

Cognition and Behavior

# Multimodal Temporal Pattern Discrimination Is Encoded in Visual Cortical Dynamics

Sam Post, William Mol, Omar Abu-Wishah, Shazia Ali, Noorhan Rahmatullah, and  Anubhuti Goel<https://doi.org/10.1523/ENEURO.0047-23.2023>

Department of Psychology, University of California, Riverside, Riverside, California 92521

## Abstract

Discriminating between temporal features in sensory stimuli is critical to complex behavior and decision-making. However, how sensory cortical circuit mechanisms contribute to discrimination between subsecond temporal components in sensory events is unclear. To elucidate the mechanistic underpinnings of timing in primary visual cortex (V1), we recorded from V1 using two-photon calcium imaging in awake-behaving mice performing a go/no-go discrimination timing task, which was composed of patterns of subsecond audiovisual stimuli. In both conditions, activity during the early stimulus period was temporally coordinated with the preferred stimulus. However, while network activity increased in the preferred condition, network activity was increasingly suppressed in the nonpreferred condition over the stimulus period. Multiple levels of analyses suggest that discrimination between subsecond intervals that are contained in rhythmic patterns can be accomplished by local neural dynamics in V1.

**Key words:** two-photon; audiovisual temporal patterns; temporal discrimination; temporal learning; visual cortical dynamics

## Significance Statement

Judging whether to stop or go through a yellow light requires determining the duration of the yellow light, and language users must produce sequences of syllables in a temporally structured manner: thus, the ability to tell time is critical. An emerging hypothesis is that local changes in neural activity can contain information about time in the subsecond range. Based on prior human experiments, we have designed a novel timing task for mice and show that mice learn to discriminate between two temporal patterns of audiovisual stimuli. Task performance is accompanied by visual cortical circuit mechanisms. By combining cutting-edge tools with simple behavior, we provide fundamental insight into the neural mechanisms of timing which will also guide future therapies for timing deficits.

Received February 9, 2023; accepted June 29, 2023.

The authors declare no competing financial interests.

Author contributions: S.P. and A.G. designed research; S.P., W.M., O.A.-W., S.A., and N.R. performed research; S.P. and A.G. analyzed data; S.P. and A.G. wrote the paper.

This research was supported by start-up funds and funds from Brain & Behavior Research Foundation.

Acknowledgments: We thank Aaron Seitz, Ladan Shams, Dean Buonomano, and Edward Zagher for discussions that helped improve the paradigm; Bart Kats for help with using the Nautilus clusters for running the SVM analysis; and Karen Kay Mullins for assistance with the mouse colony.

Correspondence should be addressed to Anubhuti Goel at [anubhuti.goel@ucr.edu](mailto:anubhuti.goel@ucr.edu).

<https://doi.org/10.1523/ENEURO.0047-23.2023>

Copyright © 2023 Post et al.

This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

## Introduction

A key aspect of sensory discrimination in learning and memory and in generating complex behavior is extracting temporal features from external stimuli. For example, one may need to keep a beat and synchronize tempo when in a band; a prey may need to jump out of the way of a predator at just the right moment; the timing of a yellow light must be predicted to decide whether to slow down or to go through it; and meaning in spoken language derives from sequences of syllables that are highly temporally structured. Based on psychophysical and pharmacological data, it is most likely that there are multiple neural mechanisms that code for the temporal structure of sensory events since they are timed over a broad range of scales, ranging from microseconds to days (Mauk and Buonomano, 2004; Buhusi and Meck, 2005; Paton and

Buonomano, 2018). However, a growing body of literature suggests that time intervals in the subsecond and second range are encoded in the emergent changing patterns or neural dynamics across many brain areas, including sensory cortex (Pastalkova et al., 2008; Carnevale et al., 2015; Gouvêa et al., 2015; Namboodiri et al., 2015; Goel and Buonomano, 2016; Soares et al., 2016; Bakhurin et al., 2017; Emmons et al., 2017; Heys and Dombeck, 2018; Tsao et al., 2018; Zhou et al., 2020; Tonoyan et al., 2022).

In the traditional hierarchical view of brain organization, the role of primary sensory cortex is to generate a reliable representation of the sensory world, and sensory representations are then decoded by higher-order areas (Hubel and Wiesel, 1962; Felleman and Van Essen, 1991; Miller and Cohen, 2001). Other studies suggest a more active involvement that shapes sensory perception (Glickfeld et al., 2013; Znamenskiy and Zador, 2013); a large body of experimental evidence has now shown that sensory areas contribute to several “higher-order” nonsensory features (Gordon and Stryker, 1996; Shuler and Bear, 2006; Niell and Stryker, 2010; Zhou et al., 2010; Brosch et al., 2011; Zelano et al., 2011; Keller et al., 2012; Gavornik and Bear, 2014a, b; Namboodiri et al., 2015), such as timing and temporal context. Although the locus of temporal predictions and sub-second and second timing has traditionally been attributed to higher-order cortical areas (Leon and Shadlen, 2003; Jazayeri and Shadlen, 2015; Licata et al., 2017) and subcortical areas (Bakhurin et al., 2017; Zhou et al., 2020; Toso et al., 2021), accumulating evidence suggests that primary visual cortex (V1) exhibits response modulation to “higher” functions such as spatiotemporal learning as well as reward prediction and attention (Gavornik and Bear, 2014a). Specifically, Shuler and Bear (2006) showed that as rats perform a visually cued timing task, V1 cortical activity rapidly modulated to predict the arrival of reward. Additionally, cholinergic function contributed to the modification of V1 activity (Chubykin et al., 2013). Namboodiri et al. (2015) used a similar task and showed that indeed cortical activity in V1 reflected the duration of a target interval.

While most studies implement timing tasks as discrete durations or time intervals, temporal structure in sensory stimuli is often organized as sequential events. Sequences may be composed of simple isochronous stimuli and intervals between stimuli, as in rhythms, or complex arrangements of varying stimulus and interval durations, such as in language and music. However, the neural dynamic regimes in the sensory cortex that contribute to processing and learning rhythmic patterns remain largely unclear. One prevailing idea is that neural oscillations allow communication between sensory and motor cortical areas, thus producing temporal predictions and entrainment to rhythms (Merchant et al., 2015). Specifically, do emergent neural dynamics in V1 contribute to learning the temporal structure of rhythmic patterns? To understand how visual cortical dynamics adapt to the temporal structure of a multimodal rhythm in a goal-directed task, we implemented a novel audiovisual (AV) timing task, temporal pattern sensory discrimination (TPSD), in awake behaving mice using two-photon calcium imaging in V1, layer 2/3 (L-2/3). In the TPSD task, mice learn to discriminate between two temporal patterns. Our paradigm

builds on previous work in temporal pattern discrimination, which suggests that multisensory stimuli enhance the discriminability of sequences (Raposo et al., 2012; Barakat et al., 2015). Examination of visual cortex as a locus of change, in an audiovisual task, was influenced by studies showing modulation of visual cortical plasticity by functional input from other brain areas such as hippocampus (Finnie et al., 2021) and auditory cortex (McIntosh et al., 1998; Zangenehpour and Zatorre, 2010; Deneux et al., 2019; Garner and Keller, 2022). Studies have also shown that audiovisual stimuli evoke multimodal plasticity in V1 (Morrell, 1972; Petro et al., 2017).

Here, we show that mice can discriminate between two temporal patterns to achieve expert status on a goal-directed task and that learning was accompanied by robust changes in visual cortical dynamics that reflected the temporal structure of the experienced rhythms. Further, using multiple analyses we show that emergent activity in V1 contributes to trial outcomes. In conclusion, this study underscores the hypothesis that intrinsic network mechanisms contribute to learning and representation of temporal patterns.

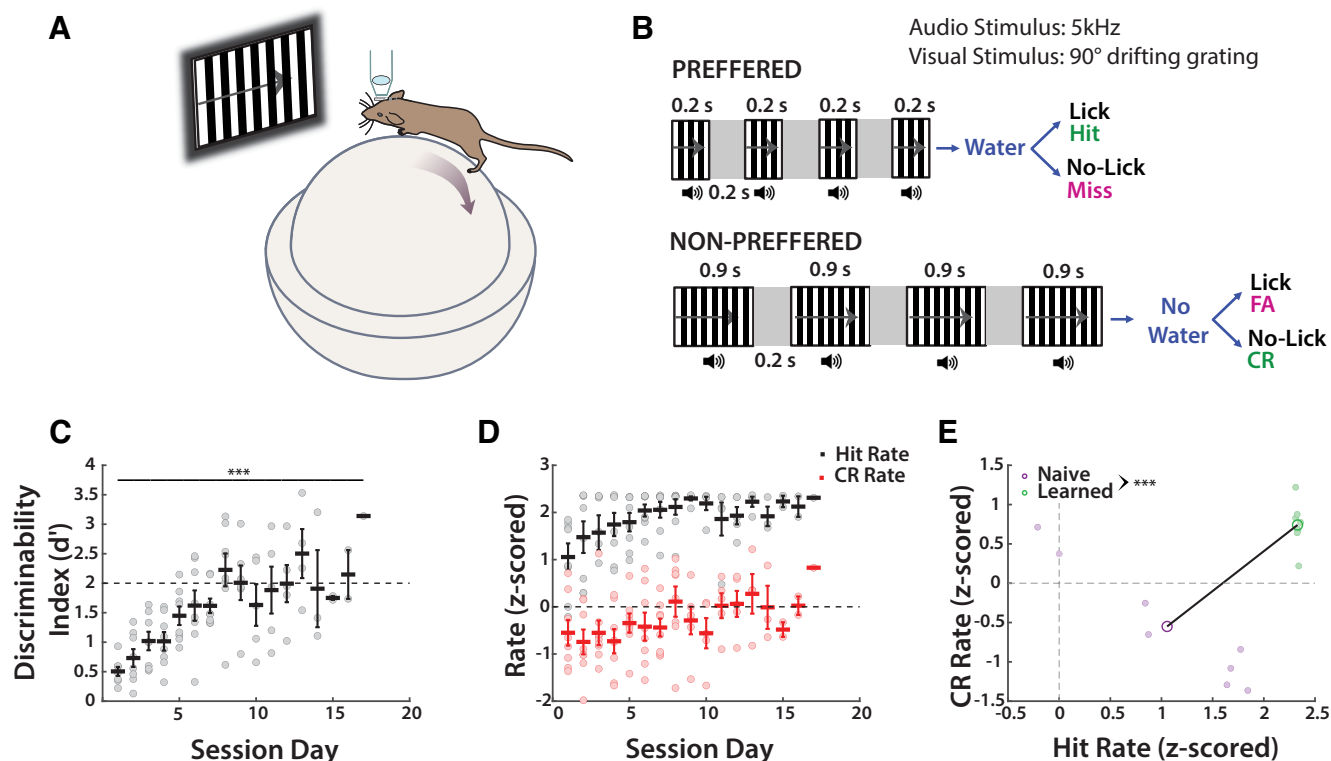
## Materials and Methods

### Experimental animals

All experiments followed the US National Institutes of Health GUIDE for the Care and Use of Laboratory Animals, under animal use protocols approved by the Chancellor's Animal Research Committee and Office for Animal Research Oversight at the University of California, Riverside (ARC #2022-0022). We used male and female FVB.129P2 (FVB) WT mice (stock #004828, The Jackson Laboratory). All mice were housed in a vivarium with a 12 h light/dark cycle, and experiments were performed during the light cycle. The FVB background was chosen because of its robust breeding.

