Title:
Reward-driven cerebellar climbing fiber activity influences both neural and behavioral learning

Authors:
Shuyang Jin and Court Hull

Affiliations:
Department of Neurobiology, Duke University, Durham, NC, USA

Email: Hull@neuro.duke.edu

Abstract

The cerebellum plays a key role in motor coordination and learning. In contrast with classical supervised learning models, recent work has revealed that CFs can signal reward-predictive information in some behaviors. This raises the question of whether CFs may also operate according to principles similar to those described by reinforcement learning models. To test how CFs operate during reward-guided behavior, and evaluate the role of reward-related CF activity in learning, we have measured CF responses in Purkinje cells of the lateral cerebellum during a Pavlovian task using 2-photon calcium imaging. Specifically, we have performed multi-stimulus experiments to determine whether CF activity meets the requirements of a reward prediction error (rPE) signal for transfer from an unexpected reward to a reward-predictive cue. We find that once CF activity is transferred to a conditioned stimulus, and there is no longer a response to reward, CFs cannot generate learned responses to a second conditioned stimulus that carries the same reward prediction. In addition, by expressing the inhibitory opsin GtACR2 in neurons of the inferior olive, and optically inhibiting these neurons across behavioral training at the time of unexpected reward, we find that the transfer of CF signals to the conditioned stimulus is impaired. Moreover, this optogenetic inhibition also impairs learning, resulting in a deficit in anticipatory lick timing. Together, these results indicate that CF signals can exhibit several characteristics in common with rPEs during reinforcement learning, and that the cerebellum can harness these learning signals to generate accurately timed motor behavior.

Introduction

Climbing fibers (CFs) play a key role in many forms of cerebellar learning, and are thought to provide instructional signals that generate synaptic plasticity and modify behavior according to the principles of supervised learning (Raymond and Medina, 2018). Recent work has revealed that CFs can also display characteristics that are typically associated with reward-prediction errors (rPEs), the instructional signal necessary for reinforcement learning, in some behaviors guided by predictions about reward (Hull, 2020; Kostadinov and Hausser, 2022). In particular, it has become clear that CFs can respond to unexpected reward, and generate learned responses to reward-predictive cues (Heffley and Hull, 2019; Heffley et al., 2018; Kostadinov et al., 2019; Larry et al., 2019; Sendhilnathan et al., 2021). However, CFs also exhibit properties that are inconsistent with leading models of reinforcement learning, such as elevated spiking in response to omitted reward (Heffley et al., 2018). Thus, while many properties of CF signaling in reward tasks are consistent with rPEs, they may also be consistent with signaling other information such as stimulus salience, valence or novelty.

To determine how CFs generate teaching signals in reward-based tasks, and what similarities these signals have with classically described rPEs, it is important to test whether they obey key principles of

that have been described in other brain structures such the ventral tegmental area (VTA) thought to be necessary for reinforcement learning. There, some dopamine (DA) neurons are thought report rPEs in reward-guided tasks (Bromberg-Martin et al., 2010), and to follow many of the formal predictions established by reinforcement learning theory (Rescorla and Wagner, 1972; Sutton and Barto, 1990; Sutton and Barto, 1998). For example, DA neurons can signal positive rPEs (increased spiking) for unexpected outcomes that are better than anticipated (surprise reward delivery), and these responses shift with learning to represent sensory stimuli that predict such outcomes (Schultz et al., 1997). In accordance with the formalism of leading reinforcement learning theories such as temporal-difference learning (TD learning), this shift requires the initial response to unexpected reward. In the VTA, this has been demonstrated using so-called 'stimulus blocking' conditions that leverage the absence of reward-driven spiking in response to a fully trained sensory cue (Steinberg et al., 2013). Because reward-driven spiking is abolished in this condition, a new reward-predictive stimulus that carries the same prediction cannot generate learned DA neuron responses.

To assess the similarity of cerebellar CF responses to those of DA neurons, and test whether CFs follow these key principles of reinforcement learning, we have used a stimulus-blocking paradigm to test the requirements for generating learned CF responses to reward-predictive sensory cues. We find that, like canonical DA neurons of the VTA, cerebellar CFs do not learn sensory stimuli that carry fully redundant reward predictions. To test the causality of reward-driven CF responses in generating learned, reward-predictive sensory responses, we have optogenetically inhibited CF activity across learning, preventing the initial response to unexpected reward. These experiments impair the transfer of CF responses to a reward-predictive auditory stimulus, suggesting a potential role of cerebellar learning in generating predictive CF activity. Finally, we also observe an impairment of learned lick-timing after inhibiting reward-driven CF activity, suggesting a role for these teaching signals in generating predictively timed movements. Together, these results reveal important links between cerebellar CF responses and the predictions of reinforcement learning theory, and indicate a key role for reward-driven CF responses in cerebellar learning.

Results

To test the properties of climbing fiber activity during reward conditioning, we used a Pavlovian conditioning task optimized for Ca^{2+} imaging in the lateral cerebellar cortex (Heffley and Hull, 2019). Previously work indicated that mice can effectively associate a visual cue with reward (Heffley and Hull, 2019), and that CF activity is transferred to the visual cue with learning. To test whether CFs can also respond to other types of reward-predictive sensory stimuli, we trained animals on a version of the same task using an auditory conditioned stimulus (CS). Specifically, mice were trained to associate a tone (200 ms, 5 kHz, 75 dB) with a delayed saccharine water reward (unconditioned stimulus- US; 700 ms delay, **Fig. 1A**).

Mice learned to associate the auditory cue with reward, developing robust licking to the auditory CS (**Fig. 1B,C**) (training = 1089.3 ± 103.6 trials, 9 ± 0.7 sessions, n = 6 mice). To assess learning, we measured the fraction of trials with predictive and reactive licking, respectively defined as licking onset occurring before or after reward delivery. Learning resulted in a significant increase in predictive licking (Pre-learning: 5.2 ± 0.1%; Post-learning: 69.5 ± 7.3%; $p = 3.52 \times 10^{-4}$, n = 6 mice, paired t-test **Fig. 1D**), and a decrease in reactive licking (Pre-learning: 92.4 ± 0.7%; Post-learning 24.5 ± 7.1%; $p = 2.36 \times 10^{-4}$, n = 6 mice, paired t-test, **Fig. 1D**). In addition, mice exhibited significantly shorter reaction times (Pre-learning, 1491.7 ± 75.6 ms; Post-learning, 601.8 ± 78.3 ms; $p = 8.16 \times 10^{-4}$, n = 6 mice, paired t-test, **Supp. Fig. 1A**) and lower miss rates after learning (Pre-learning, 29.0 ± 6.7%; Post-learning, 4.0 ± 1.8%; p = 0.02, n = 6 mice, paired t-test, **Supp. Fig. 1B**).

Using this task, we measured complex spikes (Cspk) in the dendrites of Purkinje cells, as a proxy for CF activity (Heffley et al., 2018), in lobule simplex before and after learning (**Fig. 1E,F, Supp. Fig. 1C,D**).

Consistent with previous work (Heffley and Hull, 2019), we observed a significant Cspk response to unexpected reward delivery in naive animals, best revealed in the lick-triggered average of Cspk activity ($p = 1.19 \times 10^{-7}$, paired t-test, **Fig. 1E, Supp. Fig. 1C**). In contrast, we did not observe a Cspk response to the auditory CS in naive animals (n = 436 dendrites, 6 mice, p = 0.12, paired t-test, **Fig. 1E, Supp. Fig. 1C**).

After learning, a significant Cspk response occurred within 250 ms after the auditory CS (n = 427 dendrites, 6 mice, $p = 5.51 \times 10^{-49}$, paired t-test, one-tailed; time from cue onset to peak Cspk rate: 224.43 ± 2.31 ms, **Fig. 1F, Supp. Fig 1D**). In addition, Cspks no longer occurred following the first lick after reward (p = 0.08, paired t-test, **Fig. 1F, Supp. Fig. 1D**). Because the lick-related Cspk response is absent after learning, the response in naive animals is likely not exclusively motor-related, and may involve reward anticipation or detection.

To further test the role of motor output in Cspk responses after learning, we segregated trials into subsets according to lick times and lick rates. Specifically, we averaged trials with the earliest and latest quartiles of lick onsets, and trials with the highest and lowest quartiles of lick rates (**Fig. 1G,H**). This analysis revealed that the Cspk response had the same amplitude (p = 0.21, n = 427 dendrites, 6 mice, paired t-test, **Fig. 1G**) and timing (time from cue on earliest lick trials = 212.41 ± 2.66 ms; latest lick trials: 208.27 ± 2.54 ms, p = 0.21, paired t-test, **Fig. 1G**) on trials with the earliest licks and latest licks. Similarly, Cspk responses exhibited only a slightly smaller amplitude on trials with the slowest lick rates as compared to trials with the fastest licking ($p = 2.62 \times 10^{-7}$, paired t-test, **Fig. 1H**), and similar timing across these two trial types (time from cue onset on trials with highest lick rates: 187.90 ± 2.71 ms; lowest lick rates: 185.01 ± 2.69 ms, p = 0.38, paired t-test, **Fig. 1H**). Consistent with previous work (Heffley and Hull, 2019; Heffley et al., 2018), these results suggest that the cue-associated Cspk responses after training are not primarily related to licking.

In addition to licking, mice also exhibited other body movements that could modulate Cspk activity in lobule simplex. We therefore used a piezoelectric sensor (Heffley and Hull, 2019) to measure body movement, and segregated trials with the most and least movement (**Fig. 1I**). This analysis revealed a similar amplitude ($p = 5.38 \times 10^{-5}$, paired t-test, **Fig. 1J**), and an almost identical timing of Cspk responses on trials with the least and most movements (time from cue onset on trials with least movements: 197.5 ± 2.72 ms, most movements: 201.17 ± 2.49 ms; p = 0.25, paired t-test, **Fig. 1J**).

Consistent with the previous results (Heffley and Hull, 2019), these data suggest that Cspk activity shifts from the time of licking and reward consumption to the reward-predictive auditory cue, and that the cue-related Cspk responses are not driven by motor signals due to licking or body movements in trained animals. Moreover, the Cspk response to the auditory cue is similar to previously measured responses to a learned visual cue (Heffley and Hull, 2019), suggesting that CFs can represent reward-predictive stimuli of different modalities.

