedures). Early epochs showed no pronounced pattern, while later epochs showed a large and consistent trajectory (Figure 2.13 BC). We found that the magnitude of each target's main shared trajectory increased (Figure 2.13 D) and that the similarity of each target's main shared trajectory to the final epoch increased (Figure 2.13 E), demonstrating neural trajectory consolidation. Was all the observed main shared variance due to these emergent neural trajectories? We took each trial's main shared variance contributions and subtracted the trial-averaged main shared trajectory, leaving behind main shared variation beyond that due to the average trajectory, and calculated how much main shared variance remained. The large fraction of remaining main shared variance shows that the observed coordination of population activity is not simply explained by the activity's average time-course. Interestingly, initial main shared variance showed no trial-averaged neural trajectory. The fraction of main shared variance due to the average neural trajectory increased with training (Figure 2.13 F), showing that main shared variance became increasingly temporally-structured over learning. Finally, we tested whether the main shared trajectories were sufficient for generating neuroprosthetic control. We applied them through the decoder and found that they evolved over learning to generate straight and accurate movements in late training (Figure 2.13 GH). These results show that emergent shared temporal structure produces control signals for skillful movement.

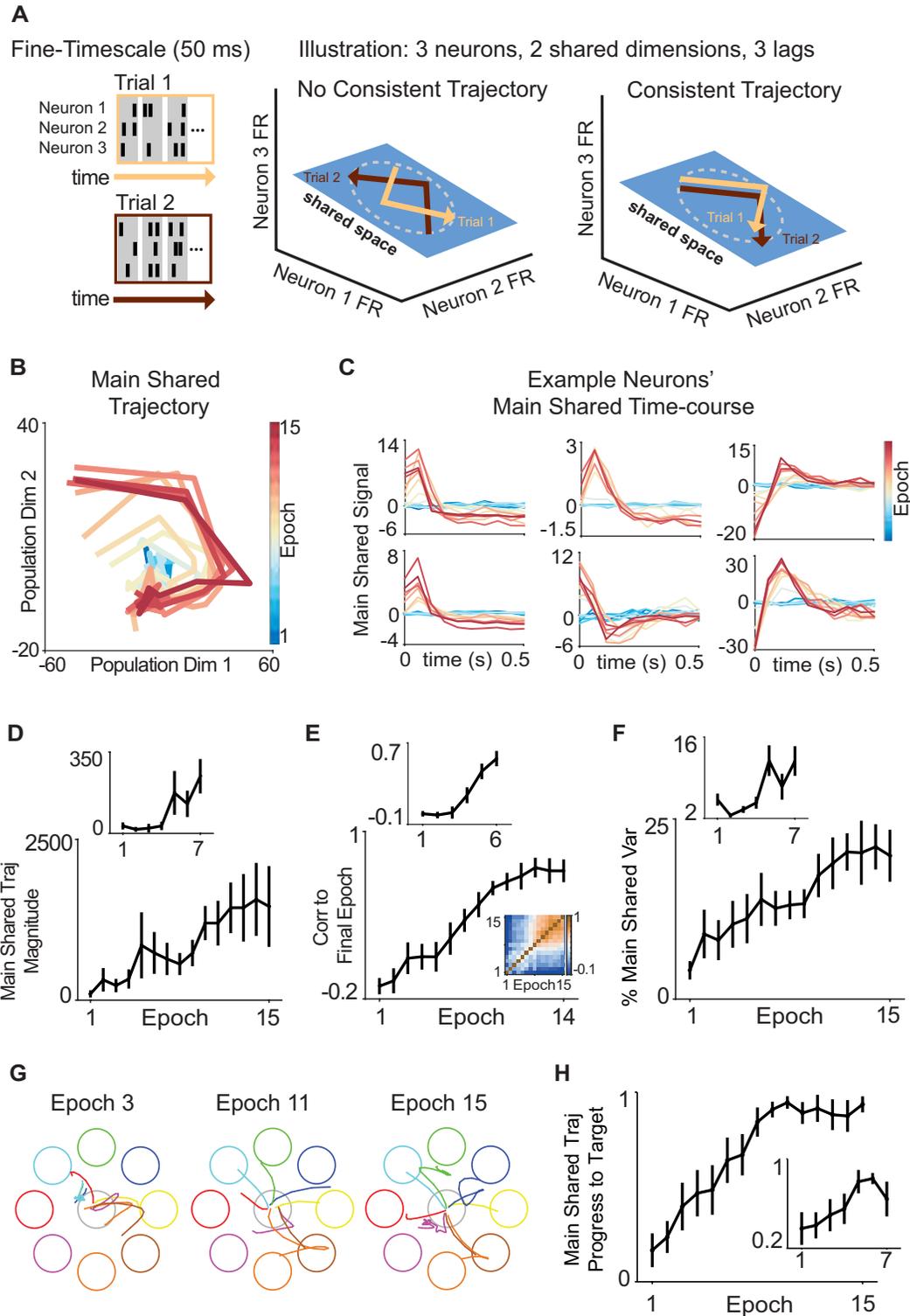


Figure 2.13: Shared neural trajectories are consolidated at fine-timescale which produce accurate movement.

Figure 2.13: **Shared neural trajectories are consolidated at fine-timescale which produce accurate movement.** **(A)** Shared variance can arise without a consistent neural trajectory (left) or with a consistent neural trajectory (right). **(B)** For one example target, every epoch’s trial-averaged main shared trajectory is plotted in the same informative 2D plane of the high-dimensional neural space (see Supplemental Experimental Procedures). **(C)** 6 example cells’ main shared time-course over training epochs for one target. **(D)** The magnitude of the main shared trajectory increased, as quantified by the squared norm (Monkey P:  $r=0.95$ ,  $p=8.8e8$ , early < late  $p=9.1e-7$ ; Monkey R:  $r=0.86$ ,  $p=1.3e-2$ , early < late  $p=1.0e-6$ ). **(E)** The correlation between each epoch’s main shared trajectory and the final epoch increased and stabilized (Monkey P:  $r=0.98$ ,  $p=3.0e-9$ , early < late  $p=2.2e-9$ ; Monkey R:  $r=0.93$ ,  $p=6.3e-3$ , early < late  $p=5.5e-8$ ). **(F)** The percent of fine-timescale main shared variance due to the trial-averaged main shared trajectory increased (Monkey P:  $r=0.96$ ,  $p=9.7e-9$ , early < late  $p=4.4e-6$ ; Monkey R:  $r=0.77$ ,  $p=4.4e-2$ , early < late  $p=7.1e-5$ ). **(G, H)** The trial-averaged main shared trajectory simulated through the decoder increased movement accuracy (Monkey P:  $r=0.92$ ,  $p=1.4e-6$ , early < late  $p=1.6e-8$ ; Monkey R:  $r=0.84$ ,  $p=1.8e-3$ , early < late  $p=5.9e-4$ ).

## 2.4 Discussion

There are substantial challenges to investigating the neural basis of classical motor skill learning, as experimenters cannot observe all cells driving movement, cannot track the precise neural-movement transformation, and cannot analyze how a complete task-relevant population coordinates over learning. In this work, we leverage the reduced preparation of the BMI paradigm to study how the direct cells driving the decoder explored activity patterns from trial-to-trial, and how they consolidated control signals underlying skillful movement. Using FA, we uncovered that private and shared signals both evolve and contribute to control differently in a task-relevant population over training.

### Private signals contribute initial neural and prosthetic variability

Previous studies have shown that neural variability reduction in motor brain areas accompanies motor adaptation (Mandelblat-Cerf et al., 2009), procedural learning (Barnes et al., 2005), skill learning (Costa et al., 2004; Kao et al., 2005; Peters et al., 2014; Santos et al., 2015), and neuroprosthetic learning (Arduin et al., 2013; Zacksenhouse et al., 2007), suggesting that the brain searches for goal-achieving patterns by modulating neural variability. We asked how a task-relevant neural population coordinates to search for goal-achieving activity patterns. By dissecting total trial-to-trial variability with FA, we found that initially large trial-to-trial variability which is private to each neuron decreases over training (Figure 2.9 B). This result extends (Zacksenhouse et al., 2007), which found that early neuroprosthetic training was accompanied by increased total firing rate variability which was unexplained by BMI kinematics. Thus, a task-relevant population can explore activity space by increasingly varying each neuron independently, suggesting that the brain possesses flexibility in finding goal-achieving patterns throughout population activity space.

How did initial decoder-timescale private variance contribute to cursor variability? We supplied shared and private variance separately as simulated inputs to the decoder and found that private variance contributed large cursor workspace exploration in early training (Figure 2.11 G) which achieved positions closer to the target, while shared variance initially had little impact on control (Figure 2.11 C). Thus, our findings suggest that each neuron in the task-relevant population uses independent patterns to explore behavioral states in early learning.

### Role of private signals in solving the credit assignment problem

Neuroprosthetic learning studies provide intriguing evidence that the brain can solve the credit assignment problem by specifically adapting the neurons which contribute to the global error signal provided by prosthetic cursor feedback (Ganguly et al., 2011; Gulati et al., 2014; Jarosiewicz et al., 2008; Koralek et al., 2013). To explain changes specific to the neurons driving errors, neuroprosthetic learning models require private neural noise to provide a beneficial exploratory signal for learning (Héliot et al., 2010; Legenstein et al., 2010). These

models posited a fixed level of exploratory neural noise, and simulations indicated that this noise can hurt final cursor performance. Thus, our observation of initially large and later reduced private trial-to-trial variability is consistent with the view that motor cortex varies neurons independently to learn their contribution to control and then reduces this variability to generate refined control.

## Shared variance consolidates for skillful neuroprosthetic control

It is known that particular neural ensembles (Cao et al., 2015; Peters et al., 2014) and spatiotemporal patterns are consolidated during natural motor learning (Costa et al., 2004; Peters et al., 2014; Santos et al., 2015) and neuroprosthetic learning (Carmena et al., 2003; Ganguly and Carmena, 2009; So et al., 2012a), but less is known about whether and how the population coordinates over learning. We asked: for skilled control, does each neuron use independent patterns, or does the population utilize covariation patterns? For both trial-to-trial variability and decoder-timescale activity, we found that the shared-to-total variance ratio increased (Figure 2.9 C, 2.10 D) and that population activity is coordinated in a stable, low-dimensional shared space (Figure 2.9 E, 2.10 F). Using decoder simulations, we observed that emergent decoder-timescale shared variance contributed more skillful control than both private variance and total activity (Figure 2.11 EF, 3.16 CD). We hypothesize that the population finds this solution by selecting particular shared inputs which produce goal-achieving activity within a characteristic manifold.

## Flexibility in acquiring neural patterns

Recent studies have asked how flexible the brain is in acquiring neural patterns for neuroprosthetic control (Hwang et al., 2013; Sadtler et al., 2014). (Sadtler et al., 2014) found that the structure of shared neural variance plays a causal role in neuroprosthetic learning; given one training session, it was easier for a subject to produce control signals within an existing BMI manifold of shared variance than outside the BMI manifold. Another study (Hwang et al., 2013) found that subjects can operate a BMI by re-purposing activity patterns associated with their physical movements. Thus, given a familiar task context, subjects more readily select neural patterns within their pre-existing repertoire, facilitating rapid control. In our study, subjects possess no starting BMI control strategy as they learn a neuroprosthetic skill de novo over days of training. Using a decoder which permitted decoder-timescale variance from both private and shared sources as input, the neural population's initial shared variance turned out to be small (Figure 2.10 C), misaligned with the decoder (Figure 2.12 B, 3.16 A), and produced neither significant cursor variability (Figure 2.11 G) nor accurate movements (Figure 2.11 C). Instead, the population initially generated private variance unconstrained to a manifold (Figure 2.10 C), and over training, shared variance increased and a BMI manifold emerged containing consolidated neural trajectories for skillful control (Figure 2.11 EF, 2.13 H). Because private variance permits subjects to observe the behavioral consequence of activity patterns outside the initial shared space, it may enable the consolidation of BMI

manifolds and increase the space of patterns and thus decoders that subjects can learn. We hypothesize this flexible neural exploration and consolidation mechanism is employed for long-term novel skill learning, distinct from the more constrained neural exploration used for faster learning.

## Refined control through matching of neural activity to the effector

To control a novel effector, subjects must discover motor control neural activity which is matched to the effector. A recent neuroprosthetic learning study found that changes in neurons' mean firing rates across movement directions correlate with the decoder's properties (Orsborn et al., 2014). In this work, we asked how neural dynamics for individual movement directions drive the decoder (Figure 2.12 A), finding that shared variance in the decoder's readout space increased over training (Figure 2.12 B). This occurs both because shared variance increases overall, and because the shared space aligns with the readout space (Figure 2.12 D). Thus, neuroprosthetic skill emerges as subjects consolidate shared variance matched to the decoder. Given that subjects could also generalize control for untrained reaches (Ganguly and Carmena, 2009), these neural changes may reflect the development of an internal model for BMI control (Golub et al., 2015; Heliot et al., 2010). How might our findings extend to acquiring novel motor skill, which requires generating novel control signals to the body? The brain's task in BMI approximates current models of the brain's function in natural motor learning. Our decoder is a linear readout of neural activity into control signals, just as a common model posits muscles perform a linear readout of motor cortical activity (Churchland et al., 2012; Kaufman et al., 2014; Lillicrap and Scott, 2013). In analogy to motor cortical control of the decoder, motor control studies suggest that motor cortex generates neural control signals for our biomechanical plant, showing that motor cortical activity contains information about muscle-level dynamics as well as kinematics (Sergio et al. 2005) and possesses statistics which reflect the biomechanics it must control (Lillicrap and Scott, 2013). Thus, our findings on private and shared neural variance in motor cortex could apply to exploring and consolidating physical movements.

## Refined control through consolidation of coordinated neural trajectories

Because we observed decoder-timescale shared variance could produce skilled cursor movements (Figure 2.11 D-F), we asked whether covariation patterns acquired a temporal sequence at a fine-timescale which supported learning. Interestingly, initial covariation patterns traverse no consistent trajectory (Figure 2.13 D). Over training as a shared space consolidates, neural trajectories consolidate within the shared space (Figure 2.13 EF) which are sufficient to drive straight and accurate cursor movements (Figure 2.13 GH). Future work will need to analyze whether these emergent temporal patterns obey dynamical rules such as those underlying natural motor control (Churchland et al., 2012; Kao et al., 2015). One

interesting observation is that a large fraction of shared variance exists outside the readout space and thus does not directly drive the cursor (Figure 3.16 A), possibly reflecting internal neural dynamics necessary to generate the control signal to the decoder. This is consistent with motor control hypotheses that large neural populations exploit their high-dimensionality such that neural activity can obey lawful dynamics while a linear readout of the activity can produce flexible movements (Churchland et al., 2012) and can prepare movements without executing them (Kaufman et al., 2014). Interpreting neural data will be aided by modeling how neural networks might learn dynamics for BMI control as is being done for motor control (Hennequin et al., 2014; Sussillo et al., 2015).

### **Biological sources of private and shared variance: noise and supporting networks**

We note that private variance in the direct cells can arise from multiple processes (Faisal et al., 2008): 1) “noisy” intrinsic spiking variance at a fixed firing rate, such as in a Poisson Process with fixed firing rate, 2) the action of private input activity changing the cell’s underlying firing rate, and 3) input activity structure which FA does not model, for example time-varying shared variance. We used Poisson simulations to ensure changing mean firing rate did not explain our observed trial-to-trial variability changes (Figure 3.15). In the latter two cases, some of the variance which is private in the population of direct cells may be shared with unobserved cells in connected networks. Thus, private variance need not be purely corruptive noise, a simplifying assumption sometimes made (Santhanam et al., 2009). Cells in M1 likely possess tens of thousands of synapses (Cragg, 1967); the input activity space is very high dimensional. Some of these inputs may be independent for each recorded cell while some may be coordinated across cells. These two classes of inputs would drive distinct statistical population variance, which FA attempts to describe with shared and private variance. There are many potential biological implementations for the observed increase in consistency and coordination in the direct cell population. Private trial-to-trial variability can decrease because variable private inputs become more consistent or their synaptic weights to direct cells weaken. Decoder-timescale shared variance can increase because anatomically shared inputs strengthen in signal or synaptic weight or because non-anatomically-shared inputs become coordinated. Our FA results suggest new hypotheses about the neural substrate of skill learning in the direct cells and the supporting network. Indeed, a growing body of research indicates that during BMI control, the brain can enter a novel, wide-scale state producing new representations in indirect cells (i.e. not connected to the BMI) in primary motor cortex (Clancy et al., 2014; Ganguly et al., 2011; Gulati et al., 2014; So et al., 2012b), primary sensory cortex (Clancy et al., 2014), across cortical areas (Wander et al., 2013), and striatum (Koralek et al., 2013, 2012). These functional changes are subserved by ensemble re-activation during slow wave sleep (Gulati et al., 2014) and corticostriatal plasticity (Koralek et al., 2012), and suggest that emergent task-relevant coordination is driven by the formation of large-scale circuits including corticostriatal loops.

## Implications for neural prostheses

Advances in neural engineering have produced algorithms to facilitate high performance even when decoders are re-trained frequently. How might decoder and neural adaptation synergize? Recent work shows closed-loop decoder adaptation (CLDA), which uses information about the task goal to fit the decoder to the subject's explored activity patterns (Dangi et al., 2013; Gilja et al., 2012; Shanechi et al., 2016), can improve performance in conjunction with neural learning (Orsborn et al., 2014). Future decoders might benefit from more detailed models of neural population dynamics and how they change with learning. Indeed, a recent algorithmic approach yielded significant performance improvement by modeling neural population dynamics underlying natural movements to decode the subject's intent while moving freely (Kao et al., 2015). Perhaps neural learning can help to generalize this approach to immobile patients, as we found coordinated neural dynamics can be consolidated over training in the absence of overt movement. Given our findings that main shared variance achieves better performance than total activity in simulations (Figure 3.16 B-D), a performance-motivated extension would be to design a decoder that is able to denoise neural observations based on learned neural dynamics (Shenoy and Carmena, 2014). In conclusion, this study uncovers the emergence and consolidation of coordinated dynamics for BMI control, highlighting the potential for neural prostheses to become more natural and automatic to the user, as well as an experimental tool for future investigations on the neural basis of skill learning.

## 2.5 Appendix A

### Supplemental Results

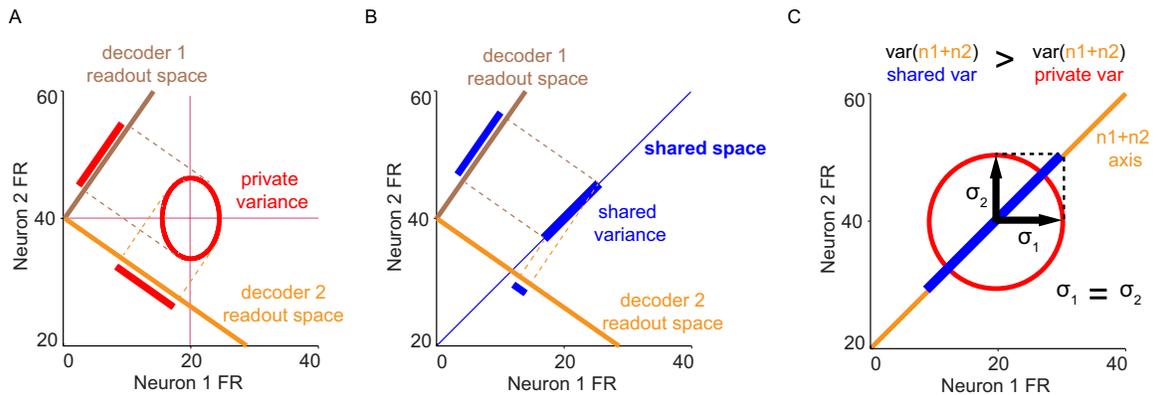


Figure 2.14: **Manners in which private and shared variance generate movement variance depending on the readout space.** (A) Private variance is high-dimensional and therefore generates variance in any readout space. (B) Low-dimensional shared variance generates significant variance only in readout spaces with which it is aligned. (C) Illustration of two neurons with equal variance. If the two neurons exhibit shared variance as opposed to private variance, they can concentrate more variance in the readout dimensions and thus generate movement variance more efficiently.

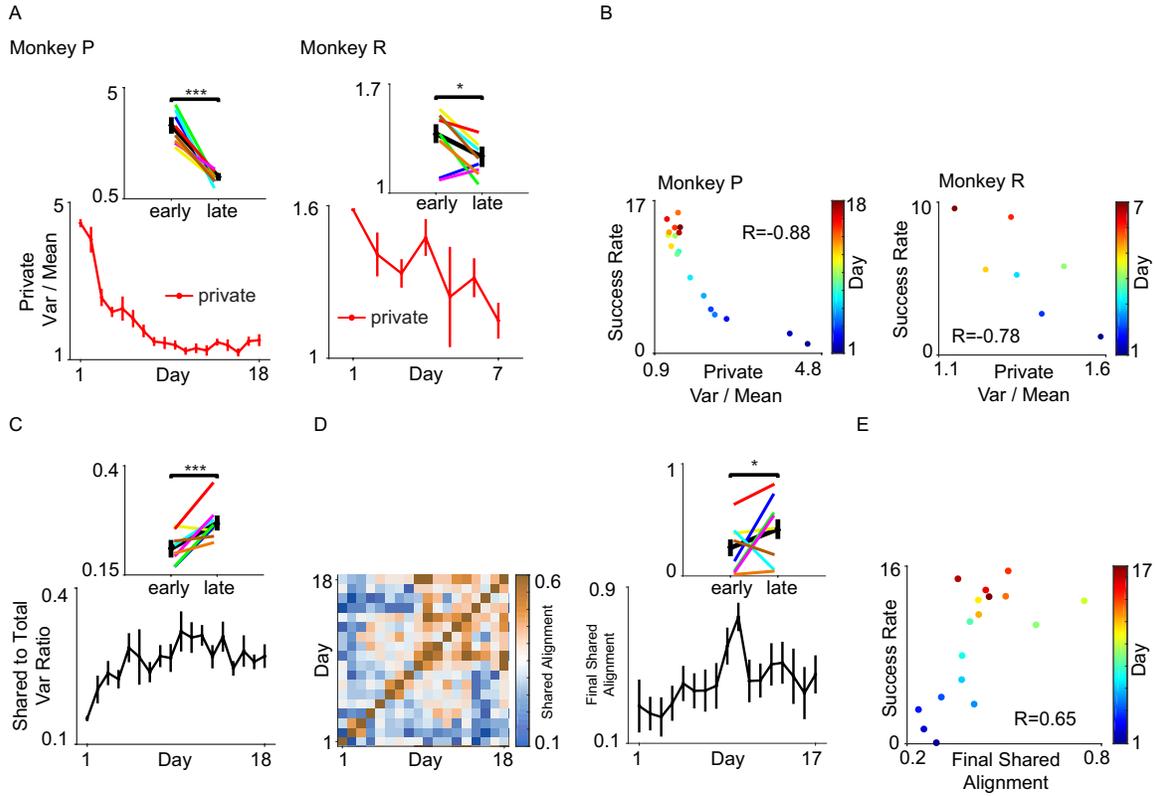


Figure 2.15: **Trial-to-trial variability over days of training.** (A) Over days, private trial-to-trial variability (private variance divided by mean rate) decreased (Monkey P:  $r = -0.77$ ,  $p = 1.7e-4$ , early > late  $p = 2.2e-8$ , 8 targets consistent; Monkey R:  $r = -0.84$ ,  $p = 1.8e-2$ , early > late  $p = 2.4e-2$ , 6 targets consistent). The window size was 0.9 sec for Monkey P and 1.3 sec for Monkey R as in Figure 4B. (B) Private trial-to-trial variability decrease correlated with success rate (Monkey P:  $r = -0.88$ ,  $p = 1.9e-6$ ; Monkey R:  $r = -0.78$ ,  $p = 3.7e-2$ ). (C) The ratio of shared variance to total variance increased (no normalization by mean rate) increased (Monkey P:  $r = 0.58$ ,  $p = 1.2e-2$ , late > early  $p = 6.5e-4$ , 7 targets consistent). (D) The shared space stabilized (Monkey P:  $r = 0.55$ ,  $p = 2.3e-2$ , late > early  $p = 4.5e-2$ , 6 targets consistent). (E) Shared space alignment to the final day correlated with success rate (Monkey P:  $R = 0.65$ ,  $p = 4.7e-3$ ).

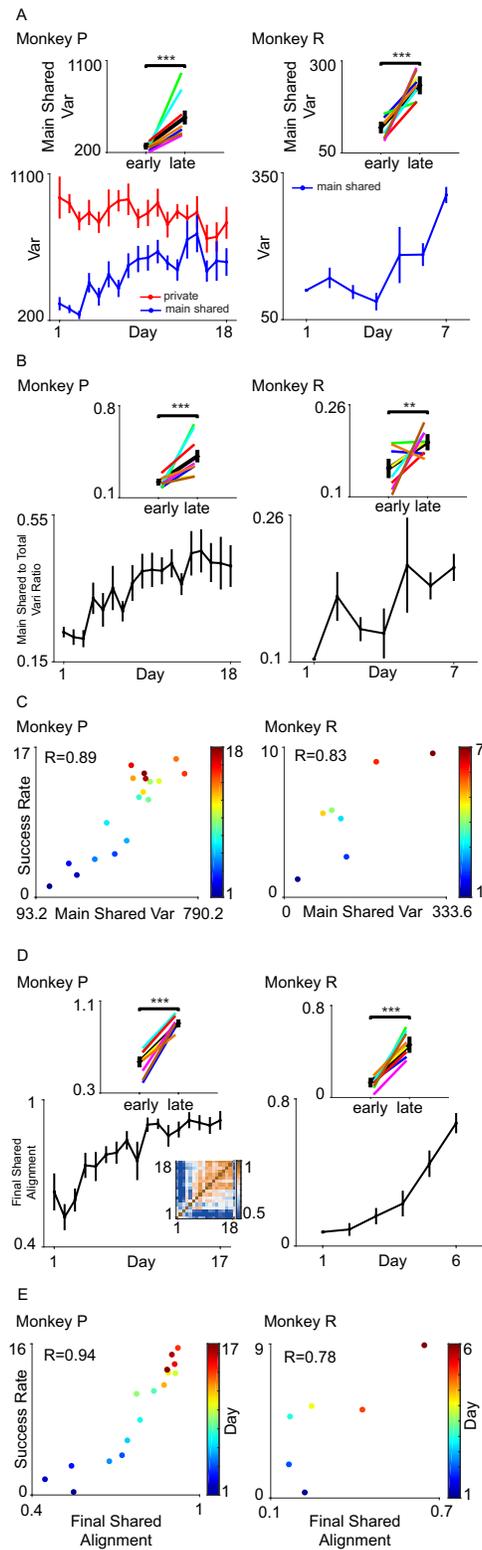


Figure 2.16: Decoder-timescale variance over days of training.

Figure 2.16: **Decoder-timescale variance over days of training.** **(A)** The shared dimensionality started at 4 and fell to 2 in late training, thus a main shared dimensionality of 2 was used. Main shared variance increased (Monkey P:  $r = 0.79$ ,  $p=1.1e-4$ , late > early  $p=1.1e-7$ , 8 targets consistent; Monkey R:  $r=0.77$ ,  $p=4.2e-5$ , late > early  $p=5.2e-3$ , 6 targets consistent). **(B)** The main shared to total variance ratio increased (Monkey P:  $r=0.88$ ,  $p=1.6e-5$ , late > early  $p=3.9e-7$ , 8 targets consistent; Monkey R:  $r=0.77$ ,  $p=4.2e-5$ , late > early  $p=5.2e-3$ , 6 targets consistent). **(C)** Main shared variance correlated with success rate (Monkey P:  $r=0.89$ ,  $p=5.4e-7$ ; Monkey R:  $r=0.83$ ,  $p=2.1e-2$ ). **(D)** Main shared covariance stabilized (Monkey P:  $r=0.89$ ,  $p=1.3e-6$ , late > early  $p=2.1e-8$ , 8 targets consistent; Monkey R:  $r=0.94$ ,  $p=5.5e-3$ , late > early  $p=4.8e-5$ , 8 targets consistent). **(E)** The main shared space alignment to the final epoch correlated with success rate (Monkey P:  $r=0.94$ ,  $p=2.3e-8$ ; Monkey R:  $r=0.78$ ,  $p=6.7e-2$ ).

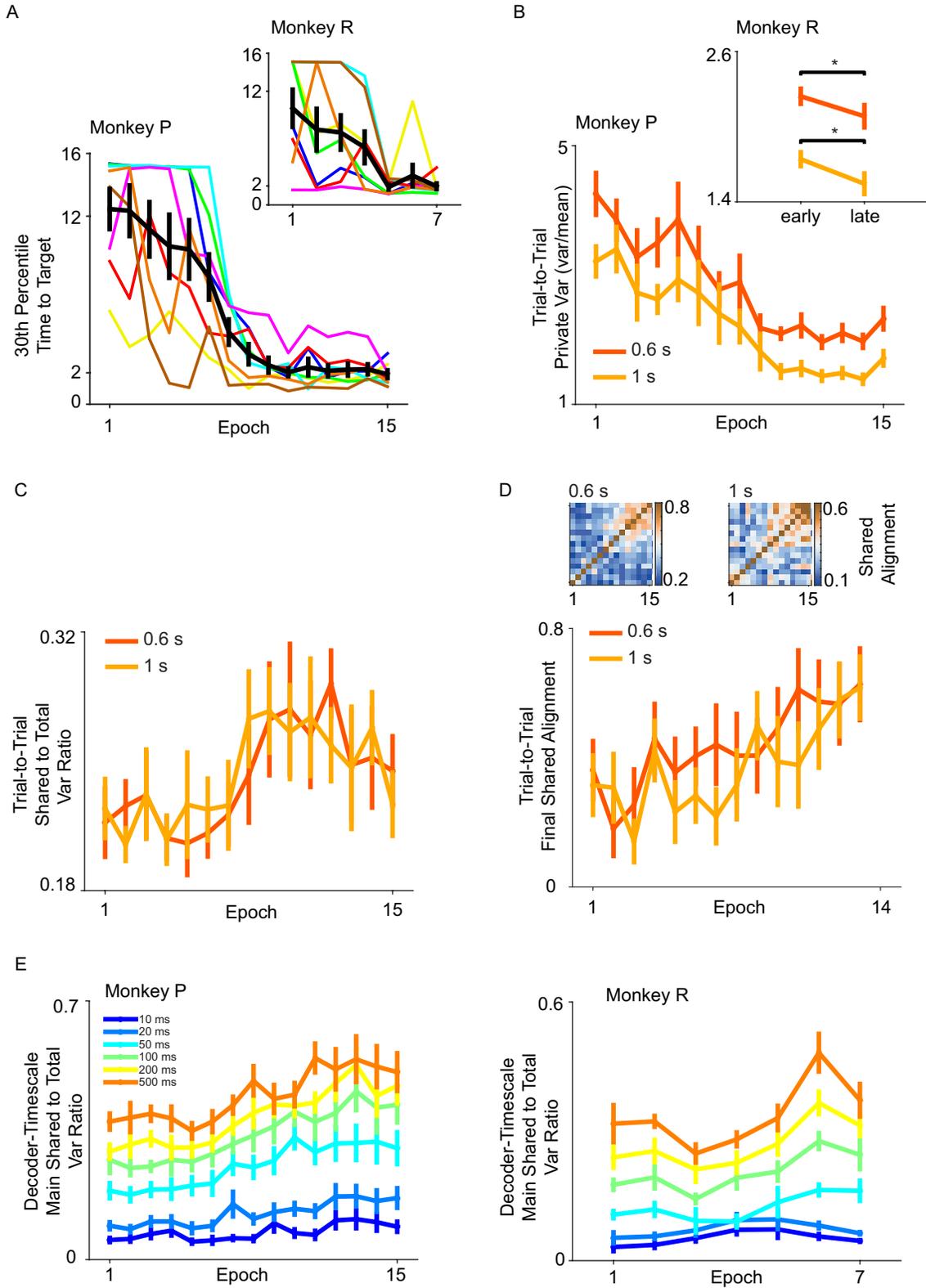


Figure 2.17: Variance results are robust to temporal parameter choice.

Figure 2.17: **Variance results are robust to temporal parameter choice.** **(A)** The 30th percentile of time to target over all targets. Minimum 30th percentile over all conditions was (Monkey P: 0.9 sec; Monkey R: 1.3 sec), which was used in Fig 4 for trial-to-trial variability. **(B)** Trial-to-trial private variability decreased for analysis window = 0.6 sec: (Monkey P:  $r=-0.91$ ,  $p=2.4e-6$ , early > late  $p=2.9e-12$ , 8 targets consistent; Monkey R:  $r=-0.45$ ,  $p=3.1e-1$ , early > late  $p=4.4e-2$ , 6 targets consistent), and for analysis window = 1 sec (Monkey P:  $r=-0.93$ ,  $p=4.3e-7$ , early > late  $p=1.2e-8$ , 8 targets consistent; Monkey R:  $r=-0.68$ ,  $p=9.2e-2$ , early > late  $p=1.4e-2$ , 6 targets consistent). **(C)** Trial-to-trial shared to total variance ratio increased for analysis window = 0.6 sec: (Monkey P:  $r=0.67$ ,  $p=6.5e-3$ , late > early  $p=1.0e-3$ , 6 targets consistent) and window = 1 sec: (Monkey P:  $r=0.60$ ,  $p=1.7e-2$ , late > early  $p=2.7e-3$ , 5 targets consistent). **(D)** The trial-to-trial shared space alignment to the final epoch increased for window = 0.6 sec (Monkey P:  $r=0.87$ ,  $p=5.4e-5$ , late > early  $p=6.7e-4$ , 7 targets consistent) and for window = 1 sec (Monkey P:  $r=0.67$ ,  $p=6.5e-3$ , late > early  $p=1.0e-3$ , 7 targets). **(E)** The decoder-timescale main shared to total variance ratio increased over a range of bin widths. Bin width = 10 ms: (Monkey P:  $r=0.72$ ,  $p=2.3e-4$ , late > early  $p=2.8e-3$ , 7 targets consistent, Monkey R:  $r=0.50$ ,  $p=2.6e-1$  N.S., late > early  $p=4.4e-2$ , 7 targets consistent), 20 ms: (Monkey P:  $r=0.89$ ,  $p=1.1e-5$ , late > early  $p=5.9e-6$ , 8 targets consistent; Monkey R: N.S.  $r=0.47$ ,  $p=2.8e-1$ , late > early  $p=0.11$ , 7 targets consistent), 50 ms: (Monkey P:  $r=0.93$ ,  $p=4.9e-7$ , late > early  $p=4.3e-5$ , 8 targets consistent; Monkey R:  $r=0.75$ ,  $p=4.9e-2$ , N.S. late > early  $p=6.6e-2$ , 8 targets consistent), 100 ms: (Monkey P:  $r=0.95$ ,  $p=7.1e-8$ , late > early  $p=2.2e-6$ , 8 targets consistent, Monkey R:  $r=0.76$ ,  $p=4.6e-2$ , late > early  $p=2.9e-2$ , 7 targets consistent), 200 ms: (Monkey P:  $r=0.92$ ,  $p=8.0e-7$ , late > early  $p=2.0e-6$ , 8 targets consistent, Monkey R:  $r=0.72$ ,  $p=6.7e-2$ , late > early  $p=2.3e-2$ , 7 targets consistent), 500 ms: (Monkey P:  $r=0.88$ ,  $p=1.6e-5$ , late > early  $p=2.2e-4$ , 8 targets consistent, Monkey R:  $r=0.58$ ,  $p=1.7e-1$  N.S., late > early  $p=5.2e-2$  N.S., 7 targets consistent).

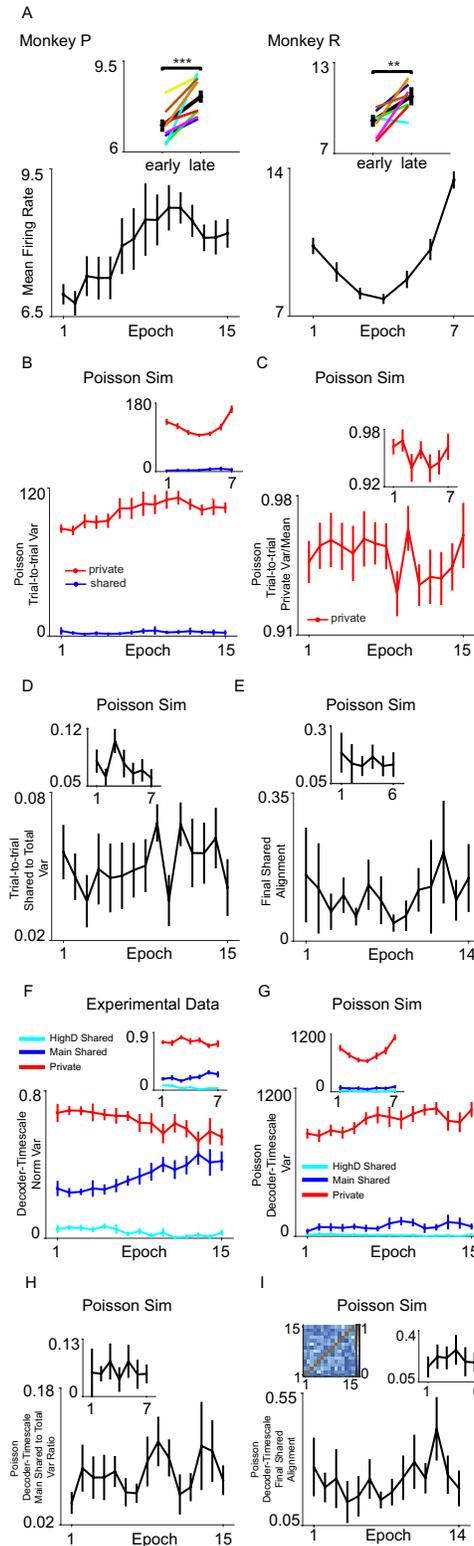


Figure 2.18: Mean firing rate changes do not explain trial-to-trial variability and decoder-timescale variance changes.

