

NEUROSCIENCE

Evidence for a neural law of effect

Vivek R. Athalye,^{1,2*} Fernando J. Santos,^{1*} Jose M. Carmena,^{2,3,4,†‡} Rui M. Costa^{1,5,†‡}

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. Seeking evidence for a neural law of effect, we found that mice learn to reenter more frequently motor cortical activity patterns that trigger optogenetic VTA self-stimulation. 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.

According to Thorndike's law of effect (1), actions that lead to reinforcements are repeated more frequently (2). Through repeated attempts, actions are shaped to more directly and reliably achieve reinforcement (3, 4), a process accompanied by the refinement of behavior-specific neural ensembles and activity patterns in motor cortices (5–9). Learning occurs because neural patterns initiating actions that lead to reinforcement are reentered more often, as supported by neural activity operant conditioning experiments (10–15).

Reinforcement is thought to rely on the activity of midbrain dopamine neurons. When animals receive reward, dopamine neurons in the ventral tegmental area (VTA) produce a spike burst that encodes the difference between the animal's expected and received rewards (16). This reward-prediction error signal is useful for optimizing reward-seeking behavior (17, 18). Indeed, phasic VTA activity constitutes a neural basis of reinforcement, as animals shape their behavior to receive electrical (19, 20) as well as optogenetic (21, 22) 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 (Fig. 1A). We recorded the activity of tens of neurons in primary motor cortex (M1) and used it to trigger optogenetic stim-

ulation of dopaminergic VTA neurons with blue light (21). Tyrosine hydroxylase (TH)-Cre mice (23) expressing channelrhodopsin-2 (ChR2 group, $n = 10$) in VTA dopaminergic cells were implanted with an optic fiber in the VTA and an electrode array in contralateral M1 layer 5 (Fig. 1B and fig. S1). To control for the effects of viral expression and shining light in the VTA, we expressed yellow fluorescent protein (YFP group, $n = 6$) in Cre-positive mice that underwent the same experimental procedure. Mice were trained to control a brain-machine interface (BMI) that transformed the activity of groups of neurons in M1 into real-time auditory feedback. When mice produced the target neural activity pattern that led to the target tone, they received a train of blue laser pulses, providing phasic stimulation of dopaminergic cells in the VTA. The self-stimulation optogenetic protocol used here has been previously shown to reinforce lever pressing (fig. S2).

This closed-loop self-stimulation paradigm (24) 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 through the "decoder," and delivers temporally precise reinforcement after target neural activity is produced. Our decoder received input from two arbitrarily selected M1 ensembles of two to four well-isolated single units (see supplementary methods and fig. S3) (14, 15). Two target neural population activity patterns (targets 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, whereas target 2 required the reverse modulation (see supplementary 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 pattern exploration along the task-relevant neural dimension—the differential modulation of ensembles 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 was used to estimate 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 (Fig. 1C). 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, four 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's and ensemble 2's modulation state for each 500-ms cycle and mapped it to one of seven auditory tones (ranging from 5 to 19 kHz). This daily calibration yielded a Gaussian-like distribution over tones during baseline and ensured that the chance rate of tone occupancy did not change over training days, despite potential day-to-day variability in neural recordings (Fig. 1D). Animals had to produce substantial ensemble modulations to achieve the targets (Fig. 1E). During the BMI training block, neural patterns close to target 1 decreased the tone, whereas neural patterns close to target 2 increased the tone (Fig. 1A). Target achievement resulted in a 1-s playback of the target tone, and only target 1 achievement resulted in phasic VTA stimulation 1.5 s after target hit, consisting of a 14-Hz train delivered for 2 s (Fig. 1F).

We trained animals on four consecutive daily sessions and quantified how reinforcement changed BMI tone distributions relative to session 1 (Fig. 2, A and B). Experimenters were blind to the type of virus injected in the VTA. ChR2 animals changed their target tone occupancy from their baseline bootstrap distribution by sessions 3 and 4, whereas YFP animals showed no preference for target 1 (Fig. 2C). With training, target 1 was occupied significantly more often in ChR2 animals and did not change in YFP animals (Fig. 2D). ChR2 animals increased preference for target 1 versus target 2 (Fig. 2E) and biased their overall distribution toward low-pitch tones close to target 1 and away from high-pitch tones close to target 2 (Fig. 2F). Interestingly, neuroprosthetic-triggered VTA stimulation did not reinforce specific overt movements (19, 20, 22) or place preference (21), suggesting that animals are not simply undergoing motor learning (fig. S4).

Given that the differential modulation between ensembles 1 and 2 shifted toward 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

¹Champalimaud Neuroscience Programme, Champalimaud Centre for the Unknown, Lisbon 1400-038, Portugal.