Stimulus-blocking prevents learned Cspk responses

To test whether the CFs exhibit other signatures of reinforcement learning signals (Cohen et al., 2012; Hollerman and Schultz, 1998; Mirenowicz and Schultz, 1994; Roesch et al., 2007; Schultz et al., 1993), we next performed a stimulus-blocking experiment. A key prediction of reinforcement learning is that the transfer of neural activity from US to CS cannot occur unless there is an initial neuronal response to the US (Schultz et al., 1997; Steinberg et al., 2013; Waelti et al., 2001). This requirement has important consequences. In particular, it means that a new CS (CS2) cannot be learned if it carries the same reward prediction as a CS (CS1) that has already been learned. This so-called ‘stimulus blocking effect’ is a hallmark of reinforcement learning systems, as it serves to appropriately bind the CS and US (Keiflin and Janak, 2015). Moreover, this property ensures that an additional stimulus that provides no new information is not learned.

To test how CFs respond to a CS that does not carry new reward-predictive information, we introduced a new CS (CS2) after mice had already learned the initial auditory CS (CS1), and the US response was extinguished. For CS2, we chose a visual cue that we have previously demonstrated can be learned by CFs in this behavioral paradigm (Heffley and Hull, 2019). We then trained animals where reward delivery was predicted by the co-presentation of CS1 and CS2 (**Fig. 2A**). After a similar number of trials (mean = 1006 $\pm$ 43.3 trials) and training sessions (mean = 8.5 $\pm$ 0.2 sessions, n = 6 mice) required for CS1 learning, we measured the conditioned licking response and climbing fiber response to CS1 alone, CS1+CS2, and CS2 alone (**Fig. 2B-G**).

We observed a predictive licking response to both CS1+CS2 and CS1 alone (CS1+CS2 reaction time 465.1 ± 29.3 ms, 86.0 ± 2.7% of trials with a predictive response, 8.9 ± 2.9% of trials with a reactive response; CS1 alone reaction time: 489.7 ± 31.2 ms, 84.3 ± 4.0% of trials with a predictive response, 10.0 ± 3.5% of trials with a reactive response; n = 6 mice; **Fig. 2 B,C, Supp. Fig. 2**). In contrast, mice licked only in response to the US when CS2 was presented alone (reaction time: 1511.1 ± 88.4 ms, 5.0 ± 1.7% of trials with a predictive response, 90.8 ± 3.2% of trials with a reactive response, **Fig. 2 D, Supp. Fig. 2**). This result is consistent with previous stimulus blocking experiments related to midbrain dopamine circuits (Schultz et al., 1997; Steinberg et al., 2013; Waelti et al., 2001) and indicates that the redundant CS2 was not learned.

Consistent with the robust association of CS1 with reward, we observed significant Cspk activity after the presentation of either CS1+CS2 ($p = 7.87 \times 10^{-65}$, paired t-test, n = 398 dendrites, 6 mice, **Fig. 2E**) or CS1 alone ($p = 1.84 \times 10^{-49}$, paired t-test, n = 385 dendrites, 6 mice, **Fig. 2F**), and no Cspk response to the first lick after reward (CS1+CS2, p = 0.68; CS1, p = 0.90; paired t-test, **Fig. 2E,F**). Notably, the amplitude of the Cspk response to CS1 (2.61 ± 0.14 Hz) was not significantly different from the response to CS1+CS2 (2.73 ± 0.12 Hz, p = 0.45, paired t-test), suggesting there was not an additive effect when both cues were presented simultaneously. Additionally, the time to peak Cspk activity was the same for CS1+CS2 and CS1 presentations (CS1+CS2 209.38 ± 2.25 ms; CS1: 224.43 ± 2.31 ms; p = 0.52, unpaired t-test). Thus, the learned climbing fiber response was unchanged as long as CS1 was presented.

In contrast, we did not observe cue driven Cspk activity when CS2 was presented alone (p = 0.21, paired t-test, n = 353 dendrites, 6 mice, **Fig. 2G**). Instead, in this condition we observed a significant response to the first lick after reward delivery ($p = 7.15 \times 10^{-34}$, paired t-test, **Fig. 2G**). These results mirror what has been demonstrated in similar stimulus-blocking experiments for VTA dopamine neurons, fulfilling one key prediction of reinforcement learning signals.

Optical IO inhibition impairs Cspk transfer from US to CS

While stimulus blocking experiments provide a correlative link between Cspk activity to the US and learned Cspk activity to the CS, they do not causally link these neural activity patterns. Thus, to test whether US driven Cspk activity is causally linked to learned, CS-driven Cspk activity, we repeated behavioral training while optogenetically inhibiting neurons in the inferior olive (IO) which are the source of the CFs to the cerebellar cortex. For these experiments, we unilaterally expressed the soma-restricted inhibitory anion-fluxing channelrhodopsin GtACR2 (Mahn et al., 2018; Sineshchekov et al., 2015) in the contralateral IO, and implanted an optical fiber over this region (**Fig. 3A, Supp. Fig. 3A,B**) (Gaffield et al., 2019; Gaffield et al., 2018). To verify the efficiency of this optogenetic strategy *in vivo*, we first performed single-photon calcium imaging of Purkinje cell dendrites in lobule simplex. During optical inhibition, we observed a marked reduction of activity during the illumination period ($p = 2.35 \times 10^{-4}$, n = 10 trials, 5 mice, paired t-test, **Fig. 3B,C**). Upon light offset, we observed a transient rebound response, followed by a return to baseline levels.

In mice expressing GtACR2 in the IO, we first measured the Cspk response to the auditory CS in naive mice (**Fig. 3D**). As previously, there was no cue-related Cspk response (n = 363 dendrites, 5 mice, p = 0.34, paired t-test, Fig. 3c), but a significant US-related response prior to lick onset ($p = 4.75 \times 10^{-77}$,

paired t-test, **Fig. 3D**). These results suggest that GtACR2 expression did not affect normal climbing fiber responses to reward.

Next, we trained animals while optically inhibiting the IO on each trial, beginning 400 ms before cue onset until 4 s after reward delivery, placing rebound activity well outside the trial structure (**Fig. 3E**). Notably, we did not observe a significant lick response to the onset or offset of optical inhibition (onset p=0.42; offset, p = 0.11, paired t-tests, n = 34 training sessions, 5 mice, **Supp. Fig. 3C**).

After training (mean = 8 ± 0.5 sessions, n = 5 mice), we measured the Cspk responses in the absence of optogenetic inhibition. In stark contrast with normal training, we observed Cspk activity to both the auditory cue ($p = 1.57 \times 10^{-79}$, n = 456 dendrites, 5 mice, paired t-test, **Fig. 3E**) and the first lick after reward delivery ($p = 4.24 \times 10^{-81}$, paired t-test, **Fig. 3E**). Moreover, the cue-associated Cspk response occurred at a significantly longer latency than in normally trained animals (GtACR2 Cspk latency = 384.72 ± 2.31 ms; control Cspk latency = 181.03 ± 1.98 ms; $p = 2.14 \times 10^{-243}$, unpaired t-test). This intermediate response suggests an incomplete shift of Cspk activity from the US to CS. Thus, we continued training without optical inhibition to test whether normal training would generate responses similar to those in control animals.

Following 8 additional days of normal training, we observed Cspk activity in response to the auditory cue ($p = 3.26 \times 10^{-96}$, paired t-test, n = 539 dendrites, 5 mice, **Fig. 3F**) that was significantly larger than it had been immediately after training with optogenetic inhibition (Post opto training: 2.45 ± 0.08 Hz; post additional training, 2.72 ± 0.09 Hz, p = 0.03; unpaired t-test, **Fig. 3F**). Moreover, peak Cspk activity occurred earlier, in closer proximity to the auditory cue (280.58 ± 2.02 ms; $p = 2.21 \times 10^{-169}$, paired t-test), and the US-associated Cspk response was largely abolished (vs post optical training, $p = 6.55 \times 10^{-14}$, unpaired t-test, **Fig. 3F**). Together, these results suggest that Cspk activity in response to the US contributes to the timing of the learned Cspk activity in response to a CS and the abolition of the Cspk activity in response to the US.

Climbing fiber responses are consistent with variable reward expectations across trials

On average, Cspk activity is consistent with learned reward expectation in this task. However, reward expectation can also vary across trials, affording another means to test the relationship between behavior and neural activity. We therefore examined how Cspk activity varies across trials, leveraging both the responses immediately after optogenetic training and those after subsequent normal training. To do so, we segregated trials into predictive (pre-reward) and reactive (post-reward) licking.

Following optogenetic training, on trials with predictive licking nearly all individual PCs (93%) had Cspks in response to the cue (n = 423 dendrites; **Fig. 4A**). Surprisingly, these cells also had a significant Cspk response prior to the first lick on these trials ($p = 5.76 \times 10^{-96}$, paired t-test, n = 423 dendrites, 5 mice, **Fig. 4B**). However, the timing of this response was less closely associated with lick onset than in naive animals (from time of peak Cspk response to lick onset in trials with predictive licks in GtACR2 trained mice: 294.17 ± 2.27 ms; trials with reactive licks in naive mice: 115.06 ± 2.40 ms; $p = 3.91 \times 10^{-315}$, unpaired, one-tailed t-test, **Fig. 4B** vs **Fig. 3D**). Moreover, the cue-aligned Cspk response was also less closely associated with the auditory cue than it was in the same animals with additional normal training (from time of cue to peak Cspk cue response in GtACR2 trained mice: 436.56 ± 2.43 ms; in additionally trained mice: 316.03 ± 2.12 ms; $p = 5.14 \times 10^{-188}$, **Fig. 4A** vs **4E**). This intermediate response suggests that most PCs show an incomplete / partial transfer of Cspk responses from the time of reward to the reward-predictive cue. In contrast, PCs that did not have Cspk responses to the auditory cue also did not show responses to the first lick after reward on predictive trials (p = 0.19, paired t-test, n = 33 dendrites, **Fig. 4A,B,** grey). Thus, on predictive lick trials, the only response was intermediate, neither tightly linked to the auditory cue or reward delivery.

Cspk responses were quite different on trials with reactive licking. First, far fewer PCs (53%) had a significant Cspk response to the cue (n = 242 dendrites, **Fig. 4C**). Moreover, the amplitude of the Cspk response in those PCs was significantly smaller as compared to trials with predictive licking (p = 1.71 ×

10^{-35}, unpaired t-test, **Fig. 4A** vs **4C**). However, both cue responsive and non-cue responsive PCs showed activity to the first lick after reward on reactive trials (cue-responsive dendrites: $p = 2.09 \times 10^{-49}$, n = 242 dendrites; non-cue-responsive dendrites: $p = 1.02 \times 10^{-33}$, n = 214 dendrites; paired t-tests, **Fig. 4D**). This Cspk activity was more closely associated with lick onset as compared with trials with predictive licking, indicating that it was associated with anticipation or licking (from time of peak Cspk activity to lick onset in trials with reactive licks: 116.74 ± 2.19 ms; with predictive licks: 294.17 ± 2.27 ms; $p = 9.95 \times 10^{-294}$, unpaired t-test, **Fig. 4B** vs **4D**). Thus, while most CFs showed a strong cue response on trials with predictive licking, they did not respond robustly as to the cue on trials with reactive licking, but instead responded as if the reward was unexpected, showing robust responses to prior to lick onset. In this way, Cspk activity is consistent with tracking reward expectation across trials.