### Go/no-go TPSD task for head restrained mice

Awake, head-restrained young adult mice (2–4 months) were allowed to run on an air-suspended polystyrene ball while performing the task in our custom-built rig (Fig. 1A). Before performing the task, the animals were subjected to handling, habituation, and pretrial phases. After recovery from headbar/cranial window surgery, mice were handled gently for 5 min every day, until they were comfortable with the experimenter and would willingly transfer from one hand to the other to eat sunflower seeds. This was followed by water deprivation (giving mice a rationed supply of water once per day) and habituation to the behavior rig. During habituation, mice were head restrained and acclimated to the enclosed soundproof chamber and allowed to run freely on the 8 cm polystyrene ball. Eventually, mice were introduced to the lickport that dispensed water (3–4  $\mu$ L) and recorded licking [custom-built at the University of California, Los Angeles (UCLA) electronics shop], followed by the audiovisual stimuli. This was repeated for 10 min per session for 3 d. Starting water deprivation before pretrials motivated the mice to lick (Guo et al., 2014). After habituation and an ~15% weight loss, mice started the pretrial phase of the training. During pretrials, mice were shown the preferred stimulus only with no punishment time associated with



**Figure 1.** Mice achieve expert status on the TPSD task ( $n = 8$ ). **A**, Schematic of mouse on polystyrene ball. **B**, Experimental paradigm is a go/no-go task composed of synchronous audiovisual stimuli. **C**,  $d'$  shows mice learn to discriminate temporal patterns (one-way ANOVA:  $F_{(1,16)} = 5.45$ ,  $p = 1.23 \times 10^{-7}$ ). **D**, Hrs and CRRs do not significantly change with learning (Hr: Kruskal–Wallis test:  $H_{(16)} = 21.74$ ,  $p = 0.1515$ ; CRr one-way ANOVA:  $F_{(1,16)} = 1.23$ ,  $p = 0.26$ ). **E**, Hr and CRr in naive and learned sessions are significantly different (Hr: two-tailed  $t$  test:  $t_{(14)} = 4.42$ ,  $p = 5.85 \times 10^{-4}$ ; CRr: two-tailed  $t$  test:  $t_{(14)} = 4.46$ ,  $p = 5.3 \times 10^{-4}$ ). Refer to Extended Data Figures 1-1, 1-2, 1-3, 1-4, which show no dependence on trial ratios or training paradigm. The Extended Data also show dependence on the stimulus for learning.

incorrect responses. This was done (1) to teach the mice the task structure and (2) to encourage the mice to lick and to remain motivated. The first day consisted of 150 trials, and subsequent days of 250 trials. The reward, as in the TPSD main task, was dispensed at 1.2 s and remained available to the mice until 2 s, at which time it was sucked away by a vacuum. The mice were required to learn to associate a water reward soon after the stimulus was presented and that there was no water reward in the intertrial interval (4 s period between trials). Initially during pretrials, the experimenter pipetted small drops of water onto the lickport to coax the mice to lick. Once the mice learned this and licked with 80% efficiency, they were advanced to the go/no-go task.

The TPSD task is a go/no-go task composed of two sequences of synchronous audiovisual stimuli (Fig. 1B). Visual stimuli are 90° drifting sinusoidal gratings and are accompanied by a synchronous 5 kHz tone at 80 dB. Within each sequence, four stimuli are presented that differ only in temporality. Our preferred sequence is composed of four stimuli of 200 ms; our nonpreferred sequence is composed of four stimuli of 900 ms.

Each set of the sequences is separated by a 200 ms period of silence accompanied by a gray screen. A water reward is dispensed at 1.2 s and remains available until 2 s, at which time it is sucked away by a vacuum. A custom-built lickport (UCLA engineering) dispensed water, vacuumed it, and recorded licking via breaks in an infrared beam. Breaks

were recorded at 250 Hz. The window in which the licking of mice counts toward a response is 1–2 s in both stimuli. A time-out period (6.5–8 s), in which the monitor shows a black screen and there is silence, is instituted if the mouse incorrectly responds. The first session was composed of 250 trials, and subsequent days of 350. Depending on the stimulus presented, the behavioral response of the animal was characterized as “Hit,” “Miss,” “Correct Rejection” (CR) or “False Alarm” (FA; Fig. 1B). An incorrect response resulted in the time-out period.

To expedite learning, we set the preferred (P)/nonpreferred (NP) stimuli ratio to 70:30 as we found that mice are more prone to licking (providing a “yes” response) than to inhibiting licking (providing a “no” response). We additionally instituted an individualized lick rate threshold to encourage learning as we found that lick rates differed significantly from mouse to mouse. Licking thresholds were calculated from lick rates for mice and shows no significant correlation between licking thresholds and learning rates (Pearson’s correlation coefficient,  $r = 0.4684$ ;  $p = 0.3012$ ). This indicates that the individualized lick rate threshold was used as a learning aid to complete the task and did not affect their learning rates or their reliance on the stimulus for task completion. To confirm that mice learned rather than took advantage of the biased 70:30 preferred to nonpreferred trial ratio, we tested mice for two additional sessions using a 60:40 ratio of preferred to

nonpreferred stimuli (Fig. 1). We retain a greater number of preferred stimuli as the total time mice encounter preferred stimuli is less than that of encountering nonpreferred stimuli within a 60:40 trial session (294 vs 588 s, respectively). Following this, mice performed a control task, during which visual and auditory stimuli were not presented. Our data show that mice did not retain learned performance, indicating that they relied on the sensory stimuli for task completion (Fig. 2).

We additionally performed experiments on mice ( $n = 2$ ) using a modified paradigm of TPSD (TPSD<sub>mod</sub>) in which the longer duration was the preferred stimulus and the shorter was the nonpreferred (Figs. 1-3). We modified the paradigm to also have the same total time between the preferred and nonpreferred stimuli (2.6 s). This paradigm entailed either three or seven synchronous audiovisual stimuli separated by 0.2 s gray screens, in which there was no sound. The preferred stimulus was three intratrial stimuli of 733 ms; the nonpreferred stimulus was seven intratrial stimuli of 200 ms. Water was dispensed at 2.3 s in the preferred stimulus. The period in which lick counted toward a decision was 2–3.2 s. Water remained available to the mice until 3.2 s. Like in the original paradigm, lick rate thresholds were individualized to mice.

Custom-written routines and Psychtoolbox in MATLAB were used to present the visual stimuli, to trigger the lick port to dispense and retract water, and to acquire data.

### Cranial window surgery

Craniotomies were performed at 6–8 weeks. Before surgery, mice were given dexamethasone (0.2 mg/kg, i.p.) and carprofen (5 mg/kg, s.c.). Mice were anesthetized with isoflurane (induction, 5%; maintenance via nose cone, 1.5–2%) and placed in a stereotaxic frame. Under sterile conditions, a 4.5-mm-diameter craniotomy was drilled over the right V1 and covered with a 5 mm glass coverslip. Before securing the cranial window with a coverslip, we injected 60–100 nl of pGP-AAV-syn-jGCaMP7f-WPRE. A custom U-shaped aluminum bar was attached to the skull with dental cement to restrain the head of the animal during behavior and calcium imaging. For 2 d following surgery, mice were given dexamethasone (0.2 mg/kg) daily.

### Viral constructs

pGP-AAV-syn-jGCaMP7f-WPRE were purchased from Addgene and diluted to a working titer of  $2e^{13}$  with 1% filtered Fast Green FCF dye (Thermo Fisher Scientific).

### In vivo two-photon calcium imaging

Calcium imaging was performed on a Scientifica two-photon microscope equipped with a Chameleon Ultra II Ti:sapphire laser (Coherent), resonant scanning mirrors (Cambridge Technologies), a 20 $\times$  objective (1.05 numerical aperture; Olympus), multialkali photomultiplier tubes (catalog #R3896, Hamamatsu) and ScanImage software (Pologruto et al., 2003). Before calcium imaging, head-restrained mice were habituated to a soundproof chamber and allowed to run freely on a polystyrene ball (Figs. 1A, 3A). Visually evoked responses of L-2/3 pyramidal (Pyr) cells from V1 were

recorded at 15 Hz in 1 field of view (FOV). Each FOV consisted of a mean of 108 Pyr cells (SD = 39.2). In each animal, imaging was performed at 150–250  $\mu$ m.

### Data analysis

#### Discriminability index and CR and Hit rates

The discriminability index ( $d'$ ) was calculated using the MATLAB function *norminv*, which returns the inverse of the normal cumulative distribution function, as follows:

$$d' = \text{norminv}(\text{fraction of hits}) - \text{norminv}(\text{fraction of FAs}).$$

If either rate reached 100% or 0%, we arbitrarily changed the value to either 99% or 1%, respectively. We did this to avoid generating z scores of infinity that would inaccurately characterize the performance of the mice.

The  $d'$  values of the best 150 trials were selected by a sliding 150 trial window; the highest value was then selected. CR rates (CRrs) and Hit rates (Hrs) use the same best 150 trial interval.

#### Licking thresholds

Licking thresholds for each mouse was determined by using the average licking in the last Pretrial session minus 1 SD.

#### Licking probabilities

Probabilities were taken by binning licks per 0.1 s window per trial per mouse. We then averaged the probability per time of each mouse to generate a distribution of probabilities based on trial session, stimulus type, and trial outcome. We use the best 150 trials from each day and each mouse as determined by the  $d'$  value.