²Department of Electrical Engineering and Computer Sciences, University of California–Berkeley, Berkeley, CA 94720, USA. ³Helen Wills Neuroscience Institute, University of California–Berkeley, Berkeley, CA 94720, USA.

⁴Joint Graduate Group in Bioengineering, University of California–Berkeley and University of California–San Francisco, Berkeley, CA 94720, USA.

⁵Departments of Neuroscience and Neurology, Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY 10032, USA.

*These authors contributed equally to this work. †These authors contributed equally to this work.

‡Corresponding author. Email: rc3031@columbia.edu (R.M.C.);

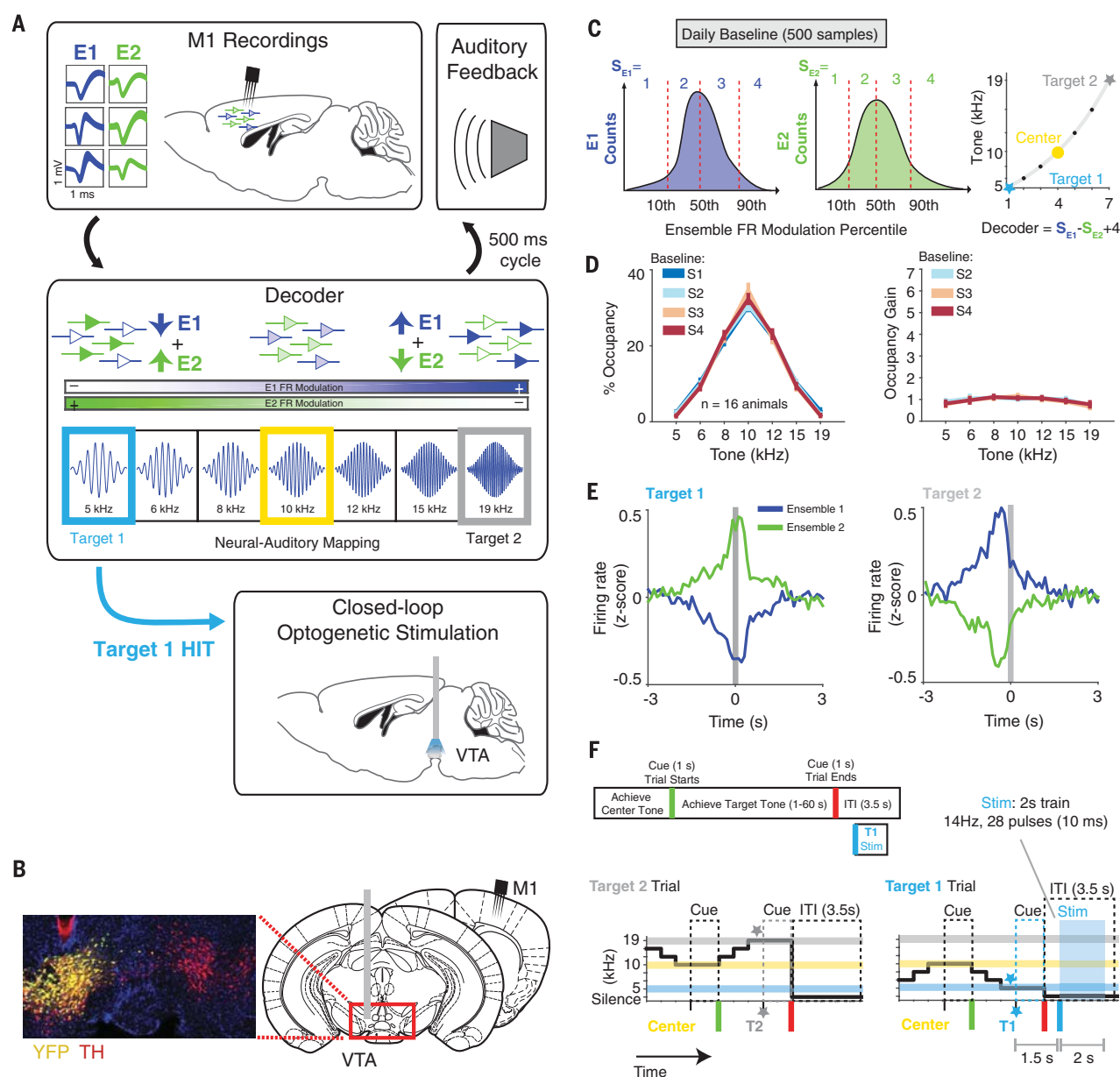
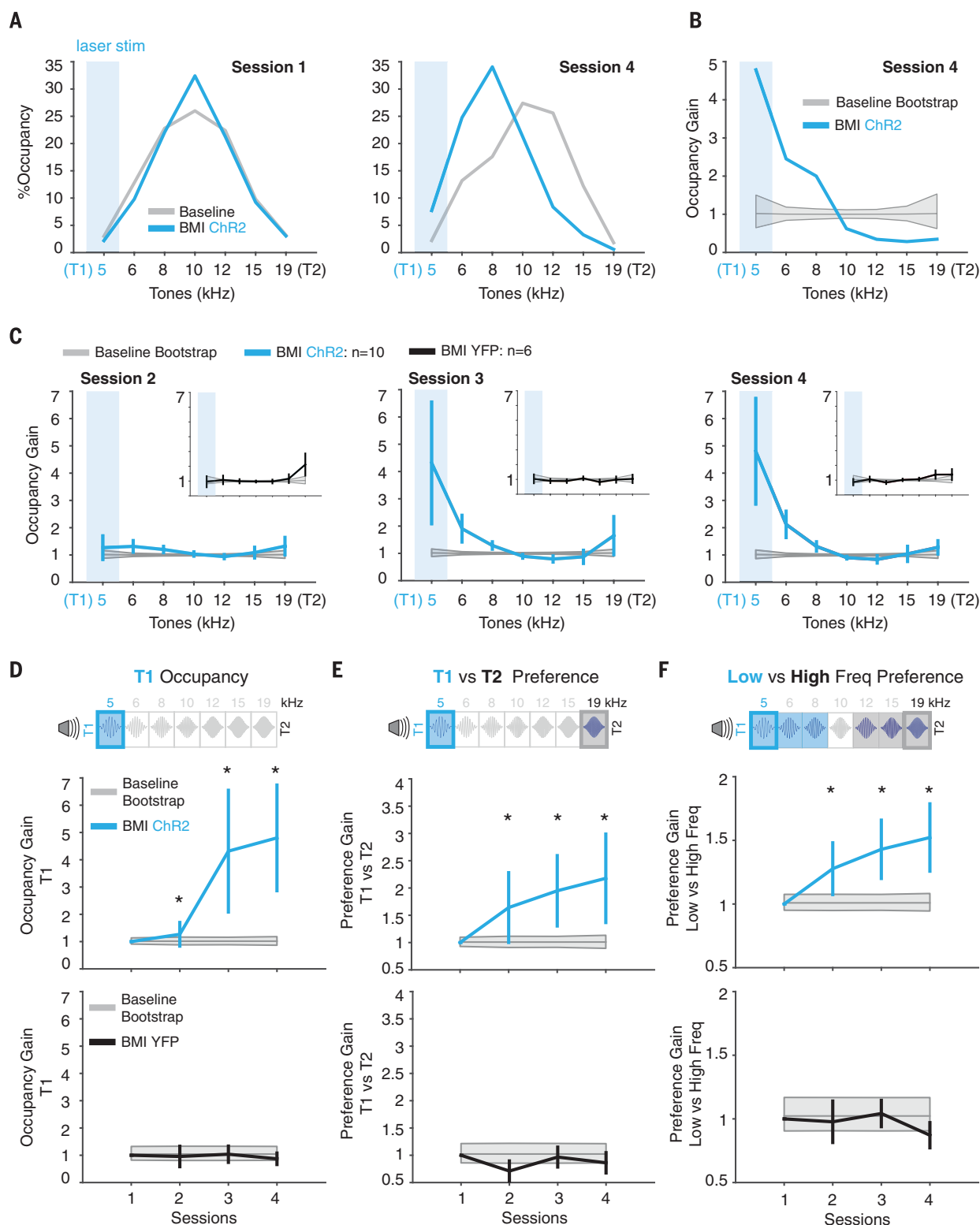