To determine whether these across-trial relationships between reward expectation and Cspk activity were influenced by optogenetic manipulation of CFs during training, we compared responses to those observed following additional normal training. In this condition, most dendrites (92%) continued to have significant cue responses on trials with predictive licking (n = 497 dendrites; **Fig. 4E**). However, when aligned to the first lick after reward this Cspk response was significantly further away from lick onset, indicating that it was now fully associated with cue presentation, and no longer intermediate between cue and reward times (from time of peak Cspk activity to lick onset in trials with predictive licks in mice with additional training: 467.78 ± 2.02 ms; $p = 1.77 \times 10^{-246}$, unpaired, one-tailed t-test, **Fig. 4F vs 4B**). Additionally, on trials with reactive licking a much larger response had developed to the cue ($p = 1.61 \times 10^{-5}$, unpaired, one-tailed t-test, **Fig. 4G vs 4C**). Thus, the learned cue response was considerably more prominent on both trial types after normal training.

Together, these data further reveal how optogenetic inhibition impairs the development of learned Cspk responses. Moreover, they demonstrate that 1) PCs preferentially receive CF input on trials where licking will be predictive, and 2) CFs only respond to licking when it is reactive, again suggesting that these responses may be related to expectation. Thus, Cspk responses appear to track expectation across trials, in addition to across learning.

Reward-associated Cspk activity influences learning

Our data suggest the Cspk activity can reflect reward expectation, but do these reward-related responses influence behavior or learning? To test this, we quantified lick responses immediately after optogenetic training, again after subsequent normal training in the same animals, and in separate animals with no optogenetic manipulation (**Fig. 5**). Following optical inhibition of IO neurons across training, we observed that licking was reactive on many trials, and frequently did not peak until well after reward delivery (**Fig 5A-B, Supp. Fig. 4**). In contrast, licking was predictive on the majority of trials following subsequent normal training, peaking before reward delivery (**Fig. 5C, Supp. Fig. 4**).

On average, following optogenetic training, mean reaction times occurred after reward delivery, favoring reactive licking ($p = 0.09$, n = 5 mice, Student's one-tailed t-test, **Fig. 5D**). These reaction times were significantly longer as compared with the same animals following subsequent normal training ($p = 0.027$, paired t-test, one-tailed, n = 5 mice) and as compared with animals trained without optogenetic inhibition ($p = 0.007$, unpaired t-test, one-tailed, n = 6 mice). There was also a trend toward higher miss rates following optogenetic training, though these were not significantly different from subsequent normal training or controls (optotrained vs subsequent normal training, $p=0.14$, paired t-test, one-tailed; vs control; $p = 0.09$, unpaired t-test, one-tailed Fig. **5E**) In contrast, the percentage of trials with predictive licking (46 ± 12%) was dramatically lower than in the same animals with subsequent normal training (74 ± 12%, $p=0.003$, paired t-test, one-tailed, **Fig. 5F**), or the control animals (69 ± 7%; $p = 0.055$, unpaired t-test, one-tailed, **Fig. 5F**). Conversely, the percentage of trials with reactive licking following optogenetic training (48 ± 11%) was markedly higher than either the same animals with subsequent normal training (24 ± 12%, $p=0.007$, paired t-test, one-tailed, **Fig. 5F**) or control animals (24 ± 7%, $p = 0.050$, unpaired t-test, one-tailed, **Fig. 5F**). The large change in the number of predictive responses without a statistically significant change in miss rates suggests an impaired learning of predictive timing, but not an impairment in the ability to initiate licks as a result of CF inhibition during training. Together,

these results support a causal role for the climbing fiber response at the time of the reward in establishing a well-timed conditioned licking response to the predictive cue.

Discussion

Our data reveal multiple features of cerebellar climbing fiber signaling that suggest a broad role in reward processing and reward-based learning. First, we find that cerebellar CFs can generate learned, reward-predictive responses to auditory stimuli in this task, extending previous findings by indicating that CFs can learn sensory predictions for additional modalities. Second, we find that these learned reward predictions exhibit a key property in common with reward prediction error (rPE) signals in the VTA, in that they cannot generate learned responses to a redundant sensory cue. This so called 'stimulus-blocking effect' has important consequences for learning, and suggests that the learned CF signals described here are not simply related to the salience of the cue, or its temporal association with reward. Third, our data also suggest a role for CF responses to unexpected reward in generating learned responses to reward-predictive cues, as optogenetically reducing CF activity impairs the transition from the US to the CS. And finally, our data also reveal a role for reward-related CF signals in learning and behavior, as optogenetic suppression of CF reward signals impairs the learning of predictively timed licking.

Previous work showed that CF responses can shift from the US to a visual CS (Heffley and Hull, 2019), and that CFs can respond to visual (Heffley et al., 2018) and auditory (Kostadinov et al., 2019) stimuli in mice trained on reward-based operant tasks. Notably, the learned CF response to an auditory cue in this study occurred with shorter latency and a larger amplitude as compared to previously measured visual responses (Heffley and Hull, 2019). We speculate that this may result from either the longer training period in this study or an enhanced salience or faster transduction time of the auditory stimulus as compared to the visual stimulus**.**

As with previous results, the US and CS-related CF responses could be disambiguated from neural activity associated with licking or body movement (Heffley and Hull, 2019; Heffley et al., 2018). Thus, these data further support the conclusion that CF responses are related to the US in the naive condition and the CS in the learned condition. However, our current results also extend previous findings by explicitly testing whether these CF responses could simply be due to salience or simple a reward-related association. Reinforcement learning requires the detection of a discrepancy between the expectation and receipt of reward (Keiflin and Janak, 2015; Schultz et al., 1993; Schultz et al., 1997; Steinberg et al., 2013). This attribute enables the learning of reward-predictive stimuli, but also prevents the learning of any stimulus that does not provide novel information about reward. Accordingly, simple pairing between stimulus and a reward is not sufficient for learning to occur (Schultz et al., 1997; Steinberg et al., 2013; Waelti et al., 2001). By demonstrating that a visual stimulus that is redundant with a previously learned, reward-predictive auditory stimulus cannot be learned, our data are in agreement with what has been shown in other reinforcement learning systems (Steinberg et al., 2013; Waelti et al., 2001). Further, these data rule out the possibility that CF responses result only from the repeated exposure to a stimulus or its temporal proximity to reward.

Importantly, there remain other properties of reinforcement learning signals that have yet to be tested in the cerebellum. Indeed, even for midbrain dopamine neurons, substantial debate continues about the complex features of dopamine signaling, and how to reconcile the varied responses of dopamine neurons with the formal predictions of modern reinforcement learning theories such as temporal-difference learning (Bromberg-Martin et al., 2010; Gershman and Uchida, 2019). Thus, while we conclude that at least some important features of CF activity are well-aligned with the predictions of reinforcement learning, much work remains to determine whether, when and how the cerebellum might participate in reinforcement learning.

Beyond the debate surrounding the computational properties of these CF responses, the circuit mechanism(s) that generate reward-driven and reward predictive CF activity also remain unclear.

Broadly, CF reward-related responses could either be computed locally in the cerebellum or inherited from upstream brain regions. In this study, our optogenetic experiments indicate a contribution of cerebellar-olivary circuits in generating learned, reward-predictive CF responses. By inhibiting CF activity to unexpected reward, we observed an impaired transfer of CF signaling from the US to the CS. Notably, this manipulation did not block the generation of a CS response, but rather reduced its amplitude and temporal specificity, while at the same time allowing the US response to remain intact. This result may be due to incomplete block of CF activity, or reflect a requirement for more widespread circuit activity in upstream regions to enable learned CF responses.

Indeed, previous studies have shown that the inferior olive receives descending inputs from multiple regions of the neocortex (Sugihara and Shinoda, 2004; Veazey and Severin, 1982; Wang et al., 2022). There is also evidence of polysynaptic input to the olive from the output of the basal ganglia (Saint-Cyr and Courville, 1982). However, there remains no clear evidence of projections from the VTA, despite evidence that the cerebellum may projects directly to the VTA (Carta et al., 2019) and SNc (Washburn et al., 2024). Thus, current evidence suggests that the cerebellum may either act in parallel with midbrain dopamine neurons or contribute to their reward-predictive responses, but may not directly inherit reward-predictions from them.

The IO also receives inputs from the cerebellum that could contribute to CF reward computations. For example, excitatory inputs from the mesodiencephalic junction (MDJ) of the midbrain (De Zeeuw et al., 1998) receive signals from the deep cerebellar nuclei (Onodera, 1984), and also may integrate these with input from cortical pyramidal tract neurons (Veazey and Severin, 1982). The IO also receives inhibitory inputs from the deep cerebellar nuclei, and has interneurons that may facilitate local computations (Bengtsson and Hesslow, 2006; De Zeeuw et al., 1998). Indeed, the inhibitory projection from the cerebellar nuclei to the IO has been shown to be necessary for suppressing behavioral responses to a redundant CS in eyelid conditioning (Kim et al., 1998). Thus, it is possible that this projection may play a role in suppressing reward responses once they are expected. However, it remains unclear if this pathway might play any role in learned CS responses, and future work will be necessary to test the circuit pathways and mechanisms responsible for such responses.

Our optogenetic experiments also suggest a key role for reward-driven CF activity in generating well timed motor output. Specifically, we find that inhibiting the IO during learning impairs the ability to generate predictively timed licking in our task. These results are consistent with recent evidence linking the lateral cerebellum with learned, predictive licking (Bina et al., 2021; Chabrol et al., 2019; Gaffield et al., 2022; Tsutsumi et al., 2019), and extend these findings by indicating a role for CF reward-related activity in learning. Here, we found that the effect of IO inhibition was specific to predictively timed learned responses, and did not affect lick rates outside of the task window. Further, these same animals learned normally in absence of IO inhibition. Thus, we conclude that reward-related CF signals can act as teaching signals for learning, as has been extensively demonstrated for traditional error driven CF signals in other tasks (Raymond and Medina, 2018). More broadly, these results are consistent with CF driven learning and cerebellar circuits contributing to predictive behaviors and anticipatory motor output (Hull, 2020). Overall, our results suggest that cerebellar CFs can exhibit several properties in common with the reinforcement learning signals that have been characterized in the dopamine system (Schultz et al., 1997; Waelti et al., 2001), and can play an important role in predictive learning.

