

Coordinated Prefrontal State Transition Leads Extinction of Reward-Seeking Behaviors

Eleonora Russo,¹ Tianyang Ma,² Rainer Spanagel,³ Daniel Durstewitz,^{1,4} Hazem Toutounji,^{1,5,6*} and Georg Köhr^{2,7*}

¹Department of Theoretical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, 68159 Mannheim, Germany, ²Physiology of Neural Networks, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, 68159 Mannheim, Germany, ³Institute of Psychopharmacology, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, 68159 Mannheim, Germany, ⁴Faculty of Physics and Astronomy, Heidelberg University, 69117 Heidelberg, Germany, ⁵School of Psychology, University of Nottingham, Nottingham NG7 2RD, United Kingdom, ⁶Institute of Neuroinformatics, University of Zurich and ETH Zurich, 8057 Zurich, Switzerland, and ⁷Department of Neurophysiology, Medical Faculty Mannheim, University of Heidelberg, 68167 Mannheim, Germany

Extinction learning suppresses conditioned reward responses and is thus fundamental to adapt to changing environmental demands and to control excessive reward seeking. The medial prefrontal cortex (mPFC) monitors and controls conditioned reward responses. Abrupt transitions in mPFC activity anticipate changes in conditioned responses to altered contingencies. It remains, however, unknown whether such transitions are driven by the extinction of old behavioral strategies or by the acquisition of new competing ones. Using *in vivo* multiple single-unit recordings of mPFC in male rats, we studied the relationship between single-unit and population dynamics during extinction learning, using alcohol as a positive reinforcer in an operant conditioning paradigm. To examine the fine temporal relation between neural activity and behavior, we developed a novel behavioral model that allowed us to identify the number, onset, and duration of extinction-learning episodes in the behavior of each animal. We found that single-unit responses to conditioned stimuli changed even under stable experimental conditions and behavior. However, when behavioral responses to task contingencies had to be updated, unit-specific modulations became coordinated across the whole population, pushing the network into a new stable attractor state. Thus, extinction learning is not associated with suppressed mPFC responses to conditioned stimuli, but is anticipated by single-unit coordination into population-wide transitions of the internal state of the animal.

Key words: alcohol; attractor states; behavioral model; change-point analysis; extinction learning; prelimbic cortex

Significance Statement

The ability to suppress conditioned behaviors when no longer beneficial is fundamental for the survival of any organism. While pharmacological and optogenetic interventions have shown a critical involvement of the mPFC in the suppression of conditioned responses, the neural dynamics underlying such a process are still largely unknown. Combining novel analysis tools to describe behavior, single-neuron response, and population activity, we found that widespread changes in neuronal firing temporally coordinate across the whole mPFC population in anticipation of behavioral extinction. This coordination leads to a global transition in the internal state of the network, driving extinction of conditioned behavior.

Received Oct. 5, 2020; revised Dec. 16, 2020; accepted Jan. 17, 2021.

Author contributions: E.R., T.M., R.S., D.D., H.T., and G.K. designed research; E.R., T.M., H.T., and G.K. performed research; H.T. contributed unpublished reagents/analytic tools; E.R. and H.T. analyzed data; E.R., H.T., and G.K. wrote the paper.

*H.T. and G.K. contributed equally to this work.

This research was supported by grants from Deutsche Forschungsgemeinschaft to D.D., G.K., and R.S. within CRC Project 1134 (Subprojects D01 and B05); Project-ID 402170461–TRR 265 (subproject A05 and A06; Heinz et al., 2020) and SPP-1665 (Du 354/8-2), by the German Ministry for Education and Research via the e:Med framework [Grants 01ZX1311A (subprojects SP7 and SP11) and 01ZX1909A]; an Heidelberg Biosciences

International Graduate School fellowship to T.M.; and the Ch. and H. Schaller Foundation fellowship to E.R. We thank Drs. Ainhoa Bilbao, Thomas Enkel, Wolfgang Kelsch, Andreas Meyer-Lindenberg, Max Scheller, Lennart Oettl, Wolfgang Sommer, and Rolf-Detlef Treede for support and stimulating discussions.

The authors declare no competing financial interests.

Correspondence should be addressed to Eleonora Russo at eleonora.russo@zi-mannheim.de or Hazem Toutounji at hazem.toutounji@nottingham.ac.uk.

<https://doi.org/10.1523/JNEUROSCI.2588-20.2021>

Copyright © 2021 the authors

Introduction

Acquiring reward drives many daily-life decisions and requires associating specific cues and actions with reinforcement. In an ever-changing environment, however, the ability to suppress previously beneficial responses that have become inappropriate or maladaptive under new circumstances is critical for the survival of an organism. The rodent prelimbic cortex (PL), together with the infralimbic cortex (IL), is implicated in extinguishing reward-seeking behavior (Quirk and Mueller, 2008; Jonkman et al., 2009; Goldstein and Volkow, 2011; Chen et al., 2013; Moorman and Aston-Jones, 2015; Riaz et al., 2019; Sharpe et al., 2019). Both pharmacological inactivation of PL and optogenetic stimulation of its inhibitory network during the presentation of conditioned stimuli facilitate extinction (Sparta et al., 2014; Caballero et al., 2019). Similarly, optogenetic stimulation of PL projections to the nucleus accumbens reduces reward seeking when reward is associated with the risk of aversive reinforcement (Kim et al., 2017).

While such manipulations highlight the role of mPFC in the extinction of reward-seeking responses, the neural dynamics driving extinction are largely unknown. At the cellular level, the acquisition of new behavioral strategies is associated with changes in PL and IL activity, with changes in PL predicting and in IL following the acquisition of the new contingencies (Rich and Shapiro, 2009). Furthermore, sudden neuronal transitions in prefrontal regions of rodents and primates signal rapid behavioral shifts during rule (rat, PL; Durstewitz et al., 2010) and reversal learning (primate, dorsolateral PFC; Bartolo and Averbeck, 2020), and mark the onset of the exploratory phase following changes in cued reward probabilities [rat, PL/anterior cingulate cortex (ACC); Karlsson et al., 2012]. Similar to human mPFC (Schuck et al., 2015), changes in prelimbic activity in rats also precede behavioral changes both for spontaneous and enforced strategy switches (Powell and Redish, 2016).

While representational switches in mPFC have been studied during the learning of new behavioral rules, it remains an open question whether similar dynamic processes are also at work during the extinction of conditioned behaviors. In fact, while rule switching does require the suppression of old stimulus–reward associations, such suppression coincides with the formation of new competing associations. In contrast, during extinction learning the loss of conditioned responses follows from the suppression of reward seeking per se. To investigate the neural dynamics underlying extinction learning, we analyzed *in vivo* multiple single-unit recordings from the rat PL area during maintenance and within-session extinction of a visually guided appetitive operant conditioning task. To enable a detailed analysis of fine-scale temporal relationships among single-unit activity, population dynamics, environmental conditions, and aspects of the behavior of the animal at a single-subject level (Gallistel et al., 2004), we combined recently developed change-point detection methods for neural activity (Toutounji and Durstewitz, 2018) with a newly developed statistical model of behavior. Crucially, the novel behavioral model provides an estimate of the learning curve of each animal by identifying sustained changes in its behavioral response. Each change in behavior is characterized by a distinct onset trial, duration, and magnitude, with the latter measuring the relative change in performance during the episode. Our analyses revealed that even when experimental conditions and behavioral responses were stable, single-unit coding in PL was not. Importantly, however, shortly before the animal started to actively suppress the previously acquired reward contingency, changes in single-unit activity became highly coordinated across the whole

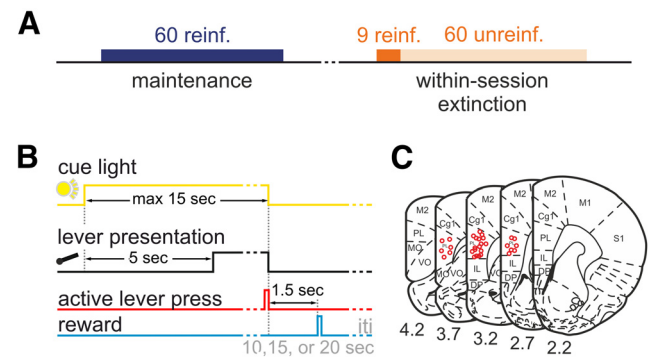


Figure 1. Behavioral paradigm and recording sites. **A**, **B**, Behavioral task (**A**) and schematics of trial timeline during reinforced trials (**B**). During reinforced trials, reward was delivered exclusively on pressing the cued lever (active lever). **C**, Histologically verified recording sites within PL of the 10 rats.

network, pushing the PL toward a new operational state that drove the extinction of reward-seeking behavior.

Materials and Methods

Subjects

Male Wistar rats (Charles River) were group housed when 8 weeks old in standard rat cages under a 12 h reversed light/dark cycle. Food and tap water were provided *ad libitum*. After tetrode implantation, rats shared a cage in groups of two, separated by a high, perforated wall (50 cm) allowing snout contact. All rats were adults (~14 weeks old) and weighed >400 g at the start of single-unit recordings (see below). All experimental procedures were performed in accordance with the European Union guidelines for the care and use of laboratory animals and were approved by the local committee (G-273/12 and G30/15; Regierungspräsidentium, Karlsruhe, Germany).

Experimental design and statistical analysis

We designed a visually guided appetitive operant conditioning paradigm to probe the extinction of reward-seeking behavior in rats (Fig. 1A,B). We chose alcohol as a reward to investigate extinction learning of both appetitive reward seeking and drug–cue association. Extinction therapy is, in fact, clinically used to treat substance use disorders, but with variable efficacy (Mellentin et al., 2017). To better understand the mechanisms leading to the extinction of conditioned responses, we developed a paradigm where the amount of alcohol administered served as a positive reinforcer, shaping conditioned behavior, but did not lead to intoxication, which would have interfered with our measurements.