Fig. 1. Closed-loop BMI paradigm for pairing specific motor cortex activity patterns with phasic VTA dopaminergic activity. (A) Schematic of the BMI paradigm. Each mouse receives a unilateral microwire array implant in the motor cortex (targeted to layer V) and a contralateral optical fiber implant in the VTA. Recorded single units are arbitrarily assigned into two ensembles (E), and the concomitant increase (up arrow) of one ensemble's activity and decrease (down arrow) in the other ensemble's activity drives the decoder to change the auditory tone produced every 500 ms. The rare, lowest-pitch tone triggers phasic optical stimulation to the VTA, whereas the rare, highest-pitch tone serves as a control. Solid triangles indicate neurons with positively modulated firing rate; open triangles indicate neurons with negatively modulated firing rate. Yellow color indicates the center-pitch tone. FR, firing rate. (B) Coronal brain slice 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) are shown. (C) BMI decoder calibration. For every session (S) 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 four activation states. The decoder's state is the difference between ensemble 1's and ensemble 2's activation state and is mapped into one of seven tones. The stars indicate target tones. (D) 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 SEM baseline distribution for each session is plotted on the left, averaged over all animals. Baseline distributions show no change from session 1, as shown on the right. (E) Ensemble 1 and 2 firing rate modulation before target 1 and target 2 hits, averaged over all recorded cells and sessions. (F) Task schematic. Trial structure is the same for target 1 and target 2, except that a target 1 hit results in phasic VTA stimulation (2-s train of 14 Hz pulses with 10-ms width). ITI, intertrial interval.