Figure 2.18: **Mean firing rate changes do not explain trial-to-trial variability and decoder-timescale variance changes.** **(A)** The mean firing rate increased over training (Monkey P:  $R=0.79$ ,  $p=5.1e-4$ , early < late  $p=7.1e-6$ , 8 targets consistent; Monkey R:  $R=0.49$ ,  $p=2.7e-1$ , early < late  $p=6.1e-3$ , 7 targets consistent). **(B)** We performed FA on a simulated data set of independent Poisson spike counts with window sizes used in Figure 4, the experimentally observed neurons' means, and 50 times the number of trials. Fitting FA on Poisson counts yields the best shared dimensionality as 0. In these plots, we used 1 shared dimension, matching our analysis of real neural data (Figure 4). Monkey P's simulated private trial-to-trial variability increased while Monkey R showed no trend (Monkey P,  $r=0.78$ ,  $p=5.1e-4$ , late > early  $p=2.3e-5$ ; Monkey R n.s. :  $r=0.33$ ,  $p=4.8e-1$ , late > early  $p=1.2e-1$ ). Both subjects' shared variance is close to 0 and showed no trend (Monkey P,  $r=0.29$ ,  $p=2.9e-1$ , early vs late  $p=4.9e-1$ ; Monkey R,  $r=-0.06$ ,  $p=9.1e-1$ , early vs late  $p=1.1e-1$ ). Because shared variance does not track changes in mean firing rate, we chose not to normalize shared variance by mean firing rate. **(C)** As mathematically expected, we observed that normalizing each simulated Poisson neuron's private variance by its mean yielded a value close to 1 with no significant trend over training. The value is expected to be less than 1, because some variance was absorbed as shared variance since we used a shared dimensionality of 1. **(D)** The trial-to-trial shared to total variance ratio showed no trend and is close to 0. **(E)** The trial-to-trial shared space alignment to the final epoch showed no trend (Monkey P:  $r=0.21$ ,  $p=4.7e-1$ , early vs late  $p=1.5e-1$ ; Monkey R:  $r=-0.55$ ,  $p=2.6e-1$ , early vs late  $p=6.9e-1$ ).

Figure 2.18: **Mean firing rate changes do not explain trial-to-trial variability and decoder-timescale variance changes.** **(F)** The experimentally observed data's decoder-timescale main shared, high-dimensional shared, and private variance, normalized by total variance. Normalized private variance decreased (Monkey P:  $r=-0.92$ ,  $p=1.4e-6$ , early > late  $p=1.3e-4$ ; Monkey R:  $r=-0.38$ ,  $p=4.1e-1$ , early > late  $p=7.6e-2$ ), and normalized high dimensional shared signal variance decreased (Monkey P:  $r=-0.80$ ,  $p=3.8e-4$ , early vs late  $p=1.2e-4$ ; Monkey R:  $r=-0.76$ ,  $p=4.8e-2$ , early > late  $p=8.7e-4$ ). Normalized main shared variance increased (Figure 5D). **(G)** Decoder-timescale shared variance changes are not explained by changes in mean firing rate. We applied FA on Poisson simulated neurons with 100 ms bins using experimentally observed average firing rates and shared dimensionality (Figure 5B). Private variance increased (Monkey P:  $r=0.76$ ,  $p=9.8e-4$ , early < late  $p=2.8e-4$ ; Monkey R:  $r=0.42$ ,  $p=3.4e-1$ , early < late  $p=2.3e-2$ ) and main shared variance was unchanged (Monkey P:  $r=0.39$ ,  $p=1.5e-1$ , early vs late  $p=1.7e-1$ ; Monkey R:  $r=0.34$ ,  $p=4.5e-1$ , early vs late  $p=4.8e-1$ ). High dimensional shared variance decreased slightly because the experimental shared dimensionality decreased (Monkey P:  $r=-0.59$ ,  $p=2.0e-2$ , early > late  $p=2.6e-4$ ; Monkey R:  $r=-0.37$ ,  $p=4.1e-1$ , early > late  $p=8.2e-3$ ). **(H)** Decoder-timescale Poisson simulated neurons' main shared variance to total variance ratio showed no trend (Monkey P:  $r=0.27$ ,  $p=3.4e-1$ , early vs late  $p=4.1e-1$ ; Monkey R:  $r=0.039$ ,  $p=9.3e-1$ , early vs late  $p=6.9e-1$ ). **(I)** Decoder-timescale Poisson simulated neurons' shared space alignment to the final epoch showed no trend (Monkey P:  $r=0.32$ ,  $p=2.6e-1$ , early vs late  $p=1.5e-1$ ; Monkey R:  $r=0.14$ ,  $p=7.9e-1$ , early vs late  $p=4.6e-1$ ).

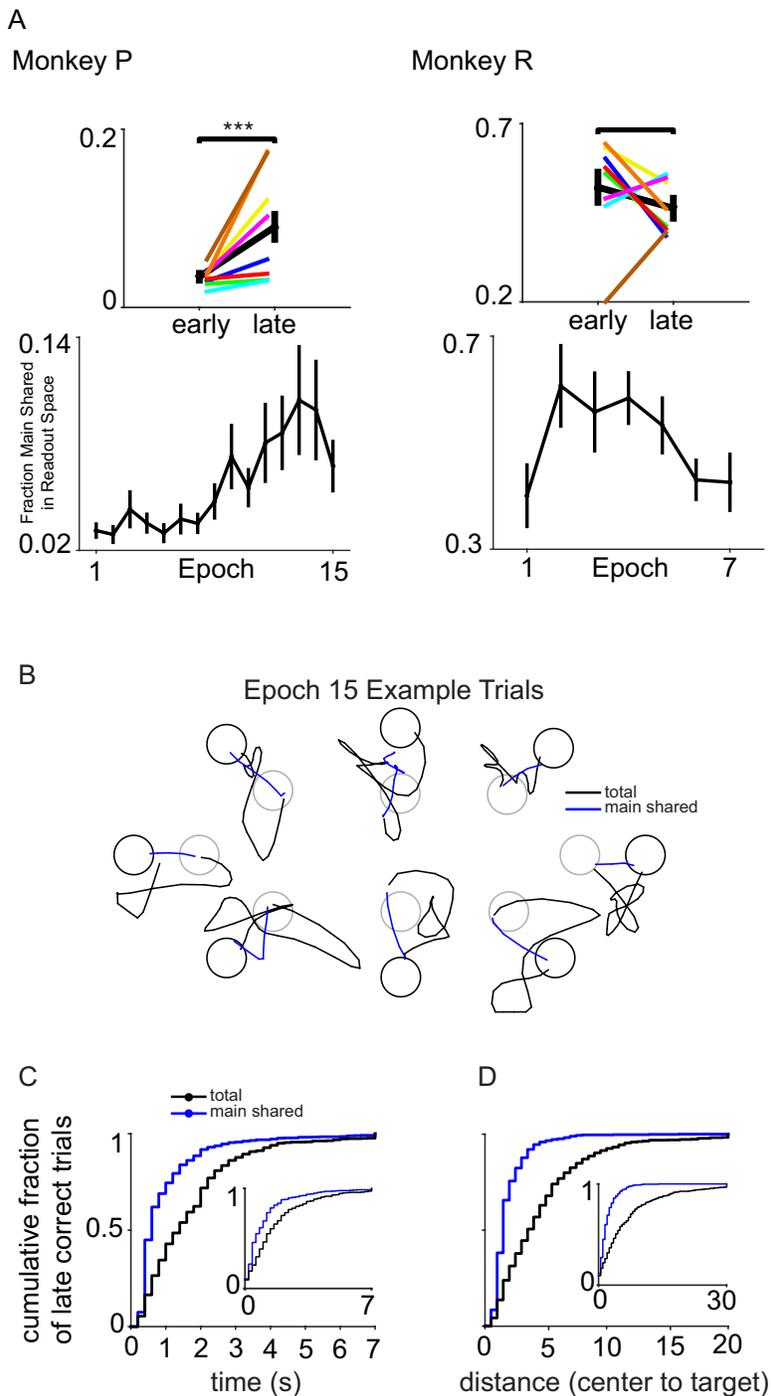


Figure 2.19: Decoder-timescale main shared variance and control.

Figure 2.19: **Decoder-timescale main shared variance and control.** **(A)** The fraction of decoder-timescale main shared variance in the readout space increased for Monkey P ( $r=0.86$ ,  $p=3.9e-5$ , late > early  $p=2.2e-4$ , 8 targets consistent) but not Monkey R ( $r=-0.28$ ,  $p=5.5e-1$ , late < early  $p=1.1e-1$ ). **(B)** Example trials driven by main shared variance and total activity from the final epoch for each target. **(C)** In late learning, successful movements driven by main shared variance are significantly faster (main shared minus total time-to-target; Monkey P: mean = -0.78 sec, main shared < total  $p=4.2e-51$ ; Monkey R: mean = -0.66 sec, main shared < total  $p=3.2e-11$ ). **(D)** In late learning, successful movements driven by main shared variance are significantly more direct (main shared minus total distance-traveled; Monkey P: mean = -2.8, main shared < total  $p=5.38e-83$ ; Monkey R: mean = -5.1, main shared < total  $p=1.8e-28$ ). For (CD), late learning trials were analyzed for which total activity and main shared variance were successful (Monkey P: epochs 8 to 15, 832 trials; Monkey R: epochs 4 to 7, 385 trials).

	Monkey P		Monkey R	
	<u>pval</u>	<u>Num Targets</u>	<u>pval</u>	<u>Num Targets</u>
<b>Trial-to-trial Variability</b>				
trial-to-trial private var ( <u>var/mean</u> )	early > late, p=6.7e-7	8	early > late, p=2.2e-2	6
trial-to-trial shared to total ratio	late > early, p=8.1e-3	7	<u>n.s.</u> : early < late	4
trial-to-trial shared align to final	late > early, p=5.8e-4	7	<u>n.s.</u> : early < late	3
<b>Decoder-timescale Variance</b>				
<u>decoder-ts</u> shared dim	early > late, p=1.4e-7	8	early > late, p=9.8e-5	7
<u>decoder-ts</u> shared main var	late > early, p=1.9e-5	8	late > early, p=1.2e-5	8
<u>decoder-ts</u> main shared to total ratio	late > early, p=2.2e-6	8	late > early, p=1.2e-4	7
<u>decoder-ts</u> main shared align to final	late > early, p=9.6e-9	8	late > early, p=2.2e-6	8
<b>FA Simulations</b>				
best progress to target, private	late > early, p=6.5e-6	7	late > early, p=3.0e-2	7
best progress to target, main shared	late > early, p=3.1e-8	8	late > early, p=9.6e-3	7
time to target	shared < private, p=1.8e-56	8	shared < private, p=4.3e-11	8
distance traveled to target	shared < private, p=3.2e-52	8	shared < private, p=2.49e-22	8
<b>Shared Variance in Readout Space</b>				
readout space main shared var	late > early, p=8.5e-9	8	late > early, p=8.3e-5	7
main shared mag ratio	late > early, p=3.1e-4	8	late > early, p=1.6e-5	8
alignment ratio	late > early, p=8.2e-6	8	<u>n.s.</u> : late > early, p=6.2e-1	3
<b>Shared Trajectories</b>				
main shared <u>traj</u> mag	late > early, p=9.1e-7	7	late > early, p=1.0e-6	7
main shared <u>traj corr</u> to final	late > early, p=2.2e-9	8	late > early, p=5.5e-8	8
<u>frac</u> main shared due to <u>traj</u>	late > early, p=4.4e-6	6	late > early, p=7.1e-5	8
main shared <u>traj</u> progress to target	late > early, p=1.6e-8	8	late > early, p=5.9e-4	8

Figure 2.20: **Summary Table.** Summary of number of targets consistent with analyzed trends.

## Chapter 3

# Evidence for a Neural Law of Effect

In the previous chapter, we studied the structure of neural variability over BMI learning. Now we consider how the brain selects the activity patterns to consolidate.

Thorndike’s law of effect states that actions that lead to reinforcements tend to be repeated more often. Accordingly, neural activity patterns leading to reinforcement are also reentered more frequently. Reinforcement relies on dopaminergic activity in the ventral tegmental area (VTA), and animals shape their behavior to receive dopaminergic stimulation. We found that mice learn to reenter more frequently motor cortical activity patterns that trigger optogenetic VTA self-stimulation, providing evidence for a neural law of effect. Learning was accompanied by gradual shaping of these patterns, with participating neurons progressively increasing and aligning their covariance to that of the target pattern. Motor cortex patterns that lead to phasic dopaminergic VTA activity are progressively reinforced and shaped, suggesting a mechanism by which animals select and shape actions to reliably achieve reinforcement.

### 3.1 Introduction

Thorndike’s law of effect (Thorndike, 1898) states that actions which lead to reinforcements are repeated more frequently (Skinner, 1938). Through repeated attempts, actions are shaped to more directly and reliably achieve reinforcement (Cohen and Sternad, 2009; Shmuelof et al., 2012), a process accompanied by the refinement of behavior-specific neural ensembles and activity patterns in motor cortices (Barnes et al., 2005; Cao et al., 2015; Costa et al., 2004; Ivezky et al., 2011; Peters et al., 2014). Learning occurs because neural patterns initiating actions that lead to reinforcement are reentered more often, as supported by neural activity operant conditioning experiments (Athalye et al., 2017; Carmena et al., 2003; Clancy et al., 2014; Fetz, 1969; Ganguly and Carmena, 2009; Koralek et al., 2012).

Reinforcement is thought to rely on the activity of midbrain dopamine neurons. When animals receive reward, dopamine neurons in the VTA produce a spike burst which encodes the difference between the animals expected and received reward (Schultz et al., 1997). This

reward-prediction error signal is useful for optimizing reward-seeking behavior (Steinberg et al., 2013; Sutton and Barto, 1998). Indeed, phasic VTA activity constitutes a neural basis of reinforcement, as animals shape their behavior to receive electrical (Corbett and Wise, 1980; Olds and Milner, 1954) as well as optogenetic (Tsai et al., 2009; Witten et al., 2011) VTA self-stimulation.

To test a “neural law of effect”, we investigated if mice would learn to reenter specific motor cortical patterns to receive dopaminergic VTA self-stimulation (Figure 4.5). We recorded the activity of tens of neurons in primary motor cortex (M1) and used it to trigger optogenetic stimulation of dopaminergic VTA neurons with blue light (Tsai et al., 2009). Th-Cre mice (Gong et al., 2007) expressing channelrhodopsin-2 (ChR2 group, N=10) in VTA dopaminergic cells were implanted with an optic fiber in VTA and an electrode array in contralateral M1 layer 5 (Figure 3.2, 3.11). To control for the effects of viral expression and light in the VTA in the absence of neuron stimulation, control animals expressing enhanced yellow fluorescent protein (YFP group, N=6) underwent the same experimental procedure. Mice were trained to control a brain-machine interface (BMI) which transformed the activity of groups of neurons in M1 into real-time auditory feedback. When mice produced the target neural activity pattern which led to the target tone, they received a train of blue laser pulses providing phasic stimulation of dopaminergic cells in the VTA. The optogenetic protocol was verified in a lever-pressing reinforcement task (Figure 3.12).

Using this closed-loop self-stimulation paradigm (Grosenick et al., 2015), we sought to measure how neural reinforcement changes the animals’ production of neural activity patterns and resulting occupancy of auditory tones.

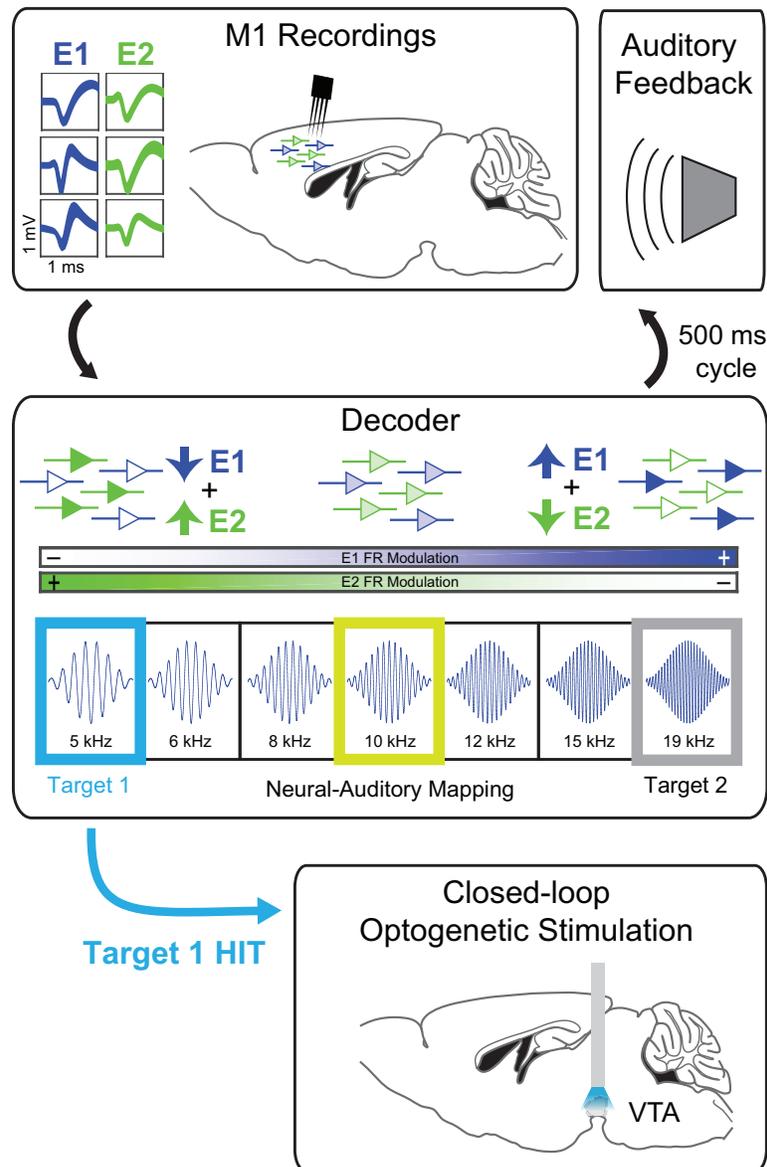


Figure 3.1: **Closed-loop BMI paradigm for pairing specific motor cortex activity patterns with phasic VTA dopaminergic activity.** Schematic of the BMI paradigm. Each mouse receives a unilateral microwire array implant in motor cortex (targeted to layer V) and a contralateral optical fiber implant in VTA. Recorded single units are arbitrarily assigned into two ensembles, and the concomitant increase of one ensemble’s activity and decrease in the other ensemble’s activity drives the decoder to change the auditory tone produced every 500 ms. The rare, lowest tone triggers phasic optical stimulation to the VTA, while the rare, highest tone serves as a control.

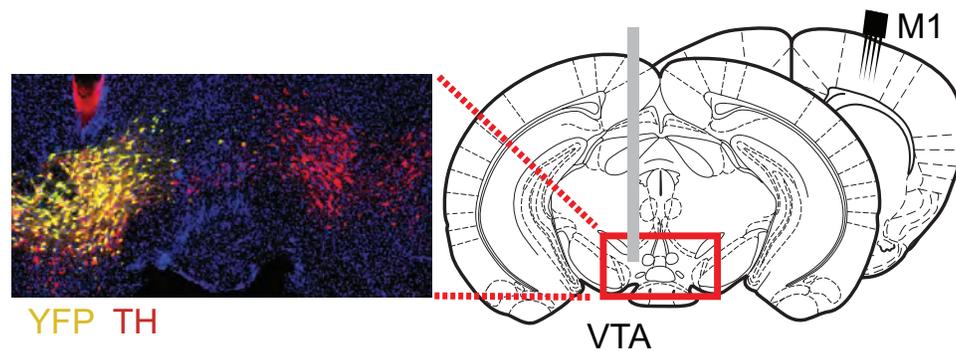


Figure 3.2: **Mouse electrode implantation in M1 and viral infection and optical fiber implantation in VTA.** Each mouse receives a unilateral microwire array implant in motor cortex (targeted to layer V) and a contralateral optical fiber implant in VTA. A coronal brain slice is shown depicting viral infection specific to the dopaminergic cells of the VTA. The immunohistochemistry labels for tyrosine hydroxylase (TH, red) and the cre-dependent fluorescent protein (YFP, yellow).

## 3.2 Methods

### Mice

All animal procedures were performed in accordance with the Champalimaud Foundation ethics committee guidelines and the Portuguese DGAV. 16 male BAC transgenic mice expressing Cre recombinase under the control of tyrosine hydroxylase (FI12) were used in these experiments, ranging in age from 3 to 5 months. Mice were housed with a 12-h light, 12-h dark cycle starting at 7AM. All experiments were done during the light cycle. Mice were housed in groups of 4 animals prior to surgery and individually after surgery.

### Electrode array implants

We stereotactically targeted M1 layer 5 for all electrode implants (from bregma: anterior 1 mm, lateral 1.6 mm, ventral 1 mm) (Franklin and Paxinos, 2008). Mice were anesthetized with isoflurane and chronically implanted unilaterally with 16 tungsten electrodes (diameter 23-35 $\mu$ m) with micro-polished tips arranged in either a 2x8 fixed array or a movable bundle of 16 wires. Electrodes were manually lowered slowly, and neural activity was monitored on all channels online to ensure proper electrode function and positioning. Electrode positioning was verified after experiments via perfusion of animals with PFA and histological confirmation of Nissl stained 70 $\mu$ m brain slices.

## Viral injections and optical fiber implants

We stereotactically targeted the VTA contralateral to the recording electrodes for all viral injections (from bregma: posterior 3.16 mm, lateral 0.48 mm, and ventral 4.4 mm) using an automated microprocessor controlled microinjection pipette (Nanoject II; Drummond, Broomall, PA) with micropipettes pulled from borosilicate capillaries. We injected the ChR2 group of mice (N=10) with 1  $\mu$ L of AAV2/1.EF1a.DIO.hChR2-eYFP (titer  $4.04 \times 10^{12}$ ) and the YFP group (N=6) with 1  $\mu$ L of AAV2/1.Ef1a.eYFP (titer  $1.85 \times 10^{12}$ ). Injections were performed with 0.2 Hz frequency pulses, with 4.6 nL injection volume per pulse. The micropipette was kept at the infection site for  $\sim 20$  min before withdrawal. Next, an optical fiber (200  $\mu$ m) core diameter, 0.22 NA) was implanted 0.2 mm above the viral injection. Following the implantation, a plastic cap was used to cover the fiber ferrule. Before further experiments, mice recovered in their home cage for 15 days to allow for viral expression.

## Histology

Final placement of the electrodes was monitored online during surgery based on neural activity, and then confirmed histologically at the end of experiments. Mice were deeply anesthetized under isoflurane and transcardially perfused with saline followed by 10 % formalin. Brains were fixed in a solution of 30 % sucrose and 10 % formalin, sectioned with a cryostat (coronal slices of 70  $\mu$ m), and Nissl stained.

## Electrophysiology

Single and multi-unit activity was recorded with a Neural Signal Processor (NSP; Blackrock Microsystems, Salt Lake City, UT). Neural activity was sorted using the online sorting application prior to each daily experimental session. Only units with a clearly identified waveform and high signal-to-noise ratio were used. For data analysis, further offline sorting of selected units was done using Plexon Offline Sorter v3 (Plexon, Inc, Dallas, TX) based on waveform characteristics, inter-spike intervals, and PCA clustering. Behavioral timestamps from the operant box were input into the NSP and synchronized with the neural data for later analyses.

## Optogenetics

A 473 nm laser was coupled to individual patchcords which mated to the chronically-implanted optical fiber. The laser intensity at the tip of the optic fiber was  $\sim 10$ mW at the start of each experiment. This was verified by measuring laser power out of an optic fiber with nearly equivalent efficiency to the implanted fiber.

## Verification of Optogenetic Protocol with Motor Reinforcement Task

We confirmed that optogenetic VTA stimulation would lead to motor reinforcement (Tsai et al., 2009; Witten et al., 2011) (Figure 3.11). Mice (N=2) received a ChR2 injection in VTA and were implanted with one optical fiber using the parameters previously listed. For behavioral training, mice were placed in an operant chamber (Med-Associates) with one inactive lever and one active lever which triggered optogenetic stimulation consisting of 10 ms pulses at 20 Hz for 2s duration (Figure 3.11). Sessions lasted for either 30 minutes or one hour. Power at fiber tip was confirmed to be  $\sim 10\text{mW}$  before each session.

## Brain-Machine Interface Task

Animals began BMI training two weeks after implantation surgery. Animals trained for 4 consecutive daily sessions in a behavioral box with speakers and controllable lights. Recorded neural data was streamed in real time to custom routines in Matlab, which translated neural activity into the appropriate auditory feedback frequency which was updated with a period of 0.5 seconds. The BMI output one of seven frequencies, ranging from 5 to 19 kHz in quarter-octave increments to match rodent discrimination thresholds (Han et al., 2007).

Animals initiated a trial by producing the middle tone (10 kHz), resulting in extended middle tone playback for 1 second, followed by auditory feedback updated every 0.5 seconds. Achievement of either of the two possible target tones (Target 1: 5 kHz; Target 2: 19 kHz) resulted in an extended playback of the target tone for 1 second, and an inter-trial interval of 3.5 seconds. The only difference between the two targets was that upon Target 1 production, a Matlab-controlled Data Acquisition I/O board triggered a 14 Hz laser train of 28 pulses with 10 ms width, delivered 0.5 seconds after the 1 second target tone feedback concluded. If animals did not achieve a target in 60 seconds, they received one second of white noise and a timeout of 10 seconds.

## Brain-Machine Interface Decoder

From the population of recorded neurons for each animal, 4-8 units were selected as input to the BMI based on waveform signal-to-noise ratio and ISI histogram (Figure 3.13). These units were arbitrarily assigned into two ensembles of 2-4 units, and the same putative units were kept in each ensemble across training days. If signal quality decreased considerably, a neighbor unit with higher signal quality was substituted. The decoder binned each unit's spiking activity with 500 ms bins and was calibrated to transform the two ensembles' activity into auditory feedback in the following way.

Each experimental session began with a baseline block for decoder calibration which determined the baseline chance rate of tone occupancy during spontaneous neural activity. Daily calibration ensured the baseline chance distribution of tone occupancy did not change over training days despite potential day-to-day variability in neural recordings (Figure 3.3).

During calibration, 500 samples of 500 ms bin spike counts were collected while the animal freely behaved in the behavioral box without receiving auditory feedback. To ensure the decoder’s input was not dominated by units with high firing rate or dynamic range, each unit’s spike counts were median-centered and range-normalized. Each ensemble’s modulation was defined as the sum of its member units’ normalized spike counts:

$$\text{Ensemble Modulation}(t) = \sum_{i \in \text{Ensemble}} \frac{x_i(t) - \text{median}(x_i)}{\text{range}(x_i)}$$

Where  $x_i(t)$  is the spike count for neuron  $i$  at time bin  $t$ , and the median and range are calculated from neuron  $i$ ’s baseline spike count distribution. For each ensemble, 4 modulation states ( $S_{E1}, S_{E2} \in \{1, 2, 3, 4\}$ ) were defined as the intervals demarcated by the 10th, 50th, and 90th percentile of the ensemble modulation distribution. The decoder calculates the differential ensemble modulation ( $S_{E1} - S_{E2} + 4 \in \{1, \dots, 7\}$ ) for each 500 ms cycle and uses it to index the auditory feedback tones (in kHz: [5 6.3 7.8 9.8 12.2 15.3 19.1]). To smooth the auditory feedback, the decoder filtered each ensemble’s modulation with a weighted moving average [ $w(0) \ w(\Delta) \ w(2\Delta)$ ] = [0.7 0.2 0.1] with  $\Delta = 500\text{ms}$  before calculating the differential ensemble modulation.

## Behavioral Data Analysis

Each behavioral session consisted of a baseline block in which the decoder was calibrated and a BMI block in which subjects controlled an auditory cursor and received stimulation upon producing the target tone. To quantify how animals’ tone production distribution changed with training, we used two normalizations. First, for each session we normalized each animal’s BMI tone distribution to the baseline tone distribution. (Note that animals did not receive tone feedback during the baseline period; applying the calibrated decoder to baseline neural activity offline resulted in the baseline tone distribution.) Second, to quantify each animal’s change from their Session 1’s performance, we divided each session’s baseline-normalized tone distribution by Session 1’s baseline-normalized tone distribution, as illustrated for one animal in (Figure 3.12B) and averaged over animals in (Figure 3.12C).

We analyzed whether tone occupancy was different as a function of tones by using a one-way repeated measures ANOVA over tones for each session and for Chr2 and YFP animals separately. Additionally, we used bootstrapping to ask if the animal-averaged normalized tone distribution was significantly different from animals performing BMI simply using their baseline tone distribution. We performed 100,000 BMI simulations for each animal and session, drawing BMI tone samples from the baseline tone distribution for that animal and session. In our simulations, we drew the same number of BMI tone samples as we measured during experimental sessions. We normalized the tone occupancy of each simulated BMI experiment 1) by that animal-sessions baseline and 2) by Session 1. We then constructed a bootstrap distribution over the 100,000 simulations. This is shown in gray for one example

animal in (Figure 3.12B). Finally, we constructed a population-level bootstrap distribution by averaging each simulation over animals, as shown in gray in (Figure 3.12C).

Next, we analyzed how the animals' occupancy and preference for Target 1 changed over learning. For each animal and session, we computed three quantities from the tone distributions, and normalized them by their baseline and by Session 1:

1. Target 1 Occupancy, which is the probability of occupying Target 1 in BMI Session X:

$$\text{Prob.in\_BMI\_SessionX}(T1)$$

Target 1 Occupancy Gain:

$$\frac{\text{Prob.in\_BMI\_SessionX}(T1)/\text{Prob.in\_base\_SessionX}(T1)}{\text{Prob.in\_BMI\_Session1}(T1)/\text{Prob.in\_base\_Session1}(T1)}$$

2. Preference for occupying Target 1 versus Target 2, which is the probability of occupying Target 1 given that a target was hit:

$$\text{Prob.in\_BMI\_SessionX}(T1|\text{target hit})$$

Preference Gain for Target 1 versus Target 2:

$$\frac{\text{Prob.in\_BMI\_SessionX}(T1|\text{target hit})/\text{Prob.in\_base\_SessionX}(T1|\text{target hit})}{\text{Prob.in\_BMI\_Session1}(T1|\text{target hit})/\text{Prob.in\_base\_Session1}(T1|\text{target hit})}$$

3. Preference for occupying low tones versus high tones, which is the probability of occupying a low tone i.e. a tone in the set of  $\{5, 6, 8\text{kHz}\}$  under the condition that the animal did not produce the center tone  $\{10\text{ kHz}\}$ , i.e. that the tone entered was in the set of  $\{5, 6, 8, 12, 15, 19\text{kHz}\}$  :

$$\text{Prob.in\_BMI\_SessionX}(\text{low}|\text{not center})$$

Preference Gain for low tones versus high tones:

$$\frac{\text{Prob.in\_BMI\_SessionX}(\text{low}|\text{not center})/\text{Prob.in\_base\_SessionX}(\text{low}|\text{not center})}{\text{Prob.in\_BMI\_Session1}(\text{low}|\text{not center})/\text{Prob.in\_base\_Session1}(\text{low}|\text{not center})}$$

As before, we used bootstrapping to ask if these quantities were significantly different from animals performing BMI simply using their baseline tone distribution. We constructed bootstrap distributions for the above quantities using the 100,000 BMI simulations which drew BMI tone samples from the corresponding animal-session's baseline distribution:

1. Target 1 Occupancy simulated from the baseline distribution:

$$\text{Prob\_in\_base\_SIM\_SessionX}(T1)$$

Bootstrap Target 1 Occupancy Gain: Bootstrap distribution of normalized values constructed from

$$\frac{\text{Prob\_in\_base\_SIM\_SessionX}(T1)/\text{Prob\_in\_base\_SessionX}(T1)}{\text{Prob\_in\_BMI\_Session1}(T1)/\text{Prob\_in\_base\_Session1}(T1)}$$

2. Preference for occupying Target 1 versus Target 2 simulated from the baseline distribution:

$$\text{Prob\_in\_base\_SIM\_SessionX}(T1|\text{target hit})$$

Bootstrap Preference Gain for Target 1 versus Target 2: Bootstrap distribution of normalized values constructed from

$$\frac{\text{Prob\_in\_base\_SIM\_SessionX}(T1|\text{target hit})/\text{Prob\_in\_base\_SessionX}(T1|\text{target hit})}{\text{Prob\_in\_BMI\_Session1}(T1|\text{target hit})/\text{Prob\_in\_base\_Session1}(T1|\text{target hit})}$$

3. Preference for occupying low tones versus high tones simulated from the baseline distribution:

$$\text{Prob\_in\_base\_SIM\_SessionX}(\text{low}|\text{not center})$$

Bootstrap Preference Gain for low tones versus high tones: Bootstrap distribution of normalized values constructed from

$$\frac{\text{Prob\_in\_base\_SIM\_SessionX}(\text{low}|\text{not center})/\text{Prob\_in\_base\_SessionX}(\text{low}|\text{not center})}{\text{Prob\_in\_BMI\_Session1}(\text{low}|\text{not center})/\text{Prob\_in\_base\_Session1}(\text{low}|\text{not center})}$$

We determined significance for each animal and session by comparing each animal-session to its corresponding baseline bootstrap distribution. At the population-level, we determined significance for ChR2 and YFP separately by comparing the animal-average for each session to its corresponding baseline bootstrap distribution of animal-averages.

We found that 5 animals (ChR2 Learners) individually showed significant Preference Gain for Target 1 versus Target 2 in both Session 3 and 4. The remaining 5 animals (ChR2 Poor Learners) as a population showed significant T1 Occupancy Gain on Sessions 3 and 4, but as a population did not show significant Preference Gain in Session 3 and 4 (Figure 3.16).

## Neural Data Analysis

We analyzed neural activity of the full neural population controlling the BMI (4-8 neurons) in the 3 second window preceding Target 1 hit using spike counts in 100 ms bins. Only animal-sessions with 9 completed trials were considered valid for analysis (we report numbers of animals later).

As in (Athalye et al., 2017), we used Factor Analysis (FA) to model the joint distribution of  $N$  neurons' spike counts  $x \in \mathbf{R}^N$  as the sum of 1) a mean rate  $\mu \in \mathbf{R}^N$ , 2) private signals with diagonal covariance  $\Psi \in \mathbf{R}^{N \times N}$ , and 3) shared signals due to a low-dimensional latent variable  $z \in \mathbf{R}^k$ ,  $k < N$ :

$$\begin{aligned} z &\sim N(0, I) \\ x|z &\sim N(\mu + Uz, \Psi) \\ x &\sim N(\mu, UU^T + \Psi) \end{aligned} \quad (3.1)$$

$U \in \mathbf{R}^{N \times k}$  are the weights (or factors) capturing how  $z$  drives  $x$

For each animal and session, we fit the model's parameters by maximizing the log-likelihood of the data with the EM algorithm for a chosen shared dimensionality  $k < N$ . The best-fitting shared dimensionality  $k$  is estimated by using cross-validated log-likelihood to determine which  $k$  best describes held-out data (Dempster et al., 1977) (Figure 3.12).

We make explicit FA's decomposition of population activity into shared signals  $x^{\text{shared}} = Uz \in \mathbf{R}^N$  and private signals  $x^{\text{private}} \in \mathbf{R}^N$

$$\begin{aligned} x^{\text{private}} &\sim N(0, \Sigma^{\text{private}}) \\ x^{\text{shared}} &\sim N(0, \Sigma^{\text{shared}}) \\ x^{\text{private}} &\perp x^{\text{shared}} \\ x &= \mu + x^{\text{shared}} + x^{\text{private}} \end{aligned} \quad (3.2)$$

$\Sigma^{\text{private}} = \Psi$  is diagonal, and  $\Sigma^{\text{shared}} = UU^T$  is low-rank, permitting variance only to occupy the "shared space," which is the column space of  $U$ .  $\text{Cov}(x) = \Sigma^{\text{total}} = \Sigma^{\text{shared}} + \Sigma^{\text{private}}$ .

Neuron  $i$ 's private variance is  $\Sigma_{ii}^{\text{private}}$ , shared variance is  $\Sigma_{ii}^{\text{shared}}$ , and total variance is  $\Sigma_{ii}^{\text{total}} = \Sigma_{ii}^{\text{private}} + \Sigma_{ii}^{\text{shared}}$ . Each neuron's shared variance to total variance ratio (SOT) is  $\frac{\Sigma_{ii}^{\text{shared}}}{\Sigma_{ii}^{\text{total}}}$ , and the neural population's average is  $\frac{1}{N} \sum_{i=1}^N \frac{\Sigma_{ii}^{\text{shared}}}{\Sigma_{ii}^{\text{total}}}$ .