To confirm that the animals perceived alcohol as a positive reinforcer, we conducted a pilot experiment and compared the conditioning effect of alcohol and saccharine. Two cohorts of male Wistar rats performed an operant conditioning paradigm, receiving reward when pressing a lever during a cue-light presentation. One cohort received a solution of saccharine (0.2% v/v in water; $n = 10$) as a reward, while the other received a solution of alcohol (10% v/v in water; $n = 11$). Behavioral training was performed as described below. Stable behavior over two successive 25-trial sessions was reached when the rats were ~12 weeks old. We found no difference in response probability between the two cohorts (74% and 73% on average for the saccharine and alcohol cohort, respectively, computed over the two stable sessions; Wilcoxon rank-sum test, $p = 0.69$). These initial results confirmed that rats reacted to alcohol as a positive reinforcer, justifying its use as such in the extinction paradigm reported here. Such an assumption is further justified by a previous study that reported the same response probability to saccharine-conditioned and alcohol-conditioned cues during both the maintenance and extinction of an operant conditioning task (Pfarr et al., 2018). The study also shows that cue-induced recall of saccharine and alcohol memories recruits comparable and largely overlapping neuronal ensembles in the infralimbic cortex.

Behavioral training. All self-administration training sessions were conducted 2 h after the beginning of the dark phase in operant chambers (interior: length, 30.5 cm; width, 24.1 cm; height, 21.0 cm; model ENV-008CT, Med Associates). These training chambers were located inside sound-attenuating cubicles containing a white noise-generating fan (model ENV-025F28). Habituation and training consisted of three steps, lasting 3–4 weeks. In step 1, only one lever was presented, and the rats underwent four to five sessions of behavioral shaping (water deprived for 20 h before the first two sessions, after which water and food were provided *ad libitum*) until they reached a maximum of 50 drops (30 μ l/drop) of 10% alcohol (v/v in water) under a fixed ratio 1 schedule for a maximum of 1 h. In step 2, the rats were trained to self-administer 10% alcohol in sessions with 60 trials on a fixed intertrial interval (ITI; 15 s) schedule for 5 d. In these sessions, the lever was presented for a maximum of 10 s and was retracted immediately following lever press. In step 3, a cue light above the active lever, an inactive lever on the opposite side, and a variable intertrial interval (10, 15, and 20 s) were introduced. The cue light was first presented for a maximum of 15 s, and 5 s after cue light onset, levers were presented for a maximum of 10 s. Response on either lever terminated the light and caused levers to retract, but only a response to the active lever was deemed successful and was followed by the delivery of alcohol. These training sessions consisted of 60 trials each. Performers with success rates >50% were selected to continue training in the intended recording chamber for approximately another week. The recording chamber (model ENV-007CT; interior: length, 30.5 cm; width, 24.1 cm; height, 29.2 cm) was higher than the training chamber to make room for an electrical swivel commutator (Dragonfly), allowing data acquisition in freely moving animals. In this chamber, one drop contained 40 μ l of 10% alcohol and was supplied in a cup by a motor-driven liquid dipper (model ENV-202 M-S), causing a delay of 1.5 s after active lever press. Head entries through the liquid dipper access opening (5.08 \times 5.08 cm) were detected by interrupting an infrared beam across the entrance (Med Associates), and the behavior was observed with a USB camera (catalog #95353, Delock). Rats with stable performance >70% on 2 subsequent days were selected for tetrode implantation.

Surgical and tetrode placement procedures. Rats were anesthetized with isoflurane (1.5–2.0%). A custom-built flexDrive (Voigts et al., 2013) containing eight tetrodes (12.5 μ m Teflon-coated tungsten wire, California Fine Wire) was unilaterally implanted with a 10° angle toward midline into the prelimbic cortex (Brodman's area 32: anteroposterior, +2.8 to +3.8 mm; mediolateral, +0.8 to +1.3 mm; dorsoventral, –2.5 to –2.6 mm). A bone screw above the cerebellum served as a ground. The craniotomy was stepwise sealed with three-component adhesive (Super-Bond C&B, MPE Dental UG) and two-component embedding resin (Technovit 5071, Kulzer). From the next day after surgery, tetrodes were advanced gradually every second day. Behavioral retraining (step 3 in recording chamber) started 1 week after surgery, and the behavioral task started a few days later when the performance was stable again (>70%). The location of the tetrodes within PL was confirmed (Fig. 1C) in fixed, 50- μ m-thick coronal sections by Nissl staining following current passing (100 μ A, 20 s) to deposit iron particles via a Prussian blue reaction (Ma et al., 2016).

Behavioral task. Each trial (Fig. 1B) started with a visual cue, followed by the presentation of two levers 5 s after cue onset, one of which, the active lever, was directly below the cue light. Only responses to the active lever were reinforced by a reward, the delivery of a drop of alcohol (40 μ l, 10% v/v in water), after a 1.5 s delay. Lever presses on the inactive lever had no consequences. The trial ended, following lever press or 10 s after lever presentation with no response, by terminating the cue and retracting the levers. Pseudorandom intertrial intervals (10, 15, or 20 s) separated the trials. We adopted a within-subject study design with the behavioral response to the visual cue and multiple single-unit activity in PL as the repeated measures and the session (maintenance and extinction) as the between-group variable. Following habituation, appetitive conditioning, tetrode implantation (Fig. 1C), and retraining, a cohort of 10 rats underwent daily maintenance sessions of 60 reinforced trials and were used as the control. The trial number was limited to 60 to ensure steady motivation throughout the session, as confirmed by the constant reaction times of the animals (two-tailed paired Wilcoxon signed-rank test

between reaction times of the first and last of 10 trials, $p = 0.4$, $n = 10$). On a later day, within-session extinction began with 9 reinforced trials followed by 60 unreinforced trials (Fig. 1A). The switch in reward contingency was not signaled to the animals, and other experimental conditions were kept constant throughout the session. The number of inactive lever presses was low and comparable in both maintenance and within-session extinction (maintenance vs within-session extinction, number of inactive lever presses: 0/1 vs 0.5/1 median/interquartile range (IQR); Wilcoxon signed-rank test, $p = 0.38$). Animals were not driven to respond merely to satisfy thirst or hunger, since neither session was performed under water or food restrictions.

Recordings. Multiple single units were simultaneously recorded using a 32-channel RHD2132 amplifier connected to an RHD2000 USB interface board (Intan Technologies). All channels were digitized with 16 bit resolution, sampled at 30 kHz, and bandpass filtered between 0.1 and 8000 Hz. The time stamps for external stimuli (cue light, lever presentation), lever presses, dipper activation, and head entries into the access opening of the liquid dipper were transmitted from the Med Associates behavioral control system (Med-PC IV software, version 4.39, Med Associates) to the Intan Technologies recording system to align behavior to neural activity.

Spike detection and sorting. After bandpass filtering between 300 and 5000 Hz (fourth-order Butterworth filter, built-in MATLAB function), the median voltage trace of all channels was subtracted from each trace to reduce noise. The day with the highest number of single units, of 3 consecutive days with >75% mean success rate during retraining, was chosen for further analysis (maintenance). The threshold for spike detection was set to 5.5 times the median absolute deviation from baseline. Detected spikes were sorted with a custom-built graphical user interface in MATLAB (provided by W. Kelsch, Central Institute of Mental Health, Mannheim, Germany) into individual cell clusters based on peak amplitude and the first three principal components of the waveform. Spike-sorting quality and unit isolation were assessed with MLIB [a MATLAB (MathWorks) toolbox for analyzing spike data by Maik Stüttgen; <https://www.mathworks.com/matlabcentral/fileexchange/37339-mlib-toolbox-for-analyzing-spike-data>]. After spike sorting, <1% of consecutive spikes in accepted clusters had an interspike interval of <2 ms. Cross-correlation analyses supported that each single unit was isolated from other units. Since we were interested in studying how changes in single-unit and population firing rates (FRs) coordinated with behavior, it was of utmost importance to ensure the stability of the recording. Even small movements in the tetrode position may result in upward or downward drift in spike amplitudes relative to the spike-detection threshold, leading to an apparent change in firing rates. To make sure that our measurements of spike amplitude remained unchanged from the beginning to the end of each recording session, we compared the average amplitude of the first and last 100 spikes for each unit. We found no substantial change in the spike amplitude, corroborating the stability of the recordings (Fig. 2A). To further confirm that a few small-amplitude variations did not impair spike detection (therefore contaminating the detected rate), we compared spike amplitude values both at the beginning and at the end of each session with the detection threshold of the corresponding tetrode. For each unit, the detection threshold was well below the spike amplitudes (Fig. 2B).