#### Data analysis for calcium imaging

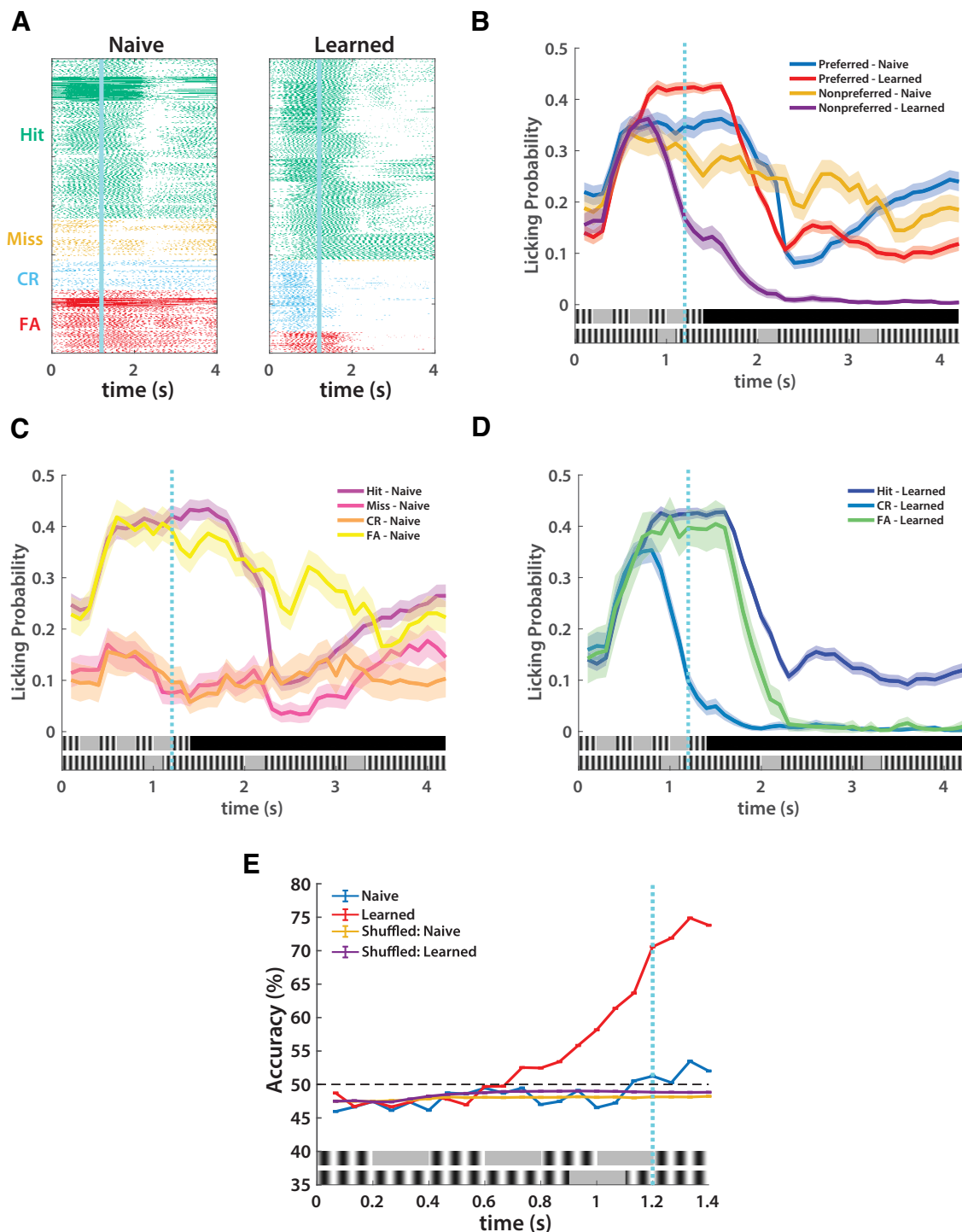
Calcium-imaging data were analyzed using suite2p (Pachitariu et al., 2017) and custom-written MATLAB routines. All data were then processed using suite2p for image registration, region of interest (ROI) detection, cell labeling, and calcium signal extraction with neuropil correction. Once suite2P had performed a rigid and nonrigid registration and then detected ROIs using a classifier, we manually selected cells using visual inspection of ROIs and fluorescence traces to ensure the cells were healthy. We then used the deconvolved spikes determined by suite2p in our subsequent analysis that used custom-written MATLAB scripts.

#### Mean network activity

We performed a bootstrap of 1000 iterations per mouse to select average activity patterns. Putative spikes were composed of either on or off times (0s or 1s). We then composed a grand distribution and used average network activity (Fig. 3F,H, Extended Data Fig. 3-1A,C). Shaded areas represent 95% confidence intervals of each respective activity curve.

#### Correlation of mean network activity with stimuli

Pearson's correlations were calculated using each bootstrapped iteration of mean network activity and a separate matrix of 0s and 1s, with 0s representing stimulus off periods and 1s representing stimulus on periods.

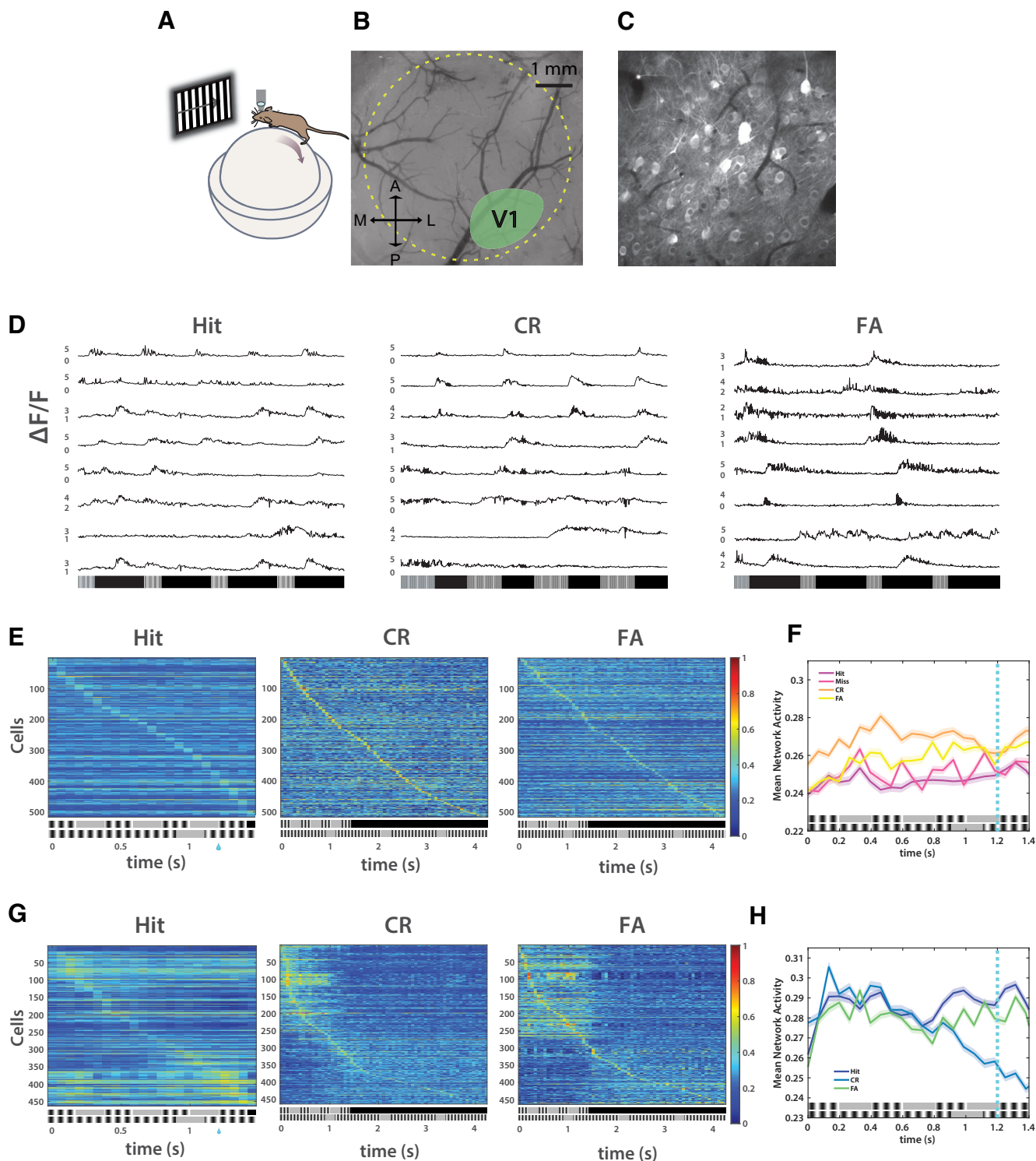


**Figure 2.** Licking profiles index learning ( $n = 8$ ). **A**, Raster plots of licking in naive and learned sessions, using the best 150 trials of all mice, as determined by the best  $d'$  for that session. **B**, Probability of a lick event by stimulus type and session. **C**, Probability of a lick event by trial outcome in naive sessions. **D**, Probability of a lick event by trial outcome in learned sessions. Miss trials are excluded as there were exceedingly few miss trials for each mouse. **E**, Accuracy of bootstrapped SVM as a function of time. Licking events per 0.067 s were the predictors, and stimulus type is the outcome. Learned session accuracy confirms learning as predictability rises above chance before the water reward at 1.2 s.

#### Time-sorted heatmaps

Heatmaps featuring sorted activity (Fig. 3E, Extended Data Fig. 6-1) were sorted using the maximum value over a given time course per unit. Units were

then displayed such that cells having a maximum value at time  $t$  were placed together; each successive grouping of cells at  $t + 1$  was placed below the previous value  $t$ .



**Figure 3.** V1 neural activity changes with learning and represents stimuli and trial outcomes ( $n = 5$ ). **A**, Schematic of a mouse on a polystyrene ball with microscope objective over V1. **B**, Example craniotomy window showing binocular V1. **C**, Example frame of video taken from binocular V1 during behavior with imaging. **D**, Raw fluorescence traces of 10 representative cells in one mouse over 3 trial outcomes in learned session. **E**, Spike-sorted mean activity of all non-lick-modulated cells, in all trials, in naive sessions. Shaded areas represent 95% confidence intervals. **F**, Mean spiking activity in naive sessions. **G**, Spike-sorted mean activity of all non-lick-modulated cells, in all trials, in learned sessions. **H**, Mean spiking activity in learned sessions. Shaded areas represent 95% confidence intervals. Refer to Extended Data Figures 3-1 and 3-2 for extended analysis on neural activity from V1 during learning. Extended Data Figures 3-3 and 3-4 show the analysis of lick-modulated cells. Refer to Extended Data Figure 3-5 for learned neural dynamics during the flipped paradigm.

### Lick-modulated cells

Lick-modulated cells were determined by using a bootstrapped support vector machine (SVM; see below for SVM methods). Hit and FA trials were compared with CR and Miss trials within the response period. The difference between these two predictors within this period should only be whether there was licking or not licking, as stimulus and the presence or absence of reward differ within predictors. The *sequentialfs* function in MATLAB, a sequential forward feature selection function, was used to identify 20 cells that contained the most predictive information per mouse per time bin per session. Upon completion, each time a cell was selected, it received a score based on the z-scored accuracy of the prediction (e.g., <50% accuracy resulted in a negative score, >50% accuracy resulted in a positive score, 50% accuracy resulted in a 0 score). All scores were then summed. The total scores of cells were then correlated with the number of times they were selected per mouse per session. Correlations that were positive and significant ( $\alpha = 0.05$ ), indicative of reliable predictability, were admitted. Cells that were >1 SD from the mean of the total scores were then selected as lick-modulated cells;  $16.83 \pm 8.5$  lick-modulated cells were found per mouse. One mouse was found to have no lick-modulated cells.