In (Figure 3.9), the SOT of indirect neurons was determined by fitting FA on all recorded neurons and then averaging the SOT over just the indirect neurons. The SOT of direct neurons was calculated by only fitting FA on the direct neurons to avoid shared variance contributions from indirect neurons. SOT analysis of direct neurons by fitting FA on direct and indirect neurons together showed similar results (Figure 3.19).

We analyzed how SOT changed over training by plotting over sessions (Figure 3.9B) and by comparing SOT pooled over each animal in early sessions (Session 1 and 2) with SOT pooled over late sessions (Session 3 and 4), and using a non-parametric one-sided rank sum test (Figure 3.9B, insets). Because we only analyzed animal-sessions with 9 completed trials, only a subset of total animals were analyzed.

For YFP ( $N=6$ ), the early-pooled data contained 10 animal-sessions (Session 1: 5 animals, Session 2: 5 animals), and late-pooled data contained 11 animal-sessions (Session 3: 4 animals, Session 4: 6 animals).

For ChR2 animals ( $N=10$ ), the early-pooled data contained 17 animal-sessions (Session 1: 10 animals, Session 2: 7 animals), and late-pooled data contained 19 animal-sessions (Session 3: 10 animals, Session 4: 9 animals). We also divided the ChR2 animals into Learners and

Poor Learners based on their behavioral performance. For ChR2 Learners (N=5), the early-pooled data contained 9 animal-sessions (Session 1: 5 animals, Session 2: 4 animals), and late-pooled data contained 10 animal-sessions (Session 3: 5 animals, Session 4: 5 animals). For ChR2 Poor Learners, the early-pooled data contained 8 animal-sessions (Session 1: 5 animals, Session 2: 3 animals), and late-pooled data contained 9 animal-sessions (Session 3: 5 animals, Session 2: 4 animals).

For each valid animal and session, we analyzed the ‘‘Covariance Gain’’ by calculating the neural population’s average SOT, dividing by Session 1’s SOT, and taking the log with base 2:  $\log_2\left(\frac{\text{SOT}_{\text{SessionX}}}{\text{SOT}_{\text{Session1}}}\right)$ . Note that we didn’t normalize SOT by a baseline-calculated SOT because too few Target 1 hits occurred during baseline. For ChR2 Learners (N=5), 4 were analyzed on Session 2, 5 on Session 3, and 5 on Session 4. For ChR2 Poor Learners (N=5), 3 were analyzed on Session 2, 5 on Session 3, and 4 on Session 4. Only 5 of 6 YFP animals were included because one animal did not complete 10 trials on Session 1. For YFP (N=6), 5 were analyzed on Session 2, 3 on Session 3, and 5 on Session 4.

We correlated Covariance Gain with each animal-session’s ‘‘Performance Gain’’ (Figure 3D), which was:  $\log_2\left(\frac{\text{Prob.in.BMI.SessionX}(T1|\text{target hit})}{\text{Prob.in.BMI.Session1}(T1|\text{target hit})}\right)$ . We didn’t normalize the Performance Gain by baseline because we didn’t normalize SOT by baseline. (However, the correlation results are virtually the same if Performance Gain is normalized by baseline.) Note that in (Figure 3.9A, left), Session 4 ChR2 data only included 8 animals. 1 animal was excluded for not completing 9 trials, and another animal was excluded because his BMI tone histogram was not saved due to a computer malfunction at the end of the experiment.

Finally, we analyzed the angle (Bjorck and Golub, 1973; Sadtler et al., 2014) between each animal-session’s shared space, a subspace of the N neurons’ population activity space  $\mathbf{R}^N$ , and the decoder axis, which is a vector in  $\mathbf{R}^N$  containing 1 for Ensemble 1 indices and -1 for Ensemble 2 indices. Projection of neural activity onto the decoder axis captures the difference between Ensemble 1 activity and Ensemble 2 activity. Thus, angle represents the degree to which shared variance projects into the decoder axis (Figure 4.15A). Angle is plotted over all sessions (Figure 4.15B), and compared between early and late sessions (Figure 4.15B inset) using valid animal-sessions as defined earlier for SOT.

We assessed the significance of the FA results for SOT and subspace angle by comparing experimental data to FA applied to a shuffled neural activity data set. The shuffled neural activity data set was generated by shuffling the time labels of each neuron’s binned spike count. This maintains each neuron’s marginal spike count distribution, but breaks the population’s joint distribution. FA was performed on 1000 shuffles of each data set to generate the FA shuffle distributions, and experimental results were compared to the shuffled data set 95% confidence interval (Figure 3.15).

## Behavioral Video Analysis

We assessed the effect of neuroprosthetic VTA stimulation on physical movement by analyzing behavioral videos. We manually annotated the head and back position of the mouse at

Target 1 hit (Figure 3.14A) and analyzed head position variance (Figure 3.14B).

We next sought to analyze movement. We processed the videos to yield binary images of white mouse on black background via background subtraction, thresholding to binary images, and erosion and dilation to remove cables. We then calculated the offset which would translate the mouse’s body at Target 1 hit to the center of the frame and the angle which would rotate the mouse’s body to orient the head up and back down. 1.5 second videos directly preceding Target 1 hit were aligned by applying this offset and rotation. We assessed the positions most commonly occupied by calculating the average of the videos (Figure 3.14C). In addition, we analyzed the spatial distribution of movement by calculating the mean pixel change via the average absolute difference between each consecutive frame (Figure 3.14D). We assessed how specific the movements were by analyzing the spread of the mean pixel change via the image’s entropy (Figure 3.14E). Finally, we assessed the trial-to-trial similarity of pixel change by calculating the mean pixel change for each trial’s video, evaluating the correlation between each trial pairs mean pixel change, and averaging over each trial-pair correlation (Figure 3.14F).

### 3.3 Results

Our closed-loop self-stimulation paradigm (Grosenick et al., 2015) provides a principled way to study neural reinforcement as it assigns chosen recorded neurons (“direct neurons”) to drive behavior, defines the transform between neural activity and behavior via the “decoder,” and delivers temporally-precise reinforcement after target neural activity is produced. Our decoder received input from two arbitrarily-selected M1 ensembles of 2-4 well-isolated single units (see methods, Figure 3.13) (Clancy et al., 2014; Koralek et al., 2012). Two target neural population activity patterns (Target 1 and 2) were specified which occur with equal frequency in spontaneous activity: Target 1 required the simultaneous positive modulation of Ensemble 1 and negative modulation of Ensemble 2, while Target 2 required the reverse modulation (see methods). The BMI provided optogenetic reinforcement of Target 1 only, permitting comparison of the two targets. Further, it provided continuous auditory feedback of neural activity patterns’ exploration along the task-relevant neural dimension - the differential modulation of Ensemble 1 and 2.

We sought to measure how neural reinforcement changes the animals’ production of neural activity patterns and resulting occupancy of auditory tones. The initial conditions of learning were established with decoder calibration to set the baseline chance rate of neural activity patterns occupying the tones. During a baseline block preceding each BMI training block, calibration estimated the distribution of Ensemble 1 and 2 modulations during spontaneous neural activity while mice freely moved in the behavioral box without receiving auditory feedback or VTA stimulation (Figure 3.3).

Each unit’s spiking activity was binned in 500 ms bins, and an ensemble’s firing rate modulation was defined as the sum of each unit’s median-centered and range-normalized spike count. For each individual ensemble, 4 modulation states were defined by the 10th,

50th, and 90th percentile of the modulation distribution from the baseline block. The decoder calculated the difference between Ensemble 1 and Ensemble 2's modulation state for each 500 ms cycle and mapped it to one of seven auditory tones (ranging from 5-19 kHz). This daily calibration yielded a Gaussian-like distribution over tones during baseline and ensured the chance rate of tone occupancy did not change over training days, despite potential day-to-day variability in neural recordings (Figure 3.3). Animals had to produce significant ensemble modulations to achieve the targets (Figure 4.6). During the BMI training block, neural patterns close to Target 1 decreased the tone, while neural patterns close to Target 2 increased the tone (Figure 4.5). Target achievement resulted in a 1 second playback of the target tone, and only Target 1 achievement resulted in phasic VTA stimulation 1.5 seconds after target hit, consisting of a 14Hz train delivered for 2 sec (Figure 3.5).

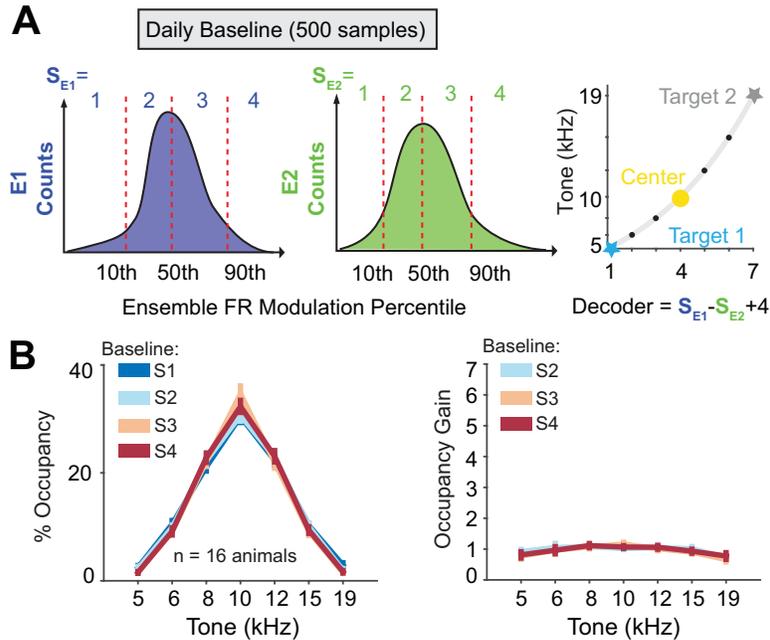


Figure 3.3: **BMI decoder calibration.** (A) Every session during the baseline period, 500 samples of 500 ms spike counts are collected from spontaneous neural activity as the mouse freely behaves in the box with no task or auditory tones. Each ensemble’s firing rate modulation is defined as the sum of the member neurons’ normalized spike counts (mean-centered, range-normalized) and then quantized into 4 activation states. The decoder’s state is the difference between ensemble 1 and ensemble 2’s activation state and is mapped into one of 7 tones. (B) BMI calibration on baseline period spontaneous neural activity results in a Gaussian-like distribution over tones, such that Target 1 (5 kHz) and Target 2 (19 kHz) are rare. The mean and s.e.m. baseline distribution for each session is plotted on the left, averaged over all animals. Baseline distributions show no significant change from session 1, as shown on the right.

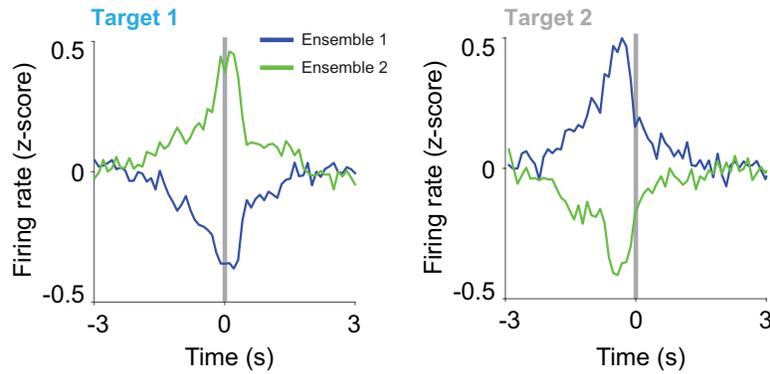


Figure 3.4: **Target neural patterns for the BMI.** Ensemble 1 and 2 firing rate modulation before Target 1 and Target 2 hits, averaged over all recorded cells and sessions. Only Target 1 resulted in VTA self-stimulation.

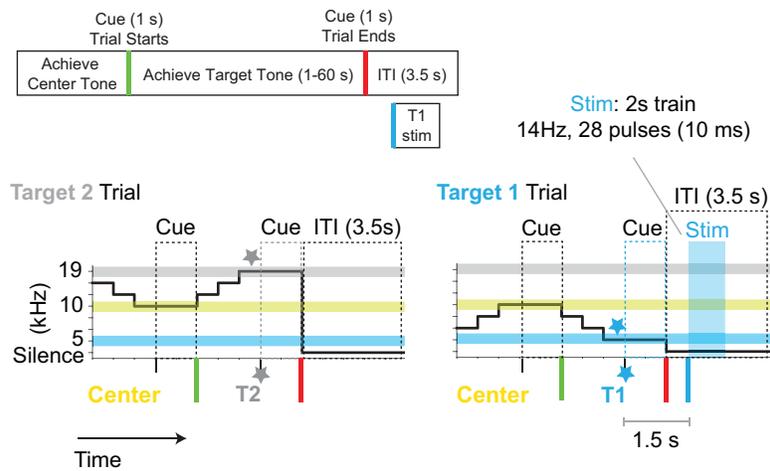


Figure 3.5: **BMI trial structure.** Subjects initiate trials by entering the center tone. If subjects do not complete a trial within 60s, a timeout white noise tone plays for 1s. When subjects hit a target, the target tone cue plays for 1s, a 0.5s pause ensues, and then subjects receive Trial structure is the same for Target 1 and Target 2, except that Target 1 hit results in phasic VTA stimulation (2s train of 14 Hz pulses with 10 ms width).

We trained animals on 4 consecutive daily sessions and quantified how reinforcement changed BMI tone distributions relative to Session 1 (Figure 3.6A,B). Experimenters were blind to the type of virus injected in VTA. ChR2 animals changed their target tone occupancy from their baseline bootstrap distribution by Session 3 and 4, while YFP animals showed no preference for Target 1 (Figure 3.7C). With training Target 1 was occupied significantly more often in ChR2 animals, and did not change in YFP animals (Figure 3.7A). ChR2 animals increased preference for Target 1 versus Target 2 (Figure 3.7B), and biased their overall distribution towards low pitch tones close to Target 1 and away from high pitch tones close to Target 2 (Figure 3.7C). Interestingly, neuroprosthetic-triggered VTA stimulation did not reinforce specific overt movements (Corbett and Wise, 1980; Olds and Milner, 1954; Witten et al., 2011) or place preference (Tsai et al., 2009), suggesting that animals are not simply undergoing motor learning (Figure 3.14).

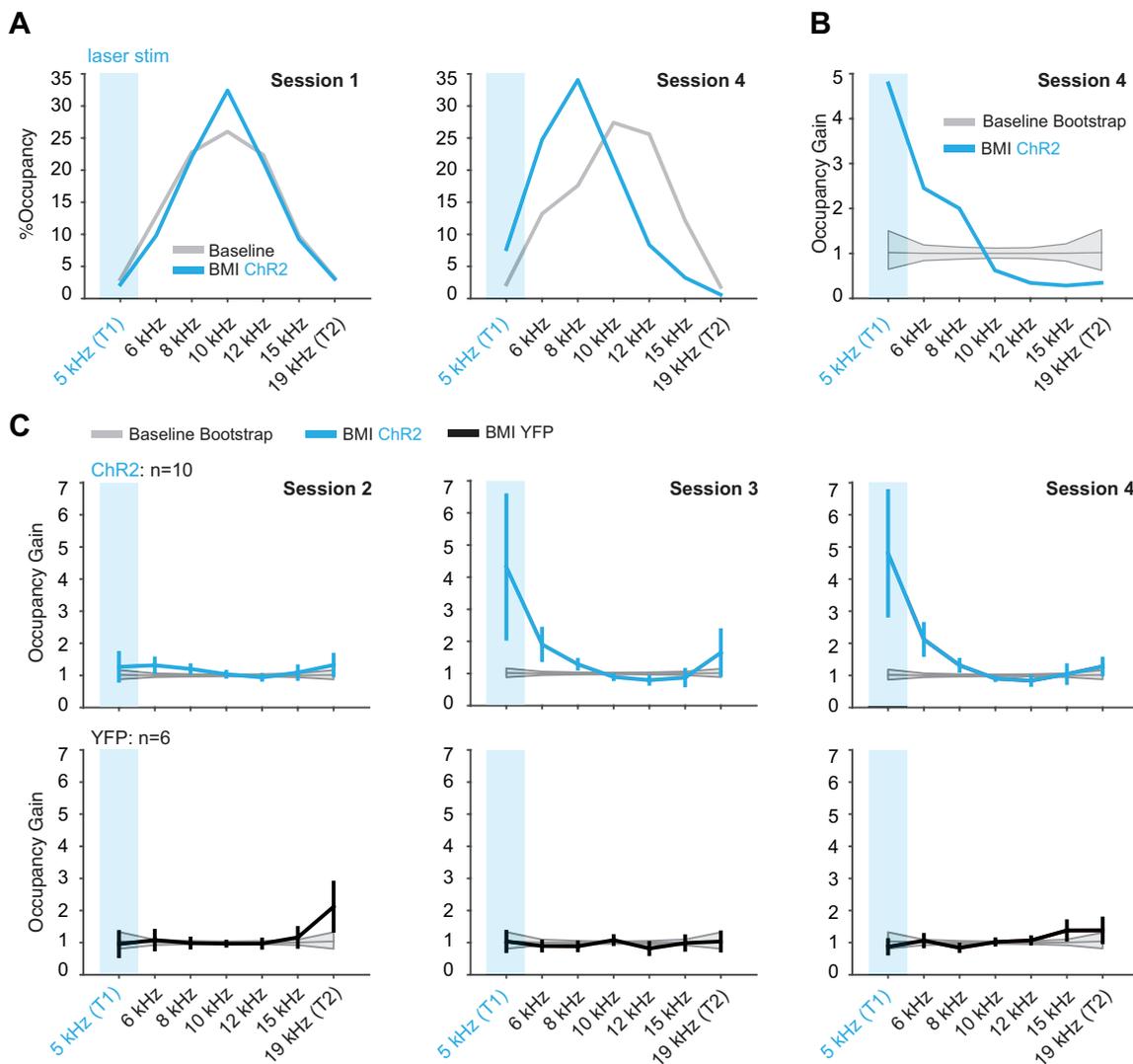


Figure 3.6: Target pattern reentrance increases during VTA optogenetic self-stimulation.

Figure 3.6: **Target pattern reentrance increases during VTA optogenetic self-stimulation.** (A) Distribution of the percent of time that each tone was occupied during baseline (gray) and BMI (cyan) blocks of Session 1 (left) and Session 4 (right) in an example mouse. (Note that no tones were actually played during the baseline block.) (B) Quantification of the behavioral changes between Session 1 and 4 for the example mouse. The Session 4 Occupancy Gain (cyan) is the Session 4 BMI distribution normalized to the Session 4 baseline distribution, then normalized to the Session 1 ratio. The 95% confidence interval of the Occupancy Gain baseline bootstrap distribution is plotted in gray. To generate the bootstrap distribution, the BMI session was simulated 10,000 times as though neural activity were drawn from that session's baseline period. (C) The Occupancy Gain is plotted over sessions 2 through 4. The 95% confidence interval of the baseline bootstrap distribution is shown in gray. The mean and s.e.m. over ChR2 animals (N=10) is shown in cyan (top) and over YFP animals (N=6) is shown in black (bottom). By Session 4, the behavioral changes were statistically different across tones for ChR2 animals but not YFP (repeated measures ANOVA, ChR2:  $F(6,48)=3.46$ ,  $p=6.4e-3$ ; YFP:  $F(6,30)=0.96$ ,  $p=0.47$ ). On Session 4, 5 kHz (Target 1) was significantly different from all tones from 8 kHz to 19 kHz (Tukey's post hoc multiple comparisons test).

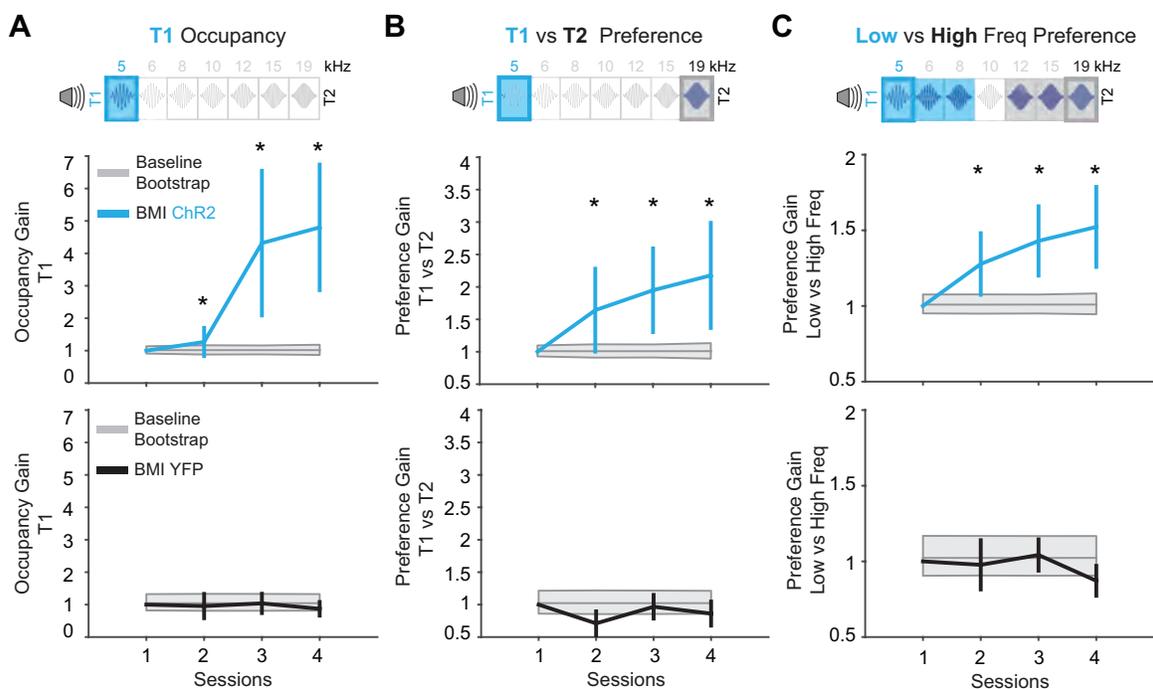


Figure 3.7: **BMI trial structure.** (A) Top: the Occupancy Gain for 5 kHz (Target 1) is plotted over sessions. The 95% confidence interval of the bootstrap distribution is plotted in gray. Middle: The average of ChR2 animals (cyan) shows a significantly larger effect than expected from bootstrap from Session 2-4 (Session 2:  $p=1.2e-3$ , Session 3:  $p<1e-5$ , Session 4:  $p<1e-5$ ). Bottom: YFP animals (black) never showed an effect significantly larger than expected from bootstrap. (B) Top: the Preference Gain for 5 kHz (Target 1) versus 19 kHz (Target 2) is plotted over sessions. This is calculated as the increase factor of the probability of hitting Target 1 given that a target was hit (see methods). The 95% confidence interval of the bootstrap distribution is plotted in gray. Middle: ChR2 animals (cyan) were significantly larger than expected from bootstrap after Session 1 ( $p<1e-5$  for Sessions 2-4). Bottom: YFP animals (black) were never significantly larger than bootstrap. (C) Top: the Preference Gain for low tones (5-8 kHz, close to Target 1) versus high tones (12-19 kHz, close to Target 2) is plotted over sessions. This is calculated as the increase factor of the probability of hitting a low tone given that the center tone was not produced. Middle: ChR2 animals (cyan) were significantly larger than expected from bootstrap after Session 1 ( $p<1e-5$  for Sessions 2-4). Bottom: YFP animals (black) were never significantly larger than bootstrap. For A-C, a star indicates that the population average is significantly larger than expected from the baseline bootstrap distribution.

Given that the differential modulation between Ensemble 1 and 2 shifted towards Target 1, we asked more generally how the joint activity of neurons involved in producing the pattern (direct neurons) was shaped by reinforcement. Because the ensembles’ simultaneous modulation triggered reinforcement, VTA stimulation might strengthen shared inputs to direct neurons and thus increase covariance over learning (Athalye et al., 2017). We used Factor Analysis (FA) to partition fine-timescale neural variance arising from two sources: private inputs to each cell which drive independent firing (private variance) and shared inputs which drive multiple cells simultaneously (shared variance). Neural variance changes were not demanded by our task, as subjects could use neural activity drawn from any distribution to ultimately hit Target 1 (Figure 3.8A).

We analyzed fine-timescale spike counts (100 ms bins) in a 3 second window preceding target hit (Figure 3.8B). FA models population spike counts  $x = \mu + x^{\text{private}} + x^{\text{shared}}$  as the sum of a mean firing rate  $\mu$ , private variation  $x^{\text{private}}$  which is uncorrelated across neurons, and shared variation  $x^{\text{shared}} = Uz$  which is driven by latent shared inputs  $z$  through the linear factors  $U$ . Because  $x^{\text{private}}$  and  $x^{\text{shared}}$  are independent, the total covariance matrix  $\Sigma^{\text{total}} = \Sigma^{\text{shared}} + \Sigma^{\text{private}}$  is decomposed into the sum of 1) a diagonal private covariance matrix  $\Sigma^{\text{private}}$  and 2) a low-rank shared covariance matrix  $\Sigma^{\text{shared}}$ . Geometrically, private variance spans all of high-dimensional population activity space for which each neuron’s activity is one dimension, while shared variance is constrained to a low-dimensional “shared space” because there are fewer shared inputs than neurons. The number of shared dimensions was fit using standard model selection (Figure 3.15) by maximizing cross-validated log-likelihood (Athalye et al., 2017; Dempster et al., 1977; Everitt, 1984; Sadtler et al., 2014; Yu et al., 2009).

We assessed neural coordination with a covariance index defined as the ratio of the shared variance to total variance averaged over neurons (SOT) (Figure 3.8C). While (Figure 3.8A-C) use 2 neurons for illustration, we emphasize that FA was applied to the joint activity of all neurons used to control the BMI (ranging from 4-8).

We then asked if learning, defined as the proportion of hits of Target 1 versus Target 2 normalized to Session 1, was correlated with the increase in covariance, defined as the SOT normalized to Session 1. The increase in covariance correlated with learning in Chr2 but not YFP animals (Figure 3.9A). This correlation became stronger as learning progressed.

These data suggest that the degree of learning related to the degree of neural variance changes. To further dissect this, we analyzed Chr2 animals and found 2 groups distinguished by their degree of learning (Figure 3.16). Each individual of the Learner group (N=5) showed statistically-significant preference for T1 versus T2 for both Session 3 and 4. The Poor Learner group (N=5) as a population showed an increase in Target 1 occupancy, but did not improve preference for Target 1 over Target 2 (Figure 3.16). Learners significantly increased their covariance index over training, while Poor Learners and YFP did not (Figure 3.9B and 3.17, 3.18A). This effect was ensemble-specific, as only neurons controlling the BMI (direct neurons) increased their covariance index, while other recorded neurons (indirect neurons) did not (Figure 3.9B and 3.19, 3.20).

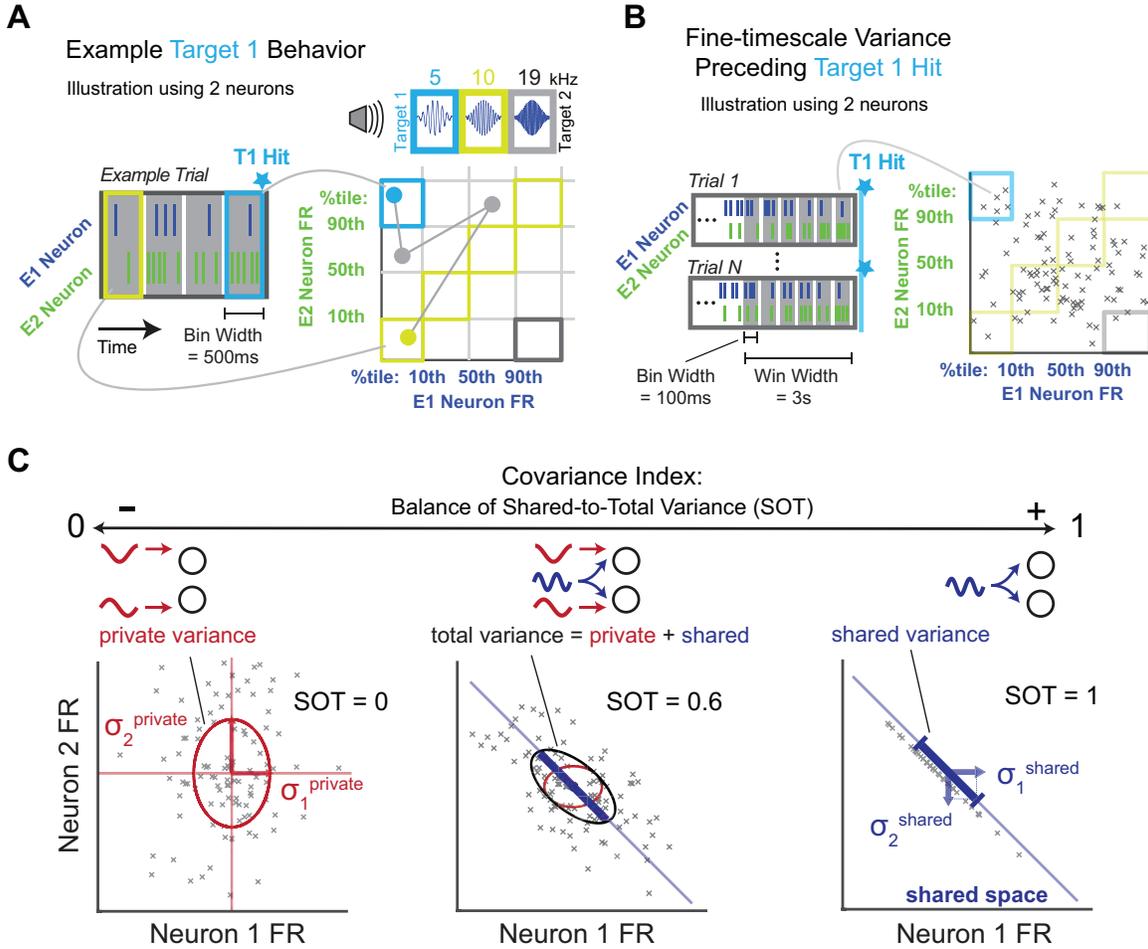


Figure 3.8: **Analysis of fine-timescale neural variance preceding target hit (A)** The decoder maps spike counts in 500 ms bins into quantizations of (Ensemble 1, Ensemble 2) space. Neural activity can take multiple routes to achieve Target 1. **(B)** We analyzed variance of spike counts at fine-timescale (100 ms bins) in a 3s window preceding target hit. **(C)** Factor Analysis was used to analyze the ratio of shared variance to total variance (SOT), which ranges from 0 to 1, for the full population controlling the BMI. A 2-neuron illustration shows a neural solution with SOT=0, 0.6, and 1.

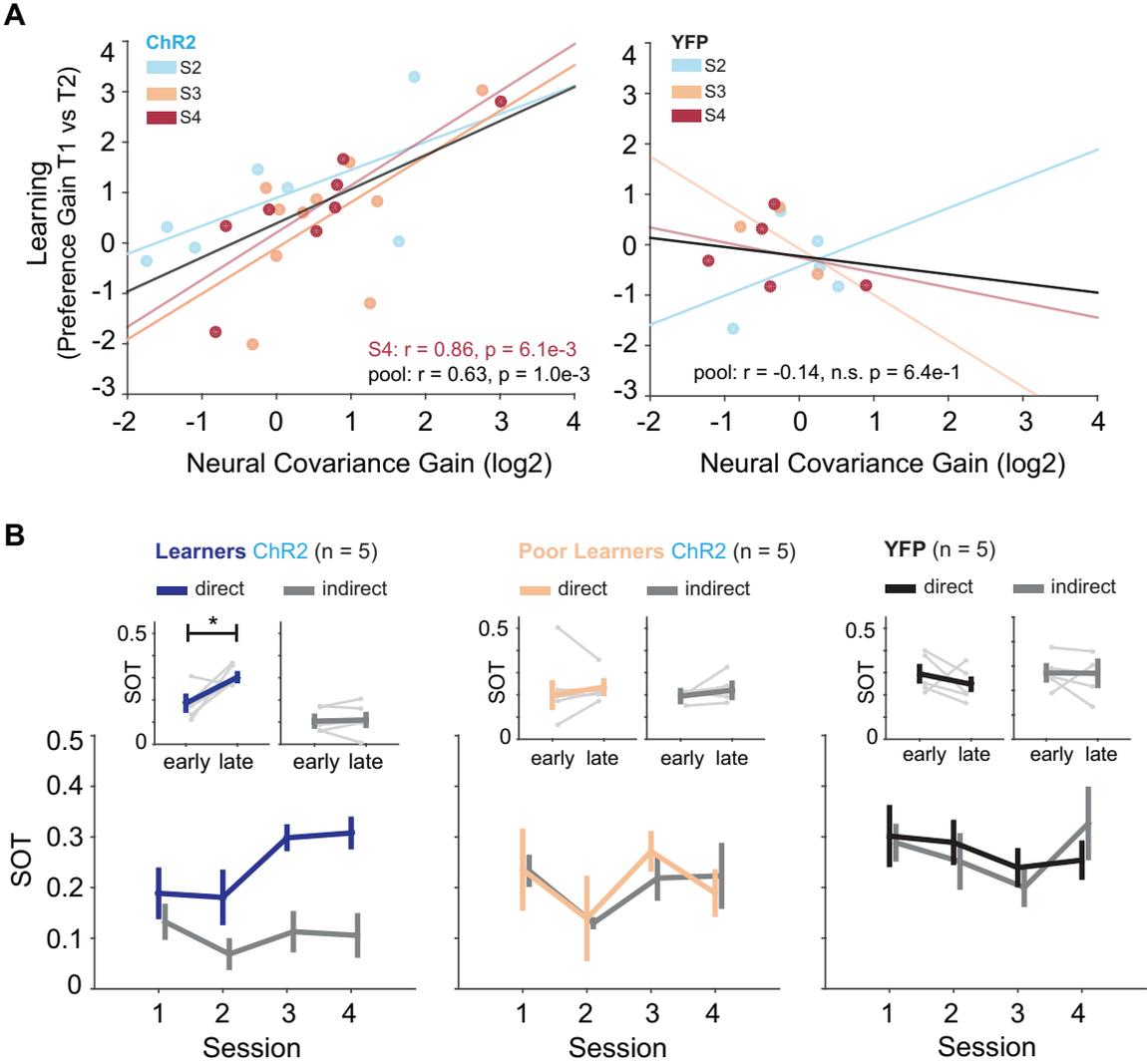


Figure 3.9: Learning correlates with increase in covariance of the neurons that produce the target pattern.

Figure 3.9: **Learning correlates with increase in covariance of the neurons that produce the target pattern.** (A) Correlation of change in shared variance before Target 1 hit (Neural Covariance Gain) with change in preference for Target 1 over Target 2 (Learning), over Session 2, 3 and 4. ChR2 animals (left) showed a significant correlation (ChR2 S4:  $r=0.86$ ,  $p=6.1e-3$ ; ChR2 pool S3, S4:  $r=0.71$ ,  $p=1.0e-3$ ; ChR2 pool S2, S3, S4:  $r=0.62$ ,  $p=9.8e-4$ ; ChR2 S3:  $r=0.60$ ,  $p=6.5e-2$ ; ChR2 S2:  $r=0.62$ ,  $p=1.3e-1$ ) while YFP animals (right) showed no correlation (YFP pool S2, S3, S4:  $r=-0.14$ , n.s.  $p=6.4e-1$ ; YFP S4:  $r=-0.32$ ,  $p=6.0e-1$ ; YFP S3:  $r=-0.69$ ,  $p=5.1e-1$ ; YFP S2:  $r=0.37$ ,  $p=5.4e-1$ ). (B) SOT of direct and indirect cells was plotted over sessions for ChR2 Learners (left,  $N=5$ ), ChR2 Poor Learners (middle,  $N=5$ ), and YFP subjects (right,  $N=5$ ). ChR2 Learners are animals in the ChR2 group who individually showed significant Preference Gain for T1 versus T2 on both Session 3 and 4. ChR2 Poor Learners constitute the remaining animals who as a population showed significant T1 Occupancy Gain on Session 3 and 4 but did not show significant Preference Gain for T1 versus T2. For direct cells, ChR2 Learners SOT increased from early (Session 1 and 2 pooled) to late training (Session 3 and 4 pooled), while ChR2 Poor Learners and YFP did not (one-sided rank sum test; ChR2 Learners: early < late  $p=1.6e-2$ , ChR2 Poor Learners: early < late n.s.  $p=2.1e-1$ , YFP: early < late n.s.  $p=8.3e-1$ ). Raw SOT pooled over all ChR2 animals also showed significant increase from early to late training (ChR2: early < late  $p=1.7e-2$ ). For indirect cells, SOT showed no change for all groups (ChR2 Learners: early < late n.s.  $p=4.3e-1$ , ChR2 Poor learners: early < late n.s.  $p=2.7e-1$ , YFP: early < late n.s.  $p=7.1e-1$ ) Traces in the insets shows the average of each animals SOT in Session 1 and 2 (early) versus the average of Session 3 and 4 (late).

Finally, we asked whether dopaminergic self-stimulation shaped the neural covariance to more easily achieve the target pattern. Only neural variance which causes differential modulation between Ensemble 1 and 2 can change the feedback tone and contribute to target achievement, corresponding to variance which is aligned with the decoders Ensemble 1 minus Ensemble 2 axis (Figure 4.15A). We analyzed the relationship between shared neural variance and the decoder by calculating the angle between the shared space and the decoder axis. The angle between the shared space and the decoder axis decreased significantly for Learners but not for Poor Learners and YFP (Figure 4.15B and 3.18B).