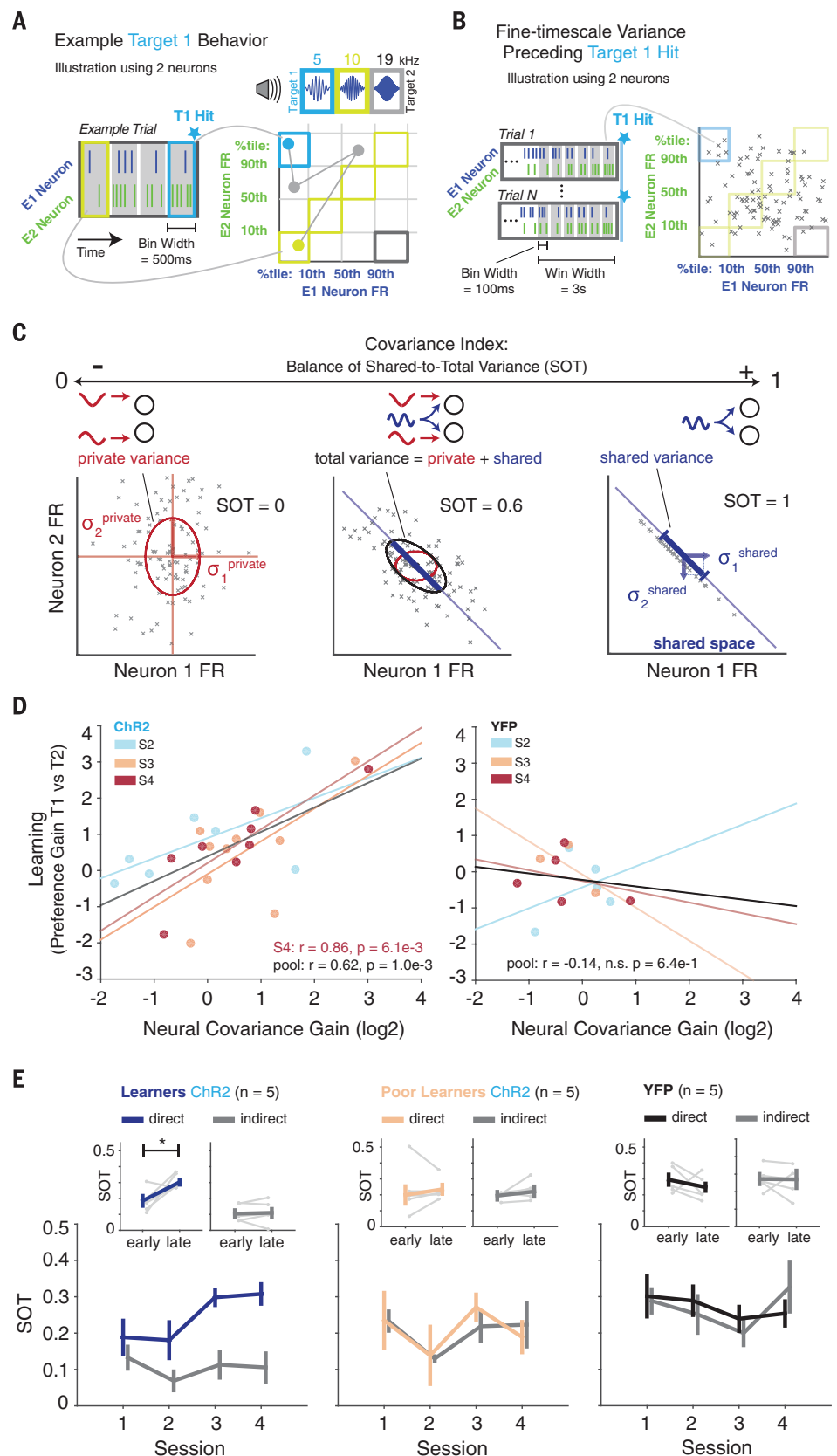
Fig. 2. Target pattern reentrance increases during VTA optogenetic self-stimulation.