Statistical analysis

All data were analyzed using built-in and custom-made MATLAB routines (MathWorks), as was model fitting. To correct for multiple comparisons, significance levels were adjusted using the procedure of Benjamini and Hochberg (1995). ANOVAs were performed using the SPSS software (IBM).

z scores. We obtained the single-unit instantaneous firing rate as follows: spike trains were first convolved with a Gaussian kernel (SD = 60 ms), and the resulting time series were time averaged within 100 ms bins. Binning was aligned to the cue onset of each trial. The z-scored response over the considered block of trials was then computed for each unit by first trial averaging (mean firing rate across a block of trials), followed by subtracting the mean and dividing by the SD of the baseline trial-averaged firing rate (2 s before either cue light or lever

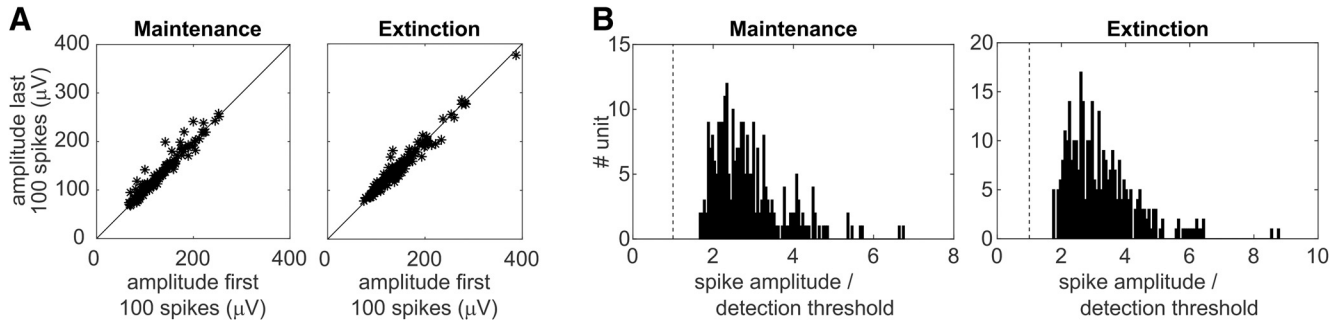


Figure 2. Recording stability. **A**, Absolute value of the average spike amplitude of the first and last 100 spikes of the session for each unit recorded in the maintenance and extinction sessions. The agglomeration of points along the diagonal rules out drifts in the amplitude of the recorded spikes during the session. **B**, Histogram of the ratio between spike amplitude and the spike detection threshold, showing that spike amplitudes are at least two times higher than the detection threshold both at the beginning and at the end of the recording session. Ratios were computed over the average of both the first and last 100 spikes of the session.

presentation). Significant units were identified as those whose absolute average z-scored response over three bins following the stimulus (cue light or lever presentation) is >1.96 (i.e., a value outside the 95% confidence interval of the standardized normal distribution). The area under the curve (AUC) of the z-scored response was similarly computed as the absolute average over three bins.

Modeling behavior. We describe the behavior of each animal by a sequence of learning episodes, each of which is defined by its onset trial, duration, and relative change in performance. Formally, the behavior of each animal is treated as a binary vector $y = y_1, \dots, y_T$, with T the number of trials in a session ($y_t = 1$ for active lever press and $y_t = 0$ for omission or inactive lever press). We thus use an inhomogeneous Bernoulli process $x = x_1, \dots, x_T$ to model response probability, as follows:

$$\Pr(y_t = 1|x_t) = x_t^{y_t} (1 - x_t)^{1-y_t}.$$

The time-varying response probability $x_t \in [0,1]$ characterizing the learning curve of the animal is, in turn, modeled as a weighted sum of B logistic sigmoids, as follows:

$$x_{t,B} = \beta + \sum_{b=1}^B \frac{w_b}{1 + e^{-\frac{t - c_b}{\tau_b}}},$$

where β is baseline response probability. Each sigmoid corresponds to a learning episode that is parametrized by a weight (or magnitude), w_b , measuring the relative change in performance during the episode, a center trial, c_b , and a time constant, τ_b , that scales with episode duration. Centers c_b correspond to B behavioral change points (behavioral $CP_{50\%}$). To highlight the neuronal mechanisms initiating behavioral extinction, we chose a drop of 10% in response probability to mark the onset (behavioral $CP_{10\%}$) of an extinction-learning episode (defined as episodes where $w_b < 0$). These drops were defined for each extinction-learning episode b as the first trial t where $1/(1 + \exp(-(t - c_b)/\tau_b)) \geq 0.1$.

To estimate the learning curve, model parameters are initialized using the Paired Adaptive Regressors for Cumulative Sum (PARCS) method for any given B (Toutounji and Durstewitz, 2018), then inferred by constrained maximization of data log-likelihood, as follows:

$$\mathcal{L}(x_B|y) = \sum_{t=1}^T y_t \log(x_{t,B}) + (1 - y_t) \log(1 - x_{t,B}).$$

Baseline and weight constraints are imposed to assure that response probability is bounded between 0 and 1. Other parameter constraints ($1 \leq c_b \leq T$ and $\tau_b > 0$) enforce model identifiability. Model selection relies on an iterative procedure where, starting from $B = 0$ (corresponding to no learning), a null model of order B is compared against the order $B+1$ alternative, using the following likelihood-ratio test:

$$\lambda(B + 1 : B) := 2(\mathcal{L}(x_{B+1}|y) - \mathcal{L}(x_B|y)) \sim \chi_3^2,$$

where the significance level is set to $\alpha = 0.05$. The number of degrees of freedom corresponds to the difference in the number of parameters between the two models.

Change-point analysis. To relate neural and behavioral dynamics during each session, we identified sudden jumps (i.e., CPs) in neuronal firing rates within certain windows of interest both across the entire population and single units (population CP/single-unit CP). Detection of multiple change points in neuronal recordings was performed using the PARCS method (Toutounji and Durstewitz, 2018). An important advantage of using PARCS for multiple change-point detections compared with other methods (Olshen et al., 2004; Cho and Fryzlewicz, 2015) is that it avoids segmenting the data which, for short time series, as is the case here (60–69 data points), may quickly deplete statistical power. The method relies on the observation that a time series containing change points can be approximated by a piecewise constant function, while its integral [i.e., the cumulative sum transformation of the time series (CUSUM)] by a piecewise linear function that bends at the CPs. For neural CPs, the method is applied to the square root-transformed spike counts in each window of interest, bringing counts closer to a Gaussian distribution and stabilizing the variance (Kihlberg et al., 1972). PARCS infers the piecewise linear approximation of the CUSUM transformation and its bending points corresponding to the CPs using adaptive regression splines (Friedman and Silverman, 1989). The significance of a CP is then decided using a test statistic that quantifies the amount of bending at the CP and a permutation bootstrap procedure (Toutounji and Durstewitz, 2018). Crucially, when considering population CPs, PARCS operates on the multivariate CUSUM transformation of the time series, assuring that population CPs are not averaged out: the test statistic takes into account the amount of bending at a candidate population CP for each individual neuron, regardless of whether the spike counts of each neuron increase or decrease following the CP. An upper bound of three on the number N of CPs per unit or population is chosen and the nominal significance level for the permutation bootstrap procedure of the method is set to $\alpha = 0.2$ to correct for the conservativeness method's (Toutounji and Durstewitz, 2018).

Relating behavioral model to population CPs. To quantify locking between population CPs and behavior, we developed a measure for comparing the likelihoods that two sets of population CPs are sampled from one behavioral response probability distribution $p^b(y_t = 1)$. This distribution is the sum of B bell-shaped curves (each peaking at one behavioral $CP_{50\%}$ c_b and of width that scales with τ_b), computed by normalizing the first differences $|\Pr(y_{t+1} = 1|x_{t+1,b}) - \Pr(y_t = 1|x_{t,b})|$ to sum up to 1. Similarly, a neural response probability distribution $p^p(\Delta FR_t \neq 0)$ is computed as the sum of N Dirac δ functions, centered at the N population CPs c_n . Weights are computed by averaging and normalizing $|\Delta FR_{c_n}|$ over the whole population, such that $p^p(\Delta FR_t \neq 0)$ sums up to 1. Given two neural response distributions, p_i^p and p_j^p , we compute the likelihood ratio, as follows:

$$\lambda(p_i^p : p_j^p) := 2 \log \frac{\sum_{n=1}^{N_i} p_i^p (\Delta FR_{c_n} \neq 0) p^b(y_{c_n} = 1)}{\sum_{n=1}^{N_j} p_j^p (\Delta FR_{c_n} \neq 0) p^b(y_{c_n} = 1)}$$

Positive $\lambda(p_i^p : p_j^p)$ indicates a stronger locking to the same behavior of the set i of population CPs, relative to the set j . We compute $\lambda(p_i^p : p_j^p)$ for each of the 10 animals, where p^b and p_i^p correspond to the response probability and neural response distributions of the extinction session for the animal, respectively, and p_j^p to the neural response distribution of each of the other nine animals ($i \neq j$).

Characteristics for comparing different CP sets. We compare different sets of single-unit CPs using the following statistics: frequency, # s.u.CP/units; relative rate change, $\rho := 2|FR_{pre\ s.u.CP} - FR_{post\ s.u.CP}| / (FR_{pre\ s.u.CP} + FR_{post\ s.u.CP})$; and sign ratio, # positive s.u.CPs/# negative s.u.CPs. The firing rates $FR_{pre\ s.u.CP}$ and $FR_{post\ s.u.CP}$ were computed over the periods of constant firing rates around the single-unit CP, defined by PARCS as the longest periods before and after a CP where no other change point was detected.

Sensitivity analysis. We computed for each unit the following statistic:

$$d = \frac{\bar{r}_2 - \bar{r}_1}{\sqrt{(\sigma_1^2 + \sigma_2^2)/2}},$$

where \bar{r}_i and σ_i^2 are the firing rate mean and variance over a block of trials, respectively (first vs last 12 maintenance and extinction trials). Positive d' indicates an increase of firing rate in block 2 compared with block 1 measured in units of average SD, and vice versa.

Comparing single-unit CP distributions across sessions. We compare the empirical distributions function (EDF) of single-unit CP occurrence during maintenance and within-session extinction within the five task windows considered using P-P (probability-probability) plots. Given the different number of trials in the two sessions (60 and 69 trials, respectively), we first linearly time warped extinction trials to fit within 60 bins, which we then used to compute the EDF of single-unit CPs of that session.