Acknowledgements:
This work was supported by grants from the following NIH institutes: NINDS 5R01NS096289 (CH), NINDS R01NS112917 (CH), NINDS R01-NS128054 (CH). Funding was also provided by the Ruth K. Broad Foundation (SJ). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We thank Drs. Stephen Lisberger, Lindsey Glickfeld, Fan Wang, and Josh Huang for input throughout the project, and Mary Anne Hughes, Ziye Xu, and Wenjuan Kong for technical support.

Author Contribution Statement:
C.H. and S.J. designed experiments. S.J. conducted experiments and analyzed data. C.H. and S.J. wrote and edited the manuscript.

Competing Interests:
The authors declare no competing interests.

Data Availability:
Due to the large size of these datasets, data that support the findings of this study are available from the corresponding author upon request.

Code Availability:
Code that supports the findings of this study are available at: https://github.com/Glickfeld-And-Hull-Laboratories/ImagingCode-Glickfeld-Hull

Figure Legends

Figure 1. Pavlovian task to associate reward with a predictive auditory cue (CS1). **A)** Schematic of experimental design. Mice were presented with an auditory tone (200 ms, 5 kHz, 75 dB) prior to reward (700 ms after tone onset). **B,C)** Summary histograms of mean lick rates across experiments in naive (**B**) and trained (**C**) mice (n=6 mice). Dashed lines represent the time of cue and reward respectively. **D)** Summary of the fraction of trials across mice with predictive licks (orange), reactive licks (gray), or no licks (misses, blue) during the trial. **E)** Peristimulus time histogram (PSTH) of complex spike (Cspk) rate aligned to cue onset in naive animals (n=436 dendrites). Inset, Cspk rate aligned to the first lick after cue presentation **F)** Same as E) but for trained mice (n=427 dendrites). **G)** PSTH of Cspk rates in trained mice for trials with the earliest and latest licks. Dots represent mean onsets for earliest (black) and latest (blue) 20% of lick bursts. **H)** Same as C) for trials with the highest (black) and lowest (blue) 20% of pre-reward lick rates. **I)** PSTHs of mean piezo voltage traces reflecting the 20% of trials with the most (black) and least (blue) animal movement **J)** Same as H) for Cspk rate on trials with the most (black) and least (blue) animal movement. All data are presented as mean +/- SEM.

Figure 2. Cspk responses after training with a stimulus blocking paradigm. **A)** Task design (left) and training sequence (right) for mice initially trained on auditory CS1, and subsequently trained on the simultaneous presentation of auditory CS1 and visual CS2. **B)** Summary histogram of mean lick rates across mice aligned to cue onset in CS1+CS2 trained animals for CS1+CS2 presentation after learning (n=6 mice). Dashed lines represent the time of cue and reward respectively. **C)** Same as B) for CS1 presentation. **D)** Same as B) for CS2 presentation. **E)** PSTH of Cspk rate aligned to cue onset in CS1+CS2 trained animals for CS1+CS2 presentation after learning (n=398 dendrites). Inset, Cspk rate aligned to the first lick after cue presentation. **F)** Same as E) but for CS1 presentation (n=385 dendrites). **G)** Same as E) but for CS2 presentation (n=353 dendrites). All data are presented as mean +/- SEM.

Figure 3. Cspk responses across training with optical inhibition of the contralateral IO. **A)** Task design for mice initially trained with GtACR2 inhibition of the IO during CS1 presentation, and subsequently trained with CS1 and no IO inhibition. **B)** Single photon imaging of bulk Cspk responses across mice (n=5 mice) during optical inhibition via fiber implant above the IO. **C)** Summary of mean Cspk response amplitude in individual mice (gray, n=5 mice) and across animals (black) in control and during optical IO inhibition. **D)** PSTH of Cspk rate aligned to cue onset for naive mice prior to GtACR2 training. Inset, Cspk rate aligned to the first lick after cue presentation (n=363 dendrites). **E)** Same as D) but in the same mice after training with IO inhibition (n=456 dendrites) **G)** Same as E) but in the same mice after additional training with CS1 (and no IO inhibition) (n=539 dendrites). All data are presented as mean +/- SEM.

Figure 4. CF responses are consistent with reward expectation across trials. **A)** PSTH of Cspk rates on trials with predictive licking for PC dendrites that were cue responsive (black, n=423 dendrites) and non-cue responsive (grey, n=33 dendrites) aligned to cue onset in animals immediately after CS1 training with GtACR2 inhibition of the IO (n=5 mice). **B)** Same as A), but aligned to lick onset. **C)** PSTH of Cspk rates on trials with reactive licking for PC dendrites that were cue responsive (black, n=242 dendrites) and non-cue responsive (grey, n=214 dendrites) aligned to cue onset in animals immediately after CS1 training with GtACR2 inhibition of the IO. **D)** Same as C), but aligned to lick onset. **E)** PSTH of Cspk rates on trials with predictive licking for PC dendrites that were cue responsive (black, n= 497) and non-cue responsive (grey, n= 42 dendrites) aligned to cue onset in animals immediately after additional CS1 training (no IO inhibition, n=5 mice). **F)** same as E), but aligned to lick onset. **G)** PSTH of Cspk rates on trials with reactive licking for PC dendrites that were cue responsive (black, n= 237 dendrites) and non-cue responsive (grey, n=302 dendrites) aligned to cue onset in animals immediately after additional CS1 training (no IO inhibition, n=5 mice). **H)** same as G), but aligned to lick onset. All data are presented as mean +/- SEM.

Figure 5. Training with IO inhibition impairs learned lick timing. **A)** Example lick raster, aligned to cue onset, for a mouse prior to CS1 training. **B)** Same as A, for the same mouse after CS1 training with GtACR2 inhibition of the IO. **C)** Same as A), for the same mouse after additional training with normal CS1 training (no IO inhibition). **D)** Summary of mean reaction times for the same mice before (naive), after GtACR2 training (post opto), and after additional CS1 training (+8 days CS1) (n=5 mice), as well as in a separate cohort that underwent only normal training (control, learned, n=6 mice). **E)** Same as D), for miss rates. **F)** Summary of the fraction of trials across mice with predictive licks (orange), reactive licks (gray), or no licks (misses, blue) during the trial for the same groups as in D). All data are presented as mean +/- SEM.

Supp. Figure 1, related to Figure 1. Summary of auditory CS1 behavioral training and example Cspk responses quantified by normalized GCaMP fluorescence ($\Delta F/F$). **A)** *Left*, mean reaction time, defined as the first lick burst following the auditory cue, across animals (n=6 mice). *Right*, summary of mean reaction times before and after learning. **B)** *Left*, Mean miss rates, defined as the fraction of trials with no licks within 1.7 s after the cue, across animals. *Right*, Summary of mean miss rate before and after learning. **C)** Normalized GCaMP responses from an example segmented PC dendrite aligned to cue onset (left) and lick onset (right) for individual trials (gray, n = 100) and averaged across trials (black). **D)** Same as C), but for an example segmented PC dendrites (n=100 trials) in the same mouse after CS1 training. All data in A) and B) are presented as mean +/- SEM.

Supp. Figure 2, related to Figure 2. Summary of trial outcome and behavioral training for stimulus blocking paradigm. **A)** Summary of the fraction of trials across mice with predictive licks (orange), reactive licks (gray), or no licks (misses, blue) during the trial (n=6 mice). **B)** Mean reaction times, defined as the first lick burst following the cue, across animals. orange=CS1 presentation, blue=CS1+CS2 presentation, green=CS2 presentation. **C)** Same as B), for mean miss rates, defined as the fraction of trials with no licks within 1.7 s after the cue, across animals (n=6 mice). All data are presented as mean +/- SEM.

Supp. Figure 3, related to Figure 3. Expression of GtACR2 in in the IO, and effect of GtACR2 activation on licking. **A)** Coronal slice showing IO expression of fusion red (red) associated with GtACR2 expression following injection of pAAV-CKIIa-stGtACR2-FusionRed and DAPI staining (blue). Dotted line represents the outline of the brain section, and solid box represents the expanded field of view in B). **B)** Expanded field of view from A) showing expression in the inferior olive. Dotted lines represent IO subdivisions; IOA: subnucleus A of the medial nucleus; IOB: subnucleus B of medial nucleus; IOBe: beta nucleus of the medial nucleus; IOC: subnucleus C of medial nucleus; IOD: dorsal nucleus; IOK: cap of Kooy of the medial nucleus. **C)** *Left*, summary histogram of mean lick rates across mice expressing GtACR2 in the IO, aligned to laser onset (laser stim = 1.5 mW, 5 s, n=5 mice). *Right,* same as for C) but aligned to laser offset. Data are presented as mean +/- SEM.

Supp. Figure 4, related to Figure 5. Lick rasters for all mice that underwent training with GtACR2 inhibition of the IO. Lick rasters, aligned to cue onset, for each mouse (individual rows) prior to training with GtACR2 inhibition of the IO (left), after CS1 training with GtACR2 inhibition of the IO (middle) and after additional training with normal CS1 training (right).

Materials and Methods

Animals

All experimental processes involving mice were approved by the Duke Institutional Animal Care and Use Committee. Mice were singly or group-housed (<= 4 per cage) on 12-hour light/dark cycles. Experiments were performed using adult mice (> P65) randomly selected from both sexes. Imaging and optogenetic experiments were carried out using Tg(PCP2-Cre)3555Jdhu mice (Jackson Labs, 010536).

General Surgical Procedures

All surgical procedures were performed with an aseptic technique. Mice were injected with dexamethasone (3 mg/kg, or 4 mg/ml SQ) three to 24 hours before surgery. Ketamine (50 mg/kg) and xylazine (5 mg/kg) were administered by intraperitoneal injection for initial anesthesia, and then isofluorane (1-3%) was used to maintain the mice being anesthetized during surgery. A heating pad (TC-1000 CWE Inc.) was used to keep the body temperature at 37° Celsius.