(A) Distribution of the percent of time that each tone was occupied during baseline (gray) and BMI ChR2 (cyan) blocks of session 1 (left) and session 4 (right) in one mouse. (No tones were actually played during the baseline block.) T1, target 1; T2, target 2. (B) Quantification of the behavioral changes between sessions 1 and 4. The session 4 occupancy gain (cyan) is the session 4 BMI distribution normalized to the session 1 baseline distribution, then normalized to the session 1 ratio. For (B) to (F), the 95% confidence interval for the baseline bootstrap distribution is plotted in gray (see supplementary methods). 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 over sessions 2 through 4. For (C) to (F), mean and SEM over ChR2 animals ($n = 10$) are shown in cyan and over YFP animals ($n = 6$) are shown in black. By session 4, the behavioral changes were statistically different across tones for ChR2 but not YFP [repeated measures analysis of variance (ANOVA): ChR2, $F_{6,48} = 3.46$, $P = 6.4 \times 10^{-3}$; YFP, $F_{6,30} = 0.96$, $P = 0.47$]. In session 4, 5 kHz (target 1) was significantly different from all tones from 8 to 19 kHz (Tukey's post hoc multiple comparisons test). (D) Top: The occupancy gain for 5 kHz (target 1) over sessions is shown. Middle: ChR2 (cyan) were significantly larger than bootstrap from sessions 2 through 4 (session 2, $P = 1.2 \times 10^{-3}$; session 3, $P < 1 \times 10^{-5}$; session 4, $P < 1 \times 10^{-5}$). Bottom: YFP (black) were never significantly



larger than bootstrap. (E) Top: The preference gain for 5 kHz (target 1) versus 19 kHz (target 2) is plotted over sessions. Middle: ChR2 (cyan) were significantly larger than bootstrap after session 1 ($P < 1 \times 10^{-5}$ for sessions 2 through 4). Bottom: YFP (black) were never significantly larger than bootstrap. (F) Top: The preference gain for low-pitch tones (5 to 8 kHz, close to target 1) versus high-pitch tones (12 to 19 kHz, close to target 2) over sessions is shown. Middle: ChR2 (cyan) were significantly larger than bootstrap after session 1 ($P < 1 \times 10^{-5}$ for sessions 2 through 4). Bottom: YFP (black) were never significantly larger than bootstrap. For (D) to (F), an asterisk indicates that the population average is significantly larger than the baseline bootstrap distribution.

Fig. 3. Learning correlates with an increase in covariance of the neurons that produce the target pattern. (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) Analysis of variance of spike counts with 100-ms bins in a 3-s window preceding target hit. "x" indicates a spike count vector at one time point. (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 two-neuron illustration shows a neural solution with SOT = 0, 0.6, and 1. (D) 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 sessions 2, 3, and 4. Chr2 animals (left) showed a significant correlation [Chr2 S4: correlation coefficient (r) = 0.86, P = 6.1×10^{-3} ; Chr2 pool S3, S4: r = 0.71, P = 1.0×10^{-3} ; Chr2 pool S2, S3, S4: r = 0.62, P = 9.8×10^{-4} ; Chr2 S3: r = 0.60, P = 6.5×10^{-2} ; Chr2 S2: r = 0.62, P = 1.3×10^{-1}], whereas YFP animals (right) showed no correlation (YFP pool S2, S3, S4: r = -0.14, n.s. P = 6.4×10^{-1} ; YFP S4: r = -0.32, P = 6.0×10^{-1} ; YFP S3: r = -0.69, P = 5.1×10^{-1} ; YFP S2: r = 0.37, P = 5.4×10^{-1}). n.s., not significant. (E) SOT of direct and indirect neurons over sessions for Chr2 learners (left, n = 5), Chr2 poor learners (middle, n = 5), and YFP subjects (right, n = 5). Chr2 learners individually showed significant preference gain for target 1 versus target 2 in both sessions 3 and 4. Chr2 poor learners constitute the remaining animals who as a population showed significant target 1 occupancy gain on sessions 3 and 4. For direct neurons, Chr2 animals' and Chr2 learners' SOT increased from early (sessions 1 and 2 pooled) to late training (sessions 3 and 4 pooled), whereas Chr2 poor learners and YFP did not (one-sided rank sum test; Chr2, early < late, P = 1.7×10^{-2} ; Chr2 learners, early < late, P = 1.6×10^{-2} ; Chr2 poor learners, early < late, n.s. P = 2.1×10^{-1} ; YFP, early < late, n.s. P = 8.3×10^{-1}). For indirect neurons, SOT showed no change for all groups (Chr2 learners: early < late, n.s. P = 4.3×10^{-1} ; Chr2 poor learners, early < late, n.s. P = 2.7×10^{-1} ; YFP, early < late, n.s. P = 7.1×10^{-1}). Traces in the insets show the average of each animal's SOT in sessions 1 and 2 (early) versus the average of sessions 3 and 4 (late). Error bars indicate mean \pm SEM. The asterisk indicates that the population average is significantly larger than the baseline bootstrap distribution.