Single-unit and behavioral CP coordination. We collected all single-unit CPs and aligned them with respect to each behavioral CP_{10%} of the corresponding animal (i.e., for animals with B extinction-learning episodes, single-unit CPs were considered B times). The aligned single-unit CPs were then binned using a three-trial bin. For statistical testing, we used permutation bootstraps to test whether single-unit CP frequencies within 10 trials from the behavioral CPs were statistically larger than what is expected by chance. We generated bootstrap histogram samples by randomly shuffling the occurrence of the single-unit CPs of each animal and repeating the alignment procedure described above. This was done by permuting the trial order but keeping co-occurring single-unit CPs of different units at the same trial. Behavioral CPs were left unchanged. The 5000 histogram samples so obtained were then used to produce an EDF over single-unit CP frequency per histogram bin. At each of the seven frequency values around the behavioral CP of the original single-unit CP histogram, a p value was assigned on the basis of the EDF of their corresponding bin. To test significance with a higher temporal resolution than a three-trial binning, we repeated the bootstrap procedure by sliding the histogram bin edges by one and then two trials. As control, we repeated the test using behavioral CPs from the extinction session but single-unit CPs from maintenance.

Predicting behavioral CPs from PL population vectors. We tested whether the firing rate of the PL population in different task windows of interest could predict changes in animal behavior. For each animal, we considered the population vector $V_t^W = [FR_t^1, \dots, FR_t^M]$ constructed from firing rates FR_t^i of units i in trial t in the task window W . On the basis of these population vectors, we then trained a support vector machine classifier with linear kernel (slack variables minimized with L^1 norm and box constraint = 1) to divide the trials occurring before the first behavioral CP from those occurring after the behavioral CP. For this analysis, if behavior contained multiple extinction-learning episodes, only the first was considered, to have a comparable chance level across animals.

Classifier accuracy was computed with a 10-fold cross-validation to avoid overfitting. Since the sample was imbalanced and the two classes (before/after behavioral CP) were not of equal size, we used the Cohen's κ coefficient to quantify classifier accuracy relative to chance level. Cohen's κ ranges between -1 and 1 and is defined as follows:

$$\kappa = \frac{p_0 - p_e}{1 - p_e},$$

with p_0 the fraction of correctly classified samples and p_e the expected probability of correct classification due to chance. Kappa values were computed on the classifier output collected over the 10 folds for each animal and each task window. For statistical testing, the significance of the κ coefficients was tested through bootstrap. Any monotonic change in firing rate can improve the performance of a classifier trained to divide temporally ordered samples. To account for this factor and test exclusively for the behavioral CP and population rate coordination, we created a bootstrapped sample by repeatedly assigning the behavioral CP to a random trial. We then trained the classifier to divide trials occurring before the shuffled behavioral CP from those occurring after it. The procedure was repeated 100 times for each animal and each task window. The obtained κ^{boot} values were then averaged per animal and task window, generating a reference set $\kappa_{\text{boot}}^W = [\kappa_1^{\text{boot}}, \kappa_i^{\text{boot}}, \dots, \kappa_{10}^{\text{boot}}]$ with $i = 1 \dots 10$ indexing the animal. Since the performance of a classifier highly depends on the number of units composing the population vector, we compared the set of original κ values with κ_{boot}^W with a one-tailed paired t test. Significance was assessed with Benjamini–Hochberg correction for multiple comparisons.

Coordination between single-unit CPs across windows. To quantify coordination between single-unit CPs detected within one window and those of the same unit detected within another, we proceeded as follows: for each single-unit CP detected in the first window, we computed the distance in trials between its occurrence and that of the nearest single-unit CP of the same unit detected in a different task window. Absolute distance values were then averaged across all units of an animal.

Results

We recorded single-unit activity within the PL region of the mPFC (Fig. 1C) while 10 adult male Wistar rats (~14 weeks old) performed a visually guided operant conditioning task during maintenance and its subsequent within-session extinction (Fig. 1A,B). Alcohol in low concentration was used as a positive reinforcer (see Materials and Methods). Response probability (rate of active lever presses) during maintenance was high, indicating that rats had learned to associate the visual cues with reward (Fig. 3A). This high response probability dropped during within-session extinction, indicating that behavior was extinguished when the visual cues were not reinforced any more (60 maintenance vs last 18 within-session extinction trials, percentage of active lever presses: $87.7 \pm 2.5\%$ vs $12.8 \pm 3.1\%$, mean \pm SEM; right-tailed Wilcoxon signed-rank test, $p = 9.8 \times 10^{-4}$; Fig. 3A). To inspect the timing of behavioral changes during within-session extinction, responses were binned in blocks of six consecutive trials. We observed a gradual reduction in response probability across the whole cohort starting at trials 16–21, and an intermittent, albeit not significant, increase at trials 40–45 (Wilcoxon signed-rank tests between consecutive blocks with Benjamini–Hochberg adjusted $p < 0.05$ and $p < 0.08$, respectively; Fig. 3B).

PL activity remains modulated by conditioned cues during extinction learning

Next, we investigated the PL response properties following cue and lever presentation by inspecting the z scores of 132 and 162 units recorded during maintenance and extinction, respectively

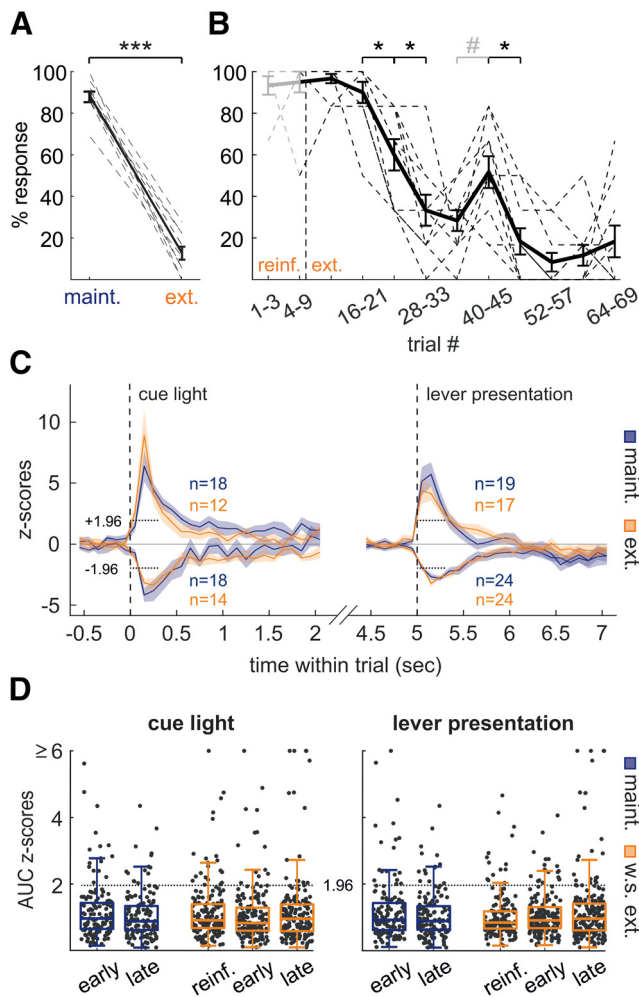


Figure 3. PL activity remains modulated by conditioned cues during extinction learning. **A, B**, Percentage of active lever presses during maintenance and the last 18 trials of extinction (**A**) and throughout within-session extinction (**B**). Dashed lines show percentages for individual animals. Solid line and error bars show the mean \pm SEM. Asterisk and hash symbols mark Benjamini–Hochberg-corrected $p < 0.05$ and $p < 0.08$, respectively. **C**, The z-scored activity of significantly responding units (number of units shown for each curve) following cue light and lever presentation (see Materials and Methods). Horizontal dotted lines mark the significance threshold and testing window. Solid lines and shading show the mean \pm SEM. **D**, AUC for z-scored single-unit response (see Materials and Methods) of all units computed on trial blocks of steady-state behavior (early/late: first/last 12 trials during maintenance and extinction; reinforced: 9 reinforced trials during within-session extinction). Boxplot whiskers extend to include points within 1.5 of the IQR. Horizontal dotted lines mark the significance threshold.

(see Materials and Methods). Figure 3C compares the average z-scored activity of significantly responding units during maintenance and significantly responding units during the unreinforced trials of within-session extinction. Excitatory as well as inhibitory rate modulations following cue light and lever presentation were comparable (uncorrected Wilcoxon rank-sum test; cue light maintenance vs extinction excitatory/inhibitory, $p = 0.56/0.31$; same for lever presentation, $p = 0.50/0.48$). To test whether the response of PL units to task stimuli is predictive of the behavioral response probability (Fig. 3A,B), we also compared the activity of all recorded units during trial blocks of stable behavior in both maintenance and within-session extinction (first/last 12 trials during maintenance and extinction; 9 reinforced trials during within-session extinction; Fig. 3D). We found no significant difference in the overall z score distributions of PL unit responses, neither when comparing different

steady-state blocks within the same session nor when comparing corresponding blocks between the two sessions (uncorrected Wilcoxon rank-sum test; cue light maintenance vs extinction early/late, $p = 0.17/0.73$; early vs late maintenance/extinction, $p = 0.32/0.47$; reinforced vs extinction early/late, $p = 0.13/0.49$; same for lever presentation, $p = 0.68/0.53$, $p = 0.45/0.80$, $p = 0.64/0.50$). These results demonstrate that, as behavior changes toward extinguishing the cue–reward association, PL remained responsive to task-related cues to a similar degree as during maintenance. Moreover, during within-session extinction, the overall response to cues remained consistent across different blocks of steady-state behaviors, whether the animal responded to the task or not. These findings are in line with previous observations showing that the overall proportion of different mPFC units responding to different task aspects remains about the same despite changes in task rules and contingencies (Ma et al., 2016).