Surgical Procedures for Imaging Experiments

The surgical procedures for craniotomy and virus injection were similar as stated by Heffley et al. (2018). Specifically, a headplate (HE Parmer) was mounted to the skull with C&B Metabond (Parkell), then a 3 mm craniotomy was prepared over lobule simplex of the right cerebellar hemisphere (~ 2.8 mm posterior and 1.4 mm lateral to lambda). To express GCaMP in Purkinje cell dendrites of lobule simplex, a glass micropipette was filled with virus AAV1.CAG.Flex.GCaMP6f.WPRE.SV40 (UPenn vector core, titer = 1.2×10^{13}) or AAV1.CAG.Flex.GCaMP7f.WPRE.SV40 (Addgene 104496, titer = 1.5×10^{13}) diluted 1:4 – 1:20 in ACSF (titer for injection: 7.5×10^{11}). In each targeted lobule, 150-200 nL virus was injected at 2-4 sites using the following depths: 500 µm, 380 µm, 250 µm, and 100 µm. The injection rate was 30-50 nL/min. For a subset of animals with a craniotomy over lobule simplex, the virus injection was performed at an approach angle of 40 degrees with the same injection rate, volume, and depths (Kostadinov et al., 2019). A glass window consisting of two 3 mm and a 5 mm coverslip (Warner Instruments No. 1) was implanted using C&B Metabond.

Surgical Procedures for Optogenetics Experiments

To express an inhibitory opsin in the inferior olive, the skull was leveled, and the following coordinates were used for the burrhole: x = 0.4 mm × (distance between bregma and lambda/6.85 mm); y = 6.85 mm × (distance between bregma and lambda/6.85 mm). Virus AAV1.αCaMKII.stGtACR2-FusionRed (Addgene 105677-AAV1, titer = 1.2×10^{13}) was injected into the brainstem at a depth of 5.8 mm and 5.4 mm with a rate of 40-60 nL/min. The injection volume for each depth was 300 nL. An optical fiber (CFMLC21U-20, 105 µm, 0.22 NA, 6 mm in length, Thorlabs; or R-FOC-BL100C-22NA, 100 µm, 0.22 NA, 6.5 mm in length, RWD Life Science Co) was then implanted using the same coordinates using C&B Metabond. Mice were then headposted with a headplate (HE Parmer). After seven to 10 days of recovery, a second surgery including craniotomy over lobule simplex and GCaMP injection was done.

After the surgery, all animals were post-operatively monitored by our lab personnel for 5 days, and cefazolin (50 mg/kg) and buprenex (0.05 mg/kg) were injected every 12 hours for 2 days. At least a full week of recovery period was allowed before further procedures.

Behavior

All mice were habituated to head restraint for at least 5 days (10 min – 1h/day) before training and imaging. Imaging mice were allowed for at least 13 days after cranial injections to allow sufficient

expression levels of GCaMP. Training and imaging sessions are shorter than 90 minutes to keep the animals engaged.

Pavlovian conditioning paradigm

Mice were water deprived for at least 3 days until body weight was reduced and stabilized at around 85% of the original weight. All behavioral training occurred during the light cycle. Animals were head-fixed facing a reward delivery tube and a monitor. To measure body movement, a piezoelectric sensor (C.B. Gitty, 41 mm 'jumbo' piezo) was placed beneath the animal, and the time course of voltages was collected through MATLAB (Mathworks).

During the stimulus-blocking experiment, mice were first trained to associate an auditory cue (5 kHz, 75 dB, 200 ms tone) with a saccharine reward (approximately 2.5 µl) that was delivered 777 ms after the cue onset. Behavioral parameters including the onset of cue presentation, time of reward delivery, and time of licks were collected through MWorks (http://mworks-project.org) and MATLAB (Mathworks). Training of the auditory cue consisted of 1089.3 ± 103.6 trials (n = 9 ± 0.7 sessions, n = 6 mice), with an inter-trial interval (ITI) of 12 – 30 s. Learning of the association between the tone and the reward was defined by a stabilized reaction time of shorter than 650 ms since cue onset and miss rates below 15%. Once the mice learned the association, the same cohort of mice was then trained under a paired-stimuli condition. In this condition, the animals were presented with both a visual cue (600 ms, high contrast black and white vertical grating) and the same tone simultaneously during training, and were trained for a similar number of trials that took for the animals to learn the tone. Behavioral responses to the paired stimuli, auditory cue only, and visual cue only were then tested respectively and orderly in the imaging sessions following training. Training of the paired stimuli consisted of 1006 ± 43.3 trials (n = 8.5 ± 0.2 sessions, n = 6 mice).

During the training sessions with optogenetic suppression of the inferior olive, a 450 nm blue laser (Optoengine, MGL-III-450) was coupled to the optical fiber using a black ceramic sleeve (R-MS-1.25, 1.25 mm, black, RWD Life Science Co). The laser was turned on 400 ms before the cue onset, and terminated 4s after the reward delivery. A laser power of <1 mW was used. The black ceramic sleeve blocked the blue light and avoided the animals' detection of the light delivery. The training with optogenetic suppression and auditory tone consisted of 742.2 ± 57.5 trials (n = 8 ± 0.5 sessions, n = 5 mice), and the training with only auditory tone consisted of 795.2 ± 64.9 trials (n = 8 ± 0.5 sessions).

Calcium Imaging

Single-photon imaging

A customized microscope (Sutter SOM) and a CMOS camera (Qimaging, Rolera em-c^2) were used for single-photon imaging. Excitation (470 nm) of the GCaMP6f or GCaMP7f was from an LED (ThorLabs, M470L3) through a 5x or 4x objective (Mitutoyo, 0.14NA), and the fluorescence signals were collected by a green filter (520 nm center wavelength, 25 mm diameter, 36 nm bandwidth, OD 6, Edmund Optics 67-030) with a 10 Hz frame rate. The field of view (FOV) was 3.5 × 3.5 mm (1002 × 1004 pixels).

Optogenetics during single-photon mesoscale imaging

In each trial, blue light was delivered to the inferior olive by a 450 nm blue laser (Optoengine, MGL-III-450) coupled to the optical fiber using a black ceramic sleeve (R-MS-1.25, 1.25 mm, black, RWD Life Science Co). To match the behavior paradigm, the laser was turned on for 5 s, and the ITI was 30 s. The imaging sessions consisted of 15 trials, and a laser power of <1.5 mW was used.

Two-photon imaging

A resonant scanning microscope (Neurolabware) and a 16x objective (Nikon CFI75 LWD 16xW 0.80NA) were used to measure *in vivo* Ca^{2+} signals. A polymer (MakingCosmetics, 0.4% Carbomer 940) was used to preserve the immersion solution during two-photon imaging. A 920 nm excitation light from a Ti:Sapphire laser (Spectra Physics, Mai Tai eHP DeepSee) was raster scanned onto the brain at a 30

Hz frame rate via a resonant galvanometer (8 kHz, Cambridge Technology). Emitted photons were collected through a green filter (510 ± 42 nm band filter (Semrock)) onto GaAsP photomultipliers (H10770B-40, Hamamatsu). The microscope was controlled using Scanbox software (Neurolabware), and the FOV was 1030 µm x 581 µm (796 x 264 pixels). The laser power onto the brain was < 40 mW when imaging Purkinje cell dendrites.

Histology

Animals were administered with a cocktail of ketamine/xylazine (200mg/kg & 30mg/kg, respectively, IP) and at least 15 minutes were allowed until breath rate was < 60 bpm. Perfusion was then done transcardially with 10 - 15 ml 1x PBS followed by 4% paraformaldehyde (PFA) in 1x PBS. Immediately after the perfusion, dissection was done to isolate the brain, and the same PFA solution was used for post-fixation for 12 – 48 hrs (4°C). 100 µm sagittal slices were prepared using a vibratome (Pelco 102). For sections that included both the cerebellum and the brain stem, the brain was secured in a 2% agar block during slicing to maintain the section intact. Slices were first stained with DAPI (EMD Millipore Corp, DAPI, Dihydrochloride, 268298-10 mg) and then mounted with mounting medium (Southern Biotech Fluoromount-G) or directly mounted with mounting medium (Southern Biotech DAPI Fluoromount-G). Slides were imaged using a fluorescence (Nikon Eclipse 80i) or confocal microscope (Leica SP8).

Data Analysis and Statistics

Data were analyzed and visualized using MATLAB (Mathworks) code. Unless stated otherwise, data were presented as mean ± S.E.M, and differences were considered statistically significant when $p<0.05$. Data distribution was assumed to be normal but was not formally tested. Analysis code was uniformly applied across experimental conditions.

Behavior

Pavlovian conditioning paradigm

Licks were counted in an ‘all-or-none’ manner using a custom lick sensor (Heffley and Hull, 2019; Heffley et al., 2018), and all lick times during each trial were extracted. Lick bursts were defined as more than 3 licks with a 500 ms window. To determine whether the animals learned the task, licks were aligned to cue onset, and when calculating the following parameters, the time of the cue onset was considered 0 ms. The reaction time was defined as the time it took for the initial lick burst (or initial lick if no lick bursts were observed in a trial) to occur, and was calculated by averaging trials with a first lick or lick burst that occurred 200 ms - 3 s. Predictive licks (conditioned licking response before reward) were defined as the first lick or lick burst between 200 ms and 700 ms. Trials with reactive licks (after reward delivery) were defined as the first lick or lick burst later than 700 ms. Trials with unconditioned licking response were defined as the first lick after 700 ms and before 1700 ms. The miss rate was calculated by dividing the number of trials without a lick within 1.7 s by the total number of trials. For the cohort of mice that were trained with IO photoinhibition, licks were also aligned to the laser offset to determine whether the licking was perturbed by the optogenetic stimulation or the rebound climbing fiber activity after laser offset. Lick rates was measured as the mean lick rates during 1.5 s immediately before and after the cessation of illumination, and was compared across the 2 windows.

Trials were segregated according to lick responses in two different ways. First, to segregate trials according to the lick response latency, times of the initial lick burst (more than 3 licks within a 300 ms window) after the cue onset in each trial were identified, and then trials were divided into quartiles. Second, to segregate trials according to the lick frequency after reward delivery, the lick frequency of each trial was first calculated by dividing the total number of licks by the total length of the lick response window (from 100 ms to 2.5 s after reward delivery), and then trials were segregated into quartiles. To segregate trials according to body movements, mean voltage amplitude during the pre-reward window (from 100 ms to 800 ms after cue onset) was first calculated for each trial, and then trials were segregated into ten percentiles.