shared inputs to direct neurons and thus increase covariance over learning (13). We used factor analysis (FA) to partition fine-time scale 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 (Fig. 3A). We analyzed fine-time scale spike counts (100-ms bins) in a 3-s window preceding target hit (Fig. 3B). FA models population spike counts $x = \mu + \alpha^{\text{private}} + \alpha^{\text{shared}}$ as the sum of a mean firing rate μ ; private variation

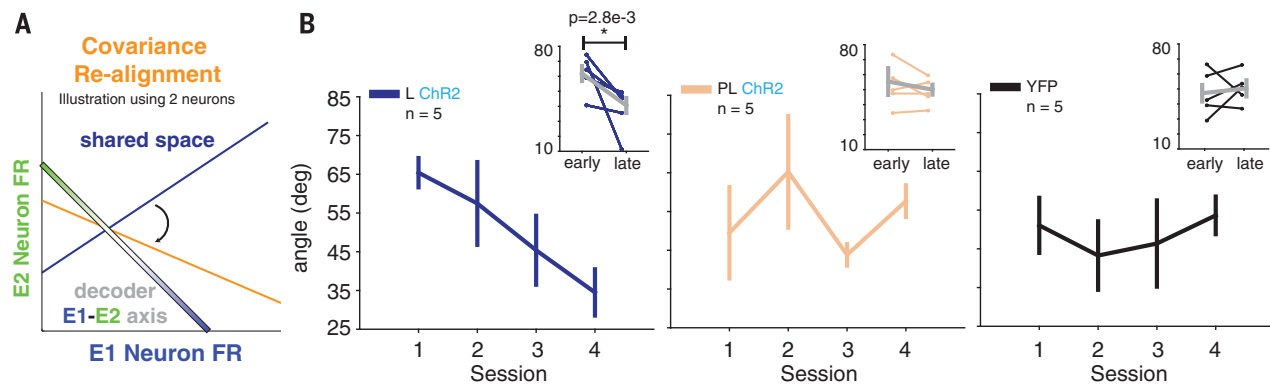


Fig. 4. Covariance of the neurons that produce the target pattern gradually aligns to the decoder. (A) Analysis of shared variance alignment with the decoder's ensemble 1 and ensemble 2 assignments by using the angle between the shared space and the decoder's "ensemble 1 minus ensemble 2" axis. Curved arrow indicates rotation of the shared space to align with the decoder. (B) The angle between shared variance and the decoder axis decreased for ChR2 learners (left) but not for poor learners (middle) and YFP (right) (one-sided

rank sum test comparing sessions 1 and 2 to sessions 3 and 4; ChR2 learner, late < early, $P = 2.8 \times 10^{-3}$; ChR2 poor learner, late < early, n.s. $P = 3.7 \times 10^{-1}$; YFP, late < early, n.s. $P = 7.5 \times 10^{-1}$). Traces in the insets show the average of each animal's angle in sessions 1 and 2 (early) versus the average of sessions 3 and 4 (late). Error bars indicate mean \pm SEM. The asterisk indicates that the population average is significantly larger than the baseline bootstrap distribution.

x^{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^{private} and x^{shared} are independent, the total covariance matrix $\Sigma^{\text{total}} = \Sigma^{\text{private}} + \Sigma^{\text{shared}}$ is decomposed into the sum of a diagonal private covariance matrix Σ^{private} and a low-rank shared covariance matrix Σ^{shared} . Geometrically, private variance spans all of the high-dimensional population activity space for which each neuron's activity is one dimension, whereas 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 by using standard model selection (fig. S5) by maximizing cross-validated log likelihood (13, 25–28).

We assessed neural coordination with a covariance index defined as the ratio of the shared variance to total variance averaged over neurons (SOT) (Fig. 3C). Although Fig. 3, A to C, uses two neurons for illustration, we emphasize that FA was applied to the joint activity of all neurons used to control the BMI (ranging from four to eight). 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 (Fig. 3D). 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 two groups distinguished by their degree of learning (fig. S6). Each individual of the learner group ($n = 5$) showed statistically significant preference for target 1 versus target 2 for both sessions 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 (fig. S6).