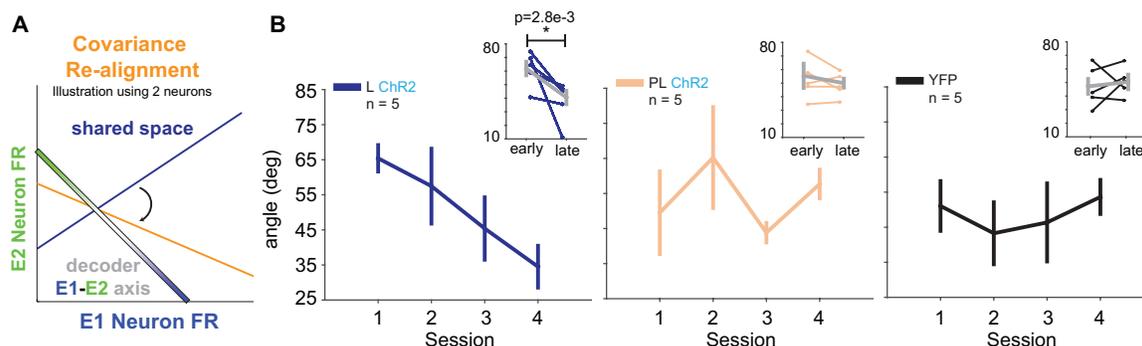


Figure 3.10: **Covariance of the neurons that produce the target pattern gradually aligns to the decoder.** (A) Analysis of shared variance alignment with the decoder’s Ensemble 1 and Ensemble 2 assignments via the angle between the shared space and the decoder’s “Ensemble 1 minus Ensemble 2” axis. (B) The angle between shared variance and the decoder axis decreased for ChR2 Learners (left), but not Poor Learners (middle) and YFP (right): (one-sided rank sum test comparing Session 1 and 2 to Session 3 and 4; ChR2 Learner: late > early  $p=2.8e-3$ ; ChR2 Poor Learner: late > early n.s.  $p=3.7e-1$ ; YFP: late > early n.s.  $p=7.5e-1$ ). Traces in the insets show the average of each animal’s angle in Session 1 and 2 (early) versus the average of Session 3 and 4 (late).

The results presented here show that mice reenter specific neural patterns that trigger dopaminergic VTA self-stimulation more often as training progresses. Dopaminergic self-stimulation not only increases the reentry of a target pattern, but further shapes the distribution of activity patterns to more directly achieve the target pattern. The covariance increased specifically between direct neurons and gradually became aligned with the decoder. Individual neuron firing properties did not correlate with learning (Figure 3.21), highlighting that it was the specific pattern that was reinforced. This reinforcement of specific neural ensembles and activity patterns extends recent work showing individual neuron conditioning via electrical self-stimulation of the nucleus accumbens (Eaton et al., 2016). Overt motor behavior was not reinforced over BMI learning though it is challenging to measure all behavioral output (Figure 3.14), highlighting that reentrance of neural patterns need not drive motor output. Nevertheless, these results may have implications for motor reinforcement,

where actions are selected more often and optimized over iterations to more directly achieve reinforcements.

### 3.4 Discussion

In the experiments presented here, subjects learned to produce neural patterns *de novo*, which leverages different mechanisms from BMI learning experiments in which subjects adapted to decoder perturbations. BMI-experienced subjects learn to control a decoder by selecting activity patterns from their existing shared space (Sadtlter et al., 2014). In contrast, our Learners initially exhibit little shared variance, and this shared variance is misaligned with the decoder. Thus, they likely select patterns from their high-dimensional private variance, gradually developing and re-aligning shared variance with learning (Athalye et al., 2017). Analysis and modeling indicate that private variance is useful for broad exploration of population activity space (Athalye et al., 2017) and for learning each neuron’s contributions to achieving a goal (Héliot et al., 2010; Legenstein et al., 2010), possibly permitting the selective increase of direct neurons’ covariation index over indirect neurons.

It is unlikely that VTA stimulation directly modulated activity and plasticity in M1 through monosynaptic projections because we stimulated the VTA contralateral to our M1 recordings. Indeed, VTA stimulation induced no observed changes in the mean firing rates of recorded neurons (Figure 3.22). Thus, M1 reinforcement is likely driven by inputs from and plasticity in broader networks, such as cortico-basal ganglia circuits. Cortico-striatal plasticity is modulated by dopamine (Shen et al., 2008; Yagishita et al., 2014) and is necessary for motor and neuroprosthetic learning (Costa et al., 2004; Koralek et al., 2012; Santos et al., 2015). Actor-critic reinforcement learning models (O’Doherty et al., 2004; Takahashi et al., 2009) suggest two sites for VTA-modulated plasticity: the dorsal striatum (actor) which contributes to selection of actions (M1 neural activity patterns), and the ventral striatum (critic) which may evaluate actions based on the value of the environmental states reached (auditory feedback). Plasticity in dorsal striatum could be mediated by glutamatergic input from contralateral M1 and dopaminergic input signaling reward from VTA (Howe and Dombeck, 2016), enabling adaptation of the policy for re-entering M1 activity patterns. Plasticity in ventral striatum (Yagishita et al., 2014) could be mediated by strong bidirectional connections with VTA, enabling adaptation of the auditory tones’ value.

In addition, VTA stimulation may have indirectly guided motor cortical plasticity. As animals acquire motor skills and consolidate cortical activity patterns, motor memories are encoded in the formation of lasting dendritic spine ensembles (Fu et al., 2012; Hayashi-Takagi et al., 2015; Peters et al., 2014; Xu et al., 2009). Further, reinforcement guides the reactivation of neurons during sleep (Gulati et al., 2017), leading to the formation of dendritic spines (Yang et al., 2014) as well as the identification of neurons responsible for achieving a target pattern (Gulati et al., 2017). Thus, our observed changes in shared variance could also reflect sleep-dependent changes in motor cortical synaptic connectivity. Recent modeling work shows that excitation-inhibition balanced spiking networks with clustered connectivity

exhibited prominent low-dimensional shared variance while non-clustered networks exhibited weak, high-dimensional shared variance (Williamson et al., 2016).

Our results provide causal evidence for a neural law of effect, describing how the brain selects and shapes neural activity patterns through neural reinforcement. As Skinner noted, selection by consequence is a mechanism driving the evolution of living things, from species to societies to behavior (Skinner, 1981). Our results help uncover how selection by consequence operates on neural activity within the brain (Costa, 2011).

### 3.5 Appendix A

#### Supplemental Results

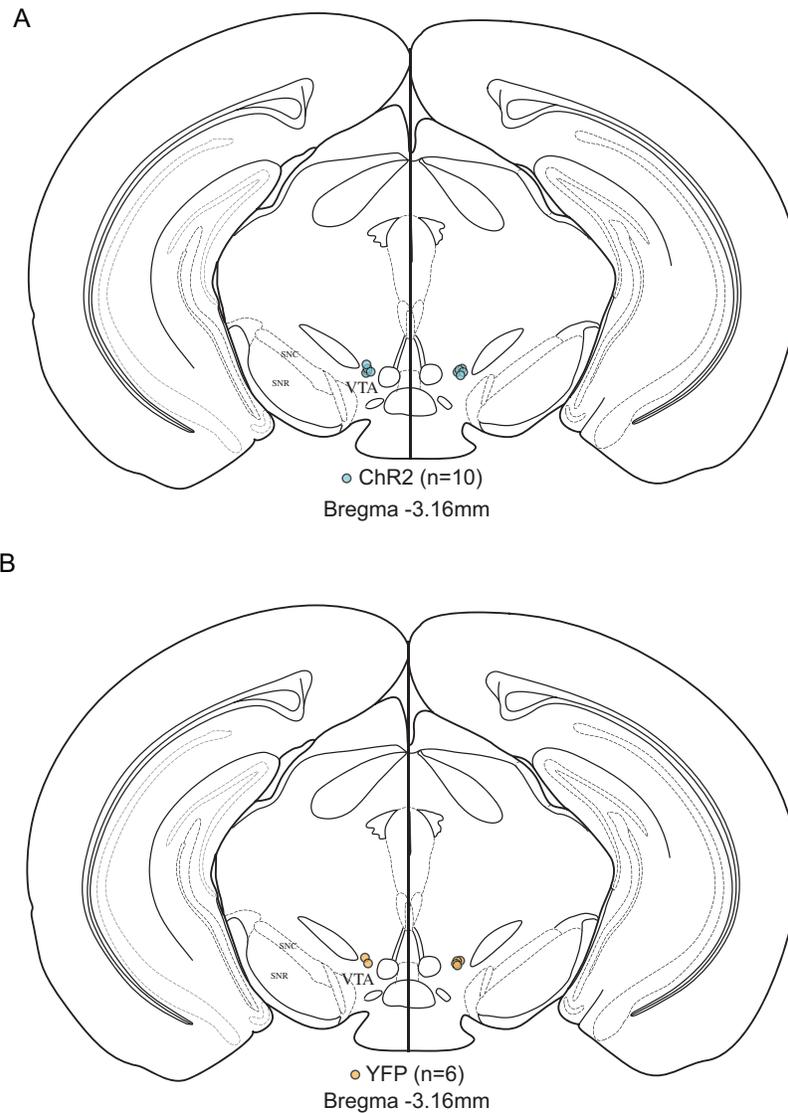


Figure 3.11: **Optical fiber placement.** Circles represent the tops of optical fibers. The implantation hemisphere was randomly assigned to each animal (left hemisphere: ChR2  $n=5$ , YFP  $n=2$ ; right hemisphere ChR2  $n=5$ , YFP  $n=4$ ). VTA, ventral tegmental area. SNc, substantia nigra pars compacta. SNr, substantia nigra pars reticulata.

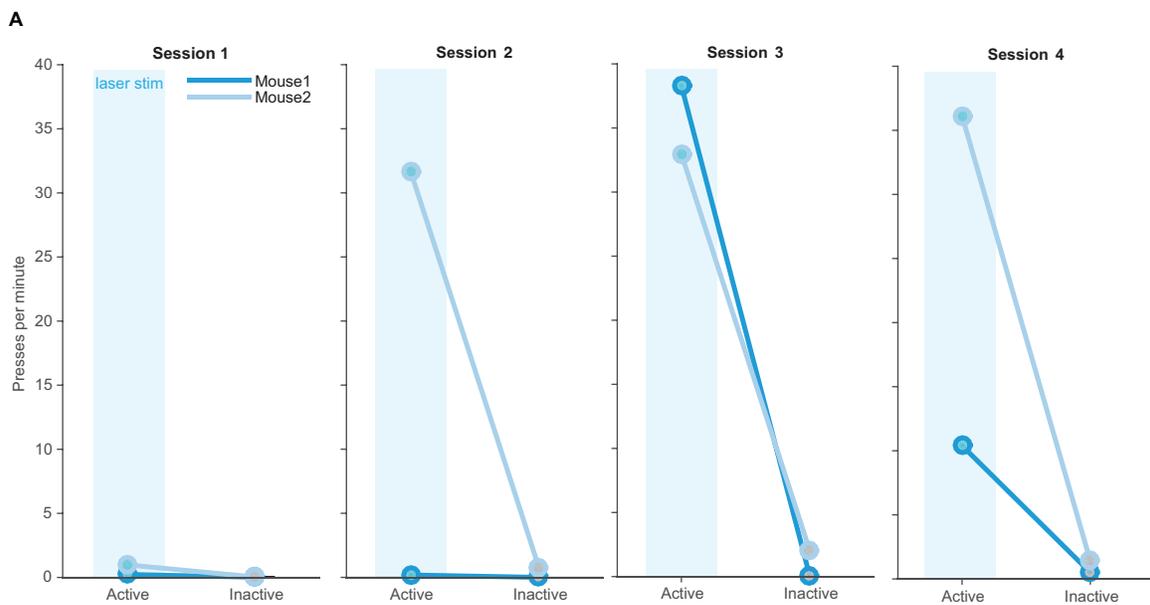


Figure 3.12: **Verification of optogenetic VTA stimulation in a motor reinforcement task.** (A) ChR2 mice (N=2) were trained on an operant lever pressing task in which active lever presses triggered optogenetic VTA stimulation while inactive lever presses resulted in no reinforcement. Performance increased over sessions.

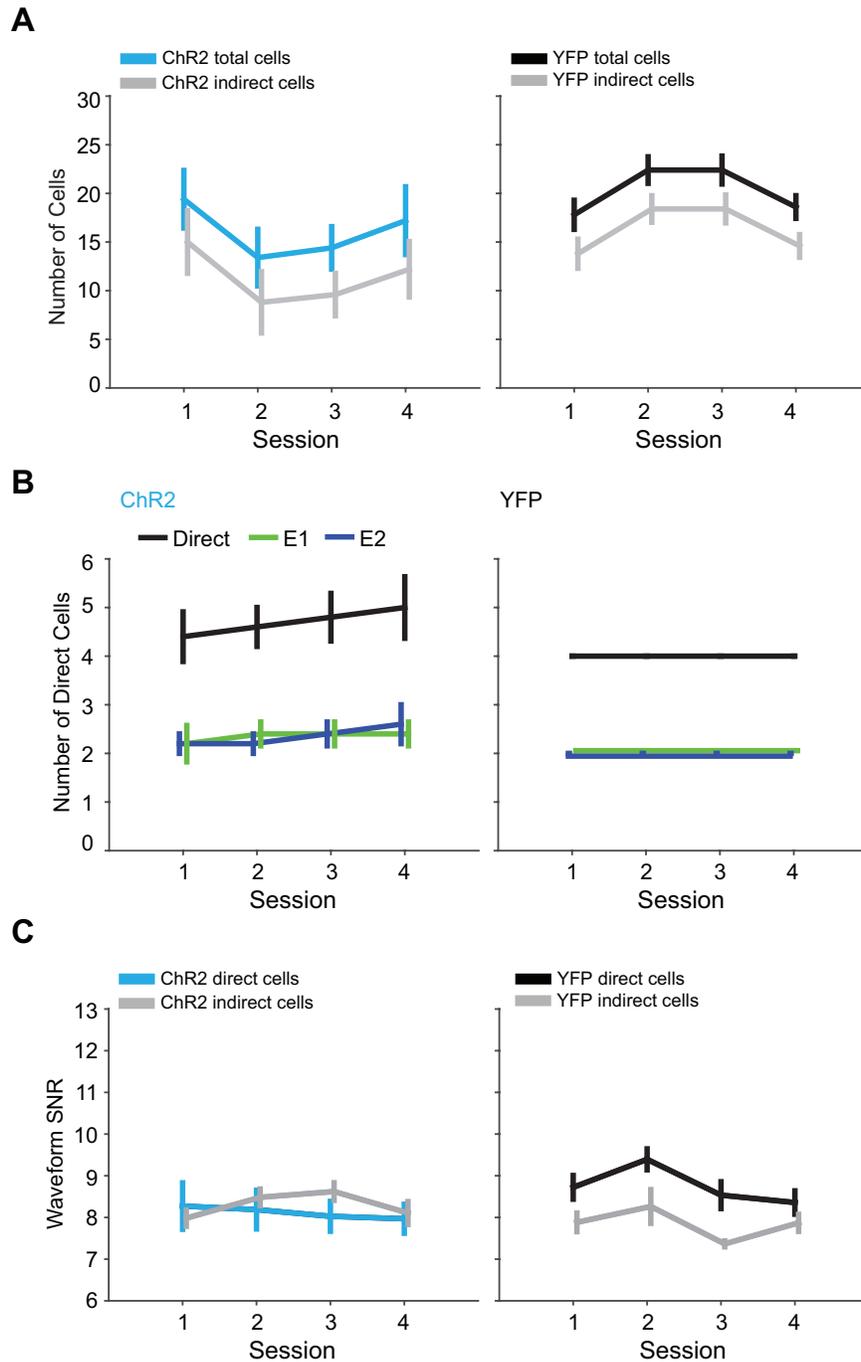


Figure 3.13: **Number of recorded neurons and SNR waveforms.** (A) The number of recorded neurons for ChR2 and YFP groups. (B) The number of direct neurons assigned to control the brain-machine interface. In each experiment, there are 2-4 E1 and E2 cells. (C) The SNR of each recorded neurons waveform, calculated as in (Suner et al., 2005), which demonstrates well-isolated single units have SNR above 4.

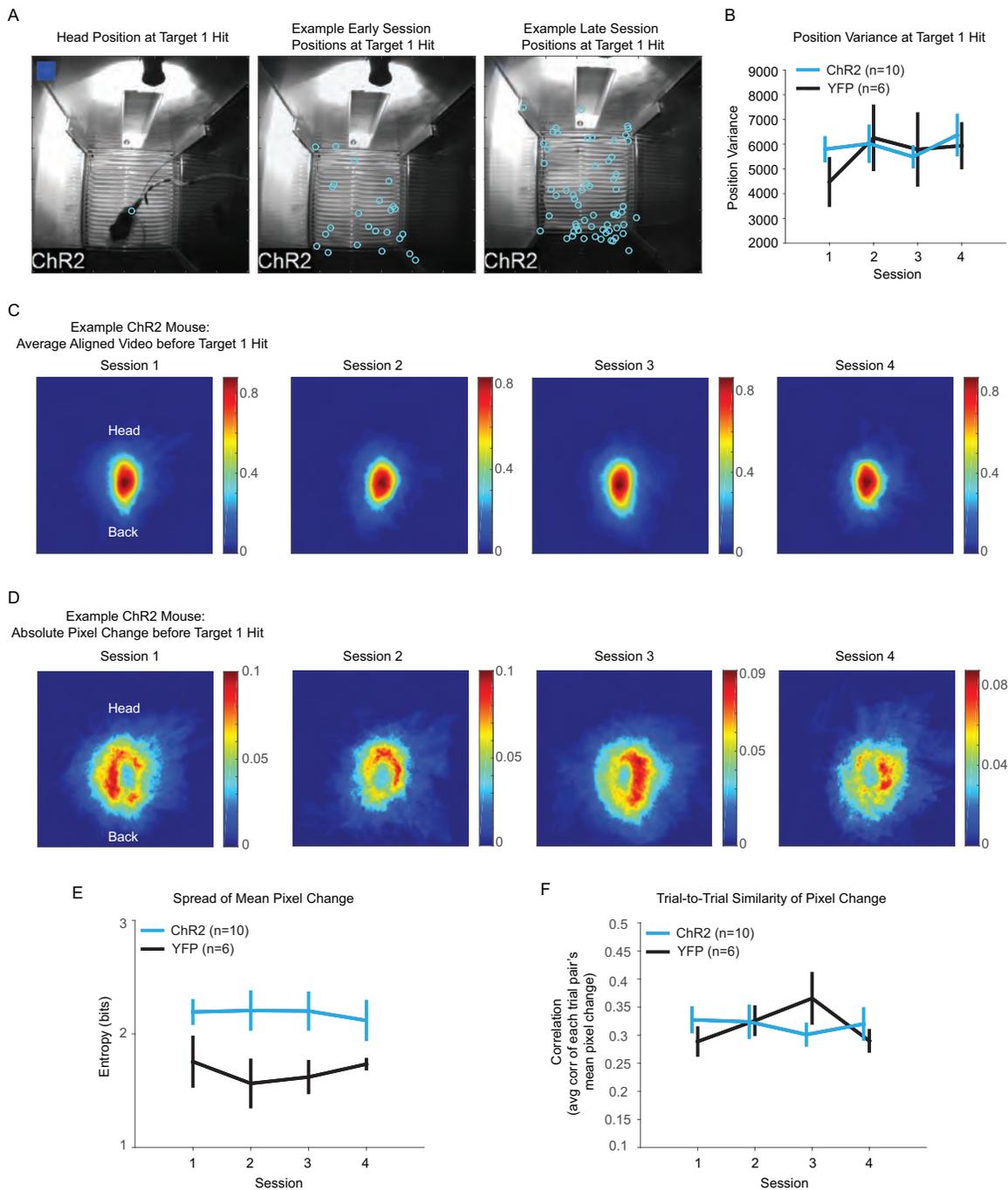


Figure 3.14: Neuroprosthetic-triggered VTA stimulation does not reinforce place preference or specific overt movements.

Figure 3.14: **Neuroprosthetic-triggered VTA stimulation does not reinforce place preference or specific overt movements.** **(A)** An example mouse's head position (left) is shown at time of Target 1 hit for an early session (session 1, middle) and a late session (session 3, right). **(B)** The variance of position at Target 1 hit does not change for ChR2 and YFP animals (rank sum test, session 1 and 2 vs session 3 and 4). There is no difference between ChR2 and YFP animals (rank sum test, ChR2 vs YFP animals pooling sessions). **(C)** 1.5 second videos preceding Target 1 hit were aligned to the mouse's position and head-back orientation at Target 1 hit (see methods) and averaged over frames and trials. An example ChR2 mouse is shown over 4 sessions. **(D)** The mean absolute pixel change in the 1.5 second videos preceding Target 1 hit is shown for an example ChR2 mouse over 4 sessions. **(E)** The spread of the mean pixel change was calculated via the entropy of the mean pixel change image. It does not change from early to late training (rank sum test, session 1 and 2 vs session 3 and 4). The spread of ChR2 mean pixel change is larger than YFP (rank sum test, ChR2 > YFP,  $p=2.6e-13$ , pooling over sessions), indicating ChR2 animals do not execute a specific movement before Target 1 hit. **(F)** The trial-to-trial similarity of the absolute pixel change (see methods) does not change for ChR2 and YFP animals (rank sum test, session 1 and 2 vs session 3 and 4). There is no difference between ChR2 and YFP animals (rank sum test, ChR2 vs YFP animals pooling sessions), indicating ChR2 animals' movements do not become more similar from trial-to-trial.

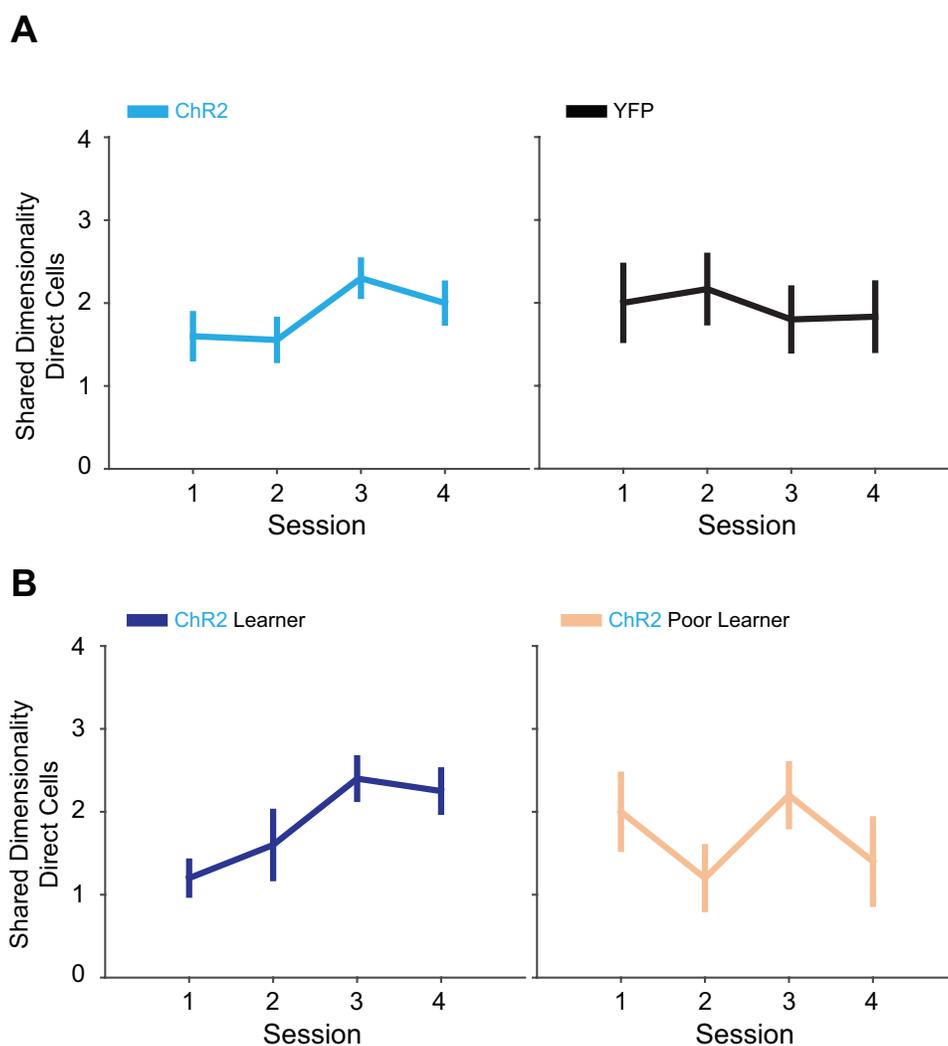


Figure 3.15: **Number of shared dimensions for FA.** The number of shared dimensions (i.e. number of factors, equivalent to the rank of the shared covariance matrix) for FA was fit using model selection by evaluating cross-validated log-likelihood. **(A)** The number of shared dimensions for ChR2 and YFP groups. **(B)** The number of shared dimensions for the ChR2 Learner and ChR2 Poor Learner groups.

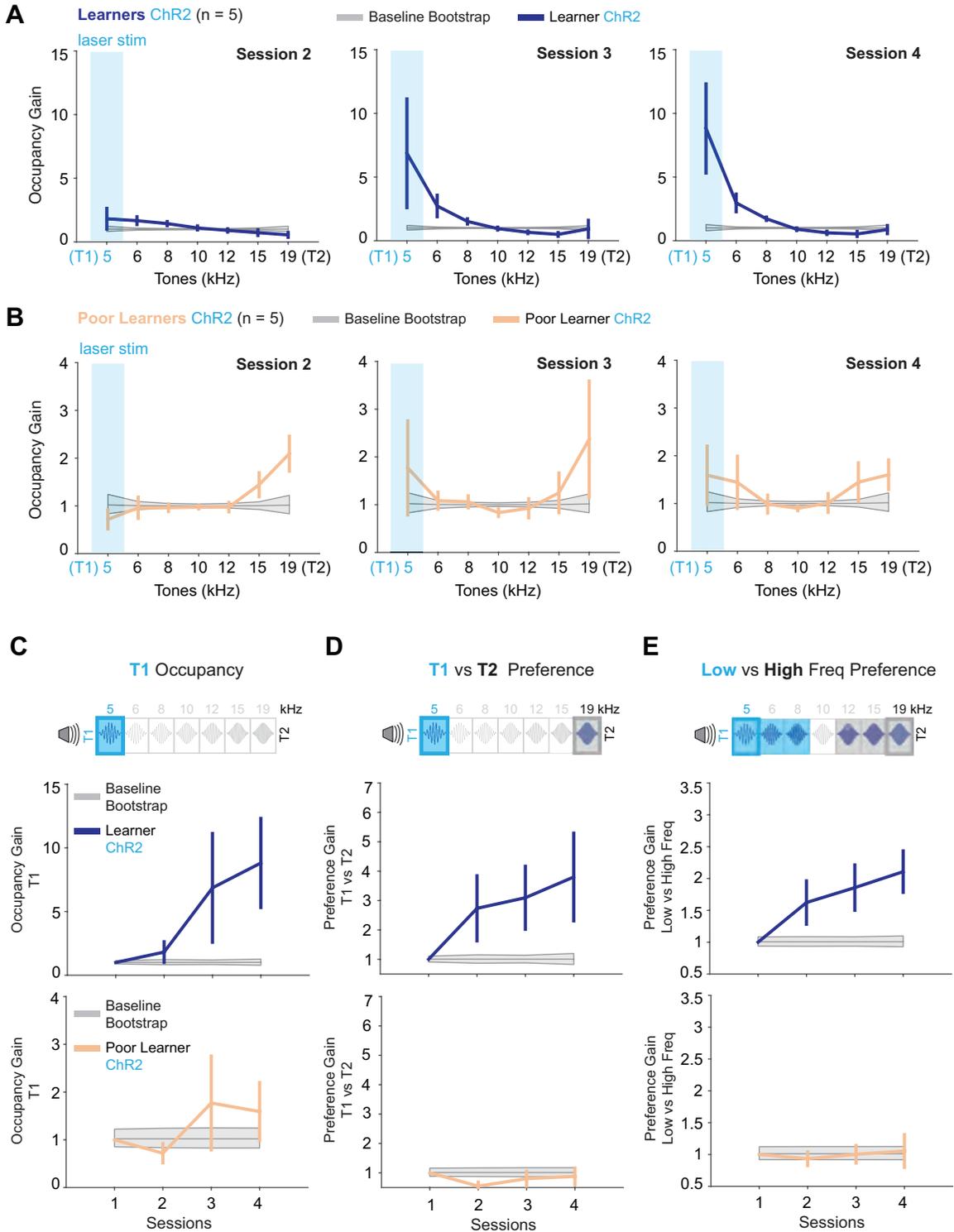


Figure 3.16: Behavior of ChR2 Learners and Poor Learners.

Figure 3.16: **Behavior of ChR2 Learners and Poor Learners.** The ChR2 group (n=10) was split into two groups based on their degree of learning: Learners (n=5) and Poor Learners (n=5). Each individual of the Learners group showed significant Preference Gain for T1 versus T2 for both Session 3 and 4. Poor Learners as a group showed significant T1 Occupancy Gain for both Session 3 and 4, but they also showed significant T2 Occupancy Gain, and thus did not show significant Preference Gain for T1 versus T2. Learners and Non-Learners' performance pooling over Session 3 and 4 was significantly different (non-parametric rank sum test: T1 Occupancy Gain  $p=1.7e-2$ ; T1 Preference Gain  $p=2.6e-4$ ; Low vs High Tones Preference Gain  $p=4.1e-3$ ). (A) Occupancy Gain for Learners over sessions 2 through 4. The 95% confidence interval of the baseline bootstrap distribution is shown in gray. (B) Occupancy Gain for Poor Learners over sessions 2 through 4. (C) The Occupancy Gain for Target 1 (5 kHz). (D) Preference Gain for Target 1 (5kHz) versus Target 2 (19kHz). (E) Preference Gain for low tones (5-8kHz, close to Target 1) versus high tones (12-19 kHz, close to Target 2).

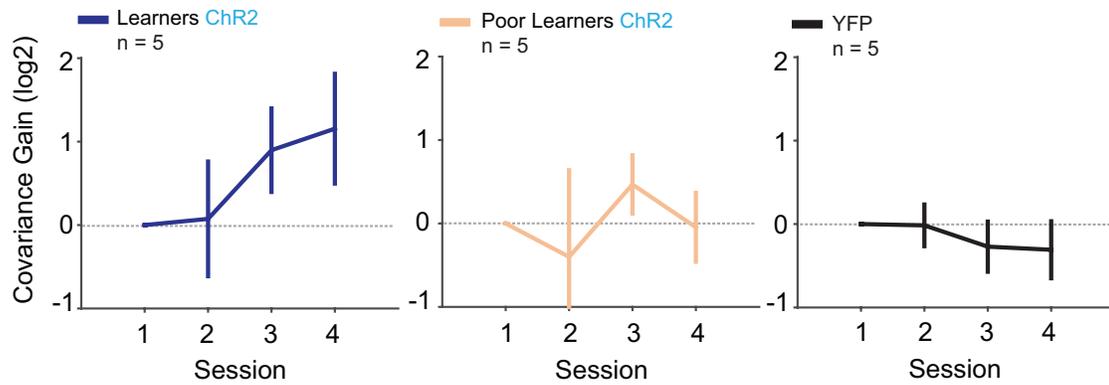
**A**

Figure 3.17: **Covariance gain over learning.** (A) The covariance gain increased for ChR2 Learners, but not for ChR2 Poor Learners and YFP. This covariance gain correlated with learning (Figure 3D).

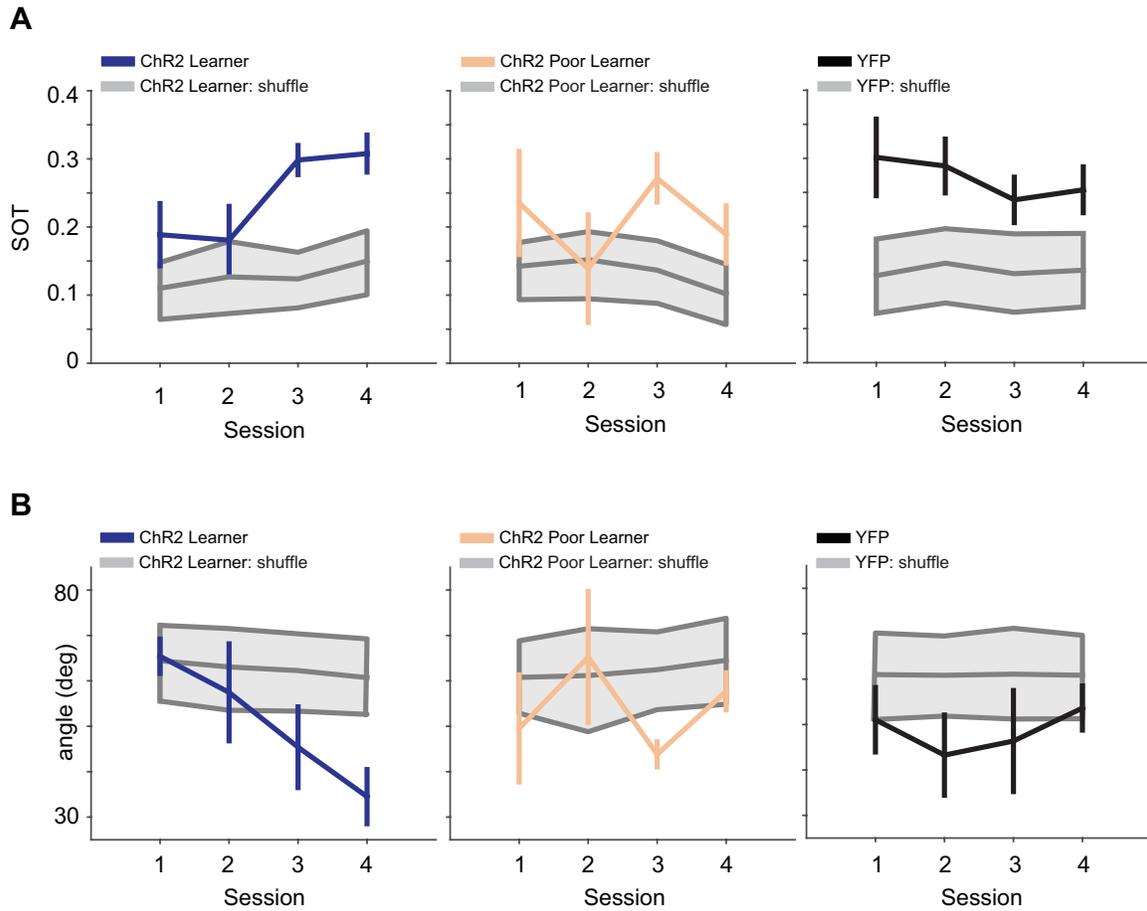


Figure 3.18: **Significance of FA results comparing experimental data to shuffled distributions.** A shuffled neural activity data set was generated by shuffling the time labels of each neuron’s binned spike count. This maintains each neuron’s marginal spike count distribution, but breaks the population’s joint distribution. FA was performed on 1000 shuffles of each data set to generate the FA shuffle distributions, and 95% confidence intervals are plotted in gray. **(A)** In ChR2 animals, SOT starts roughly at chance and increases. YFP animals are above chance over training. **(B)** In ChR2 animals and ChR2 Learners, the angle between the shared space and the decoder axis decreases below chance. In ChR2 Non-Learners and YFP animals, the angle remains roughly at chance. Colored errorbars are s.e.m. over animals. The gray shuffle distribution is the 95% confidence interval.