### Neural trajectories

We averaged the activity of each cell over each of its trials based on output (Hit, Miss, CR, FA) and then ran a principal component analysis using the MATLAB function *pca*. Trajectories of the first three components were plotted. The variance explained as follows: mean = 75.5, SD = 11.9.

### SVM

We used the SVM available in the MATLAB Machine Learning and Deep Learning toolbox via the function *fitcsvm*. We used a radial basis function as the kernel. Eighty percent of our data were applied to training the machine, and 20% were applied to testing it. Instead of training one machine, we developed a strategy wherein we performed a bootstrapped SVM per time per mouse. This allowed us to generate a distribution of accuracy percentages per time such that we could locate critical times of difference during stimulus presentation. Ten thousand machines were generated per time per mouse for the licking predictor and then were averaged as one grand distribution. One thousand machines were generated per time per mouse for the imaging predictor and then were averaged as one grand distribution. The fewer number of machines for the imaging predictor was because of computational constraints. The licking predictor consisted of binning licks per 0.067 s window per trial per mouse with either the stimulus type (preferred or nonpreferred) or trial outcome (Hit, Miss, CR, FA) as the outcome. The imaging predictor was the activity of the network with either the stimulus type (preferred or nonpreferred) or trial outcome (Hit, Miss, CR, FA) as the outcome in 0.067 s time bins. For a given mouse, the putative spiking activity of each cell was used as a feature space per a given time. With our licking data, we performed no pre-training optimization as we essentially were testing the accuracy of individual features (i.e., time bins of licking).

We performed optimization procedures on our neural data, however. We performed a fivefold cross-validation and used the built-in Bayesian optimizer in MATLAB (*bayesopt* function) to tune the hyperparameters (see Fig. 6A–F, all non-lick-modulated cells in the network as features). We also performed a feature selection procedure wherein we ran the SVM as done previously but by selecting a given number of cells as features (see Fig. 6G,H). This again entailed using the *sequentialfs* function in MATLAB to find the most predictive cells per a given interval. As an example, when we chose four cells, at each time bin for each mouse, the feature selection algorithm chose four cells that were most representative of the difference between two categories, which were then used to predict the difference between the two categories. Thus, there was a distribution of 4000 cells that were selected for that time point (4 cells by 1000 machines) and 1000 accuracy percentages (1000 machines). These machine accuracies were then averaged (see Fig. 6G,H), and the distribution of cells was sorted by time (Extended Data Fig. 6-1) to show when a given cell was most likely to be selected.

Each set of trials we performed with the SVM included all trials so as to have the most robust dataset possible. All uses of the SVM were accompanied by control trials in which outcomes were randomly shuffled.

### Statistical analyses

Statistical analysis of normality (Lilliefors test) was performed on each dataset, and, depending on whether the data significantly deviated from normality ( $p < 0.05$ ) or did not deviate from normality ( $p > 0.05$ ), appropriate nonparametric or parametric tests were performed. The statistical tests performed are mentioned in the text and the legends. For parametric two-group analyses, a Student's *t* test (paired or unpaired) was used; for parametric multigroup analyses, a one-way ANOVA was used. For nonparametric tests, we used the following: Wilcoxon rank-sum test (two groups), Kolmogorov–Smirnov test (KS; two groups), and Kruskal–Wallis test (multigroup). When multiple two-group tests were performed, a Bonferroni correction was applied to readjust the  $\alpha$  value. In the figures, significance levels are represented with the following convention: \* $p < \alpha$ ; \*\* $p < \alpha/10$ , \*\*\* $p < \alpha/100$ .  $\alpha$  Values are 0.05 unless otherwise specified. In all of the figures, we plot either the SEM or 95% confidence intervals. Graphs show either individual data points from each animal or group means (averaged over different mice) superimposed on individual data points.

### Exclusion of mice

Five WT mice were excluded from the data because the mice lost >25% body weight (a criterion we established a priori). This had adverse effects on their health that was manifested in listlessness, reduced grooming, interaction with cage mates, and, occasionally, seizures.

### Data availability

All the analyzed data reported in this study are available from the corresponding author on request. All code, including the SVM analysis used in this article, is available from the corresponding author on request.

## Results

### Mice learn to perform a multimodal temporal pattern sensory discrimination task

To examine temporal pattern learning, we have designed a novel go/no-go TPSD task (see Materials and Methods). We test our paradigm in mice as they are a robust animal model for temporally and spatially fine recording methods and for cell type-specific tagging and manipulation. Awake, head-restrained young adult mice (2–3 months of age) are habituated to run on a polystyrene ball treadmill while they perform the TPSD paradigm. Water-deprived mice are presented with two audiovisual temporal patterns (preferred and nonpreferred), as shown in the schematic in Figure 1B. Each pattern consists of four AV stimuli, where each AV stimulus lasts either 0.2 or 0.9 s and is separated by a 0.2 s gray screen. The visual stimulus consists of a drifting sinusoidal 90° grating, and the auditory stimulus consists of a 5 kHz tone. Both auditory and visual stimuli are presented concurrently; therefore, the stimuli are audiovisual. The temporal pattern with 0.2 s AV stimuli is associated with a water reward (preferred pattern), and the temporal pattern with 0.9 s AV stimuli is not (nonpreferred pattern; Fig. 1B). We quantified the performance of mice using a  $d'$  value in which  $d' = 2$  was set as a learning threshold (Fig. 1C). Mice learn to preferentially lick to the preferred pattern and to withhold licking for the nonpreferred pattern ( $12.13 \pm 3.52$  sessions to learn;  $n = 8$ ; one-way ANOVA,  $F_{(1,16)} = 5.45$ ,  $p = 1.23 \times 10^{-7}$ ). A positive  $d'$  value of 0.5 on session 1 likely resulted from mice learning to associate stimulus with reward in the pretrial task before the TPSD task. During the pretrial task, mice experience only the preferred stimulus and every trial is rewarded. This allows mice to learn to lick reliably (>80% licking) and learn the task structure–association of stimulus with water reward (see Materials and Methods). This pretrial task is similar to previous studies (Goel et al., 2018) and is a common strategy used in behavior assays (Guo et al., 2014).

Similar to other go/no-go tasks, to aid learning, we use a 7:3 preferred to nonpreferred stimulus ratio in the main task, which can artificially amplify the effect of Hit rate on the  $d'$  value of the mice. To confirm that learning is not simply a biased feature of the differential preferred/nonpreferred ratio and to examine the decision strategy of mice between learned and naive days, we compared Hrs with CRrs (Fig. 1D,E). Mice improved performance primarily by improving their CRrs in which the CRr changed from negative to positive (Fig. 1E;  $n = 8$ ; two-tailed, paired-sample Student's  $t$  test,  $t_{(14)} = 4.46$ ,  $p = 5.3 \times 10^{-4}$ ). Although significant, we find that the Hr of mice remains relatively unchanged across sessions (Fig. 1E;  $n = 8$ ; two-tailed, paired-sample Student's  $t$  test,  $t_{(14)} = 4.42$ ,  $p = 5.85 \times 10^{-4}$ ) but that their ability to inhibit licking increases across sessions, indicated by a positive CRr in learned sessions.

Licking profiles that accompany learning were more refined in expert mice (Fig. 2). We quantified the probabilities of the licking by the mice based on session day (naive vs learned), stimulus type (preferred vs nonpreferred), and trial outcome (Hit, Miss, CR, FA) as a function of time. We find that on learned days licking to the preferred stimulus is

enhanced, while licking to nonpreferred stimuli dramatically decreases before the water reward, indicative of a learned response (Fig. 2B). The most robust change in licking occurred in the nonpreferred stimulus in which mice peak in their lick probabilities in CR trials before the onset of the water reward in learned days (Fig. 2D).

### Licking profiles in learned mice predict stimulus type

To causally establish that licking is both (1) a viable measure of performance and (2) demonstrates differential learning between sessions, we developed a bootstrapped SVM, a type of binary classifier. We run our SVM 10,000 times within 0.067 s time bins using licking within a trial as the predictor and the stimulus of that trial as the outcome (see Materials and Methods). This allows us to generate a distribution of correctly predicted outcomes per time bin per mouse, which are then compared with a randomly shuffled control. We find that there is little predictability beyond chance in naive days with somewhat greater predictability following the water reward, likely attributable to increased licking at their chance encounter with the water reward (Fig. 2E). On learned days, licking becomes predictive beyond chance at 0.7 s and then accelerates beginning at 0.8 s. This suggests that mice relied on stimulus information to make a decision rather than on the presence or absence of the water reward. The high performance of the SVMs before the water reward in learned sessions establishes that mice indeed learn to discriminate temporal patterns. In addition, it identifies the decision period at ~0.7–0.8 s.

### Learning is not an artifact of behavior design

Following learning, mice underwent the following two additional protocols: (1) performing the same paradigm with a 6:4 P/NP stimulus ratio; and (2) performing the same paradigm without any visual or auditory input as a control (Control). We performed the 6:4 P/NP stimulus ratio task to confirm that learning was not a feature of the differential P/NP stimulus ratio of 7:3 in the main task, and the Control task to confirm that learning was stimulus dependent.

In the 6:4 P/NP stimulus ratio task, the  $d'$  values of mice for CRr and Hr were not significantly different from the main task (Extended Data Fig. 1-1A). Licking probabilities were similar as well, albeit there was less overall licking in the P/NP stimulus ratio of 6:4 task than in the main task (Extended Data Fig. 1-1B,C). We again used licking to predict stimulus type and found that predictability was maintained nearly identically in the P/NP stimulus ratio of 6:4 task (Extended Data Fig. 1-1D). These results confirm that learning was not a result of the P/NP stimulus ratio of the main task.

Control sessions confirmed that learning was stimulus dependent: expert performance on the task required the audiovisual stimuli and was not simply dependent on the availability of a water reward.  $d'$ , CRr, and Hr were all significantly different from the main task, showing poor performance (Extended Data Fig. 1-2A). Licking profiles showed considerable changes, both in the volume of

licking and in stimulus-dependent licking (Extended Data Fig. 1-2B,C). Using licking in Control trials to predict stimuli via the bootstrapped SVM showed chance performance throughout the trial period (Extended Data Fig. 1-2D). Thus, learning is dependent on the presence of stimulus and is not an artifact of some unknown confound.