Learners significantly increased their covariance index over training, whereas poor learners and YFP did not (Fig. 3E and figs. S7 and S8A). This effect was ensemble specific, as only neurons controlling the BMI (direct neurons) increased their covariance index, whereas other recorded neurons (indirect neurons) did not (Fig. 3E and figs. S9 and S10).

Finally, we asked whether dopaminergic self-stimulation shaped the neural covariance to more easily achieve the target pattern. Only neural variance that causes differential modulation between ensembles 1 and 2 can change the feedback tone and contribute to target achievement, corresponding to variance that is aligned with the decoder's "ensemble 1 minus ensemble 2" axis (Fig. 4A). 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 (Fig. 4B and fig. S8B).

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, which may have been strongly predicted on the basis of previous studies, 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 (fig. S11), 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 through electrical self-stimulation of the nucleus accumbens (29). Although it may be difficult to completely

rule out that very subtle movements that lead to the desired patterns of activity are being reinforced, we showed that, in this paradigm, there is no reinforcement of overt movements over BMI learning (fig. S4). Still, these results may have implications for motor reinforcement, in which actions are selected more often and optimized over iterations to more directly achieve reinforcements.

In these experiments, 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 (28). By 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 realigning shared variance with learning (13). Analysis and modeling indicate that private variance is useful for broad exploration of population activity space (13) and for learning each neuron's contributions to achieving a goal (30, 31), possibly permitting the selective increase of direct neurons' covariation index over indirect neurons. The difference between learners and poor learners could depend on the probability of the direct neurons receiving common anatomical input, or on the plasticity of common inputs to the direct 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, and most projections are unilateral. Indeed, VTA stimulation did not induce any observable changes in the mean firing rates of M1 neurons (fig. S12). 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 (32, 33) and is

necessary for motor and neuroprosthetic learning (5, 14, 34). Actor-critic reinforcement learning models (35, 36) 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 on the basis of 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 the VTA (37), enabling adaptation of the policy for reentering M1 activity patterns. Plasticity in ventral striatum (32) could be mediated by strong bidirectional connections with the 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 (8, 38–40). Further, reinforcement guides the reactivation of neurons during sleep (41), leading to the formation of dendritic spines (42) as well as the identification of neurons responsible for achieving a target pattern (41). 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, whereas nonclustered networks exhibited weak, high-dimensional shared variance (43).

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 (44). Our results help uncover how

selection by consequence operates on neural activity in the brain (45).