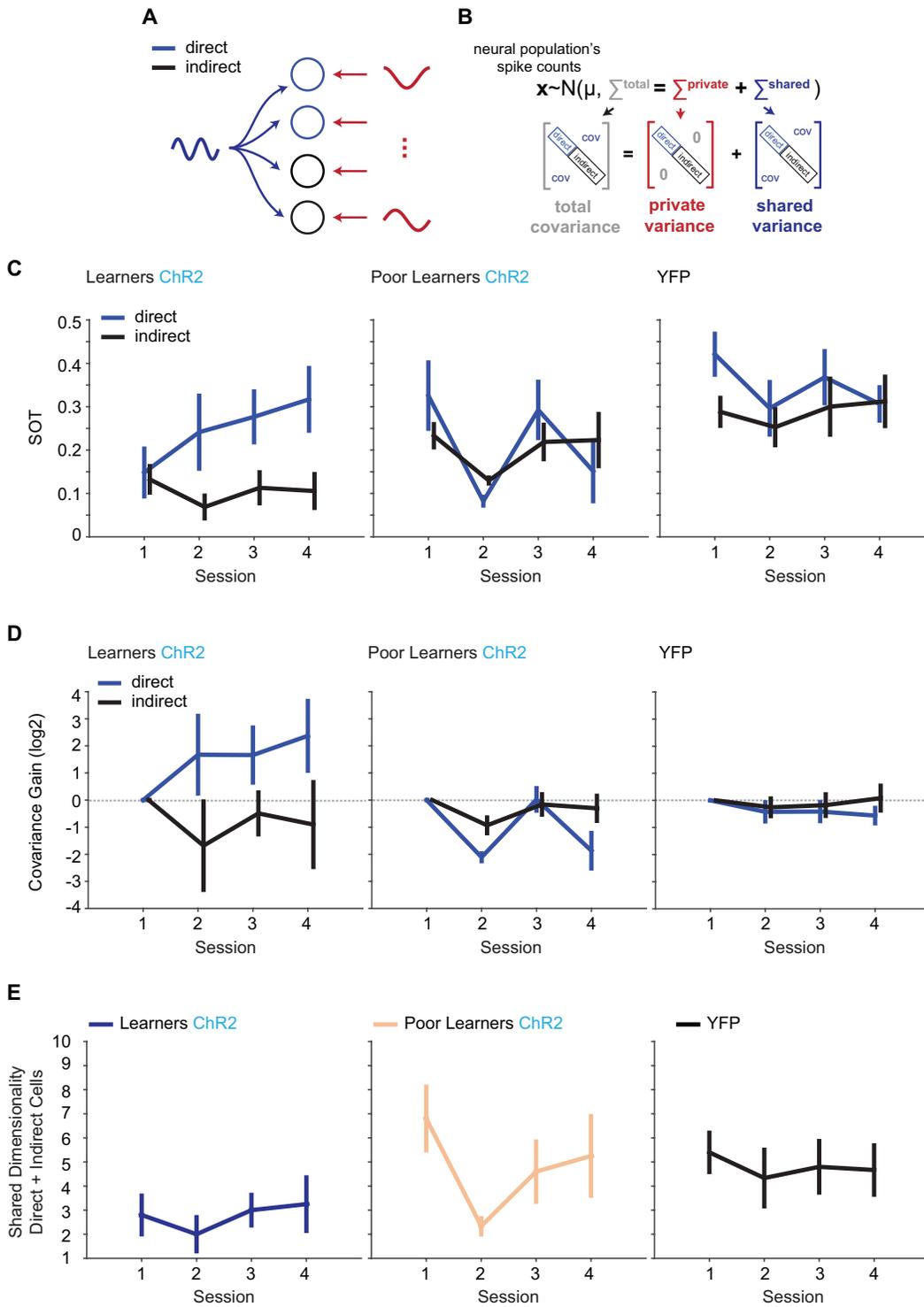


Figure 3.19: FA fit on the full population of recorded motor cortex neurons, including direct and indirect neurons.

Figure 3.19: **FA fit on the full population of recorded motor cortex neurons, including direct and indirect neurons.** (A) For these analyses, FA was fit on the full population of recorded motor cortex neurons, including direct and indirect neurons. (B) After decomposing the spike count covariance matrix into private and shared variance, we investigated how much shared and private variance the direct neurons have and how much the indirect neurons have. The direct neurons' shared variance includes variance shared with indirect neurons. (C) Trends are consistent with (Figure 3.9B), in which the direct cells' shared variance is plotted from fitting FA only on direct neurons. (D) Covariance gain trends using FA fit on the fully-recorded neural population. (E) Shared dimensionality yielded by FA model fitting for Learners, Poor Learners, and YFP.

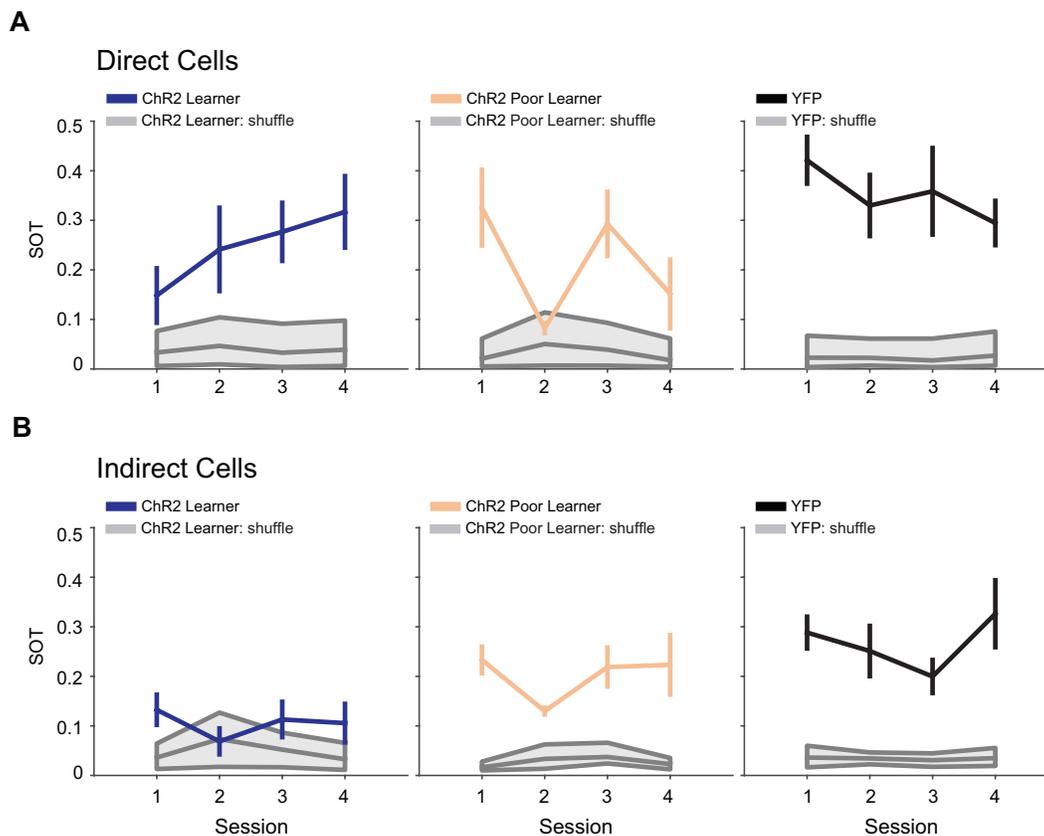


Figure 3.20: **Significance of FA fit on all neurons (direct plus indirect) compared to shuffled distributions** For these analyses, FA was fit on the full population of recorded motor cortex neurons, including direct and indirect neurons. **(A)** SOT for direct cells. The SOT is above chance for all data, other than ChR2 Non-Learners on Session 2. Only ChR2 Learners show an SOT with increasing trend over training. **(B)** SOT for indirect cells.

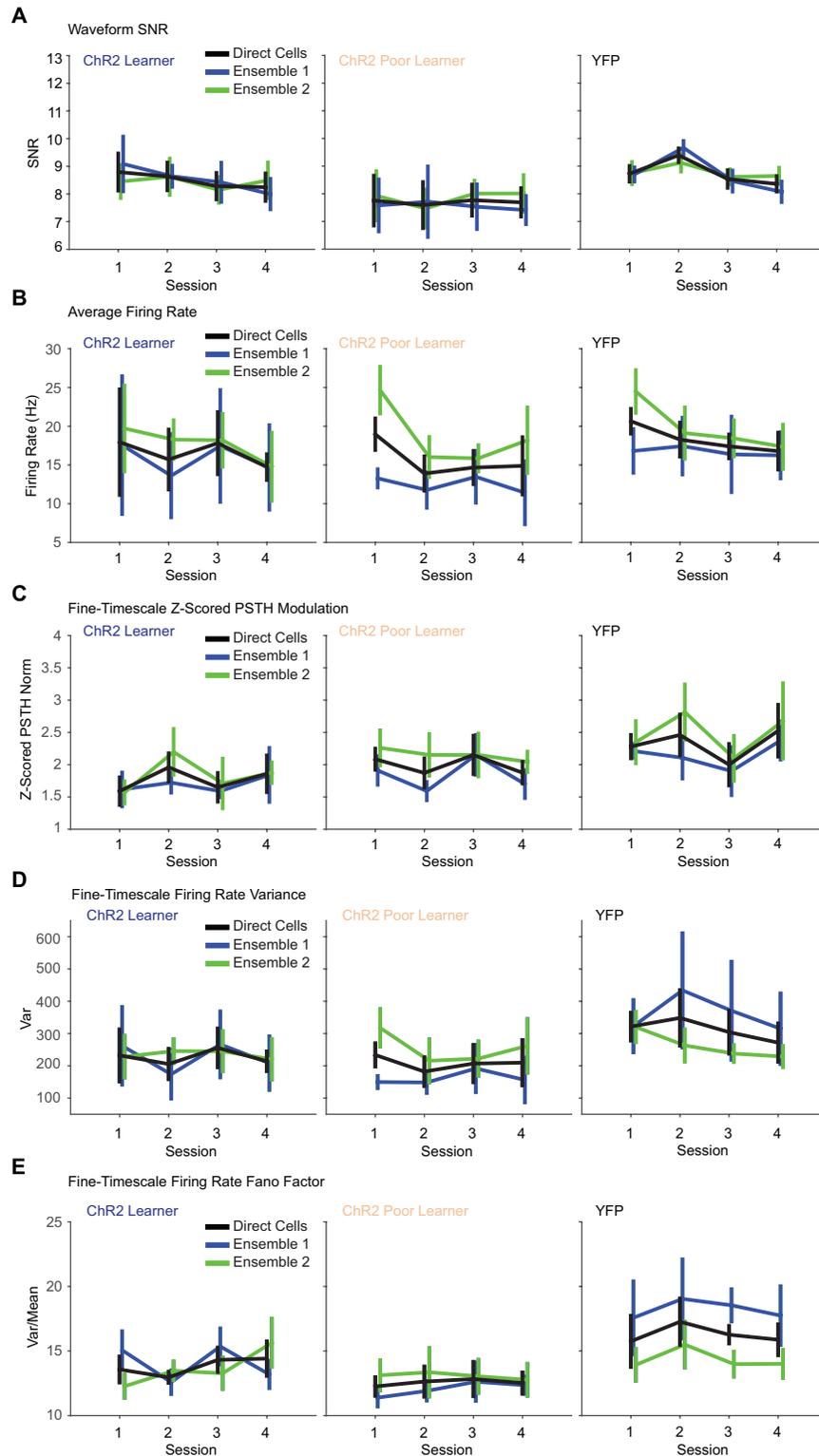


Figure 3.21: Individual neuron firing properties over training do not correlate with learning.

Figure 3.21: **Individual neuron firing properties over training do not correlate with learning.** Each analysis in this figure other than waveform SNR was performed using neural activity in the 3-second window preceding Target 1 hit binned in 100 ms bins. The s.e.m. is over animals within a session. Changes over training were analyzed with a Wilcoxon rank sum test between early and late sessions. **(A)** The SNR of neural recordings do not change over learning. **(B)** The average firing rate of direct neurons does not change over training. **(C)** The average fine-timescale modulation of direct neurons does not change over training. This was calculated as the norm of each z-scored PSTH and then averaged over each animal's neurons. **(D)** Fine-timescale firing rate variance does not change over training. **(E)** The fine-timescale firing rate Fano Factor (the variance divided by the mean using the 3-second window preceding Target 1 hit over 100 ms bins) does not change over training.

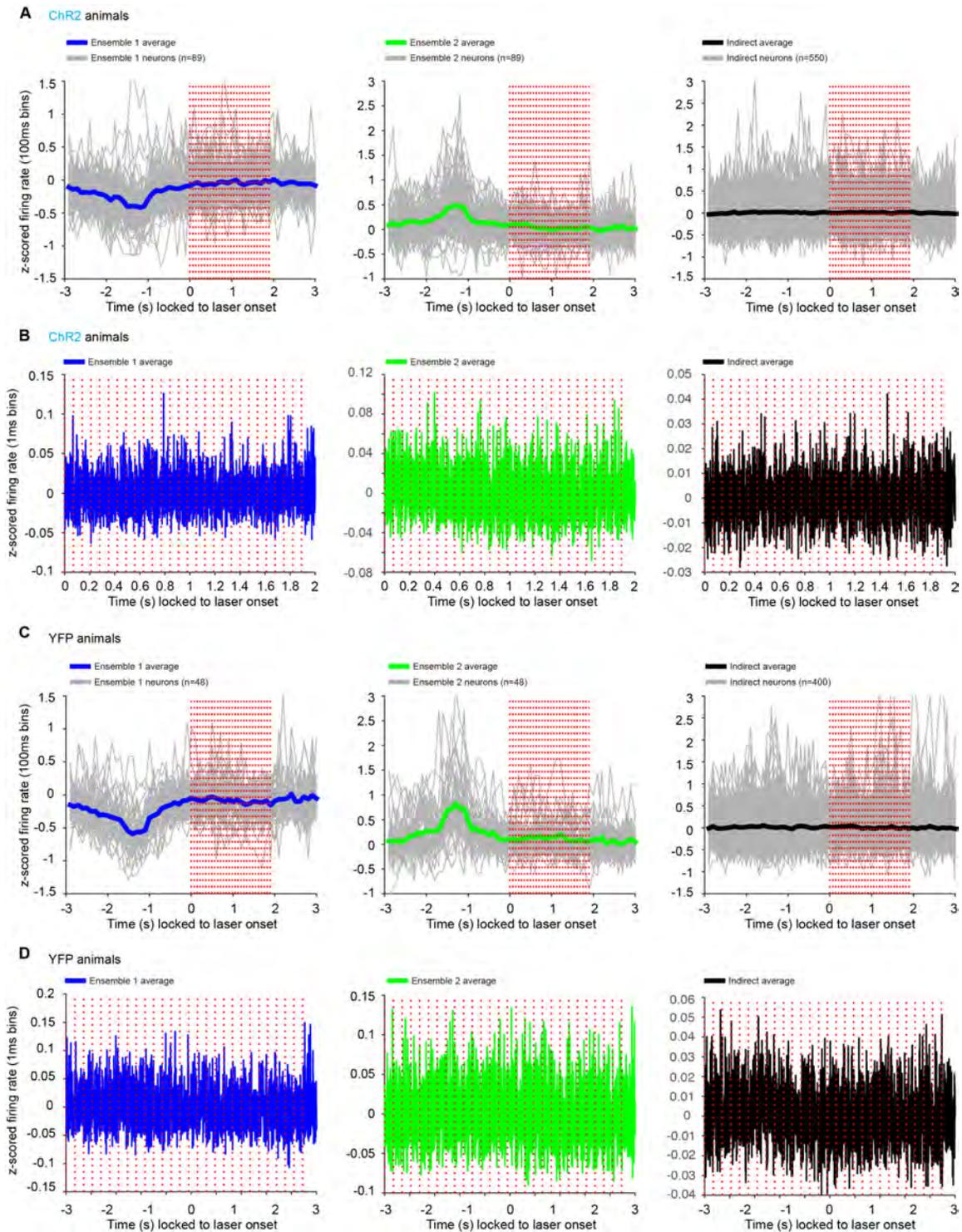


Figure 3.22: No effect of laser stimulation on average M1 firing rates.

Figure 3.22: **No effect of laser stimulation on average M1 firing rates.** In this figure, red dotted lines indicate laser pulse times. Gray lines indicate individual neuron PSTHs. Neurons recorded on separate sessions were considered separate. **(A)** In ChR2 animals, laser onset does not change the firing rate of direct or indirect cells calculated in 100 ms bins. There were n=89 Ensemble 1 neurons, n=89 Ensemble 2 neurons, and n=550 indirect neurons, pooling over animals and sessions. **(B)** In ChR2 animals, laser pulses do not trigger phasic firing responses. PSTHs calculated in 1 ms bins. **(C)** In YFP animals, there is no average firing rate change after laser onset. There were n=48 Ensemble 1 neurons, n=48 Ensemble 2 neurons, and n=400 indirect neurons, pooling over animals and sessions. **(D)** In YFP animals, there are no phasic firing responses locked to laser pulses.

## Chapter 4

# Volitional modulation of primary visual cortex activity requires the basal ganglia

In the last chapter, we observed that phasic dopaminergic activation reinforces preceding motor cortical activity patterns. The neural circuit mechanism likely relies on the dopaminergic reinforcement signals' contribution to plasticity between cortex and striatum, the input area to the basal ganglia, as this corticostriatal plasticity is necessary for BMI learning. Interestingly, the basal ganglia, through the striatum, receives input from most cortical areas and feeds back to the cortex via multiple recurrent pathways. Thus, we asked whether the basal-ganglia dependent ability to explore and consolidate activity patterns leading to reinforcement generalizes across cortex. Evidence suggests that activity in primary sensory cortices may also reflect internally driven processes, instead of purely encoding antecedent stimuli. Here, we show that rats and mice can learn to produce arbitrary patterns of neural activity in their primary visual cortex to control an auditory cursor and obtain reward. Furthermore, the ability to consolidate the neural patterns that lead to reinforcement was abolished when activity in the region of striatum that receives input from visual cortex was optogenetically inhibited. This inhibition did not affect the ability to produce the reinforced patterns after learning. These data demonstrate that cortico-basal ganglia circuits play a general role in learning to consolidate cortical activity that leads to desirable outcomes.

### 4.1 Introduction

As discussed in the previous two chapters, animals often learn new behaviors by initially exploring action space and then consolidating action patterns leading to reinforcement (Cohen and Sternad, 2009; Shmuelof et al., 2012; Skinner, 1938; Thorndike, 1898), a process accompanied by the consolidation of neural ensembles and activity patterns in motor cortices (Barnes et al., 2005; Cao et al., 2015; Costa et al., 2004; Iveczky et al., 2011; Peters

et al., 2014). BMI learning studies have shown that the motor cortices consolidate neural ensembles and activity patterns which causally lead to reinforcement (Carmena et al., 2003; Clancy et al., 2014; Fetz, 1969; Ganguly and Carmena, 2009; Koralek et al., 2012).

Critically, plasticity between cortex and striatum, the input area to the basal ganglia, is necessary for BMI learning (Koralek et al., 2012) and is modulated by dopaminergic reinforcement signals. Intriguingly, the basal ganglia, though the striatum, receives input from most cortical areas in addition to motor cortex (Hintiryan et al., 2016; Kemp and Powell, 1970; McGeorge and Faull, 1989; Webster, 1965), feeds back to the cortex via multiple recurrent pathways (Redgrave et al., 2010), and dynamically encodes action-outcome contingencies (Samejima, 2005; Tricomi et al., 2004). Thus, we asked if the basal-ganglia dependent ability to consolidate activity patterns leading to reinforcement generalizes across cortex, even to primary sensory areas such as visual cortex.

The ability to modify activity patterns based on behavioral consequences could be beneficial even in primary sensory areas. For example, attentional resources can be used to amplify activity related to salient stimuli while ignoring distractors. Many sensory areas of the cortex, including primary sensory areas, display activity that reflects task parameters, changing behavioral context, and shifting attention, suggesting that computations in these regions are influenced by ongoing internal processes (Keller et al., 2012; Martinez et al., 1999; Niell and Stryker, 2010; Shuler, 2006; Steinmetz et al., 2000; Zhang et al., 2014). These task-relevant modulations of ongoing sensory representations can emerge and evolve following repeated training or association with a salient stimulus (Makino and Komiyama, 2015). An important question is how modulatory inputs to functionally diverse cortical circuits are tuned such that their outputs contribute positively to the behavioral outcomes of an individual. Previous work has demonstrated the importance of the striatum for voluntary behavior and instrumental learning (Hikosaka et al., 1999; Yin et al., 2005, 2006, 2009). However, in addition to overt motor behaviors driven by motor cortices, cortico-striatal circuits have been theorized to also support abstract forms of learning, such as cognitive associations (Graybiel, 1997; Middleton and Strick, 1994). Furthermore, damage to basal ganglia structures in human patients, either through stroke or in diseases like Parkinsons, have been associated with deficits in sensory perception and the control of visual attention (Brown et al., 1997; Husain et al., 1997; Mercuri et al., 1997; Wright et al., 1990; Yamaguchi and Kobayashi, 1998). These data suggest that basal ganglia circuits may be involved in learning modulatory signals that influence many forms of cortical processing based on experience. However, observing and measuring these influences can be difficult, especially when their contributions to overt behavior may not be immediately apparent.

One strategy to overcome this difficulty is to use Brain-Machine Interfaces (BMIs) which directly map a subjects internally generated neural activity to the movement of an artificial effector. By explicitly defining the behavioral relevance of observable patterns of neural activity, BMI can be an important tool for studying how these patterns are generated.

Here, we asked whether neurons in the primary visual cortex, an area involved in processing low-level visual features, could learn to consolidate arbitrary activity patterns leading to reinforcement, and whether this abstract form of learning was dependent on the basal gan-

glia. To address this question, we trained rats and mice to perform a neuroprosthetic task that virtually re-routed spike activity from the primary visual cortex (V1) into the frequency of an auditory cursor. This allowed us to facilitate and observe learned modulations of V1 activity with a known relationship to behavior. Animals trained on the task successfully learned to produce this novel action by voluntarily modulating spike activity in V1. Then, using the red-shifted inhibitory opsin Jaws (Chuong et al., 2014) to inactivate striatal neurons on a trial-by-trial basis, we then investigated to what degree this instrumental learning process in V1 was also dependent on activity in dorsomedial striatum.

## 4.2 Methods

### Surgery

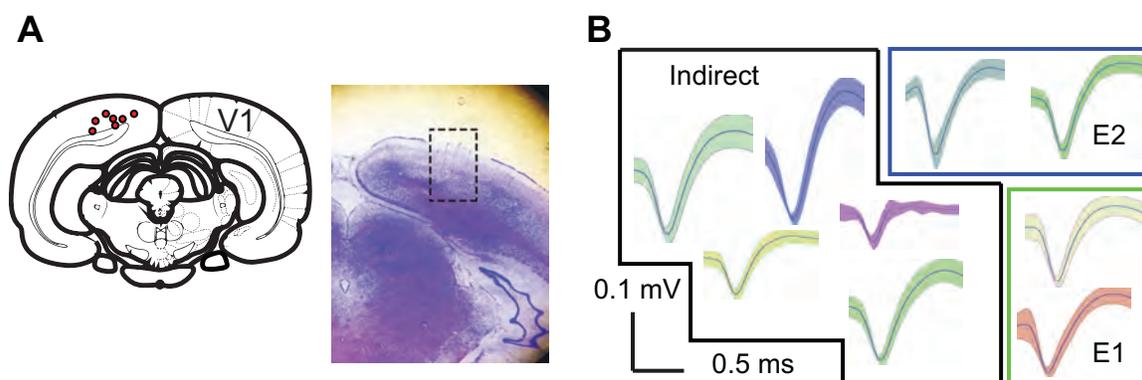


Figure 4.1: **Rat V1 histology and recordings.** (A) Left: schematic showing the location of rat V1 electrode arrays, verified by histology for 6 of 10 animals. Right image shows an example section with electrode tracks. (B) Sample V1 waveforms from an example session. Units outlined in green were chosen for Ensemble 1; units outlined in blue were chosen for Ensemble 2; and units outlined in black were recorded but had no direct relationship to cursor control (Indirect units).

All rat experiments were performed in compliance with the regulations of the Animal Care and Use Committee at the University of California, Berkeley. A total of ten singly-housed, male Long-Evans rats weighing roughly 250 grams were used for the experiments. All rats were chronically implanted with microwire arrays in V1, with 8 rats also receiving implants in the dorsomedial striatum. Each array contained 16 or 32 tungsten microelectrodes (35  $\mu$ m diameter, 250  $\mu$ m electrode spacing,  $8 \times 2$  or  $8 \times 4$  configuration; Innovative Neurophysiology, Durham, NC). Stereotactic coordinates relative to bregma and lambda were used to center the arrays (1 mm anterior of lambda, 3.5 mm lateral from the midline, and 1.25 mm ventral from the cortical surface for V1; 1.8 mm anterior of bregma, 2 mm lateral of the midline,

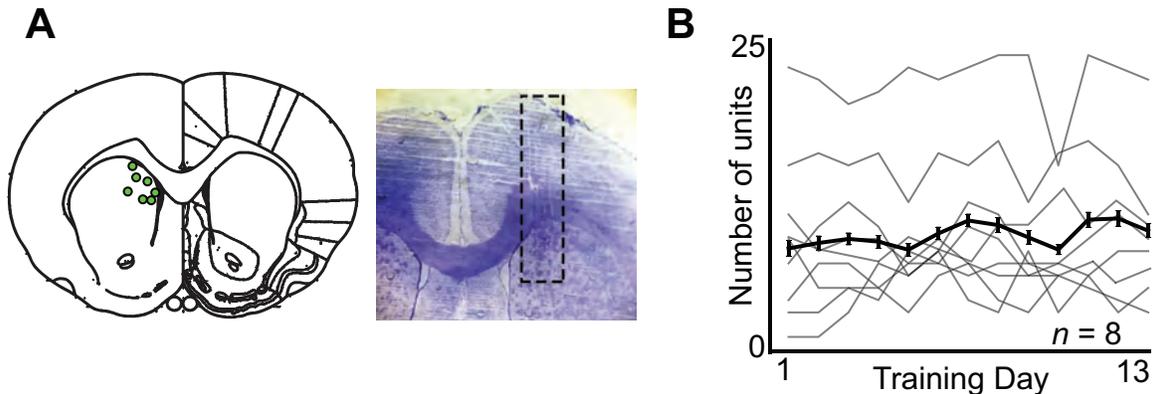


Figure 4.2: **Rat DMS histology and number of recordings.** (A) Left: schematic showing the location of rat DMS electrode arrays, verified by histology for 6 of 10 animals. Right image shows an example section with electrode tracks. (B) Total number of well-isolated DMS units for 8 animals recorded across several training sessions. Solid line shows the mean and s.e.m.; grey lines show data for individual animals.

and 5.5 mm ventral from the cortical surface for DMS). V1 implants were targeted for layer 5 neurons based on insertion depth, and this was verified histologically at the end of experiments (Figure 4.1AB). Rodents were anesthetized with Ketamine (50 mg/kg) and Xylazine (5 mg/kg) with supplemental isoflurane gas as needed. Craniectomies were sealed with cyanoacrylate and rats were allowed to recover for ten days after implantation before behavioral training. Rats were given dexamethasone treatment (0.5 mg/kg) for one week following surgery to minimize tissue damage around the implant (Zhong and Bellamkonda, 2007).

Mouse experiments were performed in accordance with the Champalimaud Centre for the Unknown Ethics Committee guidelines and approved by the Portuguese Veterinary General Board (Direco Geral de Veterinária, approval 0421/000/000/2014). A total of eight male C57BL mice of 2.5-3.5 months of age were used (four experimental mice and four control mice). Surgeries were performed under isoflurane anesthesia (1-3%). All mice were injected bilaterally with 1  $\mu$ L of viral solution in the dorsomedial striatum using coordinates relative to Bregma (0.9 mm AP,  $\pm$ 1.5 mm ML, 2 mm below brain surface). Viral solution was injected through a glass pipette by pressure (Nanojet II from Drummond Scientific, 4.6 nl pulses at a rate of 0.2 Hz). For experimental animals, the virus injected was rAAV8/CamKII-Jaws-KGC-GFP-ER2 (University of North Carolina, titer  $5.9 \times 10^{12}$ ). For control animals, the virus injected was rAAV8/CamKII-GFP (University of North Carolina, titer  $2.8 \times 10^{12}$ ). For delivery of red light, mice were implanted bilaterally with optical fibers (250  $\mu$ m diameter, NA 0.63). All mice were also implanted with 16-channel movable electrode arrays (electrode diameter 23  $\mu$ m; Innovative Neurophysiology, Durham, NC) in the right primary visual cortex using coordinates relative to Lambda (0 mm AP, 0.3 mm ML, 0.6 mm below brain surface).

## Electrophysiology

Single unit activity and local field potentials were simultaneously recorded with a Multi-channel Acquisition Processor (MAP in rats, OmniPlex in mice; Plexon Inc., Dallas, TX). Activity was sorted using an online sorting application (Plexon Inc., Dallas, TX) prior to each daily recording session. Only units with a clearly identified waveform and signal-to-noise ratio greater than 2 were used. Sorting templates were further refined using an offline sorting application (Plexon Inc., Dallas, TX). Behavioral timestamps were sent to the MAP recording system through custom Python and C++ programs and synchronized to the neural data for later analyses. Recording arrays were grounded to a screw in the occipital bone, and both arrays were also referenced locally using the online program Ref2 (Plexon Inc., Dallas, TX) to eliminate effects of volume conduction. For referencing, an electrode on each array was chosen to be subtracted from all other electrodes on that array. This was done independently for both V1 and DMS.

## 4.3 Behavioral Task

After recovering from surgery, rats were trained on the task in a dark behavioral chamber (Lafayette Instrument Company, Lafayette, IL) unless otherwise specified. During training, rats only received access to water during the task unless supplemental water was needed to maintain their body weight at a healthy level. At the start of each session, two ensembles of 2 well-isolated V1 units each were chosen for inclusion in the direct population based on SNR, interspike-interval histograms, and refractory periods. No other selection criteria were used to partition the recorded cells into each ensemble. Although these direct units were consistently well-isolated, we also ensured that many well-isolated units were included in the indirect population to enable a proper comparison. The units assigned to the direct population remained relatively constant throughout training using the stability of spike waveforms, sorting templates, and interspike intervals across sessions as a guide. After sorting and partitioning of direct and indirect units, a 15-minute baseline period was recorded in which animals received a sucrose water reward on a variable-interval schedule. During this time, spike counts were recorded for each ensemble binned into 100 ms bins, and a distribution of state values was calculated by subtracting the counts from E1 from E2 in each time bin. From here, the distribution was fit by a Gaussian mixture model (GMM) comprising between 1 and 10 Gaussian components (the exact number was chosen by finding the Akaike Information Criterion (AIC) value for each possible number of components, and choosing the number with the lowest AIC value). The probability density function (PDF) was then computed for the chosen model. By default, the neural state values (E1 E2 counts for a 100 ms bin) corresponding to the low and high targets were set at the points on the computed PDF where the area under the curve was equivalent to 1.5% and 98.5% of the total area, respectively (Figure 4.3A). However, these target values were iteratively updated by running a simulation using the data recorded in the baseline period until the probability

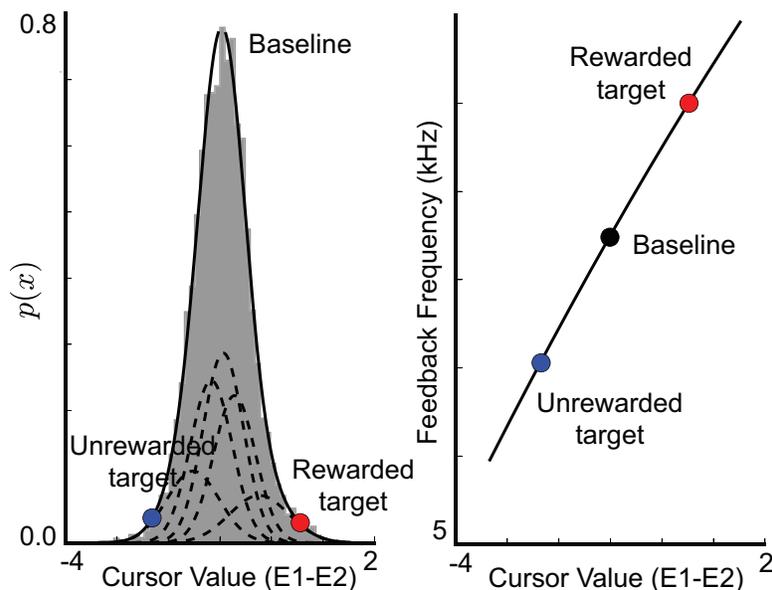


Figure 4.3: **Decoder calibration.** (A) Plot shows a histogram of binned E1-E2 values (i.e. total spike counts for all ensemble 1 neurons minus total spike counts for all ensemble 2 neurons for each 100 ms bin) during a 15-minute baseline period. A mixture of Gaussians (dashed lines) are fitted to generate the PDF (solid black line). Based on a simulation of the task using recorded baseline neural activity, thresholds are calculated for a low target (unrewarded in this plot), a high target (rewarded), and a midpoint (Baseline). (B) The line fitted to the T2, Baseline, and T1 values that is used to compute a feedback frequency based on the binned E1-E2 value.

of hitting each target was approximately 30%. Finally, using the lowest and highest target values as well as the mean of the GMM, in combination with the lowest (1000 Hz), highest (15000 Hz) and midpoint (7000 Hz) frequencies, a 2-degree polynomial function was fit to these values in order to map neural state to a frequency value. During online performance, the state value used to compute the instantaneous feedback frequency was smoothed over the most recent 10 bins, updated every 100 ms. The rodents had to then precisely modulate these neuronal ensembles to move the cursor to one of two target frequencies, one which was randomly chosen using a coin flip on a per-animal basis to be associated with a 20% sucrose solution reward (and kept consistent across training sessions). Rodents were free to reach either target, although the cursor value had to return to the middle value for a new trial to begin. A trial was marked as a miss if neither of these target states were achieved within 30 seconds of trial initiation. Recorded neural data was entered in real time to custom routines in Python and C++ that then translated those activity levels into the appropriate feedback frequency as described above and played the frequency on speakers mounted above the behavioral chamber. When a target was hit, a Data Acquisition board (National Instru-

ments, Austin, TX) controlled by a Python script triggered the operant box to supply the appropriate reward to rodents.

## Behavioral Manipulations

After initial training of at least 8 days, a contingency degradation was performed. This took place over two sessions: during the first session, animals were allowed to achieve stable performance on the task, which took approximately 30 minutes (pre-degradation). Then, reward delivery was decoupled from task performance and delivered on a variable-interval schedule that approximated the rate of reward during training conditions (CD). During this time, animals still received auditory feedback linked to their neural state. The contingency degradation continued for the remainder of the session. The next day, animals were again trained on the task under normal conditions (reinstatement). Similarly, for contingency reversal sessions, we reversed the contingency approximately 30 minutes into a training session. During sessions that involved light manipulation, we again waited for animals to achieve saturating performance in a single session, and then placed a light into the behavioral chamber in an overhead configuration.

## Optical Inhibition

For optical inhibition experiments, red light was applied to the striatum of both experimental and control groups on 50% of all trials in a session. Light was applied through a fiber-coupled LED system (Prizmatix, Givat-Shmuel, Israel). Power levels tested through the system at the optical fiber tip ranged from 4-6 mW. Red light application consisted of a square pulse that lasted the full duration of the current trial, from trial initiation until either a target or timeout was reached. Both groups were trained in this manner for 7 days. Next, the experimental animals were trained for 5 additional days in the absence of red light. After this initial training, both groups were tested to determine the role of striatal circuits in learning versus performance of the task. On day LED 50, red light was applied on 50% of all trials. On day LED late, no light was applied for the first 45 trials, after which red light was applied on 50% of all remaining trials.

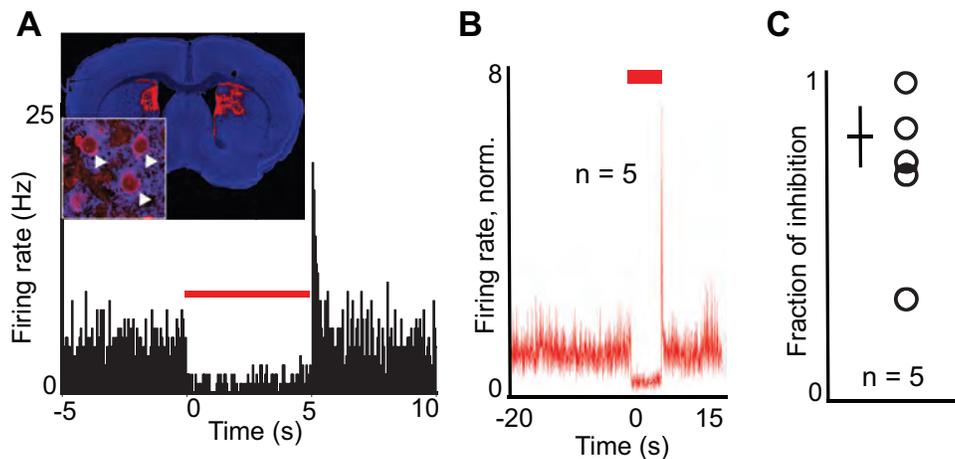


Figure 4.4: **Jaws inhibition of DMS.** (A) Jaws-mediated inhibition of DMS neurons. Inset shows a coronal section stained for neuronal cell bodies (blue) and Jaws expression (red). Zoomed inset shows a magnification with labeling apparent on single neurons (white arrows). Histogram plot shows suppression of spike activity in one example DMS neuron during Jaws-mediated inhibition (red bar). (B) Mean firing rates of all recorded DMS units during Jaws-induced inhibition with red light (red bar);  $N=5$ . Firing rates were normalized to baseline firing rate values. Shaded areas indicated s.e.m. (C) Fraction of inhibition for recorded DMS units during Jaws-induced inhibition, where a value of 1 indicates complete inhibition of all spike activity. Grey circles show individual units.

## Data Analysis

Analyses were performed in Python with custom-written routines utilizing publicly available software packages. Unit data were first binned in 1 ms time bins and digitized. Firing rate analyses were performed in relation to target achievement as indicated in figures. Unless otherwise specified, firing rates were binned into 50 ms bins for all analyses. Only two-sided statistical tests were used to determine significance. The term “early” indicates that analyses were performed using data taken from animals during the first 3 days of training (inclusive), while “late” specifies data taken from animals during the final 3 days of training. The “first trials” of a session indicates trials in the first 10 minutes, while the “last trials” of a session are defined as occurring in the last 10 minutes, unless otherwise specified in the text. Modulation depths were computed as the difference between the maximum and minimum firing rate values in a 2-second window centered around target achievement. Coherence analyses were performed using algorithms translated to Python from the Chronux toolbox (<http://chronux.org>) in conjunction with custom routines in Python. A multi-taper method was used to compute spectral estimates of spiking and LFP activity (Jarvis and Mitra, 2001; Thomson, 1982). A total of 5 tapers were used with a time-bandwidth product of 3, and estimates were computed every 50 ms with a window size of 500 ms. Coherence between spiking in LFP activity was calculated and defined as:

$$C_{xy} = \frac{|R_{xy}|}{\sqrt{|R_{xx}||R_{yy}|}}$$

where  $R_{xx}$  and  $R_{yy}$  are the power spectra and  $R_{xy}$  is the cross-spectrum. Spectral analyses were calculated relative to the delivery of reward and averaged across trials and animals. Coherence estimates can be affected by firing rate (Lepage et al., 2011) and we therefore performed a thinning procedure to equate firing rates between conditions in which rates differed (Gregoriou et al., 2009). Trial-averaged spike trains in the neuronal populations were smoothed with a moving average of 10 ms. The difference in firing rate between the populations normalized by the maximum firing rate at a given time point determined the probability that a spike would need to be removed from the population with a higher firing rate. Spikes were then removed from the population with a higher firing rate based on this probability in order to eliminate any possible influence of firing rate on coherence estimates.