In the TPSD task, the P and NP stimuli had the same number of intratrial stimuli; because each was a different duration, the total durations of the sequences were different ( $P = 1.4$  s;  $NP = 4.2$  s). To control for this potential confound, we subjected a separate cohort of mice ( $n = 2$ ) to a modified task (TPSD<sub>mod</sub>; Extended Data Fig. 1-3A). Additionally, we inverted the stimuli such that the longer intratrial stimulus was the P and the shorter was the NP to ensure that learning was not dependent on a specific duration or simply the shorter of the two durations. Mice learned the task in a mean of eight sessions, with most improvement occurring in CRr (Extended Data Fig. 1-3C–E). Licking profiles additionally verified learning with licking ramping in predictability before the water reward in learned sessions but remaining at chance level in naive sessions (Extended Data Fig. 1-3B,F–I). We performed a modified P/NP stimulus ratio as before in which the P/NP stimulus ratio in the TPSD<sub>mod</sub> task (7:3) was changed to 6:4 to rule out artifacts of experimental design. Performance remained similar to learned sessions (Extended Data Fig. 1-4A–D). Control sessions were then run in which the monitor and speakers were turned off; performance decreased as before in the original paradigm to chance levels (Extended Data Fig. 1-4E–H).

### Pyramidal cell dynamics in primary visual cortex accompany temporal pattern learning

A previous published study using sensory cortical organotypic slice cultures (Goel and Buonomano, 2016) found that information about stimulus duration is encoded in a change in pyramidal cell activity wherein the neural activity is refined to represent the learned interval. We predicted that a similar emergent neural activity contributed to learning temporal patterns *in vivo*. Therefore, to examine the pyramidal cell dynamics that are associated with TPSD, we performed two-photon calcium imaging in V1 to provide a real-time assay of neural activity during TPSD. Studies have shown that auditory inputs strongly influence neural responses in primary visual cortex (McIntosh et al., 1998; Zangenehpour and Zatorre, 2010; Deneux et al., 2019; Garner and Keller, 2022) and audiovisual stimuli evoke multimodal plasticity in V1 (Morrell, 1972; Petro et al., 2017), thus justifying V1 dynamics as a locus of change accompanying learning on the TPSD task.

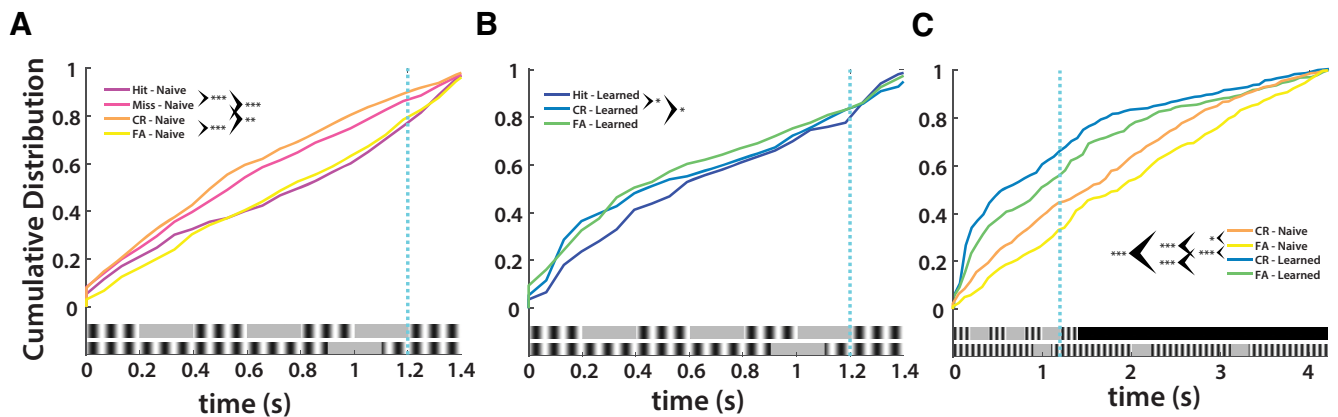
We recorded from V1 L-2/3 using two-photon calcium imaging and jRCaMP7f (half-rise time =  $27 \pm 2$  ms) with a synapsin promoter via an adeno-associated virus (AAV) vector (Dana et al., 2019; Fig. 3A–C). This indicator has been used by numerous published studies and is routinely used by researchers performing calcium imaging during behavior because of its enhanced signal-to-noise ratio and fast rise-time kinetics. Mounting evidence suggests that movement-related activity accounts for a considerable amount of variance in neural recordings, including in

primary sensory areas (Zagha et al., 2022). Further, timing and movement are linked, and therefore neural dynamics that encode time information may also code for licking. To exclusively examine neural codes for the temporal structure of the pattern, following recording we identified lick-modulated cells and removed them from subsequent analyses of neural data (see Materials and Methods; Extended Data Fig. 3-3) to distinguish sensory activity from motor and/or decision-related activity.

We find that in mice ( $n = 5$ ) during the TPSD task, there are time-dependent changes in mean network activity between naive and learned sessions. Naive sessions do not show temporal structure in either P or NP stimuli throughout the period of either (Extended Data Fig. 3-1A,C). Learned sessions show changes in activity to both stimuli. Activity is correlated until  $\sim 0.7$  s, at which there remains sustained activity in the preferred condition and suppressed activity in the NP condition (Extended Data Fig. 3-1A,C). Additionally, activity in the preferred condition is temporally coordinated with the preferred stimulus, peaking at times of intrastimulus presentation. Activity in the nonpreferred condition is likewise time locked to the preferred stimulus until 0.7 s. We suspect that the network predictively codes the preferred stimulus in both conditions, but that it is successively suppressed as it encounters the nonpreferred stimulus. Cumulative distributions of maximum spiking support this hypothesis as NP maximum spiking is significantly left shifted in learned sessions compared with naive sessions (Extended Data Fig. 3-1B,D).

Furthermore, network activity in naive sessions does not distinguish trial outcome (Hit, Miss, CR, FA; Fig. 3E,F, Extended Data Fig. 3-2A). In learned sessions, network activity indexes both stimuli and trial outcome (Fig. 3G,H, Extended Data Fig. 3-2B). In learned Hit trials, activity is temporally coordinated with the preferred stimulus (Fig. 3H). Hit learned session activity is the only condition in which activity is positively and significantly correlated with the preferred stimulus (learned Hit:  $r = 0.0134$ ,  $p = 8.33 \times 10^{-6}$ ; learned CR:  $r = -0.004$ ,  $p = 0.1892$ ; learned FA:  $r = 0.0077$ ,  $p = 0.0112$ ; naive Hit:  $r = -0.008$ ,  $p = 0.0073$ ; naive Miss:  $r = -0.0123$ ,  $p = 1.76 \times 10^{-12}$ ; naive CR:  $r = -5$ ,  $p = 0.8$ ; naive FA:  $r = -0.021$ ,  $p = 1.31 \times 10^{-12}$ ). CR trials show suppression of activity at  $\sim 0.7$  s, whereas FA trials show delayed suppression, likely leading to the incorrect response (Fig. 3G,H, Extended Data Fig. 3-2B). Delayed suppression in learned FA trials lasts until the preferred stimulus period ends at 1.4 s. We suspect that FA trial activity is predictively coded as the preferred stimulus, leading to sustained activity. CR trial activity is coded as the preferred stimulus until 0.7 s, at which point broad inhibitory activity likely suppresses the network.

We examined when cells were mostly likely to fire in naive and learned sessions. In naive sessions, maximum spiking was significantly different save for between Hits and FAs, and Misses and CRs, respectively, which indicates that network activity was not driven exclusively by sensory discrimination (Fig. 4A). Learned sessions showed significant differences between Hits and CRs, and Hits and FAs, respectively, which indicates that temporal features generate distinct network activity (Fig. 4B). Additionally,



**Figure 4.** V1 neural activity ramps earlier in learned sessions ( $n = 5$ ). **A**, Cumulative Distribution Function (CDF) of maximum spiking in naive sessions by trial outcome in preferred stimulus period (KS tests with Bonferroni correction:  $\alpha = 0.0083$ ; Hit vs Miss:  $D(0.1609)$ ,  $p = 2.58 \times 10^{-6}$ ; Hit vs CR:  $D(0.2035)$ ,  $p = 7.49 \times 10^{-10}$ ; Hit vs FA:  $D(0.062)$ ,  $p = 0.2658$ ; Miss vs CR:  $D(0.064)$ ,  $p = 0.234$ ; Miss vs FA:  $D(0.1434)$ ,  $p = 4.16 \times 10^{-5}$ ; CR vs FA:  $D(0.1899)$ ,  $p = 1.23 \times 10^{-8}$ ). **B**, CDF of maximum spiking in learned sessions by trial outcome in preferred stimulus period (KS tests with Bonferroni correction:  $\alpha = 0.0167$ ; Hit vs CR:  $D(0.1274)$ ,  $p = 9.57 \times 10^{-4}$ ; Hit vs FA:  $D(0.1339)$ ,  $p = 4.31 \times 10^{-4}$ ; CR vs FA:  $D(0.0518)$ ,  $p = 0.5518$ ). **C**, CDF of maximum spiking in naive and learned sessions by trial outcome in nonpreferred stimulus period (KS tests with Bonferroni correction:  $\alpha = 0.0083$ ; Naive CR vs Naive FA:  $D(0.124)$ ,  $p = 6.29 \times 10^{-4}$ ; Naive CR vs Learned CR:  $D(0.2612)$ ,  $p = 4.01 \times 10^{-15}$ ; Naive CR vs Learned FA:  $D(0.1873)$ ,  $p = 5.57 \times 10^{-8}$ ; Naive FA vs Learned CR:  $D(0.3369)$ ,  $p = 7.16 \times 10^{-25}$ ; Naive FA vs Learned FA:  $D(0.2478)$ ,  $p = 1.19 \times 10^{-13}$ ; Learned CR vs Learned FA:  $D(0.1058)$ ,  $p = 0.01$ ).

learned session CR and FA trials were significantly left shifted from naive session CR and FA trials, with learned CR trials the most left shifted (Fig. 4C). This demonstrates that the network in naive sessions was active throughout the NP stimulus period, whereas in learned sessions, the network had been tuned to the preferred stimulus and therefore suppressed activity in nonrelevant stimuli. Additionally, the level of suppression of activity in the NP stimulus indexes correct or incorrect responses.