Whole-trial PL population activity reflects behavioral changes during extinction learning

Despite demonstrating a consistent decrease in behavioral response probability across animals, the above analyses do not capture trial-by-trial changes and idiosyncrasies in the behavior of each animal (Gallistel et al., 2004), which may conceal relevant aspects of the relationship between PL activity and extinction learning. To address this, we developed a new statistical model of binary choice behavior that captures the response–probability dynamics of an animal (i.e., learning curve) by a weighted sum of sigmoids (see Materials and Methods). Each sigmoid corresponds to a separate learning episode, that is characterized by the following three parameters: the trial at which the sigmoid is at half height (behavioral $CP_{50\%}$), a slope that defines the duration of the episode, and a weight specifying the amount and direction of change in behavior during the episode. Statistical model selection allowed us to specify the smallest number of episodes required to explain $\geq 95\%$ of the behavioral variance of an animal. This approach revealed that the tested cohort adopted a variety of behavioral profiles that differed in the degree of abruptness of behavioral changes and in the eventual occurrence of transient reinstatements of the conditioned behavior. We found one or two extinction-learning episodes (sigmoidal curves with negative weights) in six and three animals, respectively, and two extinction-learning episodes separated by a reinstatement episode (positive weight) in one animal (Fig. 4A). We then computed the spike counts during whole trials and identified population-wide CPs from the PL units recorded from each animal using PARCS (Toutounji and Durstewitz, 2018; see Materials and Methods). PARCS identifies trials with significant changes in firing rate across the entire population, cumulating over rate changes in individual units regardless of the sign of these changes. It is important to note that this approach captures longer-lasting changes in coding (i.e., significant jumps in firing probability between periods of relative stability, and not stochastic trial-to-trial variability). Despite the varying number of units recorded per rat, two population CPs were detected in all animals. Visual inspection suggested that changes in response probabilities, as captured by the behavioral model, are often accompanied by population CPs. To quantify this relationship, we developed a statistical bootstrap procedure, based on computing a likelihood-ratio statistic, λ (see Materials and Methods). This statistic compares how strongly the change in response probability of an animal locks to its own population CPs versus an alternative set of population CPs detected in another animal. Considering all possible combinations of model-

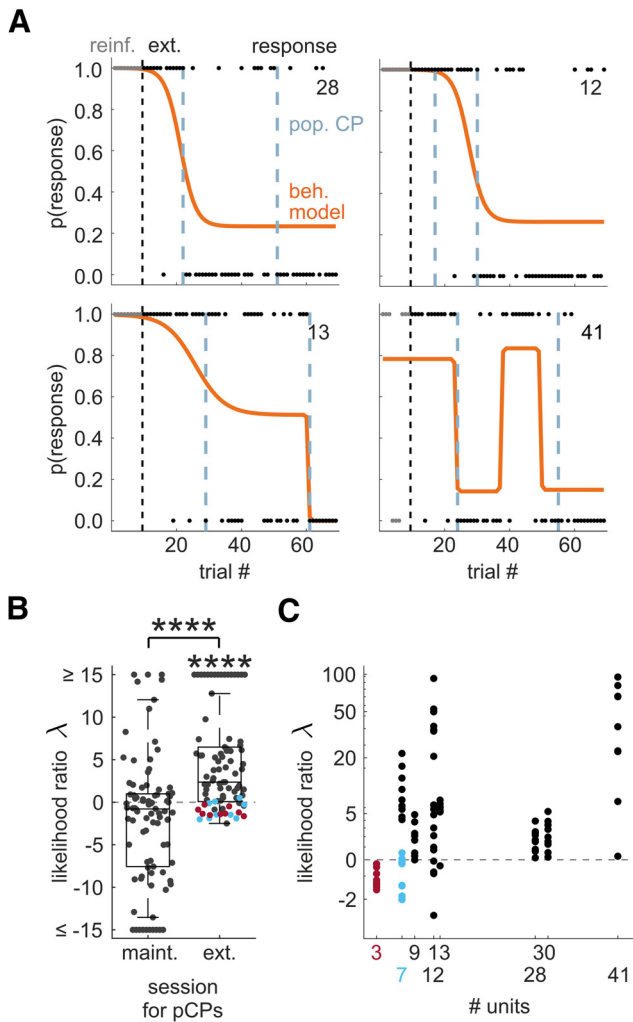


Figure 4. Whole-trial PL population activity reflects behavioral changes during extinction learning. **A**, Examples of the behavioral models (orange) of four representative animals and their respective population CPs computed over the whole-trial firing rate of the population during within-session extinction (light blue). Filled black circles indicate the trial-specific behavioral choice. Dashed line indicates the onset of extinction trials. Numbers at the top right of each panel indicate the number of recorded units. **B**, Distribution of likelihood ratio test statistic for relating the set of behavioral response models during within-session extinction to maintenance population CPs (λ_{maint} ; left) and extinction population CPs (λ_{ext} ; right). Points in magenta and cyan correspond to λ_{ext} values from two animals, which were consistently close to or < 0 , indicating a poor match between population CP and the behavior of the two. Boxplot whiskers extend to include points within 1.5 of the IQR. **C**, Number of units recorded from each rat against its corresponding likelihood ratio test statistic values λ_{ext} . Points in magenta and cyan pertain to λ_{ext} values from the corresponding rats in **B**.

estimated learning curves and population CPs, we found a substantial bias toward positive λ values during extinction (λ_{ext} ; right-tailed Wilcoxon signed-rank test, $p = 2.0 \times 10^{-10}$; Fig. 4B, right), indicating a strong match between the behavioral model and the population CPs of the same animal. As a control, we also computed another set of λ values (λ_{maint}) using maintenance population CPs and related them to the same set of learning curves estimated during within-session extinction. Contrasting the λ_{ext} distribution with the null λ_{maint} distribution (Fig. 4B) further confirmed that behavior is significantly more locked to population CPs than expected by chance ($\lambda_{\text{ext}} > \lambda_{\text{maint}}$; right-tailed Wilcoxon signed-rank test, $p = 2.0 \times 10^{-8}$). Furthermore, λ_{ext} values corresponding to an individual animal significantly correlated with the number of units recorded from that animal

(Spearman's rank correlation coefficient, $p = 1.4 \times 10^{-5}$, $r = 0.44$; Fig. 4C), indicating that the few cases with a poor match between population CP and behavior may be because of undersampling PL units, rather than to differences in the neural mechanisms underlying extinction learning across animals (Fig. 4B,C, magenta and cyan points). These analyses show that the temporal evolution of behavior during extinction learning was strongly reflected in the population dynamics of PL neurons.

PL single-unit dynamics during extinction learning is indistinguishable from that during maintenance

While population CPs result from the activity of the whole set of recorded units and may reflect the overall dynamics of the PL network, one might expect a certain degree of heterogeneity in single-unit encoding. Indeed, not all change points estimated from single-unit whole-trial spike counts (single-unit CPs) coincided with population CPs. This is because of the fact that significant, yet relatively weak, changes in the firing rate of one unit may not be shared with other units to be deemed significant at the population level. To pinpoint more exactly during which task phases extinction-related changes happened and how different units were involved in them, we identified an additional set of four single-unit CPs estimated from the spike counts of four within-trial windows of interest (Fig. 5B). The cue-light and lever-presentation windows, defined as the 0.5 s after stimulus onset, allowed monitoring PL network responses to task-related external stimuli. The delay period window, which spans the 2 s preceding lever presentation, was selected to assess potential effects of reward expectation. Finally, the ITI window, between -3 and -1 s before cue-light onset, allowed consideration of PL dynamics independently of specific task-related activity. Importantly, none of these windows included trial periods where motor responses were expected, which allowed a fair comparison between maintenance and extinction.

Irrespective of the window examined, single-unit CPs during extinction learning were distributed across the entire session (Fig. 5C). On average over all recorded units, single-unit CPs occurred in all windows with similar frequency, same relative change in firing rate ρ , and with balanced positive to negative sign ratio (see Materials and Methods for formal definitions of these three statistics; frequency: one-way ANOVA on task windows, main effect: $F_{(3,36)} = 2.0$, $p = 0.1$; Fig. 5D, right; ρ : one-way ANOVA on task windows, main effect: $F_{(3,36)} = 1.7$, $p = 0.2$; Fig. 5E, right; sign ratio: t test on the fraction of positive over negative jumps against 1, uncorrected $p > 0.05$ for all windows; Fig. 5F, right). Besides, the average PL firing rate remained constant throughout the whole session for all task windows (see Materials and Methods; one-way ANOVA for repeated measures to test the effect of the trial block on the population firing rate: ITI: $F_{(10,90)} = 0.6$, $p = 0.8$; light: $F_{(10,90)} = 0.7$, $p = 0.7$; delay: $F_{(10,90)} = 0.9$, $p = 0.5$; lever: $F_{(10,90)} = 1.9$, $p = 0.1$; Fig. 5G). Although overall changes in firing rate across units were balanced, the specific changes in single-unit firing rates resulted in a reorganization of the PL coding throughout the session: while the unit responses within the cue-light and lever-presentation phases remained unchanged on average (across the population) from the beginning to the end of the extinction session (Fig. 3D), the identity of task-responsive units varied. Approximately 21.6% of the recorded units changed the trial-averaged firing rate by > 2 SDs between the first and last 12 extinction trials of the session (sensitivity index d' ; see Materials and Methods; Fig. 5H, middle, example animal). Only 30% of the units with a significant response in the first reinforced trials of within-session