Single-photon mesoscale imaging

Lobule identity was identified based on craniotomy. Regions of interest (ROIs) were manually selected within each lobule according to the fluorescence of GCaMP signals. Normalized fluorescence ($\Delta F/F$) was measured using the cumulative fluorescence within each ROI. For the optogenetics during single-photon mesoscale imaging, baseline fluorescence (F) was calculated by averaging the fluorescence across trials during the 1s before the laser onset. To be consistent with the trial structure during the behavioral sessions with optogenetic inhibition, the $\Delta F/F$ during laser on was measured as the mean of 1.1 – 5 s after laser onset.

Two-photon imaging

The extraction of individual Purkinje cell dendrites was performed as stated by Heffley et al. (2018). Briefly, the following steps were taken: (1) Motion correction of the X-Y plane was done by registering image stacks to one reference image made by an averaged 5000-frame image. (2) Signals from individual dendrites were isolated by principal component analysis (PCA) and independent component analysis (ICA). (3) To further segment dendritic signals from the noise, a threshold of 3 - 3.5% was used when smoothing spatial filters from ICA. (4) Nearby pixels with a correlation coefficient > 0.8 were combined into a single dendrite to remove overlaps between dendrites.

To identify events used for the estimation of complex spike rates, deconvolution of the raw fluorescence was conducted on a cell-by-cell basis using MATLAB OASIS (Friedrich et al., 2017). The 'FOOPSI' method was used to account for the shape of the Ca^{2+} transients using GCaMP as a Ca^{2+} indicator. Before identifying complex spike events using deconvolution, Butterworth filter (high pass, order 3, cut off frequency 0.035 Hz) was applied to the raw fluorescence data to remove slow increase of fluorescence over time due to laser artifacts. A threshold of 3 standard deviations above the baseline was used across sessions for each experiment to account for the size of the events. Individual Purkinje cell dendrites were excluded based on the following criteria: (1) The fluorescence was above baseline consecutively for longer than 60 s. These dendrites typically exhibited a high level of fluorescence signals and a series of events that might not be related to climbing fiber activity. (2) The baseline fluorescence was below the level that allows the detection of any events. These cells were typically on a different z-plane or had a low level of GCaMP expression. F was calculated on a cell-by-cell basis, by averaging the fluorescence across trials during 50 frames (1.67s) before the cue onset. Spike rates were aligned to cue onsets or first lick after reward delivery for the mean peri-stimulus histograms (PSTHs).

To determine whether complex spike activity exhibited a significant response and whether the response latency and amplitude was significantly different across conditions, baseline activity of each cell was measured as the mean spike rates during 200 ms immediately before cue onset across trials. The response window for cue was defined as 100-300 ms after cue onset, and the response window for licking was defined as 200 ms immediately before the first lick after reward delivery. The response for each individual Purkinje cell dendrite was measured as the mean spike rates during the response window across trials, and the response latency for each individual dendrite was measured as the average time between the cue onset and the peak of Cspk rate during the response window across trials. The amplitude of Cspk response to cue was measured as the mean of peak Cspk rate across cells during the response window subtracted by the Cspk rate during baseline.

For the cohort of mice trained with optogenetic inhibition, to segregate Purkinje cell dendrites according to their responses to the cue, the baseline $\Delta F/F$ of each dendrite was measured by using the mean $\Delta F/F$ of 500 ms before the cue onset across trials, and the response amplitude was measured by using the peak $\Delta F/F$ during the window from 150 to 500 ms after the cue onset. Dendrites with a response amplitude of at least 3 standard deviations above the baseline were considered cue responsive. To

711 determine whether complex spike activity exhibited a significant response and whether the response
712 latency and amplitude was significantly different across conditions, baseline activity of each cell was
713 measured as the mean spike rates during 300 ms immediately before cue onset across trials. Consistent
714 with previous analysis, the response window for cue in the naïve condition was defined as 100-300 ms
715 after cue onset. Since the time of the peak Cspk rate varies across conditions after training, the
716 response window for cue in all other conditions was defined as 100 ms before and after the time/frame
717 of peak mean Cspk rate across cells after the cue. When all Purkinje cell dendrites were considered,
718 the response window for licking was defined as 200 ms immediately before the first lick after reward
719 delivery. When comparing the time of peak Cspk rate before the first lick after reward of cue-responsive
720 versus non-cue-responsive cells, the response window was defined as 100 ms before and after the
721 time/frame of peak mean Cspk rate before licking onset.
722
723 **References**
724

725 Bengtsson, F., and Hesslow, G. (2006). Cerebellar control of the inferior olive. Cerebellum 5, 7-14.

726 Bina, L., Romano, V., Hoogland, T.M., Bosman, L.W.J., and De Zeeuw, C.I. (2021). Purkinje cells
727 translate subjective salience into readiness to act and choice performance. Cell Rep 37.

728 Bromberg-Martin, E.S., Matsumoto, M., and Hikosaka, O. (2010). Dopamine in motivational control:
729 rewarding, aversive, and alerting. Neuron 68, 815-834.

730 Carta, I., Chen, C.H., Schott, A.L., Dorizan, S., and Khodakhah, K. (2019). Cerebellar modulation of
731 the reward circuitry and social behavior. Science 363.

732 Chabrol, F.P., Blot, A., and Mrsic-Flogel, T.D. (2019). Cerebellar Contribution to Preparatory Activity
733 in Motor Neocortex. Neuron 103, 506-519 e504.

734 Cohen, J.Y., Haesler, S., Vong, L., Lowell, B.B., and Uchida, N. (2012). Neuron-type-specific signals
735 for reward and punishment in the ventral tegmental area. Nature 482, 85-88.

736 De Zeeuw, C.I., Simpson, J.I., Hoogenraad, C.C., Galjart, N., Koekkoek, S.K., and Ruigrok, T.J.
737 (1998). Microcircuitry and function of the inferior olive. Trends Neurosci 21, 391-400.

738 Friedrich, J., Zhou, P., and Paninski, L. (2017). Fast online deconvolution of calcium imaging data.
739 PLoS Comput Biol 13, e1005423.

740 Gaffield, M.A., Bonnan, A., and Christie, J.M. (2019). Conversion of Graded Presynaptic Climbing
741 Fiber Activity into Graded Postsynaptic Ca(2+) Signals by Purkinje Cell Dendrites. Neuron 102, 762-
742 769 e764.

743 Gaffield, M.A., Rowan, M.J.M., Amat, S.B., Hirai, H., and Christie, J.M. (2018). Inhibition gates
744 supralinear Ca(2+) signaling in Purkinje cell dendrites during practiced movements. Elife 7.

745 Gaffield, M.A., Sauerbrei, B.A., and Christie, J.M. (2022). Cerebellum encodes and influences the
746 initiation, performance, and termination of discontinuous movements in mice. Elife 11.

747 Gershman, S.J., and Uchida, N. (2019). Believing in dopamine. Nat Rev Neurosci 20, 703-714.

748 Heffley, W., and Hull, C. (2019). Classical conditioning drives learned reward prediction signals in
749 climbing fibers across the lateral cerebellum. Elife 8.

750 Heffley, W., Song, E.Y., Xu, Z., Taylor, B.N., Hughes, M.A., McKinney, A., Joshua, M., and Hull, C.
751 (2018). Coordinated cerebellar climbing fiber activity signals learned sensorimotor predictions. Nat
752 Neurosci 21, 1431-1441.

753 Hollerman, J.R., and Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction
754 of reward during learning. Nat Neurosci 1, 304-309.

755 Hull, C. (2020). Prediction signals in the cerebellum: beyond supervised motor learning. Elife 9.

756 Keiflin, R., and Janak, P.H. (2015). Dopamine Prediction Errors in Reward Learning and Addiction:
757 From Theory to Neural Circuitry. Neuron 88, 247-263.

758 Kim, J.J., Krupa, D.J., and Thompson, R.F. (1998). Inhibitory cerebello-olivary projections and
759 blocking effect in classical conditioning. Science 279, 570-573.

760 Kostadinov, D., Beau, M., Blanco-Pozo, M., and Hausser, M. (2019). Predictive and reactive reward
761 signals conveyed by climbing fiber inputs to cerebellar Purkinje cells. Nat Neurosci 22, 950-962.

762 Kostadinov, D., and Hausser, M. (2022). Reward signals in the cerebellum: Origins, targets, and
763 functional implications. Neuron 110, 1290-1303.

764 Larry, N., Yarkoni, M., Lixenberg, A., and Joshua, M. (2019). Cerebellar climbing fibers encode
765 expected reward size. Elife 8.

766 Mahn, M., Gibor, L., Patil, P., Cohen-Kashi Malina, K., Oring, S., Printz, Y., Levy, R., Lampl, I., and
767 Yizhar, O. (2018). High-efficiency optogenetic silencing with soma-targeted anion-conducting
768 channelrhodopsins. Nat Commun 9, 4125.

769 Mirenowicz, J., and Schultz, W. (1994). Importance of unpredictability for reward responses in primate
770 dopamine neurons. J Neurophysiol 72, 1024-1027.

771 Onodera, S. (1984). Olivary projections from the mesodiencephalic structures in the cat studied by
772 means of axonal transport of horseradish peroxidase and tritiated amino acids. J Comp Neurol 227,
773 37-49.

774 Raymond, J.L., and Medina, J.F. (2018). Computational Principles of Supervised Learning in the
775 Cerebellum. Annu Rev Neurosci 41, 233-253.

776 Rescorla, R.A., and Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the
777 effectiveness of reinforcement and nonreinforcement. In Classical Conditioning II: Current Research
778 and Theory, B. AH, and P. WF, eds. (New York: Appleton Century Crofts), pp. 64–99.

779 Roesch, M.R., Calu, D.J., and Schoenbaum, G. (2007). Dopamine neurons encode the better option
780 in rats deciding between differently delayed or sized rewards. Nat Neurosci 10, 1615-1624.

781 Saint-Cyr, J.A., and Courville, J. (1982). Descending projections to the inferior olive from the
782 mesencephalon and superior colliculus in the cat. An autoradiographic study. Exp Brain Res 45, 333-
783 348.

784 Schultz, W., Apicella, P., and Ljungberg, T. (1993). Responses of monkey dopamine neurons to
785 reward and conditioned stimuli during successive steps of learning a delayed response task. J
786 Neurosci 13, 900-913.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science 275, 1593-1599.

Sendhilnathan, N., Ipata, A., and Goldberg, M.E. (2021). Mid-lateral cerebellar complex spikes encode multiple independent reward-related signals during reinforcement learning. Nat Commun 12, 6475.

Sineshchekov, O.A., Govorunova, E.G., Li, H., and Spudich, J.L. (2015). Gating mechanisms of a natural anion channelrhodopsin. Proc Natl Acad Sci U S A 112, 14236-14241.