### Visual cortical neural dynamics contain temporal information

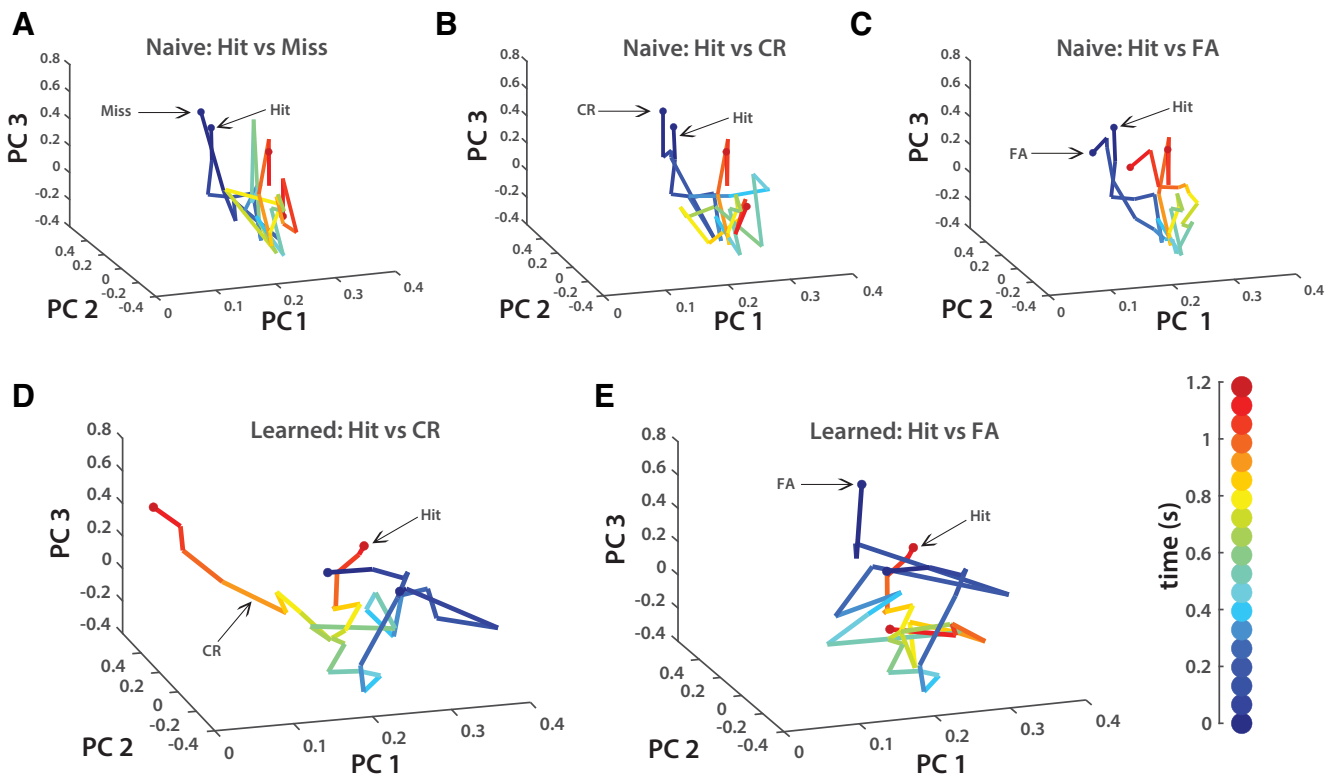
As previously discussed, we suspected that learning would be enabled through distinct patterns of network dynamics. That is, the amount of difference in neural trajectories through state space should index learning. Indeed, on naive days in which performance is poor, there is little difference in network activity regardless of stimulus type or trial outcome (Fig. 5A–C). In learned sessions, the trajectory of the network shows divergence between Hit and CR trials (Fig. 5D), but shows clustering in Hit and FA trials (Fig. 5E). Notably, the greatest divergence between Hit and CR trajectories in learned sessions begins at  $\sim 0.7$  s, the period at which licking also began to be most predictive of stimulus type in learned sessions. We suspect that this divergence does not exist between Hit and FA trials as FA trials predictively code the preferred stimulus.

However, an outstanding question is whether network divergence causes a decision to be made or whether the decision is made elsewhere, which then leads to feedback release or inhibition (I) of the network. We addressed this by using our bootstrapped SVM (0.067 s time bins, 1000 iterations/bin) to predict stimulus and trial outcome from neural activity. We found that network activity accurately predicts stimuli in learned sessions and moderately so in

naive sessions (Fig. 6A). In naive sessions, there is little predictability, suggesting that the network is not differentially tuned to stimuli or trial outcome (Fig. 6B). In learned sessions, there is high predictability between Hit and CR trials, moderate predictability between CR and FA trials, and no predictability between Hit and FA trials (Fig. 6C). The respective predictability profiles of Hit versus CR and Hit versus FA in learned sessions accord with the analysis of neural trajectories in which CR trials diverge from Hit trials, but FA trials do not (Fig. 5D,E). Because Hit and FA trials generate the same network response, evinced by predictability at chance levels, it is likely that FA trials predictively encode the preferred stimulus. Thus, in learned sessions, accurate encoding of stimuli directly contributes to trial outcome.

Not surprisingly, the neural dynamics of lick-modulated cells contained information about the trial outcome (Extended Data Fig. 3-3). We further verified this finding by comparing licking predictability with neural predictability in learned sessions (Extended Data Fig. 3-4). In Hit versus CR trials, neural predictability first rises above chance in nonlick cells at  $\sim 0.3$  s and then ramps at 0.7 s; predictability of lick-modulated cells and licking, however, occurs later at 0.7 s (Extended Data Fig. 3-4A). In Hit versus FA trials, there is no predictability in nonlick cells, lick-modulated cells, or licking (Extended Data Fig. 3-4B). These results indicate that it is necessary for the network to accurately encode temporal information before making a decision, and thus implicates sensory-driven activity in learning and trial outcome.

We additionally recorded from V1 L-2/3 in a separate cohort of mice ( $n = 2$ ) using our TPSD<sub>mod</sub> paradigm to ensure that learning was not dependent on potential artifacts within our original paradigm (Figs. 1-3). Notably, we found the network dynamics tune to the longer preferred stimulus as indicated by a right-shifted distribution in



**Figure 5.** Neural trajectories diverge only in correct trials in learned sessions ( $n = 5$ ). **A**, Neural trajectories of Hit and Miss trials, naive sessions. **B**, Neural trajectories of Hit and CR trials, naive sessions. **C**, Neural trajectories of Hit and FA trials, naive sessions. **D**, Neural trajectories of Hit and CR trials, learned sessions. **E**, Neural trajectories of Hit and FA trials, learned sessions. Vertical bar at bottom right shows color coding of time through trials.

maximum spiking in Hit trials in learned sessions compared with naive sessions (Extended Data Fig. 3-5D). CR maximum spiking was left shifted, indicating suppression of nonrelevant stimuli similar to that seen in learned sessions in the original paradigm (Extended Data Fig. 3-5C, D). Additionally, the decoder accurately predicted stimulus type from learning in learned sessions but not in naive sessions (Extended Data Fig. 3-5E).

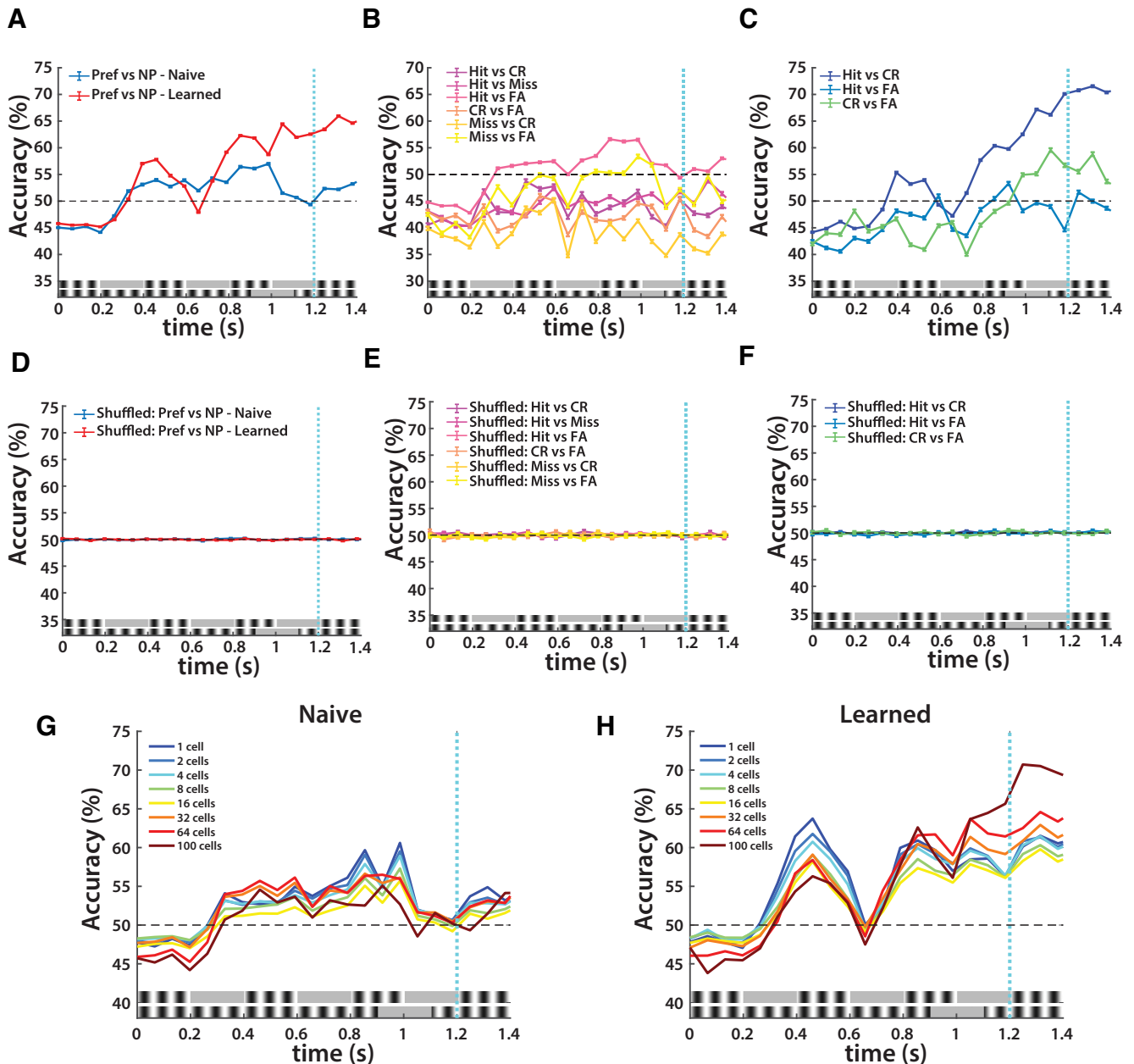
### Temporal information is encoded by intrinsic mechanisms in V1

Based on the results of the SVM using the entire network, we next characterized the dynamics by which V1 encoded temporal information. It has been shown previously that temporal information can be encoded in a variety of neural mechanisms including linear ramping, high-dimensional dynamics, and a combination of oscillators (Bakshurin et al., 2017; Zhou et al., 2020). Here we specifically addressed whether temporal coding in the TPSD task relies on dedicated or intrinsic mechanisms. We tested this by using neural data to predict stimulus outcome in learned and naive sessions as before, but by using a forward sequential feature selection algorithm for various numbers of cells, with each cell for a given mouse representing a feature. First, if temporal information is encoded in specialized functionally connected cells or circuits (i.e., through dedicated mechanisms), we predict that we would see high predictability of

stimuli using a small subset of cells in both naive and learned sessions. Additionally, we predict that the same cells would be selected throughout the stimulus period as they contain a unique ability to represent temporal information. If, however, temporal information was encoded through intrinsic mechanisms, we predict that cell selection would vary through the stimulus period, that greater numbers of cells would provide higher predictability, and that predictability would emerge over learning likely through refinement of the network.