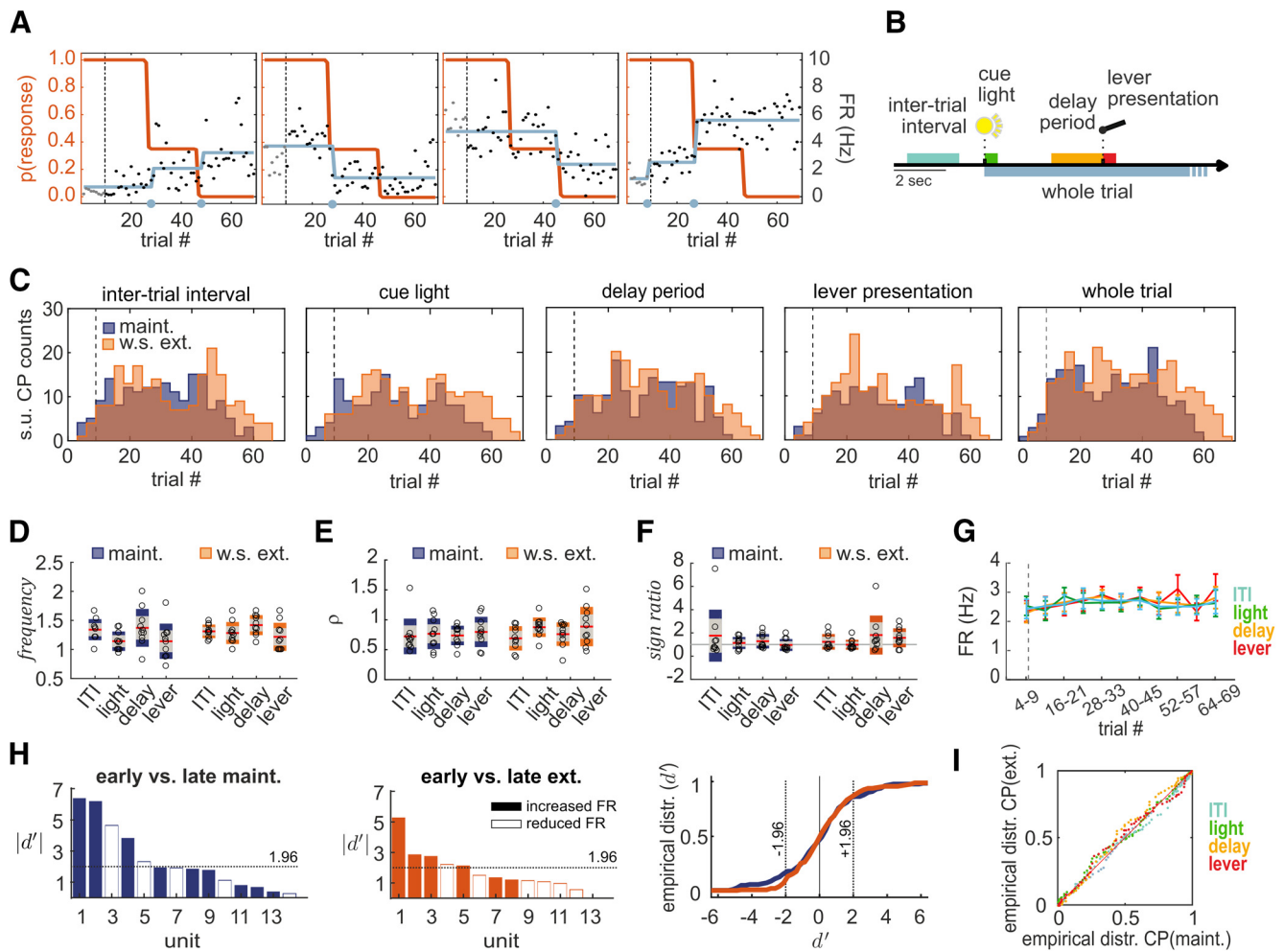


Figure 5. PL single-unit dynamics during extinction learning is indistinguishable from that during maintenance. **A**, Four examples (same animal) of single-unit whole-trial firing rates (black dots) during within-session extinction, with single-unit CPs (light blue filled circles) and the firing rate as inferred by the CP detection algorithm (light blue solid line). Behavioral model shown in orange as in Figure 4A. **B**, Five task windows of interest within which five sets of population and single-unit CPs were identified from population and single-unit firing rates. Windows are defined relative to light onset as follows: ITI, seconds -3 to -1; cue light, seconds 0 to 0.5; delay period, seconds 3 to 5; lever presentation, seconds 5 to 5.5; and whole trial, seconds 0 to 15. **C**, Distribution of single-unit CPs across maintenance and extinction trials (60 and 69 trials, respectively) for each task window, pooled from all animals. **D–F**, Number of single-unit CPs per unit (**D**), relative change in firing rate (**E**), and positive-to-negative sign ratio (**F**) computed in four task windows. Plots show the mean \pm 1.96 SEM (red and gray) and SD (blue or orange). Open circles indicate the mean for individual animals. The three quantities (**D–F**) are statistically indistinguishable when compared between sessions. Gray line in **F** marks a sign ratio of 1, where positive and negative rate changes are balanced. **G**, Population firing rate per task window over blocks of six consecutive trials during within-session extinction (trials 1–3 excluded; Fig. 3B). Dashed line indicates the onset of extinction trials. Solid lines and error bars show the mean \pm SEM. **H**, Sensitivity analysis showing increased and decreased single-unit whole-trial firing rates of a representative animal during the first and last 12 trials of maintenance (left) and extinction (middle). Empirical distribution functions (right) of the sensitivity index d' for all recorded single units from all animals in maintenance (blue) and extinction (orange) show no significant difference, despite difference in behavior. Dotted lines mark the threshold of significant change in firing rate. **I**, P-P plot comparing the empirical distribution function of single-unit CPs over maintenance and within-session extinction trials (compare **C**) for the four task windows.

extinction (40% and 20%, following cue light and lever presentation, respectively) were also responsive at the end of the session. Surprisingly, this degree of change in task-responsive units was also present during maintenance, where both experimental conditions and animal behavior were constant throughout the session (Fig. 5D–F,H, left). We found no significant difference between maintenance and extinction single-unit CPs for all windows (ITI/light/delay/lever) with regard to the distribution of CP frequency, ρ , sign ratio, and the distribution of CP occurrence across the trials of each session (frequency: repeated-measures two-way ANOVA; factors: session/window; main effects: session, $F_{(1,9)} = 1.4$, $p = 0.3$; window, $F_{(3,27)} = 5.9$, $p = 0.003$; interaction, $F_{(3,27)} = 0.5$, $p = 0.7$; Fig. 5D; ρ : repeated-measures two-way ANOVA factors: session/window; main effects: session, $F_{(1,9)} = 0.5$, $p = 0.5$; window, $F_{(3,27)} = 1.6$, $p = 0.2$; interaction, $F_{(3,27)} = 0.4$, $p = 0.7$; Fig. 5E; sign ratio: two-way

factors: session/window; main effects: session, $F_{(1,8)} = 0.13$, $p = 0.7$; window, $F_{(3,24)} = 1.6$, $p = 0.2$; interaction, $F_{(3,24)} = 1.5$, $p = 0.2$; Fig. 5F; *post hoc* Bonferroni correction applied in all ANOVAs; no significance found *post hoc*; single-unit CP trial distribution: see Materials and Methods; two-sample Kolmogorov–Smirnov test for single-unit CPs in ITI, $p = 0.4$; light, $p = 0.4$; delay, $p = 0.1$; lever, $p = 0.4$; Fig. 5, compare I, C). Moreover, similar to within-session extinction, a large fraction of units changed their trial-averaged firing rate from the start to the end of the maintenance session (30.3%; Fig. 5H, left). The distributions of the sensitivity index d' for units recorded in the maintenance versus extinction sessions were, in fact, not significantly different (two-sample Kolmogorov–Smirnov test, $p = 0.5$; Fig. 5H, right).

In summary, PL units changed their responsiveness to the task stimuli with balanced positive and negative single-unit CPs