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, 966-973.

Sugihara, I., and Shinoda, Y. (2004). Molecular, topographic, and functional organization of the cerebellar cortex: a study with combined aldolase C and olivocerebellar labeling. J Neurosci 24, 8771-8785.

Sutton, R.S., and Barto, A.G. (1990). Time-Derivative Models of Pavlovian Reinforcement. In Learning and Computational Neuroscience: Foundations of Adaptive Networks, M.G.a.J. Moore, ed. (MIT Press), pp. 497-537.

Sutton, R.S., and Barto, A.G. (1998). Reinforcement learning : an introduction (Cambridge, Mass.: MIT Press).

Tsutsumi, S., Hidaka, N., Isomura, Y., Matsuzaki, M., Sakimura, K., Kano, M., and Kitamura, K. (2019). Modular organization of cerebellar climbing fiber inputs during goal-directed behavior. Elife 8.

Veazey, R.B., and Severin, C.M. (1982). Afferent projections to the deep mesencephalic nucleus in the rat. J Comp Neurol 204, 134-150.

Waelti, P., Dickinson, A., and Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. Nature 412, 43-48.

Wang, X.L., Novello, M., Gao, Z.Y., Ruigrok, T.J.H., and De Zeeuw, C.I. (2022). Input and output organization of the mesodiencephalic junction for cerebro-cerebellar communication. J Neurosci Res 100, 620-637.

Washburn, S., Onate, M., Yoshida, J., Vera, J., Bhuvanasundaram, R., Khatami, L., Nadim, F., and Khodakhah, K. (2024). The cerebellum directly modulates the substantia nigra dopaminergic activity. Nat Neurosci 27, 497-513.

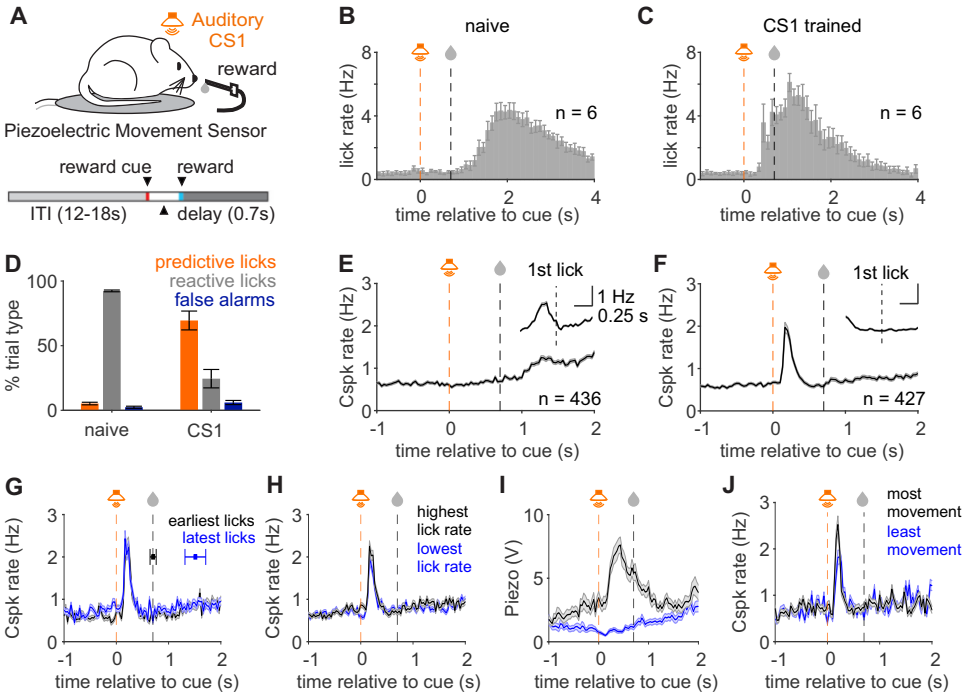

A
Auditory CS1
reward
Piezoelectric Movement Sensor
reward cue
reward
ITI (12-18s)
delay (0.7s)
B
naive
lick rate (Hz)
n = 6
time relative to cue (s)
C
CS1 trained
lick rate (Hz)
n = 6
time relative to cue (s)
D
predictive licks
reactive licks
false alarms
% trial type
naive
CS1
E
1st lick
1 Hz
0.25 s
Cspk rate (Hz)
n = 436
time relative to cue (s)
F
1st lick
Cspk rate (Hz)
n = 427
time relative to cue (s)
G
earliest licks
latest licks
Cspk rate (Hz)
time relative to cue (s)
H
highest lick rate
lowest lick rate
Cspk rate (Hz)
time relative to cue (s)
I
Piezo (V)
time relative to cue (s)
J
most movement
least movement
Cspk rate (Hz)
time relative to cue (s)

Figure 1

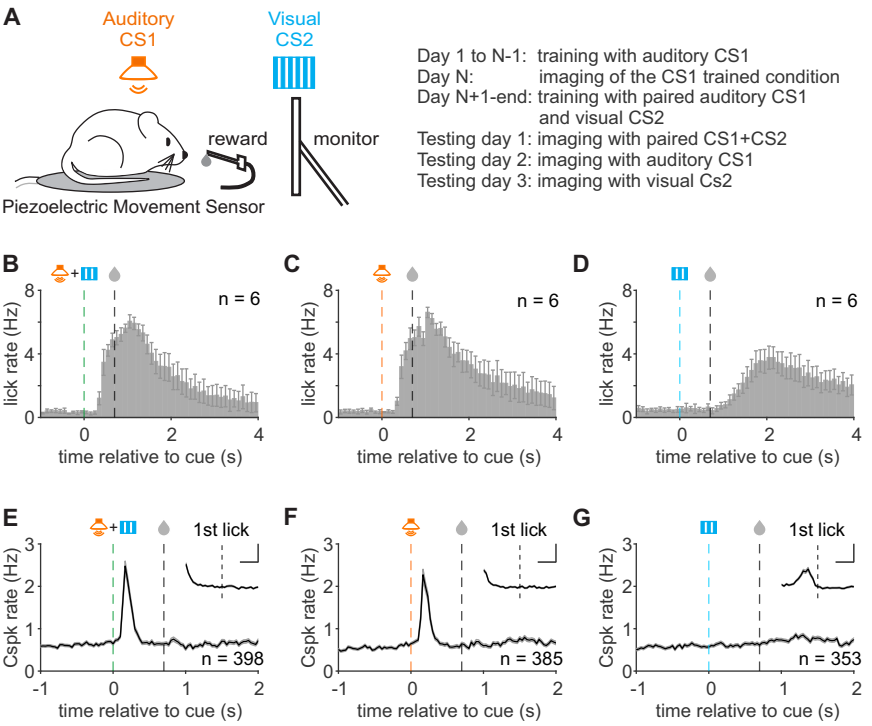

A
Auditory
CS1
Visual
CS2
reward
monitor
Piezoelectric Movement Sensor
Day 1 to N-1: training with auditory CS1
Day N: imaging of the CS1 trained condition
Day N+1-end: training with paired auditory CS1
and visual CS2
Testing day 1: imaging with paired CS1+CS2
Testing day 2: imaging with auditory CS1
Testing day 3: imaging with visual Cs2
B
C
D
lick rate (Hz)
n = 6
time relative to cue (s)
E
F
G
Cspk rate (Hz)
1st lick
n = 398
n = 385
n = 353
time relative to cue (s)

Figure 2

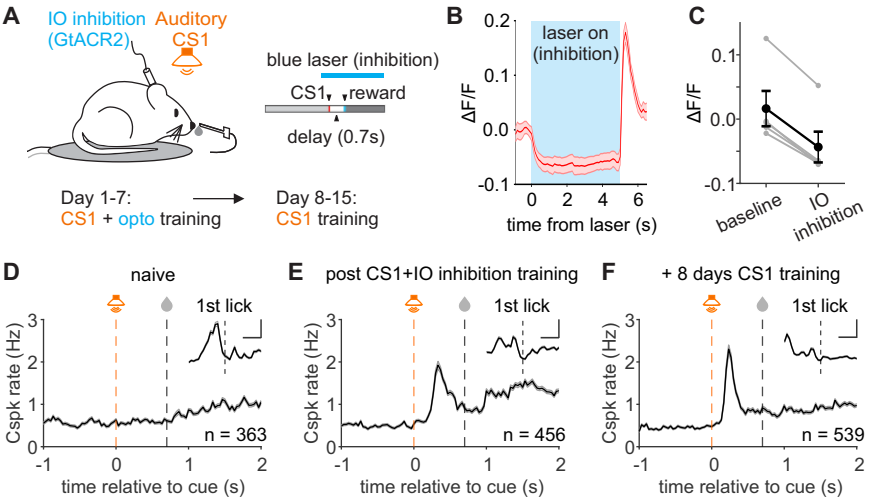

A
IO inhibition (GtACR2)
Auditory CS1
blue laser (inhibition)
CS1
reward
delay (0.7s)
Day 1-7:
CS1 + opto training
Day 8-15:
CS1 training
B
laser on (inhibition)
ΔF/F
time from laser (s)
C
ΔF/F
baseline
IO inhibition
D
naive
1st lick
Cspk rate (Hz)
n = 363
time relative to cue (s)
E
post CS1+IO inhibition training
1st lick
Cspk rate (Hz)
n = 456
time relative to cue (s)
F
+ 8 days CS1 training
1st lick
Cspk rate (Hz)
n = 539
time relative to cue (s)