We find that there is only high predictability in learned sessions and that cells that are selected vary over the stimulus period, indicating that in V1 timing is achieved without dedicated timing mechanisms but rather through changing patterns of intrinsic neural dynamics (Fig. 6G,H, Extended Data Fig. 6-1). Additionally, predictability and selectivity in learned sessions is stimulus dependent. These results suggest that temporal information in V1 does not invariantly rely on dedicated mechanisms.

However, we find that systematically there is higher predictability earlier in the stimulus period in fewer cells in the learned sessions. The high early predictability ramps immediately following 0.2 s, which is the first period at which the P and NP stimuli differ. This result suggests that in V1, temporal information can be encoded by a small subset of cells. However, as the stimulus period continues, and a decision is reached before the arrival of water, there is ramping in predictability in all cell selections, and in



**Figure 6.** SVM performance as a function of time ( $n = 5$ ). **A**, Spiking activities per mouse per trial in naive and learned sessions were predictors, and outcome was the stimulus type. **B**, Spiking activity per mouse per trial in naive sessions was the predictor, and outcome was the trial outcome (Hit, Miss, CR, FA). **C**, Spiking activity per mouse per trial in learned sessions was the predictor, and outcome was the trial outcome (Hit, CR, FA). Miss trials were again omitted because of their small sample size. **D**, Control for **A**. Spiking activities per mouse per trial in naive and learned sessions were predictors, and outcome was the randomly shuffled stimulus type. **E**, Control for **B**. Spiking activities per mouse per trial in naive sessions was the predictor, outcome was the randomly shuffled trial outcome. **F**, Control for **C**. Spiking activities per mouse per trial in learned sessions were predictors, and outcome was the randomly shuffled trial outcome. **G**, SVM decoding accuracy of spiking activity in naive sessions with varying numbers of cells selected. A forward sequential feature selection algorithm was used to select predictive cells (see Materials and Methods). **H**, Same as **G**, but in learned sessions. Extended Data Figure 6-1 provides evidence for cell selectivity during learning.

greater numbers of cells there is systematically greater predictability. Notably, at 0.4–0.5 s, in which stimuli are visually and aurally the same, there remains high predictability that differentiates P and NP stimuli, which implicates temporal information as what is most saliently encoded in V1.

We suspect that a small subset of cells at ~0.2 s indexes the temporal difference between stimuli whose activity then propagates throughout the network as the stimulus period continues and more sensory information is received. This leads to network-level tuning, which causes the network itself to be more predictive than a

small subset of cells. However, as this ramping in network-level predictability occurs following the decision period, it is unknown whether this tuning is caused by the stimulus, by the early activity of a small subset of cells that generate a particular neural trajectory, or by top-down areas amplifying relevant functional populations in V1. Thus, a combination of distinct mechanisms may be responsible for timing within the stimulus period.

## Discussion

Using a go/no-go audiovisual timing task, we have demonstrated that mice learned to perform the TPSD task successfully, as assessed through an increase in the discriminability index and a refinement of licking profiles. Learned performance was attributable to changes in response to the nonpreferred temporal pattern in which licking was suppressed early into the stimulus period. Similar results were seen in neural activity in V1 L-2/3 in which activity was suppressed in the nonpreferred stimulus but was released in a temporally defined manner in the preferred stimulus. Whereas naive sessions showed decided overlap in neural trajectories, learned sessions showed trajectories that indexed stimulus type and trial outcome, suggesting that distinct functional populations developed with learning. Neural activity was also used to decode stimulus type and trial outcome in learned sessions but not in naive sessions, which indicates that V1 undergoes synaptic plasticity changes to support temporal learning. Additionally, we found that subsecond temporal encoding relies on intrinsic temporal mechanisms. Early decoding predictability using a small subset of cells suggests that the state of the entire network does not index temporal information but that it is contained within a few cells. As the stimulus period progresses, the dynamics of the entire network become more predictive than a small subset of cells, indicating that the network state does index elapsed time. However, how this transition occurs and whether it is achieved locally or via top-down inputs require further investigation.

Because of the complexity of understanding temporal processing, time has been categorized into distinct types such as sensory versus motor timing and interval versus pattern timing, as well as distinguishing between different timescale ranges (Paton and Buonomano, 2018). For example, many tasks in interval timing require the identification of isolated segments of time such as in waiting to cross a street or identifying a single musical note. While interval timing can be studied as its own entity, it is also important in pattern timing. Pattern timing is composed of intervals and contains a temporal structure (Paton and Buonomano, 2018). For instance, to understand language, one must recognize the overall prosody of speech, as well as the pauses between words. The timescale in which interval and pattern timing occur is on the order of tens of milliseconds to a few seconds, although it is unknown whether the neural mechanisms of interval and rhythmic timing are shared and whether intrinsic timing mechanisms that contribute to interval timing also contribute to learning of temporal patterns (Hardy and Buonomano, 2016). Although early theories of timing proposed centralized mechanisms dedicated entirely

to temporal perception, it has since been established that different neural mechanisms are involved in processing time at different timescales (Paton and Buonomano, 2018). However, it has yet to be determined whether the mechanisms of temporal perception in the range of seconds and subseconds are distributed across brain regions or whether local networks within different regions can intrinsically encode time, albeit through a diversity of network dynamics (Zhou et al., 2020, 2022). Primary visual cortical circuits show robust plasticity to spatiotemporal features in stimuli and predict temporal associations (Chubykin et al., 2013; Gavornik and Bear, 2014a,b; Fiser et al., 2016; Garrett et al., 2020). Our data show robust V1 dynamics that encode temporal information about the experienced stimuli as shown by the accuracy of the decoder. As is the case in many behaviors, perturbation experiments are typically performed to establish causality between neural activity and behavior. However, using machine-learning algorithms to show a requirement of neural activity is increasingly used by many groups (Bakurin et al., 2017; Zhou et al., 2020; Lazar et al., 2021; Toso et al., 2021). While not the same as a perturbation experiment, a bootstrapped SVM allowed us to determine whether neural activity contained information about stimulus type and trial type. Importantly, we find that information in the TPSD task is learned through intrinsic changes in V1 dynamics. Further, depending on the time during the trial duration, learned information about the temporal patterns was encoded in a small selection of cells indicative of “time cells,” a medium-sized selection indicative of oscillators, as well as a large selection of cells suggestive of a change in network state. In conclusion, our data show the contribution of multiple mechanisms that allow learning and representation of time intervals.

There is a growing consensus that movement-related and arousal-related signals as well as sensory and cognitive processes may be contained in the same evolving neural activity (Zagha et al., 2022). However, identifying and dissociating neural circuits and dynamics that contribute to sensory, motor, arousal, and other cognitive processes is still a challenge. We used two strategies to address this challenge: (1) the design of the TPSD task, which consisted of a separation of stimulus onset from the response window, thus attributing the very early neural activity to sensory processes (temporal discrimination); and (2) we exclusively examined neural activity that contributed to encoding time by removing lick-modulated cells from our analysis.

Complex interplay of synaptic excitation (E) and I allows cortical neurons to adaptively respond to sensory stimuli, discriminate between stimuli, and integrate sensory inputs (Isaacson and Scanziani, 2011; Ferguson and Cardin, 2020). Converging evidence across many studies and model systems shows that selectivity to the interval of a stimulus duration is the result of dynamic shifts in E-I balance (Edwards et al., 2007; Kostarakos and Hedwig, 2012; Goel and Buonomano, 2016). Consistent with previous *in vitro* work in cortical slices (Goel and Buonomano, 2016), our data suggest that network suppression is one potential mechanism that drives learning. Encoding of intervals and patterns in sensory cortical circuits can result from changes in the E/I ratio at temporally defined periods (Goel and

Buonomano, 2016) and by multiple interneuron populations (Cardin, 2018). V1 L-2/3 is composed primarily of Pyr cells, which are synapsed by parvalbumin (PV) cells at the cell body and axonal hillock (Gonchar and Burkhalter, 1997). Somatostatin (SST) cells synapse onto the dendrites of pyramidal cells, thus providing more fine-tuned inhibition. Both PV and SST cells have been implicated in short-term plasticity, which is one of the mechanisms proposed to drive sub-second sensory timing (Buonomano, 2000; Motanis et al., 2018; Seay et al., 2020). PV cells provide reliable inhibition within a short window, which can help to constrain pyramidal cell firing to temporally defined windows (Pouille and Scanziani, 2001; Cardin et al., 2010). We speculate that inhibition, driven by PV cell activity, which is broadly tuned because of their anatomic connectivity, can drive temporal encoding of rhythmic patterns. Indeed, in a recent study, PV neurons were shown to be important in mediating reward timing (Monk et al., 2020). However, SST cells also modulate cortical output. SST neurons not only provide dendritic pyramidal cell inhibition but also control PV cell output (Atallah et al., 2012). Therefore, a complex interaction between SST and PV cells determines the balance between somatic and dendritic inhibition on pyramidal cells (Xue et al., 2014), thus contributing to temporal encoding (Cardin, 2018). However, it is important to emphasize that circuits in V1 are not the only contributors to the TPSD task and that multiple areas such as auditory cortex, ACC, and other downstream areas are likely involved.

Our results thus far suggest that the emergence of complex neural dynamics in V1 accompanies temporal pattern learning. An important hallmark of learning is being sensitive to and remembering the temporal structure of events so that we can make predictions and guide our future decisions. As a result, it is not surprising that disruptions in timing and timed performance are associated with a number of neurologic disorders such as Parkinson's disease, schizophrenia, and autism spectrum disorder. Our study opens the door to future studies probing the mechanistic substrates of excitation and inhibition that fine-tune temporal pattern learning and offers insights into translational studies of time perception.