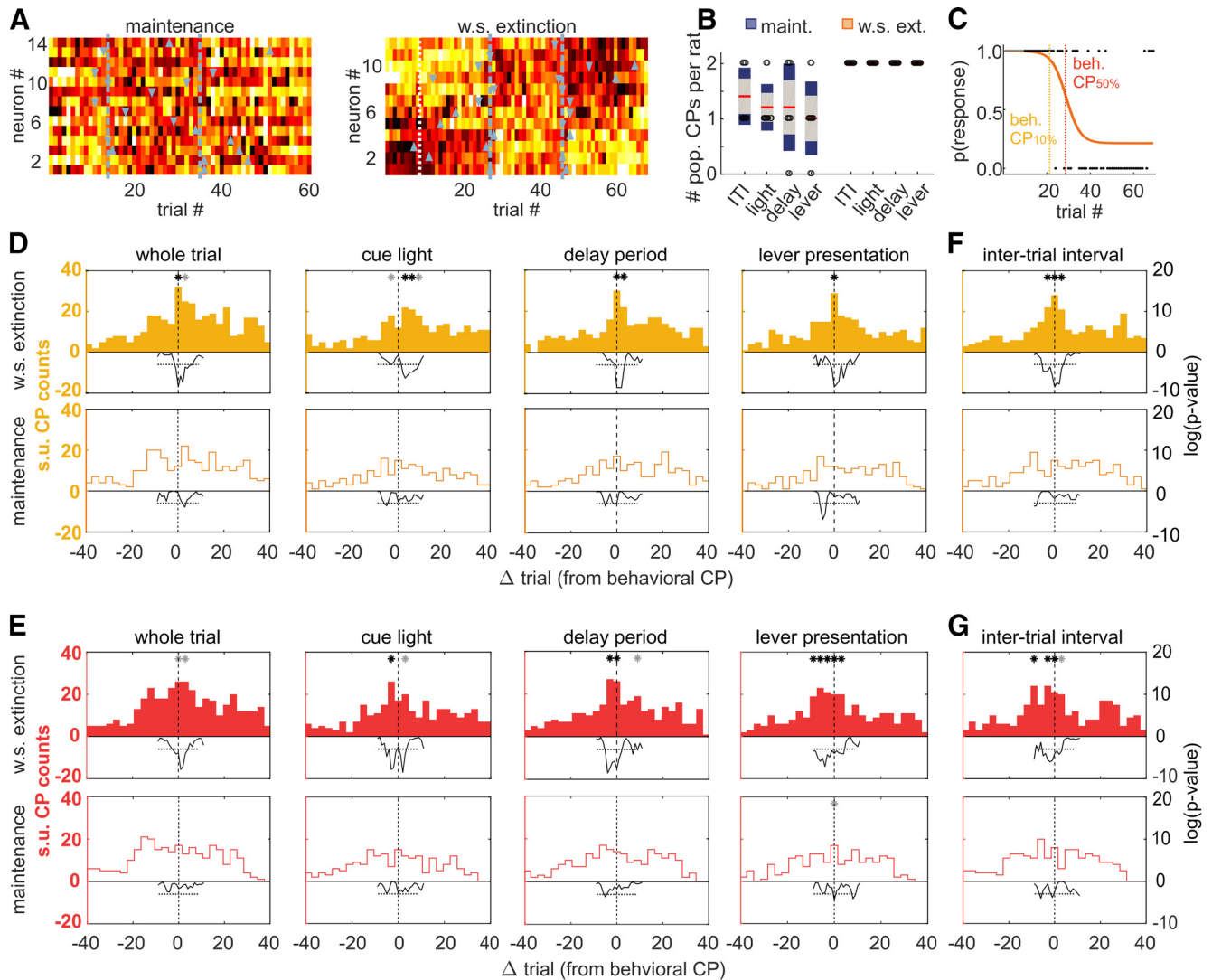


Figure 6. PL baseline rate and task-evoked responses change in anticipation of behavioral extinction. **A**, The z-scored whole-trial response of all recorded units from one representative animal during maintenance (left) and within-session extinction (right), overlaid with population CPs (blue dashed lines) and single-unit CPs (blue triangles). Triangle directions indicate whether the CP results from an increase or decrease in the firing rate of the corresponding unit. The z scores are shown with the same scale in both sessions. Dashed white line indicates the onset of extinction trials. **B**, Number of population CPs per animal. Plots show the mean \pm 1.96 SEM (red and gray) and SD (blue or orange). Open circles indicate the numbers for individual animals. **C**, Onset (yellow) and center (red) of an extinction-learning episode for one representative animal. Behavioral CP10% and behavioral CP50% correspond to 10% and 50% drops in response probability, respectively. **D**, **E**, Single-unit CP distributions for different task windows (whole trial, cue light, delay period, and lever presentation) pooled across animals and aligned with respect to behavioral CP10% (**D**) and CP50% (**E**) from the within-session extinction of each animal. Single-unit CPs of the extinction session coordinated at extinction onset in all windows (top), while those of the maintenance session showed no significant coordination when aligned to the extinction onset of the extinction session (bottom). Statistical tests performed via bootstrap (see Materials and Methods). The p values assigned to each trial lag (center of the bin) are reported on logarithmic scale for visibility. The Benjamini–Hochberg correction for multiple comparisons was performed only on the p values of the seven bins of the displayed histogram. Asterisks mark $p < 0.05$ (black) and $p < 0.1$ (gray) after correction. Horizontal dotted lines mark the log (0.05) threshold over the tested window. **F**, **G**, same as **D** and **E**, respectively, on single-unit CPs computed on the ITI window.

during both maintenance and extinction. This reorganization occurred to the same extent in both sessions and therefore was not induced by changes in experimental conditions.

PL baseline rate and task-evoked responses change in anticipation of behavioral extinction

The above analysis revealed that, during prolonged stretches of time (every session lasted ~ 30 min), PL single units changed their responsiveness to stimuli. Such variation in firing patterns also occurred during maintenance and the number of single-unit CPs was comparable in the two sessions. Therefore, we wondered whether the observed match between population CPs and change in behavioral response probability during extinction was because of a coordination of single-unit CP occurrences, locked

to extinction-learning episodes rather than to an overall increase or decrease in firing rates. Visual inspection of the timing of single-unit CPs relative to population CPs of each animal computed over the whole-trial windows during both sessions lent initial support to this hypothesis (Fig. 6A, exemplary animals). Furthermore, we found that, despite the similarity in single-unit CP-related statistics between the two sessions, the number of population CPs per rat at the population level was significantly higher during extinction, regardless of the task window considered (two-way ANOVA; main effect of session: $F_{(1,72)} = 67.8$, $p = 5.7 \times 10^{-12}$; main effect of task window: $F_{(3,72)} = 7.0$, $p = 0.6$; interaction: $F_{(3,72)} = 0.7$, $p = 0.6$; Fig. 6B). Since population CPs are more likely to emerge when more units undergo relatively simultaneous firing-rate changes, this result further

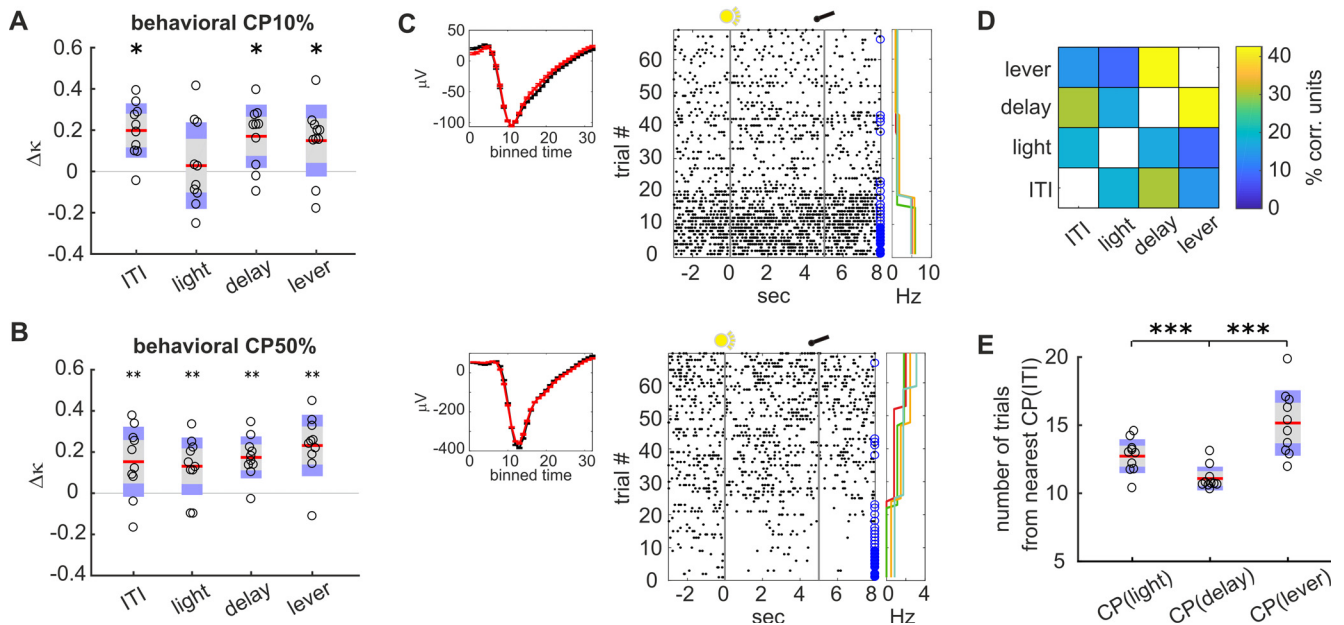


Figure 7. Reorganization in PL activity is predictive of behavioral extinction in all task windows. **A, B**, Classifier performance in predicting the behavioral state of the animal defined with respect to behavioral CP_{10%} (**A**) and CP_{50%} (**B**) values from population firing rates during four task windows (Fig. 5B). Significance was assessed via bootstrap (see Materials and Methods). Differences between data and bootstrapped Cohen's κ are reported by showing the mean \pm 1.96 SEM (red and gray) and SD (purple). Open circles indicate differences for individual animals. Population rates are predictive of extinction onset in the ITI, delay-period, and lever-presentation windows. **C**, To the left, raster plots of two representative units from the same animal, showing rate progression across extinction (bottom to top). Filled and open blue circles mark trials with reinforced and unreinforced lever presses, respectively. Vertical gray lines indicate cue-light onset and lever presentation. Single-unit firing rates based on CP detection in four task windows are color coded as in Figure 5B. To the right, average spike waveform for the first (black) and last (red) 100 spikes of the session, confirming that the observed rate changes could not be ascribed to recording artifacts (Fig. 2). **D**, Fraction of single units for which the evolution of firing rates within one window significantly correlates with that within a second window. **E**, Firing-rate changes during ITI are most coordinated with those occurring during the delay period. Absolute distance in trials between the occurrence of a single-unit CP in ITI and the closest single-unit CP of the same unit in the cue-light, delay-period, and lever-presentation windows. Plots show the mean \pm 1.96 SEM (red and gray) and SD (purple). Open circles mark values for individual animals.

suggests a lack of coordination in single-unit CPs during maintenance.

To test this hypothesis further, we considered for each animal the lag between its single-unit CPs and the onset of an extinction-learning episode. Extinction onset was defined as the trial corresponding to a decrease by 10% (behavioral CP_{10%}) in response probability during an extinction learning episode (see Materials and Methods; Fig. 6C). We found that single-unit CPs computed over the whole trial locked with zero lag to behavioral CP_{10%} (see Materials and Methods; Fig. 6D). A similar match was also found for single-unit CPs occurring during the cue-light, delay-period, and lever-presentation windows (Fig. 6D). To further confirm the temporal coordination between single-unit CPs and the learned behavior, we performed, as a control, the same analysis but matching behavioral CP_{10%} values during extinction to the single-unit CPs of maintenance instead. In addition to an increase in single-unit CP occurrence probability at the center of the maintenance session, which explains the nonuniform distribution of maintenance single-unit CPs around behavioral CP_{10%} values, no fine-tuned coordination was found (Benjamini–Hochberg-corrected bootstrap test; see Materials and Methods; Fig. 6D). Behavioral extinction thus coincided with a coordinated change in PL single-unit firing rates. Such coordination preceded the change in behavior. In fact, while the behavioral CP_{10%} values indicate a decrease in response probability from a baseline of 10%, 7 of 10 animals still responded to all lever presentations until, and including, the first behavioral CP_{10%} mark. As a further confirmation, we found in all four task windows considered that the single-unit CP probability significantly increased a few trials before behavioral CP_{50%} values (Fig. 6E).