Figure 3

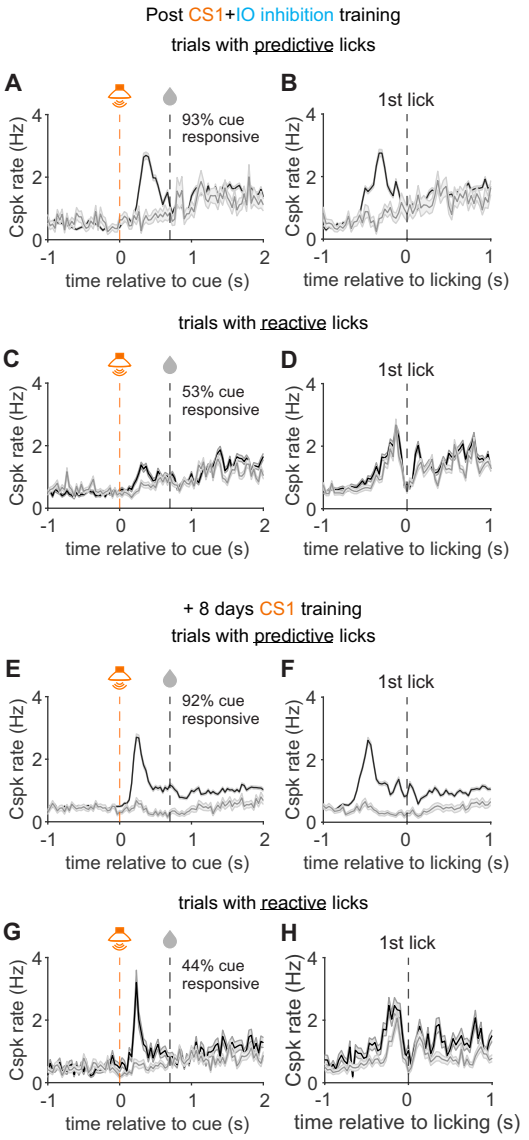

Post CS1+IO inhibition training
trials with predictive licks
A
B
93% cue responsive
1st lick
Cspk rate (Hz)
time relative to cue (s)
time relative to licking (s)
trials with reactive licks
C
D
53% cue responsive
1st lick
+ 8 days CS1 training
trials with predictive licks
E
F
92% cue responsive
1st lick
trials with reactive licks
G
H
44% cue responsive
1st lick

Figure 4

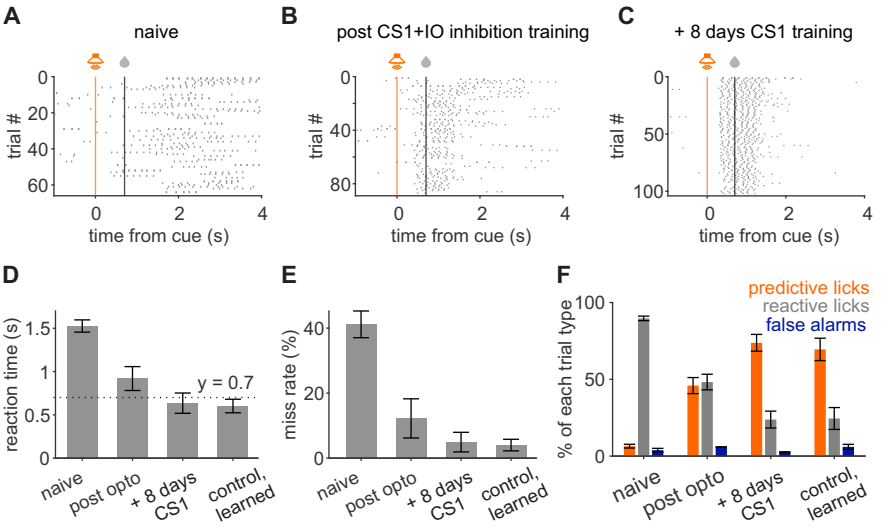
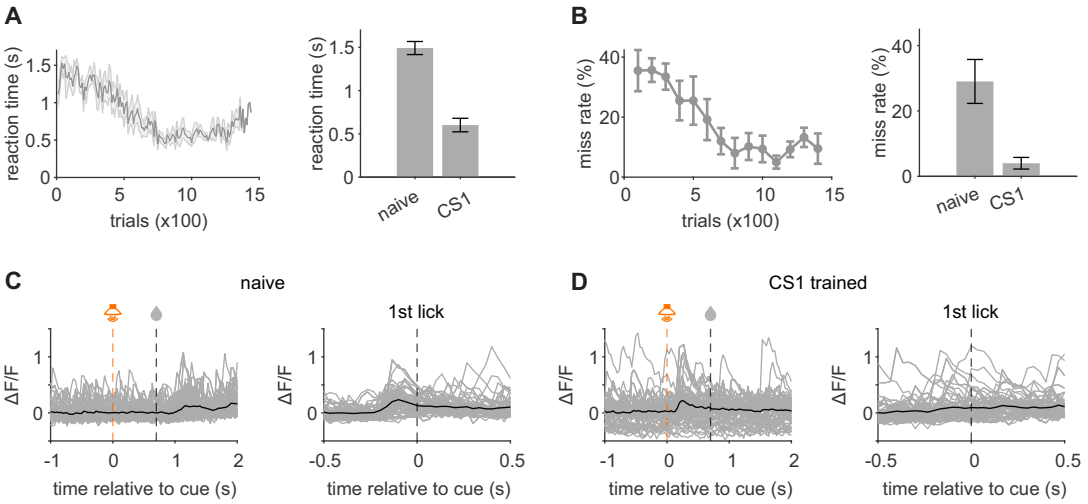
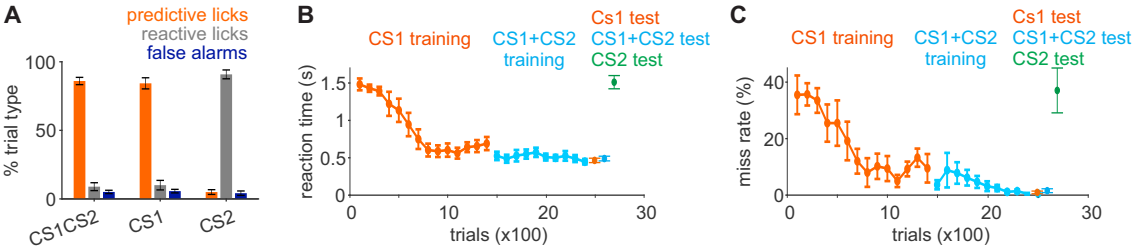

A
naive
trial #
time from cue (s)
B
post CS1+IO inhibition training
trial #
time from cue (s)
C
+ 8 days CS1 training
trial #
time from cue (s)
D
reaction time (s)
y = 0.7
naive
post opto
+ 8 days CS1
control, learned
E
miss rate (%)
naive
post opto
+ 8 days CS1
control, learned
F
% of each trial type
predictive licks
reactive licks
false alarms
naive
post opto
+ 8 days CS1
control, learned

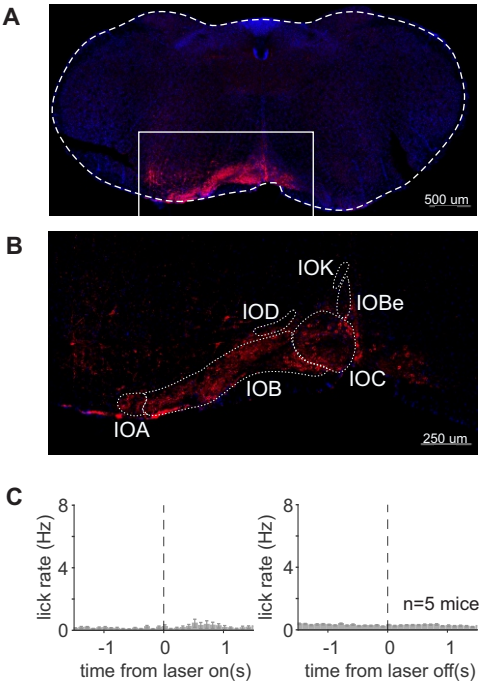
Figure 5



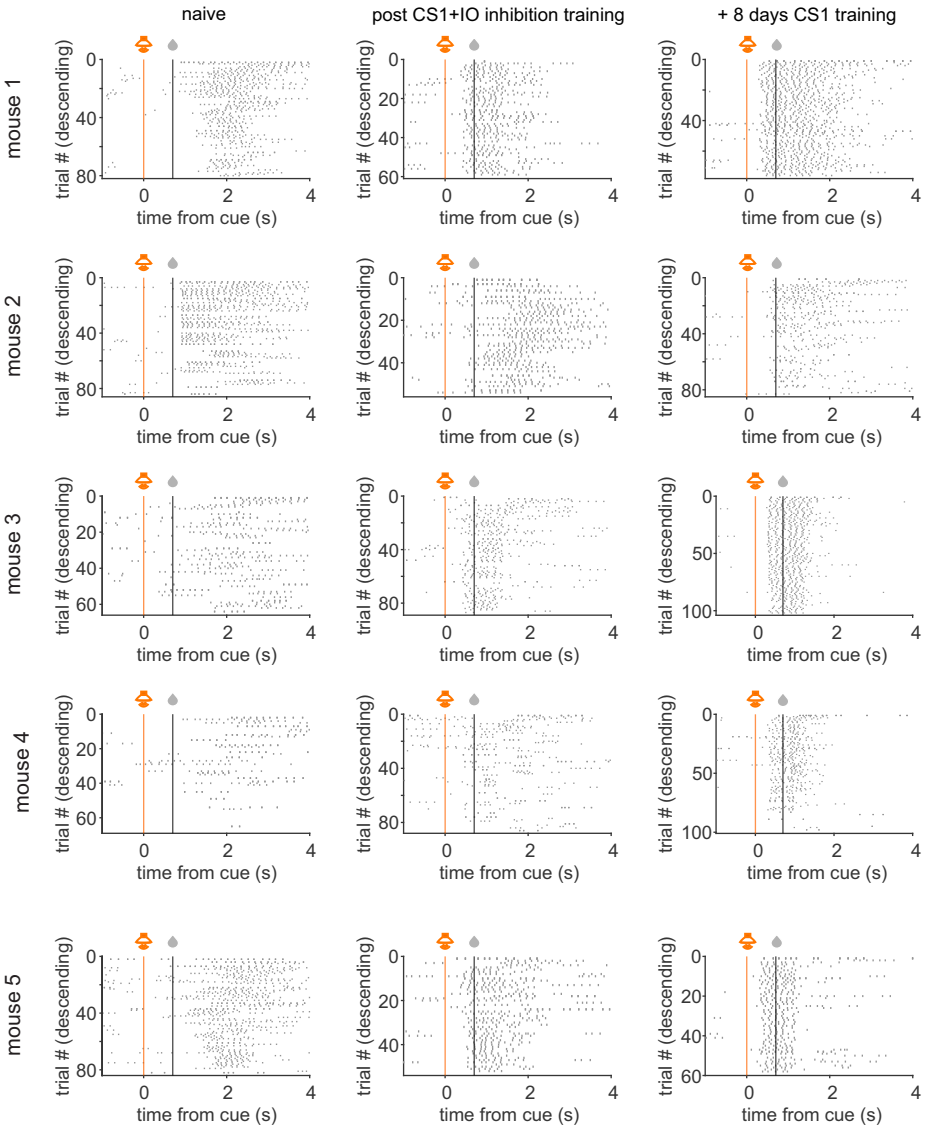
Supplemental Figure 1



Supplemental Figure 2



Supplemental Figure 3



Supplemental Figure 4