## References

- Atallah BV, Bruns W, Carandini M, Scanziani M (2012) Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron* 73:159–170.
- Bakhurin KI, Goudar V, Shobe JL, Claar LD, Buonomano DV, Masmanidis SC (2017) Differential encoding of time by prefrontal and striatal network dynamics. *J Neurosci* 37:854–870.
- Barakat B, Seitz AR, Shams L (2015) Visual rhythm perception improves through auditory but not visual training. *Curr Biol* 25:R60–R61.
- Brosch M, Selezneva E, Scheich H (2011) Representation of reward feedback in primate auditory cortex. *Front Syst Neurosci* 5:5.
- Buhusi CV, Meck WH (2005) What makes us tick? Functional and neural mechanisms of interval timing. *Nat Rev Neurosci* 6:755–765.
- Buonomano DV (2000) Decoding temporal information: a model based on short-term synaptic plasticity. *J Neurosci* 20:1129–1141.
- Cardin JA (2018) Inhibitory interneurons regulate temporal precision and correlations in cortical circuits. *Trends Neurosci* 41:689–700.
- Cardin JA, Kumbhani RD, Contreras D, Palmer LA (2010) Cellular mechanisms of temporal sensitivity in visual cortex neurons. *J Neurosci* 30:3652–3662.
- Carnevale F, de Lafuente V, Romo R, Barak O, Parga N (2015) Dynamic control of response criterion in premotor cortex during perceptual detection under temporal uncertainty. *Neuron* 86:1067–1077.
- Chubykin AA, Roach EB, Bear MF, Shuler MG (2013) A cholinergic mechanism for reward timing within primary visual cortex. *Neuron* 77:723–735.
- Dana H, Sun Y, Mohar B, Hulse BK, Kerlin AM, Hasseman JP, Tsegaye G, Tsang A, Wong A, Patel R, Macklin JJ, Chen Y, Konnerth A, Jayaraman V, Looger LL, Schreier ER, Svoboda K, Kim DS (2019) High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nat Methods* 16:649–657.
- Deneux T, Harrell ER, Kempf A, Ceballos S, Filipchuk A, Bathellier B (2019) Context-dependent signaling of coincident auditory and visual events in primary visual cortex. *Elife* 8:e44006.
- Edwards CJ, Leary CJ, Rose GJ (2007) Counting on inhibition and rate-dependent excitation in the auditory system. *J Neurosci* 27:13384–13392.
- Emmons EB, De Corte BJ, Kim Y, Parker KL, Matell MS, Narayanan NS (2017) Rodent medial frontal control of temporal processing in the dorsomedial striatum. *J Neurosci* 37:8718–8733.
- Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1:1–47.
- Ferguson KA, Cardin JA (2020) Mechanisms underlying gain modulation in the cortex. *Nat Rev Neurosci* 21:80–92.
- Finnie PSB, Komorowski RW, Bear MF (2021) The spatiotemporal organization of experience dictates hippocampal involvement in primary visual cortical plasticity. *Curr Biol* 31:3996–4008.e6.
- Fiser A, Mahringer D, Oyibo HK, Petersen AV, Leinweber M, Keller GB (2016) Experience-dependent spatial expectations in mouse visual cortex. *Nat Neurosci* 19:1658–1664.
- Garner AR, Keller GB (2022) A cortical circuit for audio-visual predictions. *Nat Neurosci* 25:98–105.
- Garrett M, Manavi S, Roll K, Ollerenshaw DR, Groblewski PA, Ponvert ND, Kiggins JT, Casal L, Mace K, Williford A, Leon A, Jia X, Ledochowitsch P, Buice MA, Wakeman W, Mihalas S, Olsen SR (2020) Experience shapes activity dynamics and stimulus coding of VIP inhibitory cells. *Elife* 9:e50340.
- Gavornik JP, Bear MF (2014a) Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nat Neurosci* 17:732–737.
- Gavornik JP, Bear MF (2014b) Higher brain functions served by the lowly rodent primary visual cortex. *Learn Mem* 21:527–533.
- Glickfeld LL, Histed MH, Maunsell JH (2013) Mouse primary visual cortex is used to detect both orientation and contrast changes. *J Neurosci* 33:19416–19422.
- Goel A, Buonomano DV (2016) Temporal interval learning in cortical cultures is encoded in intrinsic network dynamics. *Neuron* 91:320–327.
- Goel A, Cantu DA, Guilfoyle J, Chaudhari GR, Newadkar A, Todisco B, de Alba D, Kourdougli N, Schmitt LM, Pedapati E, Erickson CA, Portera-Cailliau C (2018) Impaired perceptual learning in a mouse model of Fragile X syndrome is mediated by parvalbumin neuron dysfunction and is reversible. *Nat Neurosci* 21:1404–1411.
- Gonchar Y, Burkhalter A (1997) Three distinct families of GABAergic neurons in rat visual cortex. *Cereb Cortex* 7:347–358.
- Gordon JA, Stryker MP (1996) Experience-dependent plasticity of binocular responses in the primary visual cortex of the mouse. *J Neurosci* 16:3274–3286.
- Gouvêa TS, Monteiro T, Motiwala A, Soares S, Machens C, Paton JJ (2015) Striatal dynamics explain duration judgments. *Elife* 4:e11386.
- Guo ZV, Hires SA, Li N, O'Connor DH, Komiyama T, Ophir E, Huber D, Bonardi C, Morandell K, Gutnisky D, Peron S, Xu NL, Cox J, Svoboda K (2014) Procedures for behavioral experiments in head-fixed mice. *PLoS One* 9:e88678.
- Hardy NF, Buonomano DV (2016) Neurocomputational models of interval and pattern timing. *Curr Opin Behav Sci* 8:250–257.

- Heys JG, Dombeck DA (2018) Evidence for a subcircuit in medial entorhinal cortex representing elapsed time during immobility. *Nat Neurosci* 21:1574–1582.
- Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* 160:106–154.
- Isaacson JS, Scanziani M (2011) How inhibition shapes cortical activity. *Neuron* 72:231–243.
- Jazayeri M, Shadlen MN (2015) A neural mechanism for sensing and reproducing a time interval. *Curr Biol* 25:2599–2609.
- Keller GB, Bonhoeffer T, Hübener M (2012) Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron* 74:809–815.
- Kostarakos K, Hedwig B (2012) Calling song recognition in female crickets: temporal tuning of identified brain neurons matches behavior. *J Neurosci* 32:9601–9612.
- Lazar A, Lewis C, Fries P, Singer W, Nikolic D (2021) Visual exposure enhances stimulus encoding and persistence in primary cortex. *Proc Natl Acad Sci U S A* 118:e2105276118.
- Leon MI, Shadlen MN (2003) Representation of time by neurons in the posterior parietal cortex of the macaque. *Neuron* 38:317–327.
- Licata AM, Kaufman MT, Raposo D, Ryan MB, Sheppard JP, Churchland AK (2017) Posterior parietal cortex guides visual decisions in rats. *J Neurosci* 37:4954–4966.
- Mauk MD, Buonomano DV (2004) The neural basis of temporal processing. *Annu Rev Neurosci* 27:307–340.
- McIntosh AR, Cabeza RE, Lobaugh NJ (1998) Analysis of neural interactions explains the activation of occipital cortex by an auditory stimulus. *J Neurophysiol* 80:2790–2796.
- Merchant H, Grahn J, Trainor L, Rohrmeier M, Fitch WT (2015) Finding the beat: a neural perspective across humans and non-human primates. *Philos Trans R Soc Lond B Biol Sci* 370:20140093.
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202.
- Monk KJ, Allard S, Hussain Shuler MG (2020) Reward timing and its expression by inhibitory interneurons in the mouse primary visual cortex. *Cereb Cortex* 30:4662–4676.
- Morrell F (1972) Visual system's view of acoustic space. *Nature* 238:44–46.
- Motanis H, Seay MJ, Buonomano DV (2018) Short-term synaptic plasticity as a mechanism for sensory timing. *Trends Neurosci* 41:701–711.
- Namoodiri VM, Huertas MA, Monk KJ, Shouval HZ, Hussain Shuler MG (2015) Visually cued action timing in the primary visual cortex. *Neuron* 86:319–330.
- Niell CM, Stryker MP (2010) Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* 65:472–479.
- Pachitariu M, Stringer C, Dipoppa M, Schröder S, Rossi LF, Dalglish H, Carandini M, KD H (2017) Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv* 061507. <https://doi.org/10.1101/061507>.
- Pastalkova E, Itskov V, Amarasingham A, Buzsáki G (2008) Internally generated cell assembly sequences in the rat hippocampus. *Science* 321:1322–1327.
- Paton JJ, Buonomano DV (2018) The neural basis of timing: distributed mechanisms for diverse functions. *Neuron* 98:687–705.
- Petro LS, Paton AT, Muckli L (2017) Contextual modulation of primary visual cortex by auditory signals. *Philos Trans R Soc Lond B Biol Sci* 372:20160104.
- Pologruto TA, Sabatini BL, Svoboda K (2003) ScanImage: flexible software for operating laser scanning microscopes. *Biomed Eng Online* 2:13.
- Pouille F, Scanziani M (2001) Enforcement of temporal fidelity in pyramidal cells by somatic feed-forward inhibition. *Science* 293:1159–1163.
- Raposo D, Sheppard JP, Schrater PR, Churchland AK (2012) Multisensory decision-making in rats and humans. *J Neurosci* 32:3726–3735.
- Seay MJ, Natan RG, Geffen MN, Buonomano DV (2020) Differential short-term plasticity of PV and SST neurons accounts for adaptation and facilitation of cortical neurons to auditory tones. *J Neurosci* 40:9224–9235.
- Shuler MG, Bear MF (2006) Reward timing in the primary visual cortex. *Science* 311:1606–1609.
- Soares S, Atallah BV, Paton JJ (2016) Midbrain dopamine neurons control judgment of time. *Science* 354:1273–1277.
- Tonoyan Y, Fomaciaci M, Parsons B, Buetti D (2022) Subjective time is predicted by local and early visual processing. *Neuroimage* 264:119707.
- Toso A, Reinartz S, Pulecchi F, Diamond ME (2021) Time coding in rat dorsolateral striatum. *Neuron* 109:3663–3673.e6.
- Tsao A, Sugar J, Lu L, Wang C, Knierim JJ, Moser MB, Moser EI (2018) Integrating time from experience in the lateral entorhinal cortex. *Nature* 561:57–62.
- Xue M, Atallah BV, Scanziani M (2014) Equalizing excitation-inhibition ratios across visual cortical neurons. *Nature* 511:596–600.
- Zagha E, Erlich JC, Lee S, Lur G, O'Connor DH, Steinmetz NA, Stringer C, Yang H (2022) The importance of accounting for movement when relating neuronal activity to sensory and cognitive processes. *J Neurosci* 42:1375–1382.
- Zangenehpour S, Zatorre RJ (2010) Crossmodal recruitment of primary visual cortex following brief exposure to bimodal audiovisual stimuli. *Neuropsychologia* 48:591–600.
- Zelano C, Mohanty A, Gottfried JA (2011) Olfactory predictive codes and stimulus templates in piriform cortex. *Neuron* 72:178–187.
- Zhou S, Masmanidis SC, Buonomano DV (2020) Neural sequences as an optimal dynamical regime for the readout of time. *Neuron* 108:651–658.e5.
- Zhou S, Masmanidis SC, Buonomano DV (2022) Encoding time in neural dynamic regimes with distinct computational tradeoffs. *PLoS Comput Biol* 18:e1009271.
- Zhou X, de Villers-Sidani E, Panizzutti R, Merzenich MM (2010) Successive-signal biasing for a learned sound sequence. *Proc Natl Acad Sci U S A* 107:14839–14844.
- Znamenskiy P, Zador AM (2013) Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination. *Nature* 497:482–485.