The signal-to-noise ratio for each recorded waveform was quantified as:

$$\text{SNR} = \frac{A}{2 \cdot \text{SD}_{\text{noise}}}$$

where  $A$  is the peak-to-peak voltage of the mean waveform and  $\text{SD}_{\text{noise}}$  is the standard deviation of the residuals from each waveform after the mean waveform has been subtracted (Suner et al., 2005). Units included in the analysis had a minimum SNR of 2.

For logistic regression analyses, we used functions from the publicly available python package scikit-learn ([www.scikit-learn.org](http://www.scikit-learn.org)). Regression was performed using a window of spike activity 500 ms prior to target hits, binned into 50 ms bins. L2 Regularization was

done using cross-validation to determine the optimum regularization parameter. 3-fold cross validation was performed 5 times using left out data to compute accuracies, and the average of all 5 results was taken to be the final accuracy value. Chance rates were taken as the accuracy of the analyses using shuffled data. To determine significance values, a permutation test was used that compared the accuracy of the prediction using the original dataset compared to dataset in which target identities for all trials were shuffled. Neural activity was considered to be significantly predictive of target choice if the accuracy of the prediction using the original dataset outperformed the accuracy of the shuffled version on 95% of 500 iterations.

## 4.4 Results

### A V1-driven closed-loop brain-machine interface

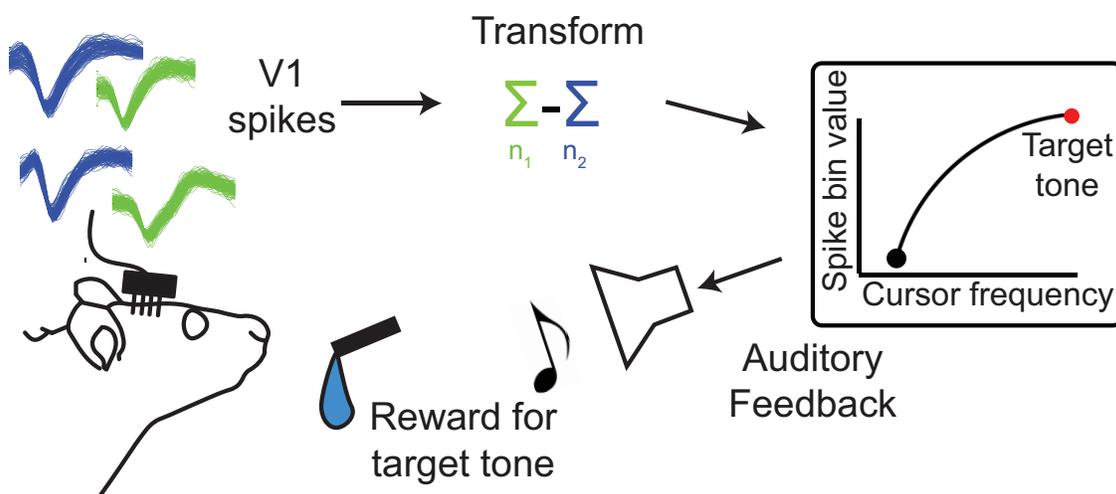


Figure 4.5: **V1 BMI paradigm.** Activity of well-isolated V1 units (top left) were used to generate auditory tones using a differential transform (top right). Animals were rewarded for producing a target tone (red). A second tone (black) at the opposite end of the frequency range terminated the trial but was not rewarded.

We adapted our previously developed neuroprosthetic task for rodents (Koralek et al., 2012) in order to directly study neural reinforcement learning in V1 (Figure 4.5). Briefly, 10 rats (and 8 mice, see later) were implanted chronically with microwire electrode arrays positioned in V1 layer 5 (L5), allowing us to isolate and record individual units (Figure 4.5A, Figure 4.1AB). In addition to V1, 8 rats were also implanted with electrode arrays in the dorso-medial striatum (DMS) near the projection target of V1 (Hintiryan et al., 2016; Khibnik et al., 2014; McGeorge and Faull, 1989) (Figure 4.2). During the course of the experiment,

animals were placed in a totally dark or lighted behavioral chamber and allowed to move freely while listening to auditory feedback that reported their neural state in real time. Each day, two neural ensembles, consisting of two well-isolated units each, were randomly chosen to directly control the continuous auditory cursor (direct units), while the remaining units recorded in V1 had no defined relationship to cursor control (indirect units, Figure 4.1B). Activity of the two direct-unit ensembles had an opposing relationship, such that spikes produced by Ensemble 1 (E1) moved the cursor closer to the rewarded frequency, while spikes in Ensemble 2 (E2) moved the cursor away from the rewarded frequency and towards the unrewarded frequency. Significant modulations were required to achieve the targets (Figure 4.6).

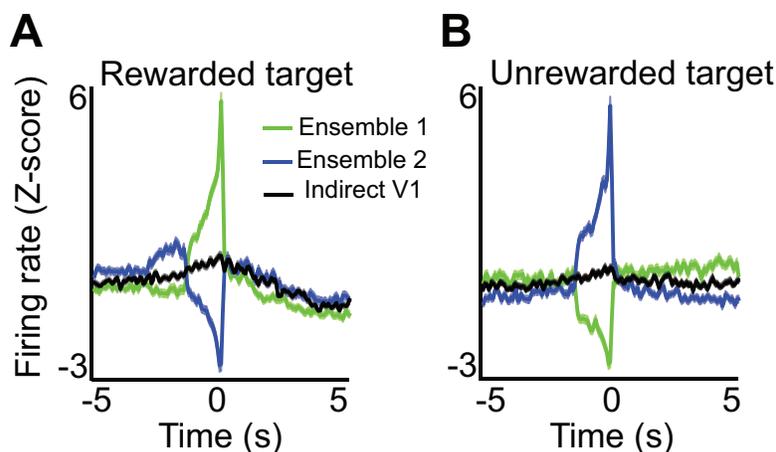


Figure 4.6: **Average Ensemble 1 and 2 modulations preceding target hit.** (A) Average z-scored firing rates of V1 neurons arbitrarily assigned to Ensemble 1 (green), Ensemble 2 (blue), or unassigned (indirect; black), time-locked to rewarded targets. Shaded areas show s.e.m. (B) Same as (A), but time-locked to the unrewarded target. Increased activity in Ensemble 1 units moved the tone frequency in the opposite direction as increased activity in Ensemble 2 units.

The highest and lowest possible tones were randomly assigned to be rewarded or unrewarded for each animal, and this association remained constant for the duration of training. Prior the start of every session, a baseline distribution of neural states (binned E1 - E2 spike counts) was used to initialize the target values such that the chance rate of attaining each target (rewarded and unrewarded) within a trial period was approximately 30% (Figure 4.3; see Methods). Animals were given 30 seconds to reach either target; otherwise the trial was considered a miss and animals received a white noise burst followed by a time-out.

## Rodents consolidate V1 activity patterns leading to reinforcement

Over the course of 7-12 training sessions (average 9.11 sessions), rats learned to perform the task well above chance level (Figure 4.7).

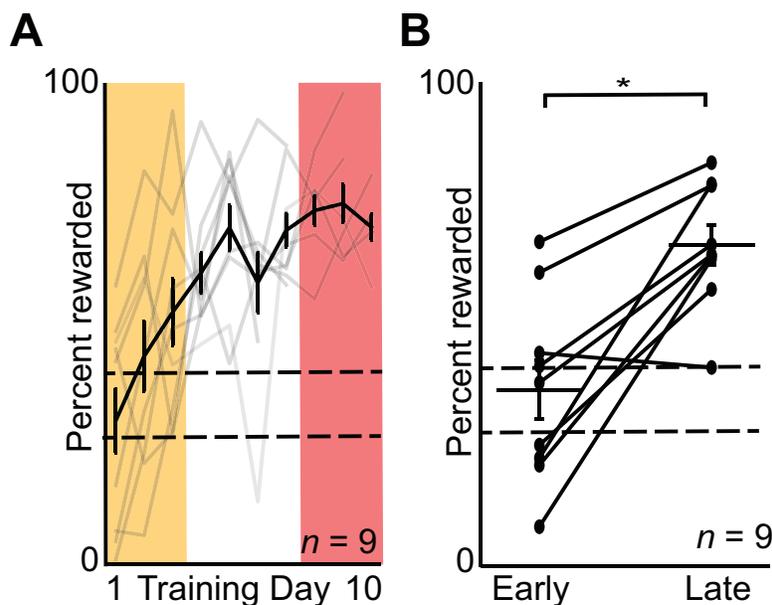


Figure 4.7: **Task performance over training.** (A) Timecourse of learning across training days. Bold line shows the mean and s.e.m. across 9 rats; grey lines show learning curves for animals individually. One animal only completed 4 sessions; data for this animal has been excluded from this plot. Dashed lines bound the range of chance performance levels. Orange highlighted region shows data range classified as the early learning phase (first 3 days) for all animals; red region shows range for late learning phase (last 3 days). (B) Comparison of performance between early learning phase and late learning phase. (N=9 animals, early mean = 39.6% rewarded; late mean = 67.2%;  $P=0.00162$  (paired t-test))

Animals in the late phase of learning, considered here the final 3 training sessions, exhibited significant improvements in the percentage of rewarded targets compared to their performance in the early phase (during the first three sessions of training) (Figure 4.7; mean of the first three days for all animals = 39.7%; mean of the last 3 days = 67.2%;  $P = 0.00162$ ). Additionally, over this same interval, latencies to rewarded targets decreased significantly, while changes in latencies to unrewarded targets were non-significant (Figure 4.8).

We observed that simply pairing particular auditory tones with reward was not sufficient to drive V1 activity. After 5 days of performance above chance level, we decoupled auditory tones from neural activity mid-session by playing back the sequence of tones generated in the first part of the session. Although reward was still delivered in tandem with the rewarded tone, modulation of the direct unit ensembles was markedly decreased when animals' neural

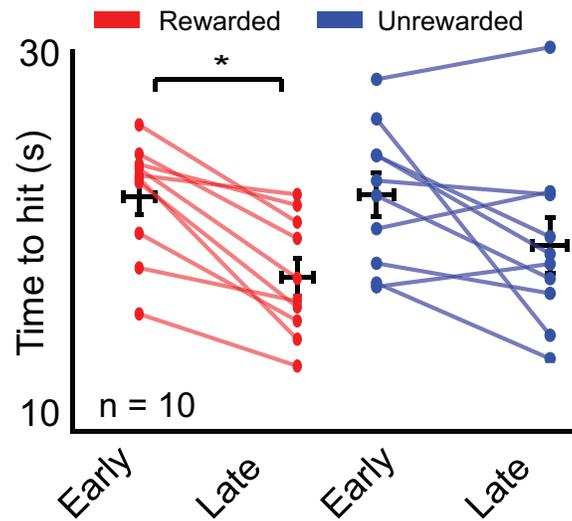


Figure 4.8: **Time to target hit.** Latencies to rewarded vs unrewarded targets for early training (first three days of training) compared to late training (last three days of training). Mean rewarded targets, early =21.4 sec; mean rewarded late = 18.0 sec;  $P=.0156$  (paired t-test). Mean unrewarded targets, early=20.7 sec; unrewarded late=18.1;  $P=0.068$  (paired t-test).

activity was not driving the cursor (Figure 4.9). This suggests that the learning we observed was not an effect of classical conditioning, and instead resulted from an intentional modulation of V1 activity.

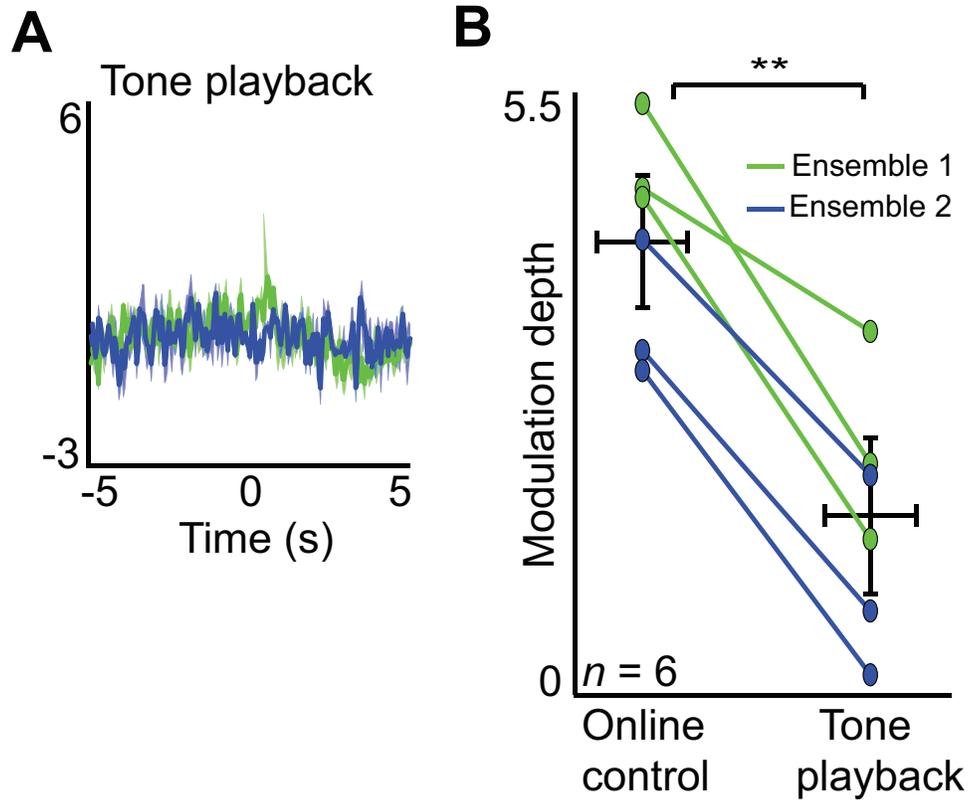


Figure 4.9: **Neural modulations during passive tone playback.** (A) Ensemble 1 (green) and Ensemble 2 (blue) modulations during passive playback of tones decoupled from ongoing neural activity, time-locked to the rewarded target tone. (B) Modulation depth of Ensemble 1 and Ensemble 2 during online control compared to tone playback. Mean modulation depth online control = 4.34; mean tone playback = 1.74;  $P=0.00035$  (paired t-test). Black bars show mean and s.e.m.

Because the chance rate for each target was reset at the start of each session to approximately 30%, increases in performance seen as animals progressed from early to late phases reflected greater improvements within single training sessions across days. We quantified this by comparing performance in the first 10 minutes of a session (first trials) and the last 10 minutes (last trials) (Figure 4.10A). This suggests that the learned ability to control V1 activity was retained between training sessions, even though animals needed to perform some *de novo* learning to adjust to the initial parameters of the transform in any given day. We observed that in late learning, over the course of single training sessions, animals acquired a strong preference for the rewarded target relative to the unrewarded target (Figure 4.10CD). This was reflected in the shift of the distribution of auditory cursor values in the direction corresponding to the rewarded target tone (Figure 4.10B). Additionally, we observed that auditory feedback was necessary for learning: sessions in which the feedback tones were muted resulted in no significant difference between the number of rewarded and unrewarded targets ( $P = 0.738$ , Figure 4.11), even though these no-feedback sessions were conducted after several days of successful normal training. These data demonstrate that closed-loop neurofeedback-based reinforcement training can be used to condition the activity of neurons in the primary visual cortex.

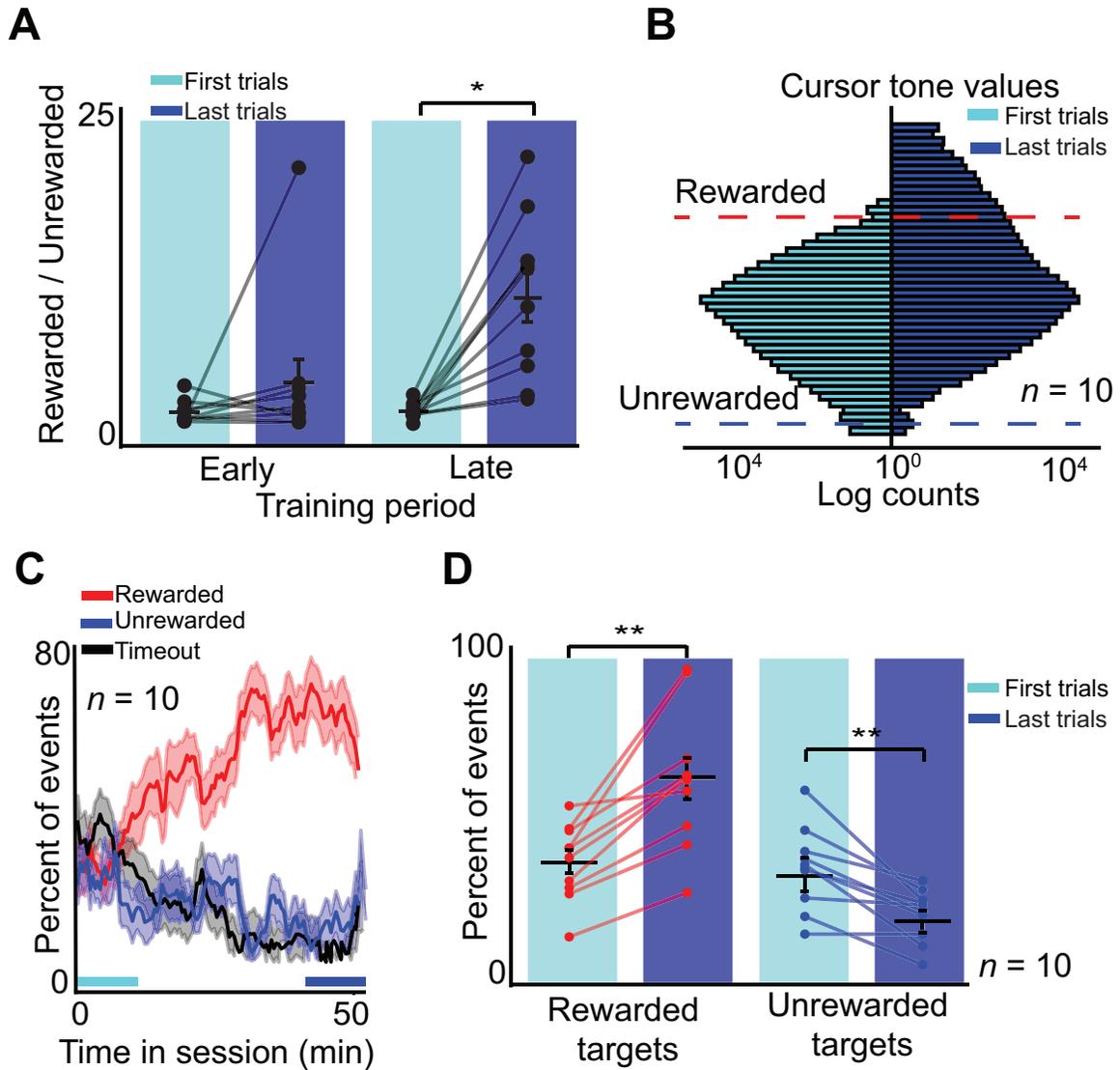


Figure 4.10: Within-session learning increases over training sessions.

Figure 4.10: **Within-session learning increases over training sessions.** **(A)** Comparison of within-session improvements during the early learning period (first 3 days) of training relative to the late learning period (last 3 days) for each animal ( $N = 9$  animals), expressed as the ratio of rewarded to unrewarded targets. Data to compute the ratio for first trials (shaded in cyan) was averaged over the first 10 mins of each session, while data for the last trials (shaded in blue) was averaged over the last 10 mins. During the early learning period, the mean rewarded/unrewarded ratio for first trials was 1.12, last trials session = 3.848;  $P = 0.275$  (paired t-test). For the late period, mean ratio early in session = 1.195, mean late in session = 11.667;  $P = 0.00130$  (paired t-test). **(B)** Mean distribution of cursor values for all animals for the first trials (first 10 mins) compared to the last trials (last 10 mins) of sessions during the late period. Cyan bars show the initial distribution, based on baseline activity, used to set the task parameters, while blue bars show the distribution at the end of the training session for the last trials. Dashed lines show the thresholds corresponding to the rewarded and unrewarded targets. **(C)** Timecourse of mean within-session task learning during the late period of training for all animals. Shaded areas represent s.e.m. Chance rates for rewarded and unrewarded targets were set at approximately 30% at the start of each training session. Comparisons in performance were made during over first 10 mins of each session (first trials, cyan bar) and over the last 10 minutes (last trials, blue bar) of each session.  $N = 10$  animals, mean of 3 sessions per animal. **(D)** Quantification of rewarded and unrewarded target hits for the first trials compared to the last trials; data same as c.  $N = 10$  animals, mean of the late period (last 3 sessions) for each animal. Paired t-test between first trials and last trials for rewarded targets:  $P = 0.00055$ ; mean early = 36.9%, mean late = 59.1%. Paired t-test between first trials and last trials for unrewarded targets:  $P = 0.00145$ ; mean early = 33.0%, mean late = 20.6%. Red and blue indicate the rewarded and unrewarded targets, respectively (\*\* indicates  $P < 0.001$ ). Black lines show mean and s.e.m.

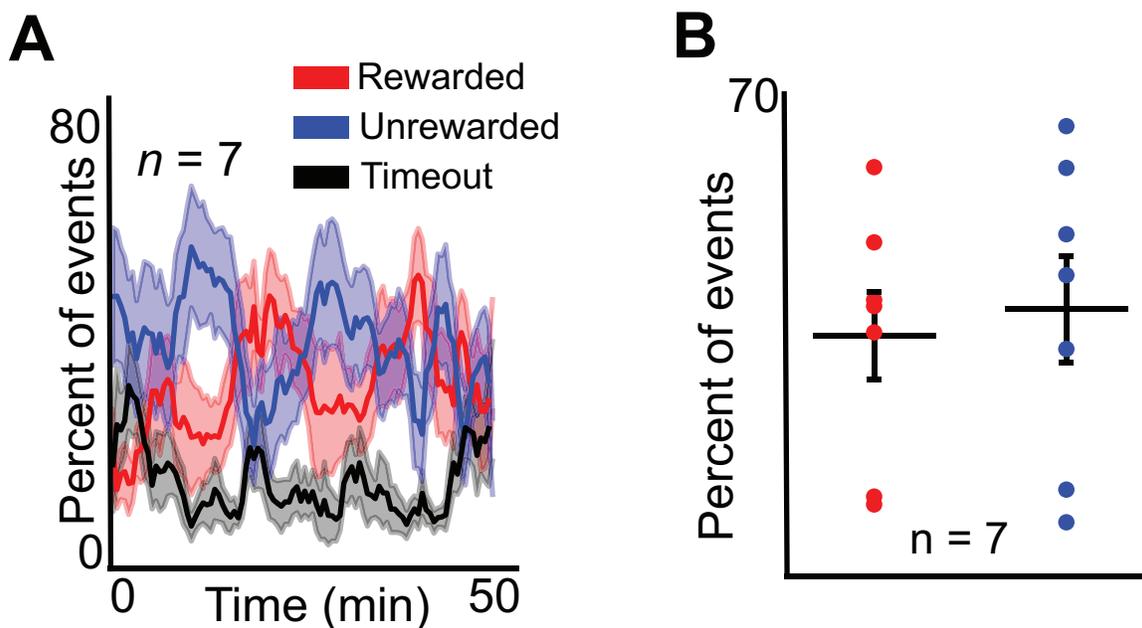


Figure 4.11: **Animals do not learn without auditory feedback.** (A) Timecourse of performance when animals were not given auditory feedback about cursor position during training. (B) Comparison of rewarded and unrewarded target hits during sessions without auditory feedback. Data is the mean over the last 10 minutes of each session.  $N = 7$  animals (mean of 1.5 sessions per animal; range 1-2 sessions). Mean rewarded targets = 38.1%; mean unrewarded = 41.7%;  $P = 0.738$  (unpaired t-test). Horizontal black line shows mean and vertical bars show standard error.

## V1 modulation is sensitive to task contingency

We next investigated the sensitivity of performance to changing task conditions. In late learning animals were able to quickly shift their neural activity to produce the rewarded tone frequency once the auditory feedback began (Figure 4.10CD). We asked whether animals' behavior was habitual and therefore insensitive to changes in action-reward contingencies; or goal-directed, in which performance remains sensitive to changing task contingencies (Dias-Ferreira et al., 2009). To test if performance of the V1-controlled task fit either of these regimes, we degraded the task contingency so that animals received randomly timed rewards irrespective of target hits, but at a similar rate. Similar to our observations of abstract skill learning in M1 (Koralek et al., 2012), task performance during degradation dropped significantly ( $N = 8$  animals; 2 sessions per animal. Pre-training = 64.5% rewarded, contingency degradation = 43.1%;  $P = 0.0038$ ) (Figure 4.12E), but returned to pre-manipulation levels once the reward contingency was reinstated (reinstatement = 61.4% rewarded;  $P = 0.289$ ). These results suggest that modulation of V1 activity was intentional and goal-directed. To test whether the association between neural activity and rewarded cursor movement could be flexibly adapted to a new task contingency, we reversed the transform after animals had achieved saturating performance. This manipulation caused the cursor to move in the opposite direction for a given spike rate modulation that what animals had initially learned. Although this manipulation caused an initial decrease in performance, animals were able to learn the reversed behavior when given sufficient training time (Figure 4.12F), showing that animals could dynamically adapt to changes in the relationship between neural patterns, feedback, and reward.

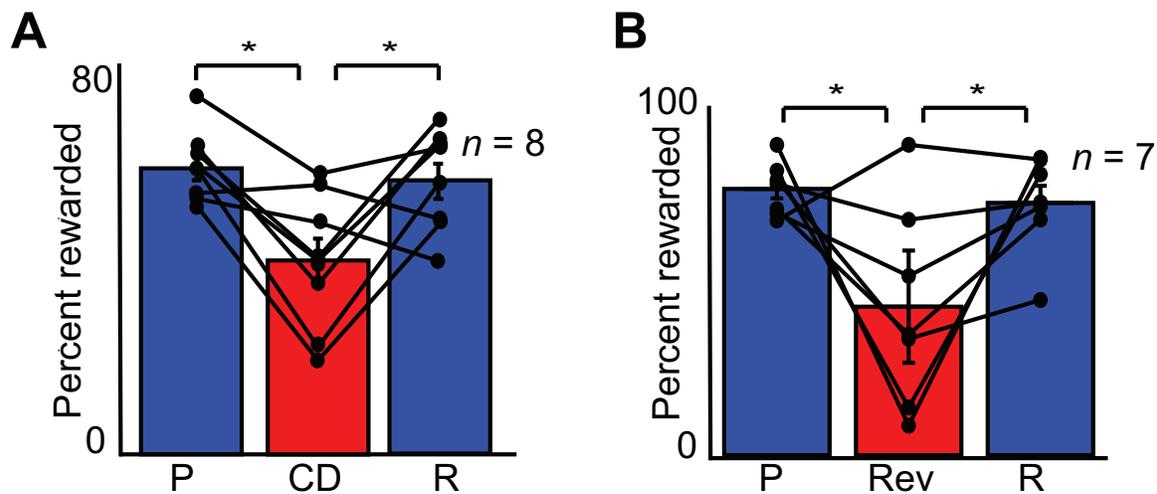


Figure 4.12: **V1 modulation is sensitive to task contingency.** (A) Percentage of rewarded trials for contingency degradation sessions. Bars show means, lines show individual animals. N=8 animals, mean of 2 sessions per animal. P: pre-degradation, mean=64.5%. CD: peri-degradation (reward decoupled from cursor), mean = 43.1%. R; reinstatement of reward, mean = 61.4%. Paired t-test between pre- (P) and peri- (CD) degradation:  $P = 0.0038$ . Paired t-test between CD and reinstatement (R):  $P = 0.0283$ . Paired t-test between pre-degradation (P) and reinstatement (R):  $P = 0.289$ . (B) Quantification of performance in contingency reversal sessions; N = 7 animals; average of 1.57 sessions per animal (range 1-2). P: pre-reversal (mean = 85.8%); Rev: perireversal (mean = 47.8%); R: recovery of performance with the decoder still reversed (mean = 81.2%). Paired t-test between P and Rev:  $P = 0.0435$ ; paired t-test between Rev and R:  $P = 0.447$ ; paired t-test between P and R:  $P = 0.689$ .

## Changes in visual context affect performance of a V1-driven task

Neurons in the primary visual cortex are known to respond to visual stimuli. The observation that animals can learn to successfully modulate V1 neurons in total darkness indicates that this activity is at least partially independent of visual input. One possibility is that over the course of learning, E1 and E2 units become decoupled from bottom-up sources of visual input; for example, visually-driven activity from the lateral geniculate nucleus of the thalamus. If this were true, then we can expect trained animals to be able to perform the task under any light condition. To test this, we trained animals both in light and total dark conditions. Interestingly, no significant difference was observed in performance at the end of a training session (last trials) for sessions performed entirely in the dark compared to in the light (train light = 74.0%, train dark = 63.5%,  $P = 0.205$ , Figure 4.13AB), suggesting that learning can occur both in the presence and absence of visual stimuli. However, changing the context within a training session, i.e. from dark to light after the animals had learned the task in darkness, or vice-versa, had a significant negative impact on performance (train dark = 87.4%, test light = 42.6%,  $P = 0.0309$ ; train light = 77.9% test dark = 35.4%,  $P = 0.043$ ) (Figure 4.13C-E). Changing the light conditions affected the mean spike rates of all V1 neurons. (Figure 4.14; mean rate in dark = 4.13 Hz; mean rate in light 6.18 Hz). These results suggest that a modulatory input can learn to generate rewarded patterns of activity in direct units under stable network conditions, but that changing the state of the network (for example, by adding or removing visually-evoked activity) is disruptive and requires compensatory re-learning.

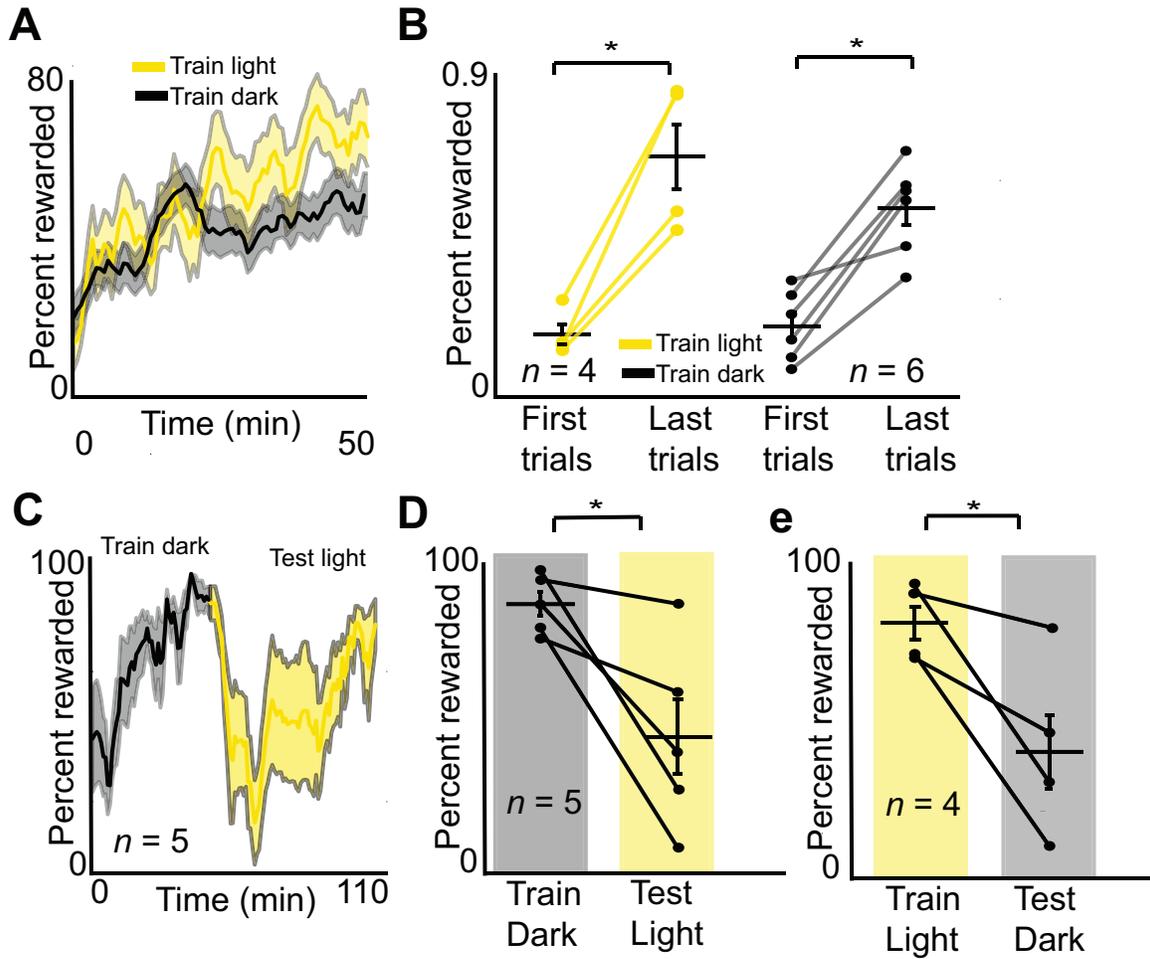


Figure 4.13: V1 modulation is sensitive to changes in visual context.

Figure 4.13: **V1 modulation is sensitive to changes in visual context.** (A) Timecourse showing the mean percentage of rewarded trials within all sessions under lighted conditions (yellow) compared to dark conditions (black). Shaded areas represent s.e.m. (B) Quantification of data in A, using the first trials and last trials in a session. Train light: N = 4 animals; average of 7.5 sessions per animal (range = 5 to 10 sessions). Paired t-test between first trials and last trials:  $P = 0.0113$ ; mean first trials = 37.8%; mean last trials = 74.0%. Train dark: N = 6 animals; 8 sessions per animal. Paired t-test between first trials and last trials:  $P = 0.00167$ ; mean early = 0.396; mean late = 0.635. Unpaired t-test between last trials for light sessions and last trials for dark sessions:  $P = 0.205$  (\* indicates  $P < .05$ ; black crosses show mean and s.e.m.). (C) Timecourse showing the mean percentage of rewarded trials when animals learned a decoder under dark conditions, and were switched to a lighted condition mid-session (“train dark, test light”). (D) Mean percentage of rewarded trials when animals were trained in dark, and then tested in the light (same data as c). N = 5 animals, mean of 1.8 sessions per animal (range 1-2). Data taken from last 15 mins of dark training and first 15 mins of light testing. Mean train dark: 87.4%, mean test light 42.6%;  $P = 0.0309$  (paired t-test). Error bars show s.e.m.; horizontal lines show mean (\* indicates  $P < 0.05$ ). (E) Mean percentage of rewarded trials when animals were trained in the light and tested in the dark. N = 4 animals, mean of 1.5 sessions per animal (range 1-2). Mean train light: 77.9%. Mean test dark: 35.4%.  $P = 0.043$  (paired t-test).

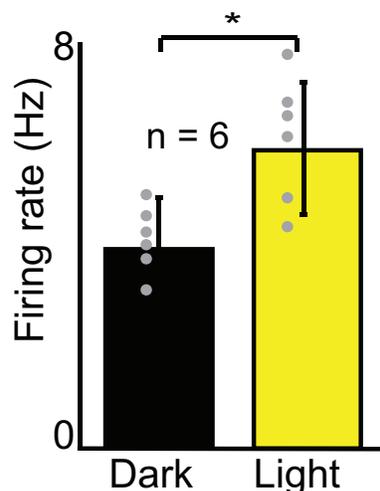


Figure 4.14: **V1 firing rates are sensitive to visual context.** Mean spike rates for all V1 units when the light conditions were changed midsession. N = 6 rats. Paired t-test between light and dark conditions:  $P = 0.0314$ .

## Learning to modulate V1 neurons is associated with changes in neural dynamics

Next, we analyzed the neural dynamics associated with learning goal-directed modulations in V1. Correlations between E1 units, whose combined positive activity modulations moved the cursor in the rewarded direction, significantly increased over the course of the session ( $N = 10$  animals, normalized change 4.42,  $P = 0.042$ ) (Figure 4.15AB), suggesting that training resulted in increased coordination between these units. No such change was observed between E2 units or between E1 and E2 units. In 72 out of 102 sessions, the relationship between performance and the E1 unit correlation was positive (mean Pearson correlation coefficient = 0.187). Of these sessions, 55.72% exhibited a significant ( $P < 0.05$ ) correlation. An example session is shown in (Figure 4.15C). We also observed that the correlation between E1 units was significantly greater in a 1-second window prior to rewarded target hits, relative to unrewarded targets or timeouts (Figure 4.15D). These data suggest that coordination between E1 units was important for success.