REFERENCES AND NOTES

1. E. L. Thorndike, *Psychol. Rev.* **2**, 1–107 (1898).
2. B. F. Skinner, *The Behavior of Organisms: An Experimental Analysis* (Appleton-Century, Oxford, 1938).
3. R. G. Cohen, D. Sternad, *Exp. Brain Res.* **193**, 69–83 (2009).
4. L. Shmuelof, J. W. Krakauer, P. Mazzoni, *J. Neurophysiol.* **108**, 578–594 (2012).
5. R. M. Costa, D. Cohen, M. A. L. Nicolelis, *Curr. Biol.* **14**, 1124–1134 (2004).
6. T. D. Barnes, Y. Kubota, D. Hu, D. Z. Jin, A. M. Graybiel, *Nature* **437**, 1158–1161 (2005).
7. V. Y. Cao et al., *Neuron* **86**, 1385–1392 (2015).
8. A. J. Peters, S. X. Chen, T. Komiyama, *Nature* **510**, 263–267 (2014).
9. B. P. Olveczky, T. M. Otchy, J. H. Goldberg, D. Aronov, M. S. Fee, *J. Neurophysiol.* **106**, 386–397 (2011).
10. E. E. Fetz, *Science* **163**, 955–958 (1969).
11. J. M. Carmena et al., *PLOS Biol.* **1**, E42 (2003).
12. K. Ganguly, J. M. Carmena, *PLOS Biol.* **7**, e1000153 (2009).
13. V. R. Athalye, K. Ganguly, R. M. Costa, J. M. Carmena, *Neuron* **93**, 955–970.e5 (2017).
14. A. C. Koralek, X. Jin, J. D. Long II, R. M. Costa, J. M. Carmena, *Nature* **483**, 331–335 (2012).
15. K. B. Clancy, A. C. Koralek, R. M. Costa, D. E. Feldman, J. M. Carmena, *Nat. Neurosci.* **17**, 807–809 (2014).
16. W. Schultz, P. Dayan, P. R. Montague, *Science* **275**, 1593–1599 (1997).
17. R. S. Sutton, A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, 1998).
18. E. E. Steinberg et al., *Nat. Neurosci.* **16**, 966–973 (2013).
19. J. Olds, P. Milner, *J. Comp. Physiol. Psychol.* **47**, 419–427 (1954).
20. D. Corbett, R. A. Wise, *Brain Res.* **185**, 1–15 (1980).
21. H.-C. Tsai et al., *Science* **324**, 1080–1084 (2009).
22. I. B. Witten et al., *Neuron* **72**, 721–733 (2011).
23. S. Gong et al., *J. Neurosci.* **27**, 9817–9823 (2007).
24. L. Groenewick, J. H. Marshel, K. Deisseroth, *Neuron* **86**, 106–139 (2015).
25. A. P. Dempster, N. M. Laird, D. B. Rubin, *J. R. Stat. Soc. B* **39**, 1–38 (1977).
26. B. S. Everitt, *An Introduction to Latent Variable Models* (Chapman and Hall, London, 1984).
27. B. M. Yu et al., *J. Neurophysiol.* **102**, 614–635 (2009).
28. P. T. Sadtler et al., *Nature* **512**, 423–426 (2014).
29. R. W. Eaton, T. Libey, E. E. Fetz, *J. Neurophysiol.* **117**, 1112–1125 (2017).
30. R. Legenstein, S. M. Chase, A. B. Schwartz, W. Maass, *J. Neurosci.* **30**, 8400–8410 (2010).
31. R. Hélot, K. Ganguly, J. Jimenez, J. M. Carmena, *IEEE Trans. Syst. Man Cybern. B Cybern.* **40**, 1387–1397 (2010).
32. S. Yagishita et al., *Science* **345**, 1616–1620 (2014).
33. W. Shen, M. Flajolet, P. Greengard, D. J. Surmeier, *Science* **321**, 848–851 (2008).
34. F. J. Santos, R. F. Oliveira, X. Jin, R. M. Costa, *eLife* **4**, e09423 (2015).
35. Y. Takahashi, G. Schoenbaum, Y. Niv, *Front. Neurosci.* **2**, 86–99 (2008).
36. J. O'Doherty et al., *Science* **304**, 452–454 (2004).
37. M. W. Howe, D. A. Dombeck, *Nature* **535**, 505–510 (2016).
38. M. Fu, X. Yu, J. Lu, Y. Zuo, *Nature* **483**, 92–95 (2012).
39. T. Xu et al., *Nature* **462**, 915–919 (2009).
40. A. Hayashi-Takagi et al., *Nature* **525**, 333–338 (2015).
41. T. Gulati, L. Guo, D. S. Ramanathan, A. Bodepudi, K. Ganguly, *Nat. Neurosci.* **20**, 1277–1284 (2017).
42. G. Yang et al., *Science* **344**, 1173–1178 (2014).
43. R. C. Williamson et al., *PLOS Comput. Biol.* **12**, e1005141 (2016).
44. B. F. Skinner, *Science* **213**, 501–504 (1981).
45. R. M. Costa, *Curr. Opin. Neurobiol.* **21**, 579–586 (2011).

ACKNOWLEDGMENTS

We thank A. Castro for the motor reinforcement experiments; A. Koralek, P. Khanna, and R. Neely for helpful discussions; and A. Vaz for animal colony management. This work was supported by a NSF Graduate Research Fellowship to V.R.A.; grants from the NSF (CBET-0954243 and EFRI-M3C 1137267) and Office of Naval Research (N00014-15-1-2312) to J.M.C.; and grants from the European Research Area ERA-NET, European Research Council (COG 617142), and Howard Hughes Medical Institute (IEC 55007415) to R.M.C. Data presented in this paper can be found on figshare at <https://doi.org/10.6084/m9.figshare.5687101.v1>.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6379/1024/suppl/DC1
Materials and Methods
Figs. S1 to S12
References (46–49)

8 August 2017; accepted 8 January 2018
10.1126/science.aao6058

Evidence for a neural law of effect

Vivek R. Athalye, Fernando J. Santos, Jose M. Carmena and Rui M. Costa

Science **359** (6379), 1024-1029.
DOI: 10.1126/science.aao6058

How to select and shape neural activity

When we learn a new skill or task, our movements are reinforced and shaped. Learning occurs because the neural activity patterns in the movement control-related brain regions that are rewarded are repeated. But how does this reinforcement work? Athalye *et al.* developed a closed-loop self-stimulation paradigm in which a target motor cortical activity pattern resulted in the optogenetic stimulation of dopaminergic neurons. With training, mice learned to reenter specific neuronal activity patterns, which triggered self-stimulation and shaped their neural activity to be closer to the target pattern.

Science, this issue p. 1024

ARTICLE TOOLS

<http://science.sciencemag.org/content/359/6379/1024>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2018/02/28/359.6379.1024.DC1>

REFERENCES

This article cites 45 articles, 11 of which you can access for free
<http://science.sciencemag.org/content/359/6379/1024#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)