Interestingly, in the late learning phase, we observed an increase in LFP power in V1 prior to rewarded target hits (Figure 4.15E). Similar changes in ongoing oscillatory activity have previously been associated with top-down processing in visual cortices (Engel et al., 2001), which is one potential mechanism by which animals may be performing the task. To further explore this possibility, we then calculated the coherence between spikes in direct (combined E1 and E2) units and local field potentials (LFP) in V1, time locked to rewarded targets. Previous reports have found that attention alters alpha-band (approx. 5-15 Hz) coherence in the deep layers of visual cortex (Buffalo et al., 2011). We found that the alpha-band spike-field coherence (SFC) of direct units increased from early to late phases of learning (Figure 4.15GH).

This effect was stronger for E1 than for E2 units (Figure 4.16). This increase was only observed during task performance but not when animals were engaged in passive behavior (Figure 4.16F). Indirect neurons did not show this effect (Figure 4.16G), suggesting that these learning-related dynamics were specific to units directly involved in cursor control. However, a relatively constant fraction of indirect neurons in each training session did show predictive power for target choice (Figure 4.20EF), suggesting that there is a broader network of neurons in V1 that may have supported learning in the direct units.

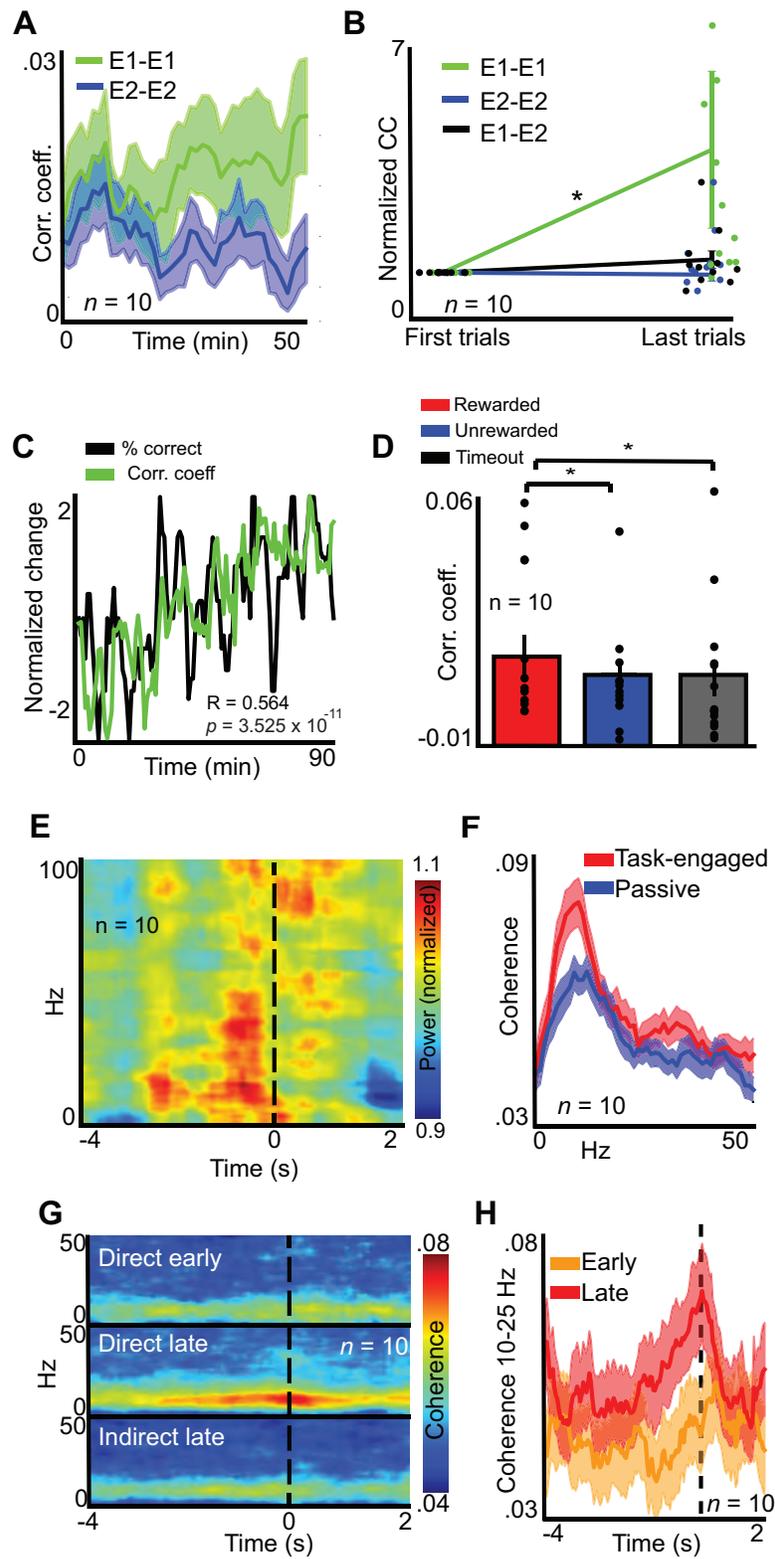


Figure 4.15: Changes in neural dynamics over learning.

Figure 4.15: **Changes in neural dynamics over learning.** **(A)** Mean pairwise correlations between units within Ensemble 1 (green) or within Ensemble 2 (blue) during training. Correlations were calculated using 1 ms bins. **(B)** Change in normalized correlation coefficients (cc) from the first trials to the last trials within sessions. CC calculated between units within Ensemble 1 (mean change = 4.418;  $P = 0.0417$ ; paired t-test), within Ensemble 2 (mean change = 0.942;  $P = 0.775$ ; paired t-test), or between units in ensembles 1 and 2 (mean change = 1.361;  $P = 0.491$ ; paired t-test).  $N = 10$  rats; mean of 4.4 sessions per rat (range 3-6). **(C)** Example data from one session showing the relationship between changes in correlation of Ensemble 1 units (green) and task performance (percent of correct trials, black). **(D)** Mean correlation between Ensemble 1 units during a 2 second window prior to target hits or timeouts. Mean cc prior to rewarded targets = 0.0285; mean cc prior to unrewarded targets = 0.0241 ( $P = 0.00167$ ; paired t-test). Mean cc prior to timeouts: 0.0240. Comparison between rewarded targets and timeouts:  $P = 0.0362$ ; comparison between unrewarded targets and timeouts:  $P = 0.549$ . **(E)** Mean spectrogram of V1 LFP time-locked to rewarded target hits. **(F)** Spike-field coherence between direct units and V1 LFP for late learning periods during task performance (red) compared to nonengaged passive behavior (blue). Shaded area represents s.e.m. **(G)** Spike-field coherograms showing the evolution of coherence over time for LFPs in V1 and spikes from direct (combined Ensemble 1 and Ensemble 2) units in during early training periods (days 1-3; top plot), late training periods (last 3 days, middle plot), and indirect units (no direct relationship to cursor control) for late periods (bottom plot) time-locked to rewarded targets. **(H)** Coherence between direct units and V1-LFP in the 10-25 Hz band for early training compared to late training. Solid lines show the mean for 10 animals and shaded areas represent s.e.m.

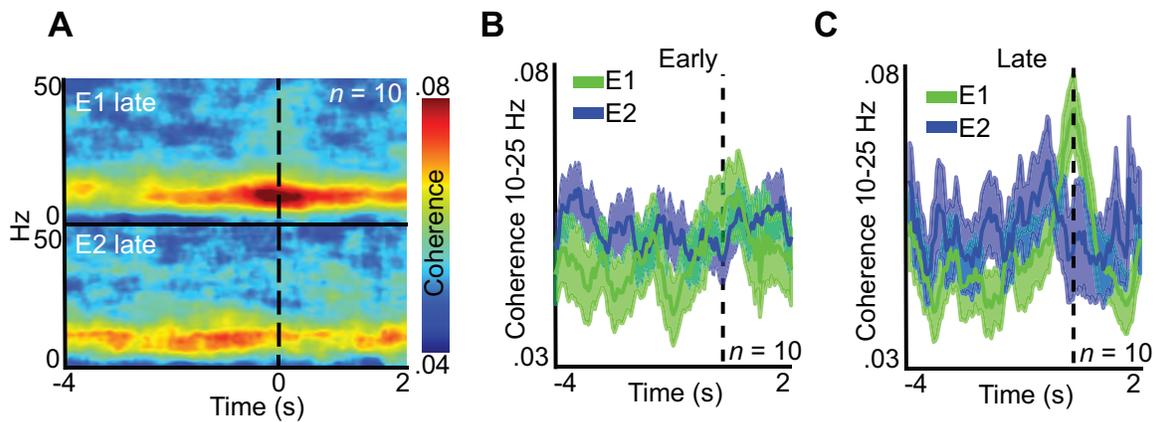


Figure 4.16: **Comparing E1 and E2 neural dynamics over learning.** (A) Comparison between spike-field coherence between E1 or E2 units and V1 LFP in late training. (B) Spike-field coherence between E1 or E2 units time-locked to rewarded targets and isolated between 10 and 25 Hz for early training periods. (C) Coherence as in (B) for late training periods.

### **Striatal activity becomes task-related with learning**

To address whether the dorso-medial striatum (DMS) plays a role in V1-based reinforcement learning, we next examined whether activity in an area of DMS that receives input from V1 changed with learning. Electrical stimulation of V1 produced a reliable spike response recorded in DMS with a delay of approximately 6 ms, suggesting a direct projection as shown in previous reports (Allen Institute for Brain Science, 2015; Faull et al., 1986; Hintiryan et al., 2016; Khibnik et al., 2014; McGeorge and Faull, 1989) (Figure 4.17AB). Conversely, stimulation of DMS while recording in V1 did not produce an observable response (not shown). In late learning, DMS units exhibited a strong modulation time-locked to the rewarded target that was not present in the early phase (Figure 4.17CD). This was accompanied by increases in beta and gamma LFP power over the same time interval (Figure 4.17F). A linear regression analysis revealed that over the course of training days, population activity of recorded DMS units became increasingly predictive of direct unit activity in V1 (Figure 4.17E), suggesting that extended training of V1 activity increasingly recruited the striatum. In accordance with this observation, field-field coherence between V1 LFP and DMS LFP was also increased in late learning around the time of target achievement (Figure 4.17GH), suggesting a possible role for the striatum in learning to produce rewarded modulations in V1.

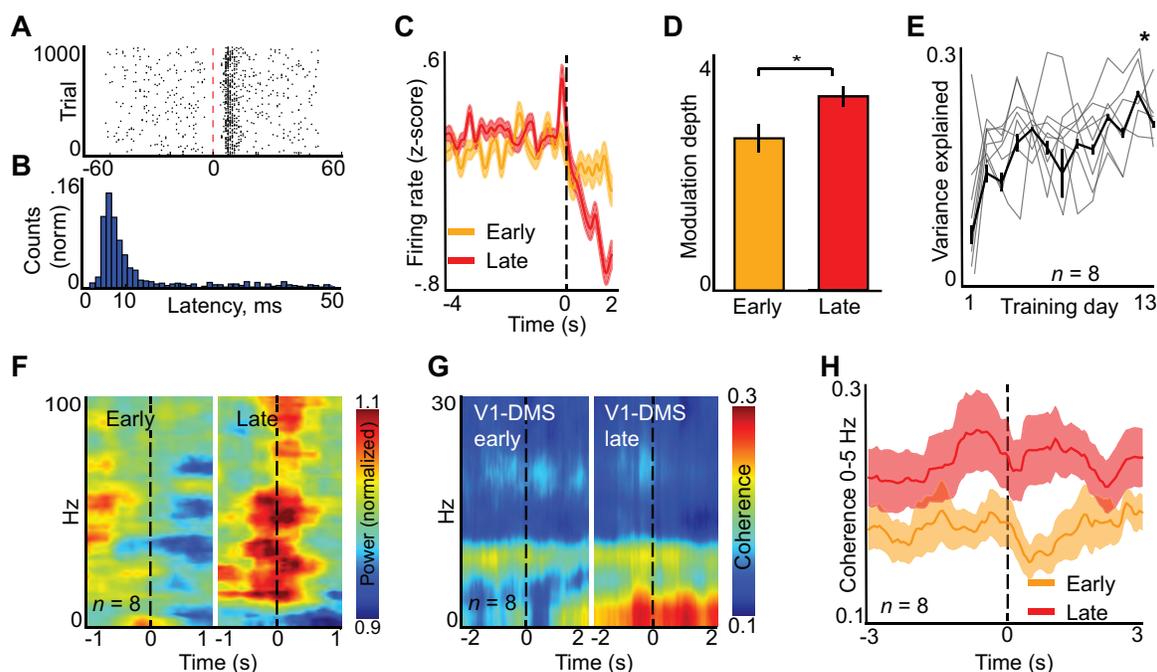


Figure 4.17: **Striatal activity becomes task-related with learning.** (A) Raster plot of an example DMS unit time-locked to V1 ICMS. (B) histogram of evoked spike counts from unit shown in (A). (C) Mean z-scored firing rate of units recorded in the dorsal medial striatum (DMS) time-locked to rewarded targets for early compared to late learning. (D) Modulation depth of DMS units in a 1-second window surrounding rewarded target hits. Mean for early periods = 2.872; mean for late periods = 3.691;  $P = 0.00015$ . (E) Proportion of variance of direct unit spikes in V1 explained by DMS unit spikes in a 500 ms window prior to rewarded target hits, computed using a linear regression analysis on individual training days. Solid black lines show mean and s.e.m.; grey lines show data for individual animals. Mean of first two days = 0.0788; mean of last 2 days = 0.276;  $P = 0.0196$  (paired ttest). (F) Mean spectrogram of DMS LFP time-locked to rewarded target hits for all animals during the early period (sessions 1-3), left), compared to the late period (last 3 sessions, right). (G) Field-field coherograms between V1 LFP and DMS LFP for early (left) compared to late training (right) sessions time-locked to rewarded targets. (H) Field-field coherence between V1 and DMS in the 0-5 Hz band in for early and late training, time locked to rewarded targets. Solid lines show mean for all animals; shaded areas represent s.e.m.

### **Striatal activity is critical for learning to modulate V1 activity**

Next, we asked whether DMS activity was necessary for learning to volitionally modulate V1 neural activity. Mice were injected bilaterally with the red-shifted inhibitory opsin Jaws (rAAV8/CamKII-Jaws-KGC-GFP-ER2) (Chuong et al., 2014) into the area of DMS which receives input from V1, and implanted chronically with optical fibers targeting DMS, and with recording electrodes in L5 of V1 (Figure 4.18A). Red light stimulation through the optical fiber decreased activity in optogenetically-identified DMS units (Figure 4.4), but had no effect on spike rates in V1 (not shown). Mice were trained on the same task as rats as described above and in (Figure 4.5). During the first 7 days of training, Jaws-injected mice and GFP-injected controls randomly received red light stimulation on 50% of trials for the whole duration of the trial (see Methods; LED 50, Figure 4.18A, bottom). Control GFP animals were able to learn the task and improved with several days of training (N = 4 animals, mean of first 3 days = 39.7%, mean of last 3 days = 69.7%, P = 0.00072) while Jaws animals did not learn (N = 4 animals, mean first 3 days LED 50 = 29.3%, mean of last 3 days LED 50 = 33.3%, P = 0.183) (Figure 4.18CD). This appeared to be a deficit in acquisition of the task, and not just poor performance in LED on trials, as no difference in success rate was observed between performance during LED on trials compared to LED off trials in any of these sessions (Figure 4.18B). However, Jaws animals were able to learn the task and improve over the course of 5 subsequent training days if no LED stimulation was provided and activity DMS was unimpaired (mean performance after LED off training = 62.1%, P = 0.00131), (Figure 4.18CD). Interestingly, V1 LFP power in the gamma band (25-60Hz) during sessions with Jaws inhibition was markedly reduced around the time of target hits, while gamma power during LED-off sessions in which animals successfully learned the task was similar to that observed in rats after learning (Figure 4.18EF, 4.19A). Together, these results suggest that inhibition of DMS activity prevents animals from learning to generate arbitrary patterns of V1 activity in order to obtain a desirable outcome.

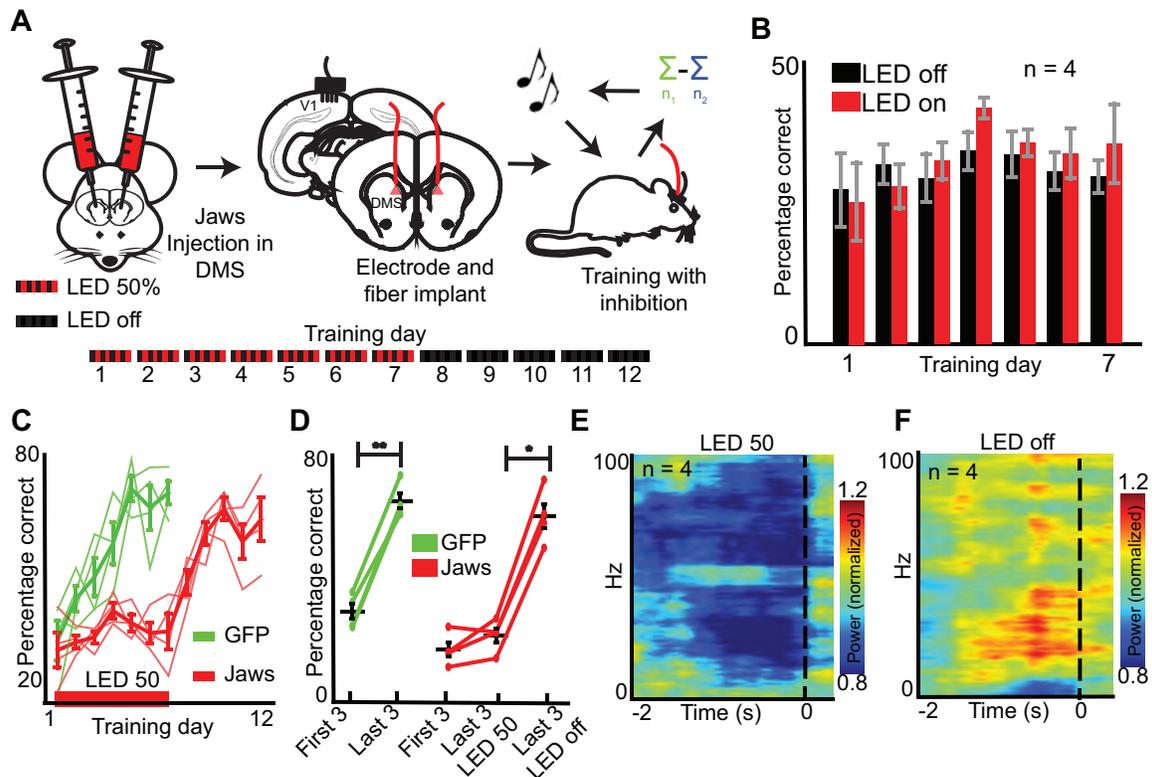


Figure 4.18: **Striatal inhibition impairs learning.** (A) Schematic showing the replication of rat experiments in Jaws-injected mice. Left: mice were injected bilaterally with rAAV8/CamKII-Jaws-KGC-GFP-ER2. Middle: After Jaws expression stabilized, mice were implanted unilaterally with an electrode array in V1 L5 and bilaterally in DMS. Right: animals were trained on the same task as rats (see Figure 4.5) with the addition of optogenetic inhibition. Bottom: timecourse of experiments. LED 50% indicates that Jaws was activated via red LED light in 50% of trials. LED off indicates that Jaws was not activated. (B) Success rates for LED-on VS LED-off trials in Jaws-injected animals across training sessions. (C) Percentage of rewarded trials for Jaws compared to control GFP mice over the course of several days of training. Red bar indicates sessions where the LED was active on 50% of trials for both groups (LED 50). (D) Quantification of performance across days for Jaws-injected animals (red; N = 4) and GFP injected animals (green, N = 4). Each time period is a mean over 3 sessions. Mean GFP, first 3 sessions = 39.7%, Mean GFP, last 3 sessions = 69.7%; P = 0.00072 (paired t-test). Mean Jaws, first 3 sessions = 29.3%, mean Jaws, last 3 sessions LED 50 = 33.3%; P = 0.183. Mean jaws, last 3 sessions LED off = 62.1%. Paired t-test between Jaws last 3 sessions LED 50 and Jaws LED off: P = 0.00131. (E) Mean spectrogram of V1 LFP in Jaws animals time-locked to rewarded targets for LED-50 sessions (E) and LED-off sessions (F). LED-50 sessions include both trials with and without Jaws activation.

## Optogenetic inhibition of DMS does not impair learned performance

In our task, the parameters used to translate neural activity to auditory tones were recalculated each day such that at the start of every session animals were required to undergo some de novo learning in order to adapt to the new calibration. Despite this, animals were able to retain some memory of training from previous days to perform better over the course of each session as training progressed (Figure 4.10A; comparison between early and late phases). Interestingly, we observed that returning Jaws animals to an LED on 50% condition after several successful days of LED off training impaired their ability to learn during that session, even though they performed well above chance during the previous session and during following session when the LED was not active (N = 4 animals, performance with LED off, pre = 63.3%, LED 50 = 38.9%, LED off, post = 59.6%, 4.19B).

This could suggest that striatal activity is required for animals to learn the initial parameters set at the beginning of the session; however, it could also indicate that striatal inhibition was interfering with task performance rather than acquisition. To disambiguate these two possibilities, we allowed animals to achieve saturating performance on the task within a single session (Train LED off, Figure 4.19CD), and then continued to train animals with the LED turned on in every trial for the remainder of the session (Test LED on, Figure 4.19CD).

Interestingly, inhibition of the striatum after within-session learning had taken place did not impair animals' ability to perform the task, and they continued to perform well above chance level (Figure 4.19CD), and in an indistinguishable manner from the LED off trials (training with LED off = 59.6%, testing with LED on = 64.4%;  $P = 0.657$ ). These data suggest that striatal inhibition prevents animals from learning to modulate neural patterns in V1 through instrumental conditioning, but not from executing these patterns after learning has taken place.

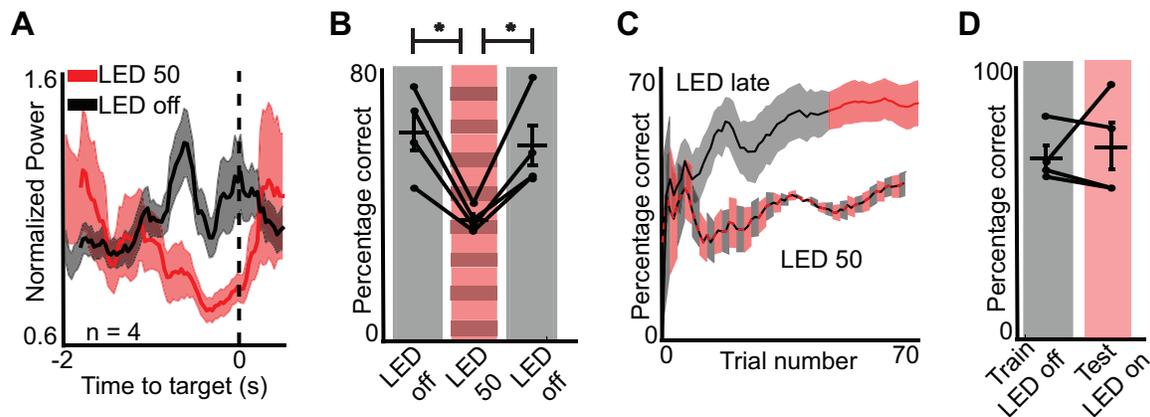


Figure 4.19: **Striatal inhibition impairs within-session learning but does not impair execution.** (A) Mean gamma power (25-60 Hz, solid lines) in V1 time-locked to rewarded targets during LED-50 vs LED-off sessions. Shaded areas show s.e.m. (B) Mean performance of trained Jaws-injected animals after several days of training. Data is plotted in the order that the training sessions occurred. Black bars indicate sessions without LED activation. Striped red bars indicate sessions where the red LED was active on 50% of trials. N = 4 animals. Mean LED off, first session = 63.3%; mean LED 50 % (second session) = 38.9%; mean LED off, last session = 59.6%. Paired t-test between LED off, first session and LED 50%:  $P = 0.0122$ . Paired t-test between LED 50% and LED off, second last session:  $P = 0.0259$ . Paired t-test between LED off, first and last sessions:  $P = 0.502$ .

## 4.5 Discussion

Here we demonstrate that animals can learn to modulate spike activity in the primary visual cortex in a goal-directed manner using an abstract virtual task. These data demonstrate that feedback-based reinforcement learning can modulate activity on the scale of a few neurons, even in a primary sensory region which is strongly driven by external sensory input. Because we observed that successful performance occurred in both the presence and absence of light, it would appear that the learned modulation of V1 units in our task is an internally-driven process. Taken along with a body of previous work describing brain-machine interface learning in other diverse cortical areas (Cerf et al., 2010; Clancy et al., 2014; Fetz, 2007; Prsa et al., 2017; Schafer and Moore, 2011; Shibata et al., 2011), these results suggest that this type of neural reinforcement ability may be a common feature that tunes the activity of cortical circuits more generally.

In the realm of motor control, the cortico-basal ganglia circuit has been hypothesized to perform a selection function in which competing cortical motor programs are either maintained or released from inhibitory control (Costa, 2011; Redgrave et al., 2011). A similar function has also been postulated to operate in the realm of abstract cognition, by which various cognitive patterns may be selected that are appropriate for the current behavioral context, and have previously led to positive outcomes (Graybiel, 1997). These models propose an inhibitory feed-forward projection from basal ganglia output nuclei (globus pallidus internal (GPi) and substantia nigra pars reticulata (SNr) that can activate cortical programs when inhibition is transiently released from the thalamus. Interestingly, although basal ganglia outputs are known to target frontal cortical areas and even higher-order visual areas like area TE in the primate (Middleton and Strick, 1994, 1996), we are not aware of any direct projections from the basal ganglia that target V1-projecting thalamic nuclei.

Despite this, we observed that activity in the striatum was necessary for neural reinforcement learning in the primary visual cortex. The effect of striatum inhibition may induce a learning deficit in a cortical region other than V1 whose input modulates Ensemble 1 and Ensemble 2 activity in the absence of visual stimulation. Frontal cortical areas, such as the cingulate cortex (Cg) in the rodent, are known to powerfully and directly influence processing in V1 to select and amplify representations of behaviorally relevant stimuli (Zhang et al., 2014). Thus, disrupting cortico-basal ganglia circuit function might impair learning of a top-down modulatory signal that generates rewarded activity in V1. Perhaps analogously, Parkinson’s patients with abnormal basal ganglia function have been shown to be impaired in voluntary and sustained control of visual attention in the absence of eye movements (Wright et al., 1990; Yamaguchi and Kobayashi, 1998). Alternatively, the learning process we observed could have recruited structures upstream of V1, such as the superior colliculus, that also form connections with the basal ganglia (McHaffie et al., 2005). Our results do not rule out these or other possibilities; rather, they simply demonstrate that the basal ganglia are necessary to learn to modulate activity in V1, and that the input for this circuit is the striatum.

In our experiments, we observed that animals performed in a goal-directed manner:

performance remained sensitive to changing task contingencies, even after many days of training (Figure 4.12). The projection of the primary visual cortex to the striatum lies along the most medial-dorsal aspect, adjacent to the ventricle wall (Khibnik et al., 2014). This lies well within the dorsomedial division of the striatum, which is known to be necessary for and to facilitate goal-directed behaviors (Yin et al., 2005, 2009), as opposed to the dorsolateral division which is required for habitual action (Redgrave et al., 2010; Wickens et al., 2007; Yin et al., 2006). One possibility is that segregation of V1 projections in the dorsomedial division of the striatum favors goal-directed learning and behavior in V1. However, it is also possible that the daily recalibration of task parameters or simply insufficient training time prevented behavior from becoming habitual. Previous work utilizing a similar task design but controlled by neurons in M1 also observed that animals behaved in a goal-directed manner (Clancy et al., 2014; Koralek et al., 2012).

From our analyses, we observed that learning-related changes in neural dynamics, such as changes in correlations and spike-field coherence (Figure 4.15) were largely restricted to the direct population consisting of units from Ensemble 1 and Ensemble 2. Absolute modulation depth of indirect (non- E1 or E2) neurons in V1 remained low relative to direct units (Figure 4.20AB), suggesting that the learning we observed operated primarily on the small scale of a few neurons. Furthermore, the modulation depth of task-irrelevant indirect neurons declined over the course of training (Figure 4.20C), echoing similar results observed across mice and monkeys using calcium imaging and electrophysiology techniques (Clancy et al., 2014; Ganguly et al., 2011; Prsa et al., 2017). It has been reported that the activity of single cells in sensory cortex is sufficient to drive a percept (Houweling and Brecht, 2008), suggesting that cortical circuits may be optimized to operate on these microscales. However, a closer analysis of indirect unit activity showed that many single units as well as the full population of indirect cells contained activity that was predictive of target choice (Figure 4.20EF). These results suggest a subtle role for this population in the learning and execution of the task; however, as we only capture a small slice of the neural population using our technique, it is difficult to conclusively say much about the relative contributions of this population as a whole. Future work will no doubt involve closer study of how nearby neural populations support learning in a particular subset of cells, perhaps by using new approaches (such as calcium imaging) that allow for stable recording of large neural populations.

An important goal of BMI research is often to decode movement parameters with high accuracy in order to translate a subjects' existing motor control repertoire into the movement of a complex artificial effector. In these cases, using high channel count recordings in motor cortices is an effective strategy due to the rich encoding of movement parameters in areas such as M1 and PMd; however, BMI tasks have successfully been implemented in other cortical regions as well (Carmena et al., 2003; Cerf et al., 2010; Clancy et al., 2014; Fetz, 2007; Hwang et al., 2013; Musallam et al., 2004; Schafer and Moore, 2011; Shibata et al., 2011). In this study, our goal was not to optimize performance or control of an effector, but rather develop a task that would enable us to study learning. The learning rates we observed for a 1-dimensional BMI task using low numbers ( $\sim 1-10$ ) of neurons in rodents were consistent with what we and others have previously reported (Arduin et al., 2013;

Clancy et al., 2014; Koralek et al., 2012; Prsa et al., 2017). It is important to note our task is different in several respects from BMI paradigms designed with the goal of maximizing performance and control in a therapeutic context, and we believe that comparing learning rates or performance metrics observed in this study with those reported across the BMI field as a whole would be misleading. Instead, our goal was to use a BMI paradigm as a method of operantly conditioning neural activity in V1 directly in order to study the learning process and the neural dynamics associated with it. This enabled us to define the final output layer of neurons directly responsible for controlling a virtual action as well as their relationship to task performance, and allowed us to observe their activity relative to each other, other V1 neurons, and activity in the dorsomedial striatum.

Although neurons in the primary visual cortex are thought to represent low-level visual features early in the visual processing stream, we observed that V1 neurons could learn to produce rewarded activity patterns that were independent of visual stimulation when spike activity was used as a control signal for a closed-loop brain-machine interface task. While here we focus on learning in the primary visual cortex, the dynamics of striatal activation, cortico-striatal dynamics over the course of learning, and the necessity of the striatum in the learning process is similar to what has been observed in a variety of tasks that engage diverse cortical regions (Barnes et al., 2005; Corbit and Janak, 2010; Koralek et al., 2012; Pasupathy and Miller, 2005; Shohamy et al., 2004; Yin et al., 2009). These results suggest that cortico-striatal projections, as part of larger cortico-basal ganglia circuits, play a generalizable role in shaping cortical activity based on ongoing experience and behavioral outcomes.

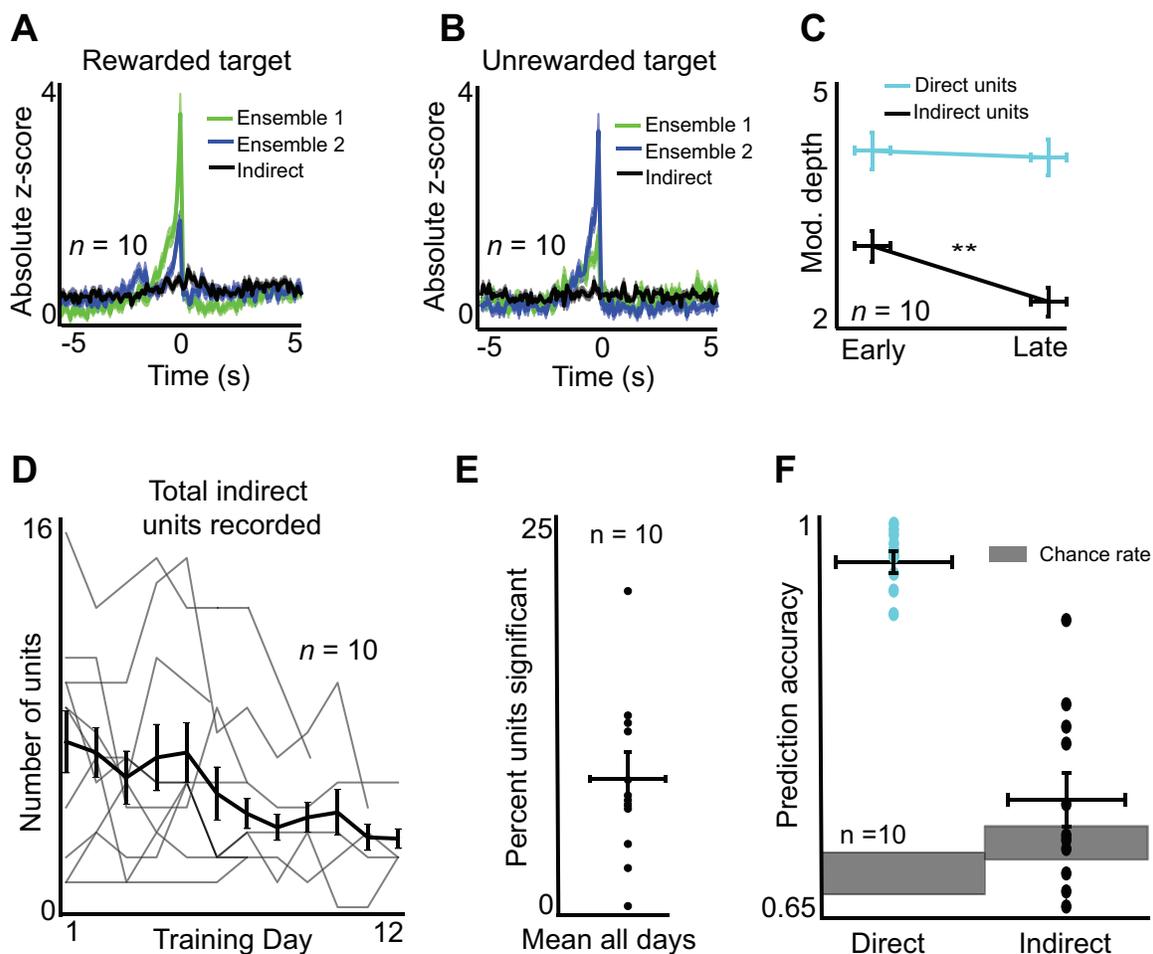


Figure 4.20: **Indirect neuron activity decreases modulation depth but predicts target choice.** (A-B) Absolute z-scored firing rates for E1, E2 and indirect units, time-locked to rewarded (A) and unrewarded (B) target hits. (C) Mean modulation depth of direct (pooled E1 and E2) units and indirect units in early (sessions 1-3) and late (sessions 8-10) training. Data taken during a 3-second window prior to rewarded target hits. Mean direct early = 3.988, mean direct late = 3.942,  $P = 0.743$  (unpaired t-test). Mean indirect early = 3.135, mean indirect late = 2.718,  $P = 0.0218$  (unpaired t-test). (D) Total number of indirect units recorded across training sessions. Black lines show mean and s.e.m.; grey lines show values for individual animals. (E) Mean percent of single indirect units with significant predictability for target choice over all training days. Significance determined by permutation test with  $P < 0.05$ . Mean = 10.782%. Black lines show mean and s.e.m. for all animals; dots show mean for each animal individually. (F) Prediction of target choice using population activity. Direct (E1 and E2) units or indirect units were used to predict target choice using a logistic regression analysis. Spike data was taken in a 500 ms window prior to target hits. Grey bars show chance accuracy rates calculated by permutation test. Mean direct = 0.964 ; mean indirect = 0.729.

# Chapter 5

## Conclusion

High performance BMIs depend on machine learning to adapt parameters for decoding neural activity, but also on the brain learning to reliably produce desired neural activity patterns. The neuroscience of how the brain explores and consolidates neural activity patterns leading to desired outcomes is not well-understood and underlies neuroprosthetic skill learning and perhaps skill learning more broadly.

### 5.1 Thesis contributions

1. **Emergence of Coordination.** We apply Factor Analysis to analyze neural coordination during BMI control by partitioning neural activity variance arising from two sources: private inputs to each neuron which drive independent, high-dimensional variance, and shared inputs which drive multiple neurons simultaneously and produce low-dimensional covariance. We found that initially, each neuron explores activity patterns independently. Over days of learning, the population's covariance increases, and a manifold emerges which aligns to the decoder. Strikingly, this low-dimensional activity drives skillful control of the decoder.
2. **Neural Law of Effect.** We find evidence for a neural law of effect, following Thorndike's law of effect, in which target neural activity patterns leading to neural reinforcement are reentered. This reinforcement results in covariance increase specifically between task-relevant neurons and the emergence of a manifold aligned to the decoder.
3. **Generalization of Neural Reinforcement to Visual Cortex** We find that the basal ganglia-dependent mechanism of neural exploration and consolidation generalizes to primary visual cortex, an area thought to primarily encode visual stimulus, suggesting a role for reinforcement learning through out cortex.

Together, these results contribute to our understanding of how the brain solves the reinforcement learning problem of learning neuroprosthetic skill, suggesting a computational role

for high-dimensional private neural variance and exploration, low-dimensional shared neural variance and consolidated control, and 1) dopaminergic midbrain and 2) striatum activity and neural reinforcement.

## 5.2 Future work

### Neural Exploration and Consolidation in Neurons Not Driving the Neuroprosthetic

In Chapter 2, we analyzed the neural variability of neurons driving the neuroprosthetic. How does exploration proceed in the surrounding recorded neurons (indirect neurons)? Do they consolidate patterns related to the neuroprosthetic? We would expect indirect neurons to initially explore, as the brain has not discovered the relevant neurons. With learning, we might expect the indirect neurons to consolidate task-related patterns, but not as prominently as neurons driving the neuroprosthetic.

### Network Credit Assignment

In Chapter 3, we observed only the neurons driving the neuroprosthetic (direct neurons) increased covariance. How does the brain perform network credit assignment to change neurons which causally drive output? Experimental work has implicated corticostriatal plasticity, reinforcement, sleep (Ganguly et al., 2011; Gulati et al., 2017; Koralek et al., 2013). Computationally, private variance (to determine individual neurons' contributions to behavioral output) and reinforcement are essential components (Hliot et al., 2010; Legenstein et al., 2010). The BMI learning paradigm which specifies the neurons driving behavioral output makes this question experimentally tractable.

### Neural Circuits Underlying Neural Reinforcement Learning

In Chapter 3, we saw that VTA activation reinforced antecedent cortical activity patterns. This is a starting point for more detailed studies evaluating the neural circuit mechanisms underlying BMI learning. Actor-critic models suggest dorsal striatum as involved in policy learning and ventral striatum in learning the value of behavioral states.

### Sensitivity of Neural Exploration to Experienced Errors

When the brain experiences large errors as during early learning, it makes sense for the brain to explore neural activity space for different activity patterns. When the brain experiences small errors, perhaps the brain does not need to explore as much. Future experiments can perturb the decoder by different amounts to determine the sensitivity of neural exploration to behavioral error.

## Laws of Neural Dynamics

In Chapter 2, we observed the emergence of low-dimensional temporal patterns which underlie skillful control. Do these temporal patterns obey dynamical rules such as those underlying natural motor control (Churchland et al., 2012; Kao et al., 2015)? Interestingly, a large fraction of neural variance is null to the decoder and thus causes no change in behavioral output. This neural activity might reflect internal neural dynamics necessary to generate control signals to the decoder.

## Comparison of BMI Learning across Brain Areas

In Chapter 3, mice learned to control a BMI with motor cortex (M1) and in Chapter 4, rats and mice learned to control a BMI with visual cortex (V1). Chapter 4 suggests reinforcement learning can proceed similarly in M1 and V1. Do analyses of neural dynamics reveal differences in learning properties across M1 and V1? How does this compare to other brain areas?

## Neural Constraints on Learning

(Sadtler et al., 2014) found that for BMI-experienced subjects learning perturbed decoders within a day, subjects can learn patterns within their existing manifold of covariance and can not learn patterns outside. In contrast, in Chapter 2, we saw that during de novo BMI learning, each neuron can explore independently, suggesting the brain possesses flexibility in acquiring neural patterns. Future experiments can test whether periods of large private variance indeed permit the brain to learn to control a broad range of decoders. In addition, future experiments can address constraints on the temporal sequence of activity neural populations can acquire.

## 2-learner BMI Systems

The basic science and engineering principles of designing a 2-learner BMI system in which the machine and brain synergistically learn are not well understood. This work has revealed neural signals related to exploration (private variance) and consolidated control (shared variance). 2-learner BMI systems could adaptively weight the balance of neural exploration and consolidated activity on the prosthetic output. If the neural population exhibits large private variance, the decoder could permit private variance into the decoder to permit behavioral exploration and facilitate learning. If the neural population exhibits large shared variance, the decoder could denoise population activity by only using shared variance as input to the decoder. This proposal is one possible way of enhancing the synergy of machine and brain learning.

# References

Aflalo, T., Kellis, S., Klaes, C., Lee, B., Shi, Y., Pejsa, K., Shanfield, K., Hayes-Jackson, S., Aisen, M., Heck, C., et al. (2015). Decoding motor imagery from the posterior parietal cortex of a tetraplegic human. *Science* (80-. ). 348, 906910.

Ajiboye, A.B., Willett, F.R., Young, D.R., Memberg, W.D., Murphy, B.A., Miller, J.P., Walter, B.L., Sweet, J.A., Hoyen, H.A., Keith, M.W., et al. (2017). Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: a proof-of-concept demonstration. *Lancet* 389, 18211830.

Allen Institute for Brain Science (2015). Allen Mouse Brain Atlas. Allen Mouse Brain Atlas.

Anderson, K.D. (2004). Targeting recovery: priorities of the spinal cord-injured population. *J. Neurotrauma* 21, 13711383.

Arduin, P., Fregnac, Y., Shulz, D.E., Ego-Stengel, V.V., Fre, Y., Shulz, D.E., Frgnac, Y., Shulz, D.E., and Ego-Stengel, V.V. (2013). Master Neurons Induced by Operant Conditioning in Rat Motor Cortex during a Brain-Machine Interface Task.pdf. 33, 83088320.

Athalye, V.R., Ganguly, K., Costa, R.M., and Carmena, J.M. (2017). Emergence of Coordinated Neural Dynamics Underlies Neuroprosthetic Learning and Skillful Control. *Neuron* 93, 955970.

Awan, N.R., Lozano, A., and Hamani, C. (2009). Deep brain stimulation: current and future perspectives. *Neurosurg. Focus* 27, E2.

Barnes, T.D., Kubota, Y., Hu, D., Jin, D.Z., and Graybiel, A.M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature* 437, 11581161.

Bensmaia, S.J., and Miller, L.E. (2014). Restoring sensorimotor function through intracortical interfaces: progress and looming challenges. *Nat. Rev. Neurosci.* 15, 313325.

Bouton, C.E., Shaikhouni, A., Annetta, N. V., Bockbrader, M.A., Friedenberg, D.A., Nielson, D.M., Sharma, G., Sederberg, P.B., Glenn, B.C., Mysiw, W.J., et al. (2016). Restoring cortical control of functional movement in a human with quadriplegia. *Nature* 0, 113.

Brandman, D.M., Cash, S.S., and Hochberg, L.R. (2017). Review: Human Intracortical recording and neural decoding for brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 4320, 11.

- Briggman, K.L., and Kristan, W.B. (2008). Multifunctional Pattern-Generating Circuits. *Annu. Rev. Neurosci.* 31, 271294.
- Brown, L.L., Schneider, J.S., and Lidsky, T.I. (1997). Sensory and cognitive functions of the basal ganglia. *Curr. Opin. Neurobiol.* 7, 157163.
- Cao, V.Y., Ye, Y., Mastwal, S., Ren, M., Coon, M., Liu, Q., Costa, R.M., and Wang, K.H. (2015). Motor Learning Consolidates Arc-Expressing Neuronal Ensembles in Secondary Motor Cortex. *Neuron* 86, 13851392.
- Carmena, J.M., Lebedev, M.A., Crist, R.E., O'Doherty, J.E., Santucci, D.M., Dimitrov, D.F., Patil, P.G., Henriquez, C.S., and Nicolelis, M.A.L. (2003). Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol.* 1, 193208.
- Cerf, M., Thiruvengadam, N., Mormann, F., Kraskov, A., Quiroga, R.Q., Koch, C., and Fried, I. (2010). On-line, voluntary control of human temporal lobe neurons. *Nature* 467, 11041108.
- Chapin, J.K., Moxon, K. a, Markowitz, R.S., and Nicolelis, M. a (1999). Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nat. Neurosci.* 2, 664670.
- Chase, S.M., Kass, R.E., Schwartz, A.B., Sm, C., Re, K., and Behavioral, S.A.B. (2012). Behavioral and neural correlates of visuomotor adaptation observed through a brain-computer interface in primary motor cortex. 624644.
- Chuong, A.S., Miri, M.L., Busskamp, V., Matthews, G.A.C., Acker, L.C., Srensen, A.T., Young, A., Klapoetke, N.C., Henninger, M.A., Kodandaramaiah, S.B., et al. (2014). Non-invasive optical inhibition with a red-shifted microbial rhodopsin. *Nat. Neurosci.* 17, 11231129.
- Churchland, M.M., Yu, B.M., Cunningham, J.P., Sugrue, L.P., Cohen, M.R., Corrado, G.S., Newsome, W.T., Clark, A.M., Hosseini, P., Scott, B.B., et al. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat. Neurosci.* 13, 369378.
- Churchland, M.M., Cunningham, J.P., Kaufman, M.T., Foster, J.D., Nuyujukian, P., Ryu, S.I., and Shenoy, K. V (2012). Neural Population Dynamics During Reaching. *Nature* 487, 120.
- Clancy, K.B., Koralek, A.C., Costa, R.M., Feldman, D.E., and Carmena, J.M. (2014). Volitional modulation of optically recorded calcium signals during neuroprosthetic learning. *Nat. Neurosci.* 17, 807809.
- Clark, G.M., Clark, J.C.M., and Furness, J.B. (2013). The Evolving Science of Cochlear Implants. *JAMA* 310, 12251226.
- Cohen, R.G., and Sternad, D. (2009). Variability in motor learning: relocating , channeling and reducing noise. *Exp. Brain Res.* 193, 6983.
- Collinger, J.L.J., Wodlinger, B., Downey, J.E.J., Wang, W., Tyler-Kabara, E.C., Weber, D.J., McMorland, A.J.C., Velliste, M., Boninger, M.L., and Schwartz, A.B. (2013). High-performance neuroprosthetic control by an individual with tetraplegia. *Lancet* 381, 557564.
- Corbett, D., and Wise, R.A. (1980). Intracranial self-stimulation in relation to the ascending dopaminergic systems of the midbrain: a moveable electrode mapping study. *Brain Res.* 185, 115.

- Corbit, L.H., and Janak, P.H. (2010). Posterior dorsomedial striatum is critical for both selective instrumental and Pavlovian reward learning. *Eur. J. Neurosci.* 31, 13121321.
- Costa, R.M. (2011). A selectionist account of de novo action learning. *Curr. Opin. Neurobiol.* 21, 579586.
- Costa, R.M., Cohen, D., and Nicoletis, M.A.L. (2004). Differential Corticostriatal Plasticity during Fast and Slow Motor Skill Learning in Mice. *Curr. Biol.* 14, 11241134.
- Cragg, B.G. (1967). The density of synapses and neurones in the motor and visual areas of the cerebral cortex. *J. Anat.* 101, 639654.
- Dangi, S., Orsborn, A.L., Moorman, H.G., and Carmena, J.M. (2013). Design and Analysis of Closed-Loop Decoder Adaptation Algorithms for Brain-Machine Interfaces. *Neural Comput.* 25, 16931731.
- Dangi, S., Gowda, S., Moorman, H.G., Orsborn, A.L., So, K., Shanechi, M.M., and Carmena, J.M. (2014). Continuous closed-loop decoder adaptation with a recursive maximum likelihood algorithm allows for rapid performance acquisition in brain-machine interfaces. *Neural Comput.* 26, 18111839.
- Eaton, R.W., Libey, T., and Fetz, E.E. (2016). Operant conditioning of neural activity in freely behaving monkeys with intracranial reinforcement. *J. Neurophysiol.* jn.00423.2016.
- Everitt, B.S. (1984). *An Introduction to Latent Variable Models* (London: Chapman & Hall).
- Faisal, A.A., Selen, L.P.J., and Wolpert, D.M. (2008). Noise in the nervous system. *Nat. Rev. Neurosci.* 9, 292303.
- Faull, R.L.M., Nauta, W.J.H., and Domesick, V.B. (1986). The visual cortico-striato-nigral pathway in the rat. *Neuroscience* 19, 11191132.
- Fetz, E.E. (1969). Operant conditioning of cortical unit activity. *Science* (80-. ). 163, 955958.
- Fetz, E.E. (2007). Volitional control of neural activity: implications for braincomputer interfaces. *J. Physiol.* 579, 571579.
- Fetz, E.E., and Baker, M. a (1973). Operantly conditioned patterns on precentral unit activity and correlated responses in adjacent cells and contralateral muscles. *J. Neurophysiol.* 36, 179204.
- Fetz, E.E., and Finocchio, D. V. (1971). Operant Conditioning of Specific Patterns of Neural and Muscular Activity. *Science* (80-. ). 174, 431435.
- Fu, M., Yu, X., Lu, J., and Zuo, Y. (2012). Repetitive motor learning induces coordinated formation of clustered dendritic spines in vivo. *Nature* 483, 9295.
- Ganguly, K., and Carmena, J.M. (2009). Emergence of a stable cortical map for neuroprosthetic control. *PLoS Biol.* 7.
- Ganguly, K., and Carmena, J.M. (2010). Neural correlates of skill acquisition with a cortical brain-machine interface. *J. Mot. Behav.* 42, 355360.
- Ganguly, K., Dimitrov, D.F., Wallis, J.D., and Carmena, J.M. (2011). Reversible large-scale modification of cortical networks during neuroprosthetic control. *Nat. Neurosci.* 14, 662667.
- Georgopoulos, A.P., Schwartz, A.B., and Kettner, R.E. (1986). Neuronal population coding of movement direction. *Science* (80-. ). 233, 14161419.

- Gilja, V., Nuyujukian, P., Chestek, C.A., Cunningham, J.P., Yu, B.M., Fan, J.M., Churchland, M.M., Kaufman, M.T., Kao, J.C., Ryu, S.I., et al. (2012). A high-performance neural prosthesis enabled by control algorithm design. *Nat. Neurosci.* 15, 17521757.
- Gilja, V., Pandarinath, C., Blabe, C.H., Nuyujukian, P., Simeral, J.D., Sarma, A. a, Sorice, B.L., Perge, J. a, Jarosiewicz, B., Hochberg, L.R., et al. (2015). Clinical translation of a high-performance neural prosthesis. *Nat. Med.* 21, 68.
- Golub, M.D., Yu, B.M., and Chase, S.M. (2015). Internal models for interpreting neural population activity during sensorimotor control. *Elife* 4, e10015.
- Gong, S., Doughty, M., Harbaugh, C.R., Cummins, A., Hatten, M.E., Heintz, N., and Gerfen, C.R. (2007). Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *J. Neurosci.* 27, 98179823.
- Graybiel, A.M. (1997). The basal ganglia and cognitive pattern generators. *Schizophr. Bull.* 23, 459469.
- Graybiel, A.M. (2008). Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.* 31, 359387.
- Grosenick, L., Marshel, J.H., and Deisseroth, K. (2015). Closed-loop and activity-guided optogenetic control. *Neuron* 86, 106139.
- Gulati, T., Ramanathan, D.S., Wong, C.C., and Ganguly, K. (2014). Reactivation of emergent task-related ensembles during slow-wave sleep after neuroprosthetic learning. *Nat. Neurosci.* 17, 11071113.
- Gulati, T., Guo, L., Ramanathan, D.S., Bodepudi, A., and Ganguly, K. (2017). Neural reactivations during sleep determine network credit assignment. *Nat. Neurosci.* 2017, 35.
- Hayashi-Takagi, A., Yagishita, S., Nakamura, M., Shirai, F., Wu, Y.I., Loshbaugh, A.L., Kuhlman, B., Hahn, K.M., and Kasai, H. (2015). Labelling and optical erasure of synaptic memory traces in the motor cortex. *Nature* 525, 333338.
- Hliot, R., Ganguly, K., Jimenez, J., and Carmena, J.M. (2010). Learning in Closed-Loop Brain Machine Interfaces: Modeling and Experimental Validation. *IEEE Trans. Syst. Man, Cybern. B Cybern.* 40, 13871397.
- Hennequin, G., Vogels, T.P., and Gerstner, W. (2014). Optimal control of transient dynamics in balanced networks supports generation of complex movements. *Neuron* 82, 13941406.
- Hikosaka, O., Nakahara, H., Rand, M.K., Sakai, K., Lu, X., Nakamura, K., Miyachi, S., and Doya, K. (1999). Parallel neural networks for learning sequential procedures. *Trends Neurosci.* 22, 464471.
- Hinterberger, T., Veit, R., Wilhelm, B., Weiskopf, N., Vatine, J.J., and Birbaumer, N. (2005). Neuronal mechanisms underlying control of a brain-computer interface. *Eur. J. Neurosci.* 21, 31693181.
- Hintiryan, H., Foster, N.N., Bowman, I., Bay, M., Song, M.Y., Gou, L., Yamashita, S., Bienkowski, M.S., Zingg, B., Zhu, M., et al. (2016). The mouse cortico-striatal projectome. *Nat. Neurosci.* 19, 11001114.

- Hochberg, L.R., and Anderson, K.D. (2012). BCI Users and Their Needs. In *Brain-Computer Interfaces: Principles and Practice*, J. Wolpaw, and W. EW, eds. (Oxford University Press)
- Hochberg, L.R., and Cochrane, T. (2013). Implanted Neural Interfaces. In *Neuroethics in Practice*, pp. 235250.
- Hochberg, L.R., Serruya, M.D., Friehs, G.M., Mukand, J.A., Saleh, M., Caplan, A.H., Branner, A., Chen, D., Penn, R.D., and Donoghue, J.P. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* 442, 164171.
- Hochberg, L.R., Bacher, D., Jarosiewicz, B., Masse, N.Y., Simeral, J.D., Vogel, J., Haddadin, S., Liu, J., Cash, S.S., van der Smagt, P., et al. (2012). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature* 485, 372375.
- Howe, M.W., and Dombeck, D.A. (2016). Rapid signalling in distinct dopaminergic axons during locomotion and reward. *Nature* 535, 505510.
- Humphrey, D.R., Schmidt, E.M., and Thompson, W.D. (1970). Predicting Measures of Motor Performance from Multiple Cortical Spike Trains. 170, 758762.
- Husain, M., Shapiro, K., Martin, J., and Kennard, C. (1997). Abnormal temporal dynamics of visual attention in spatial neglect patients. *Nature* 385, 154156. Hwang, E.J., Bailey, P.M., and Andersen, R.A. (2013). Volitional control of neural activity relies on the natural motor repertoire. *Curr. Biol.* 23, 353361.
- Jarosiewicz, B., Chase, S.M., Fraser, G.W., Velliste, M., Kass, R.E., and Schwartz, A.B. (2008). Functional network reorganization during learning in a brain-computer interface paradigm. *Proc. Natl. Acad. Sci.* 105, 1948619491.
- Jarosiewicz, B., Sarma, A.A., Bacher, D., Masse, N.Y., Simeral, J.D., Sorice, B., Oakley, E.M., Blabe, C., Pandarinath, C., Gilja, V., et al. (2015). Virtual typing by people with tetraplegia using a self-calibrating intracortical brain-computer interface. *Sci. Transl. Med.* 7, 111.
- Kao, J.C., Nuyujukian, P., Ryu, S.I., Churchland, M.M., Cunningham, J.P., and Shenoy, K. V (2015). Single-trial dynamics of motor cortex and their applications to brain-machine interfaces. *Nat. Commun.* 6, 7759.
- Kao, M.H., Doupe, A.J., and Brainard, M.S. (2005). Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature* 433, 638643. Kaufman, M.T., Churchland, M.M., Ryu, S.I., and Shenoy, K. V (2014). Cortical activity in the null space: permitting preparation without movement. *Nat. Neurosci.* 17, 440448.
- Keller, G.B., Bonhoeffer, T., and Hbener, M. (2012). Sensorimotor Mismatch Signals in Primary Visual Cortex of the Behaving Mouse. *Neuron* 74, 809815. Kemp, J.M., and Powell, T.P.S. (1970). The cortico-striate projection in the monkey. *Brain* 93.
- Kennedy, P.R., and Bakay, R.A.E. (1998). Restoration of neural output from a paralyzed patient by a direct brain connection. *Neuroreport* 9, 17071711. Kennedy, P.R., Bakay, R. a, Moore, M.M., Adams, K., and Goldwithe, J. (2000). Direct control of a computer from the human central nervous system. *IEEE Trans. Rehabil. Eng.* 8, 198202.
- Khibnik, L.A., Tritesch, N.X., and Sabatini, B.L. (2014). A direct projection from mouse primary visual cortex to dorsomedial striatum. *PLoS One* 9.

- Kim, S.-P., Simeral, J.D., Hochberg, L.R., Donoghue, J.P., and Black, M.J. (2008). Neural control of computer cursor velocity by decoding motor cortical spiking activity in humans with tetraplegia. *J. Neural Eng.* 5, 455476.
- Koralek, A.C., Jin, X., Long II, J.D., Costa, R.M., and Carmena, J.M. (2012). Corticostriatal plasticity is necessary for learning intentional neuroprosthetic skills. *Nature* 483, 331335.
- Koralek, A.C., Costa, R.M., and Carmena, J.M. (2013). Temporally Precise Cell-Specific Coherence Develops in Corticostriatal Networks during Learning. *Neuron* 79, 865872.
- Kral, A., and Sharma, A. (2012). Developmental neuroplasticity after cochlear implantation. *Trends Neurosci.* 35, 111122.
- Legenstein, R., Chase, S.M., Schwartz, A.B., and Maass, W. (2010). A Reward-Modulated Hebbian Learning Rule Can Explain Experimentally Observed Network Reorganization in a Brain Control Task. *J. Neurosci.* 30, 84008410.
- Leuthardt, E.C., Schalk, G., Moran, D., and Ojemann, J.G. (2006). The emerging world of motor neuroprosthetics: A neurosurgical perspective. *Neurosurgery* 59, 113.
- Lillicrap, T.P., and Scott, S.H. (2013). Preference Distributions of Primary Motor Cortex Neurons Reflect Control Solutions Optimized for Limb Biomechanics. *Neuron* 77, 168179.
- Mahmoudi, B., and Sanchez, J.C. (2011). A symbiotic brain-machine interface through value-based decision making. *PLoS One* 6.
- Makino, H., and Komiyama, T. (2015). Learning enhances the relative impact of top-down processing in the visual cortex. *Nat. Neurosci.* 18, 11161122.
- Mandelblat-Cerf, Y., Paz, R., and Vaadia, E. (2009). Trial-to-trial variability of single cells in motor cortices is dynamically modified during visuomotor adaptation. *J. Neurosci.* 29, 1505315062.
- Martinez, A., Anllo-Vento, L., and Sereno, M. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nat. Neurosci.* 2, 364369.
- McGeorge, A.J., and Faull, R.L.M. (1989). The organization of the projection from the cerebral cortex to the striatum in the rat. *Neuroscience* 29, 503537.
- McHaffie, J.G., Stanford, T.R., Stein, B.E., Coizet, V., and Redgrave, P. (2005). Subcortical loops through the basal ganglia. *Trends Neurosci.* 28, 401407.
- Mercuri, E., Atkinson, J., Braddick, O., Anker, S., Cowan, F., Rutherford, M., Pennock, J., and Dubowitz, L. (1997). Basal ganglia damage and impaired visual function in the newborn infant. *Arch Dis Child Fetal Neonatal Ed* 77, F111-4.
- Merzenich, M.M., Byers, C.L., White, M., and Vivion, M.C. (1980). Cochlear implant prostheses: strategies and progress. *Ann. Biomed. Eng.* 8, 361368.
- Middleton, F.A., and Strick, P.L. (1994). Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. *Science* (80-. ). 266, 458461.
- Middleton, F.A., and Strick, P.L. (1996). The temporal lobe is a target of output from the basal ganglia. *Proc Natl Acad Sci U S A* 93, 86838687.
- Musallam, S., Corneil, B.D., Greger, B., Scherberger, H., and Andersen, R.A. (2004). Cognitive Control Signals for Neural Prosthetics. *Science* (80-. ). 305, 258263.

- Naselaris, T., Kay, K.N., Nishimoto, S., and Gallant, J.L. (2011). Encoding and decoding in fMRI. *Neuroimage* 56, 400410.
- Niell, C.M.C.C.M., and Stryker, M.P.M. (2010). Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. *Neuron* 65, 472479.
- ODoherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R.J. (2004). Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning. *Science* (80-). 304, 452454.
- Olds, J., and Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J. Comp. Physiol. Psychol.* 47, 419427.
- Olveczky, B.P., Otchy, T.M., Goldberg, J.H., Aronov, D., and Fee, M.S. (2011). Changes in the neural control of a complex motor sequence during learning. *J. Neurophysiol.* 106, 386397.
- Orsborn, A.L., and Carmena, J.M. (2013). Creating new functional circuits for action via brain-machine interfaces. *Front. Comput. Neurosci.* 7, 157.
- Orsborn, A.L., Dangi, S., Moorman, H.G., and Carmena, J.M. (2012). Closed-loop decoder adaptation on intermediate time-scales facilitates rapid BMI performance improvements independent of decoder initialization conditions. *IEEE Trans. Neural Syst. Rehabil. Eng.* 20, 468477.
- Orsborn, A.L., Moorman, H.G., Overduin, S.A., Shanechi, M.M., Dimitrov, D.F., and Carmena, J.M. (2014). Closed-loop decoder adaptation shapes neural plasticity for skillful neuroprosthetic control. *Neuron* 82, 13801393.
- Pandarath, C., Nuyujukian, P., Blabe, C.H., Sorice, B.L., Saab, J., Willett, F.R., Hochberg, L.R., Shenoy, K. V., and Henderson, J.M. (2017). High performance communication by people with paralysis using an intracortical brain-computer interface. *Elife* 6, 127.
- Pasupathy, A., and Miller, E.K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature* 433, 873876.
- Perlmutter, J.S., and Mink, J.W. (2006). Deep Brain Stimulation. *Annu. Rev. Neurosci.* 29, 229257.
- Peters, A.J., Chen, S.X., and Komiyama, T. (2014). Emergence of reproducible spatiotemporal activity during motor learning. *Nature* 510, 263267.
- Prsa, M., Galinanes, G.L., and Huber, D. (2017). Rapid Integration of Artificial Sensory Feedback during Operant Conditioning of Motor Cortex Neurons. *Neuron* 93, 929939.e6.
- Ramanathan, D.S., Gulati, T., and Ganguly, K. (2015). Sleep-Dependent Reactivation of Ensembles in Motor Cortex Promotes Skill Consolidation. *PLoS Biol.* 13, 125.
- Redgrave, P., Rodriguez, M., Smith, Y., Rodriguez-Oroz, M.C., Lehericy, S., Bergman, H., Agid, Y., DeLong, M.R., and Obeso, J.A. (2010). Goal-directed and habitual control in the basal ganglia: implications for Parkinsons disease. *Nat. Rev. Neurosci.* 11, 760772.
- Redgrave, P., Vautrelle, N., and Reynolds, J.N.J. (2011). Functional properties of the basal ganglia re-entrant loop architecture: selection and reinforcement. *Neuroscience* 198, 138151.

- Sadtler, P.T., Quick, K.M., Golub, M.D., Chase, S.M., Ryu, S.I., Tyler-Kabara, E.C., Yu, B.M., and Batista, A.P. (2014). Neural constraints on learning. *Nature* 512, 423426.
- Samejima, K. (2005). Representation of Action-Specific Reward Values in the Striatum. *Science* (80-. ). 310, 13371340.
- Santhanam, G., Yu, B.M., Gilja, V., Ryu, S.I., Afshar, A., Sahani, M., and Shenoy, K. V. (2009). Factor-Analysis Methods for Higher-Performance Neural Prostheses. *J. Neurophysiol.* 102, 13151330.
- Santos, F.J., Oliveira, R.F., Jin, X., and Costa, R.M. (2015). Corticostriatal dynamics encode the refinement of specific behavioral variability during skill learning. *Elife* 4, e09423.
- Schafer, R.J., and Moore, T. (2011). Selective Attention from Voluntary Control of Neurons in Prefrontal Cortex. *Science* (80-. ). 332.
- Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. *Science* (80-. ). 275, 15931599.
- Serruya, M.D., Hatsopoulos, N.G., Paninski, L., Fellows, M.R., and Donoghue, J.P. (2002). Instant neural control of a movement signal. *Nature* 416, 141142.
- Shanechi, M.M., Orsborn, A.L., and Carmena, J.M. (2016). Robust Brain-Machine Interface Design Using Optimal Feedback Control Modeling and Adaptive Point Process Filtering. *PLoS Comput. Biol.* 12, e1004730.
- Shen, W., Flajolet, M., Greengard, P., and Surmeier, D.J. (2008). Dichotomous Dopaminergic Control of Striatal Synaptic Plasticity. *Science* (80-. ). 321, 848851. Shenoy, K., and Carmena, J. (2014). Combining decoder design and neural adaptation in brain-machine interfaces. *Neuron* 84, 665680.
- Shibata, K., Watanabe, T., Sasaki, Y., and Kawato, M. (2011). Perceptual Learning Incepted by Decoded fMRI Neurofeedback Without Stimulus Presentation. *Science* (80-. ). 334, 1413 LP-1415.
- Shmuelof, L., Krakauer, J.W., and Mazzoni, P. (2012). How is a motor skill learned? Change and invariance at the levels of task success and trajectory control. *J. Neurophysiol.* 108, 578594.
- Shohamy, D., Myers, C.E., Grossman, S., Sage, J., Gluck, M.A., and Poldrack, R.A. (2004). Corticostriatal contributions to feedbackbased learning: converging data from neuroimaging and neuropsychology. *Brain* 127.
- Shuler, M.G. (2006). Reward Timing in the Primary Visual Cortex. *Science* (80-. ). 311, 16061609.
- Skinner, B.F. (1938). *The behavior of organisms: an experimental analysis* (Oxford, England: Appleton-Century).
- Skinner, B.F. (1981). Selection by Consequences. *Science* (80-. ). 213, 501504.
- So, K., Koralek, A.C., Ganguly, K., Gastpar, M.C., and Carmena, J.M. (2012a). Assessing functional connectivity of neural ensembles using directed information. *J. Neural Eng.* 9, 26004.
- So, K., Ganguly, K., Jimenez, J., Gastpar, M.C., and Carmena, J.M. (2012b). Redundant information encoding in primary motor cortex during natural and prosthetic motor control. *J. Comput. Neurosci.* 32, 555561.

- Stavisky, S.D., Kao, J.C., Ryu, S.I., and Shenoy, K. V. (2017). Motor Cortical Visuomotor Feedback Activity Is Initially Isolated from Downstream Targets in Output-Null Neural State Space Dimensions. *Neuron* 114.
- Steinberg, E.E., Keiflin, R., Boivin, J.R., Witten, I.B., Deisseroth, K., and Janak, P.H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nat Neurosci* 16, 966973.
- Steinmetz, P.N., Roy, A., Fitzgerald, P.J., Hsiao, S.S., Johnson, K.O., and Niebur, E. (2000). Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature* 404, 187190.
- Suminski, A.J., Tkach, D.C., Fagg, A.H., and Hatsopoulos, N.G. (2010). Incorporating Feedback from Multiple Sensory Modalities Enhances Brain-Machine Interface Control. *J. Neurosci.* 30, 1677716787.
- Suner, S., Fellows, M.R., Vargas-Irwin, C., Nakata, G.K., and Donoghue, J.P. (2005). Reliability of signals from a chronically implanted, silicon-based electrode array in non-human primate primary motor cortex. *IEEE Trans. Neural Syst. Rehabil. Eng.* 13, 524541.
- Sussillo, D., Churchland, M.M., Kaufman, M.T., and Shenoy, K. V (2015). A neural network that finds a naturalistic solution for the production of muscle activity. *Nat. Neurosci.* 18, 10251033.
- Sutton, R.S., and Barto, A.G. (1998). *Reinforcement Learning: An Introduction* (Cambridge: MIT Press).
- Swanson, L.W. (2000). Cerebral hemisphere regulation of motivated behavior. *Brain Res.* 886, 113164.
- Takahashi, Y., Schoenbaum, G., and Niv, Y. (2009). Silencing the critics: Understanding the effects of cocaine sensitization on dorsolateral and ventral striatum in the context of an Actor/Critic model. *Front. Neurosci.* 2, 8689.
- Taylor, D.M., Tillery, S.I.H., and Schwartz, A.B. (2002). Direct cortical control of 3D neuroprosthetic devices. *Science* (80-. ). 296, 18291832.
- Thorndike, E.L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *Psychol. Rev.* 2, 1107.
- Tricomi, E.M., Delgado, M.R., and Fiez, J.A. (2004). Modulation of Caudate Activity by Action Contingency. *Neuron* 41, 281292.
- Tsai, H.-C., Zhang, F., Adamantidis, A., Stuber, G.D., Bonci, A., de Lecea, L., and Deisseroth, K. (2009). Phasic Firing in Dopaminergic Neurons Is Sufficient for Behavioral Conditioning. *Science* (80-. ). 324, 10801085.
- Tsu, A.P., Burish, M.J., GodLove, J., and Ganguly, K. (2015). Cortical neuroprosthetics from a clinical perspective. *Neurobiol. Dis.* 83, 154160.
- Tumer, E.C., and Brainard, M.S. (2007). Performance variability enables adaptive plasticity of crystallized adult birdsong. *Nature* 450, 12401244.
- Wander, J., Blakely, T., Miller, K.J., Weaver, K.E., Johnson, L.A., Olson, J.D., Fetz, E.E., Rao, R.P.N., and Ojemann, J.G. (2013). Distributed cortical adaptation during learning of a brain-computer interface task. *Proc. Natl. Acad. Sci.* 110, 1081810823.
- Webster, K.E. (1965). The cortico-striatal projection in the cat. *J. Anat.* 99, 329337.

- Wessberg, J., Stambaugh, C.R., Kralik, J.D., Beck, P.D., Laubach, M., Chapin, J.K., Kim, J., Biggs, S.J., Srinivasan, M. a, and Nicolelis, M. a (2000). Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature* 408, 361365.
- Wickens, J.R., Horvitz, J.C., Costa, R.M., and Killcross, S. (2007). Dopaminergic mechanisms in actions and habits. *J. Neurosci.* 27, 81818183.
- Williamson, R.C., Cowley, B.R., Litwin-Kumar, A., Doiron, B., Kohn, A., Smith, M.A., and Yu, B.M. (2016). Scaling Properties of Dimensionality Reduction for Neural Populations and Network Models. *PLoS Comput. Biol.* 12, 127.
- Witten, I.B., Steinberg, E.E., Lee, S.Y., Davidson, T.J., Zalocusky, K.A., Brodsky, M., Yizhar, O., Cho, S.L., Gong, S., Ramakrishnan, C., et al. (2011). Recombinase-driver rat lines: Tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron* 72, 721733.
- Wright, M.J., Burns, R.J., Geffen, G.M., and Geffen, L.B. (1990). Covert orientation of visual attention in Parkinsons disease: An impairment in the maintenance of attention. *Neuropsychologia* 28, 151159.
- Wu, H.G., Miyamoto, Y.R., Castro, L.N.G., Olveczky, B.P., and Smith, M.A. (2014). Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nat. Neurosci.* 17, 312321.
- Wu, W., Gao, Y., Bienenstock, E., Donoghue, J.P., and Black, M.J. (2006). Bayesian Population Decoding of Motor Cortical Activity Using a Kalman Filter. *Neural Comput.* 18, 80118.
- Xu, T., Yu, X., Perlik, A.J., Tobin, W.F., Zweig, J.A., Tennant, K., Jones, T., and Zuo, Y. (2009). Rapid formation and selective stabilization of synapses for enduring motor memories. *Nature* 462, 915919.
- Yagishita, S., Hayashi-Takagi, A., Ellis-Davies, G.C.R., Urakubo, H., Ishii, S., and Kasai, H. (2014). A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science* (80-. ). 345, 16161620.
- Yamaguchi, S., and Kobayashi, S. (1998). Contributions of the dopaminergic system to voluntary and automatic orienting of visuospatial attention. *J. Neurosci.* 18, 18691878.
- Yin, H.H., Ostlund, S.B., Knowlton, B.J., and Balleine, B.W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *Eur. J. Neurosci.* 22, 513523.
- Yin, H.H., Knowlton, B.J., and Balleine, B.W. (2006). Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. *Behav. Brain Res.* 166, 189196.
- Yin, H.H., Mulcare, S.P., Hilario, M.R.F., Clouse, E., Holloway, T., Davis, M.I., Hansson, A.C., Lovinger, D.M., and Costa, R.M. (2009). Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. *Nat. Neurosci.* 12, 333341.
- Yizhar, O., Fenno, L.E., Davidson, T.J., Mogri, M., and Deisseroth, K. (2011). Optogenetics in Neural Systems. *Neuron* 71, 934.
- Yu, B., Cunningham, J., Santhanam, G., Ryu, S., Shenoy, K., and Sahani, M. (2009). Gaussian-Process Factor Analysis for Low-Dimensional Single-Trial Analysis of Neural Population Activity (vol 102, pg 614, 2009). *J. Neurophysiol.* 102, 614635.

Zacksenhouse, M., Lebedev, M.A., Carmena, J.M., ODoherty, J.E., Henriquez, C., and Nicolelis, M.A.L. (2007). Cortical Modulations Increase in Early Sessions with Brain-Machine Interface. PLoS One 2, e619.

Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J.P., Chang, W.-C., Jenvay, S., Miyamichi, K., Luo, L., and Dan, Y. (2014). Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing. Science 345, 660665.