Interestingly, also during the ITI, when the animal was not actively engaged in the task, PL firing rate changed in anticipation of behavioral extinction (Fig. 6F,G). To further confirm whether behavioral changes could be predicted from the activity of PL neurons even before the beginning of a trial, we trained a classifier on PL population spike counts during the ITI window preceding trial t to predict whether the animal will still be committed ($t <$ behavioral CP) or not ($t >$ behavioral CP) to the task on that trial. We measured the classifier performance using Cohen's κ (where $\kappa = 1$ corresponds to a perfect prediction, $\kappa = 0$ to chance and $\kappa = -1$ to complete mismatch; see Materials and Methods). We found that, based on population vectors constructed from the ITI, $\kappa = 0.45 \pm 0.08$ when considering behavioral CP_{50%} values and $\kappa = 0.49 \pm 0.07$ when considering behavioral CP_{10%} values. To confirm that the observed effect was because of behavioral extinction and not, more generally, to random monotonic changes in firing rate across the session, we constructed a bootstrap replica where unaltered PL population vectors were used to predict the occurrence of randomly generated behavioral CPs (see Materials and Methods). PL activity during the ITI could predict the original behavioral CPs significantly better than the bootstrapped replicas (one-tailed paired t test, $p = 5.2 \times 10^{-4}$; Fig. 7A). A similar result was obtained when considering population spike counts during the delay-period and lever-presentation windows (one-tailed paired t test, Benjamini–Hochberg adjusted over the four task windows, $p < 0.012$ for all task windows except cue light; Fig. 7A, B). PL population activity during the cue-light window was at chance level in predicting behavioral CP_{10%} (Fig. 7A), but above chance when predicting behavioral CP_{50%} values (one-tailed paired t test, Benjamini–Hochberg adjusted over the four task windows, $p < 0.01$ for all task windows; Fig. 7B). This was in line

with the observed lags in temporal coordination between single-unit CPs during the cue-light window and either behavioral CP_{10%} or behavioral CP_{50%} values (Fig. 6, compare *D* and *E*, respectively), suggesting that changes in PL response to cue light occurred between drops of 10% and 50% in the response probability of an animal.

The previous analyses showed that behavioral CPs were anticipated by a coordinated change in firing rate that affected all task windows. Moreover, the ability to predict behavior from population vectors did not differ between task windows (repeated-measures one-way ANOVA on $\Delta\kappa$ computed on the four windows; main effect: $F_{(3,27)} = 1.3$, $p = 0.3$; Fig. 7*B*). These results suggest that behavioral extinction corresponds to a global reorganization of network activity across all task phases, rather than to a modulation of unit responses to specific conditioned stimuli. Figure 7*C* shows raster plots of two exemplary units from the same animal, depicting how changes in baseline firing rate across the whole trial occurred in coordination with behavioral extinction, both in stimulus-responsive (Fig. 7*C*, bottom) and nonresponsive (Fig. 7*C*, top) units. Firing-rate changes during the four task windows of interest (analyzed pairwise) correlated in ~22% of the recorded units (Spearman's correlation, Benjamini–Hochberg adjusted, $p < 0.05$; Fig. 7*D*). Notably, despite the ITI window being positioned between the lever-presentation and cue-light windows of two successive trials, the firing rate during the ITI was significantly correlated for more units with that during the delay-period window (30%) than that during the cue-light (19%) and lever-presentation (14%) windows. Since changes in firing rate were better coordinated between the ITI and the delay period than between the ITI and any other task phase, we expected that single-unit CP occurrence was comparably better aligned between the ITI and the delay-period windows than between the ITI and other task phases. To evaluate this hypothesis, we tested whether the lag between the occurrence of single-unit CPs during the ITI and single-unit CPs during other task windows was comparable for the three windows (see Materials and Methods). As expected from Figure 7*D*, single-unit CPs during the ITI were most coordinated with those during the delay-period window (repeated-measures one-way ANOVA with Greenhouse–Geisser correction: main effect, $F_{(1,10,3)} = 17.2$, $p = 0.001$, *post hoc* Bonferroni correction; ITI/light vs ITI/delay, $p = 0.0008$; ITI/lever vs ITI/delay, $p = 0.001$; ITI/light vs ITI/lever, $p = 0.07$; Fig. 7*E*). This agrees with previous observations that rate changes in accordance with reward expectations were particularly prominent within delay phases, during which the animal neither had to process specific sensory stimuli nor had to initiate specific responses (Watanabe, 1996; Leon and Shadlen, 1999). In our extinction paradigm, conditioned stimuli were identical for each trial, with only the active lever being cued throughout the session. Thus, it is possible that animals familiar with the task encoded reward expectation in PL neurons, not only during the delay period, but also during ITI.

In summary, PL encoded strategy changes by coordinating single-unit CPs in anticipation of behavioral changes. Single-unit CPs mark longer-term transitions in the firing rate of a unit, and hence the population-wide coordination of such events indicates a complete reorganization of the neural activity pattern. Such dynamics are compatible with a transition between two attractor states in the neural state space (Wang, 2002). Consistent with this idea, rate changes were not limited to any particular task

phase, or even to the proper trial periods, but were consistent across all phases including the ITI, and hence marked a global transition in the prefrontal state, possibly driven by updates in reward expectancy. In the new operational state, task events were still reflected in PL activity but did not trigger behavioral responses.

Discussion

The rodent mPFC is an anatomically and functionally heterogeneous structure consisting, among other regions, of the PL and IL. Seminal studies on fear extinction have suggested a dichotomous role of these two regions, where PL controls the expression of conditioned responses while IL mediates their extinction (for review, see Quirk and Mueller, 2008). In recent years, however, lesion (Fragale et al., 2016), pharmacological inactivation (Ramanathan et al., 2018; Caballero et al., 2019), and optogenetic stimulation (Sparta et al., 2014; Marek et al., 2018) studies challenged this view, providing increasing evidence that PL is a critical locus for extinction of reward-seeking behaviors and conditioned fear (for review, see Moorman et al., 2015). Using a within-session extinction paradigm, we studied PL dynamics leading to the extinction of conditioned reward-seeking behavior in rats. Critically, by using newly developed model-based statistical tools, we were able to identify and parametrize trial-to-trial changes in both behavior and mPFC dynamics, and to highlight fine temporal relations between them. This analysis revealed that a widespread single-unit coordination guides extinction learning by pushing PL toward a new operational state.

The methodology developed here to study learning has a few advantages to conventional analysis methods. On the behavioral side, several experimental paradigms have shown that behavior undergoes a sudden switch within few trials from low to high performance (Gallistel et al., 2004; Durstewitz et al., 2010). Such sudden transitions cannot be accounted for by standard reinforcement learning models (Sutton and Barto, 1998), which can only produce gradual changes in performance. Our approach, on the other hand, infers the learning curve from the observed behavior by considering the latter an instantiation of an underlying stochastic process. Unlike other similar model-based statistical methods (Smith et al., 2004; Deliano et al., 2016; Pelánek, 2017), our behavioral model allows us to explicitly infer the number, onset, duration, and magnitude of several learning episodes. On the neural side, using the PARCS method (Toutounji and Durstewitz, 2018) allows us to identify multiple neural change points both in single units and across the population while avoiding to average out change points by collapsing the data into a single peristimulus time histogram. Faced with a small number of trials per animal, a crucial practical advantage of using PARCS relative to other methods (Olshen et al., 2004; Cho and Fryzlewicz, 2015) is that it avoids segmenting the data when estimating multiple CPs. Segmentation, in fact, reduces the number of data points available for statistical testing, thus limiting the statistical power of the test (Toutounji and Durstewitz, 2018). Furthermore, change points are nonlinear phenomena that may indicate bifurcations between internal attractor states because of gradual changes in internal parameters (Wang, 2002), which cannot be accounted for by conventional analyses based on linear methods.

During across-session extinction, mPFC responsiveness to conditioned stimuli persists over subsequent unreinforced days (rat, PL/IL: Moorman and Aston-Jones, 2015). We showed here that during within-session extinction as well, the population

response to conditioned task stimuli remained equally strong, even when the animal stopped acting on them. This sustained responsiveness did not correspond, however, to stability in single-unit coding. While, on the one hand, some single units in mPFC can maintain their response pattern across days (rat, PL: Powell and Redish, 2014; mouse, ACC: Brebner et al., 2020), on the other hand, we found that the majority of units significantly changed their average firing rate and their stimulus responsiveness within tens of minutes. Changes in unit coding were widespread along the session and occurred under stable experimental conditions and stable behavioral responses as well as during within-session extinction. In fact, changes in single-unit coding per se do not necessarily imply changes in population coding, where coding properties might be preserved by redundancies in the ensemble representation (Narayanan et al., 2005; Puchalla et al., 2005; Hirokawa et al., 2019) or within neural trajectories (Mante et al., 2013; Enel et al., 2016). The comparable degree of rate changes during periods of distinct cognitive demands, such as maintenance and within-session extinction, may then suggest that transient representations are an intrinsic feature of prefrontal dynamics, rather than the result of specific computations.

Upon first inspection, we found no discernible difference between PL activity during maintenance and within-session extinction with respect to different single-unit rate change statistics. A more detailed analysis revealed, however, a strong temporal coordination of rate updates across the entire population during behavioral extinction, which was not present when no learning was required. These two effects, the reorganization of PL activity regardless of learning and the population-level temporal coordination specific to the learning phase, resonate with recent findings on PL plasticity during sleep in response to rule learning (Singh et al., 2019), suggesting that within PL similar neural mechanisms may underlie both the formation and fading of response–reward associations. In fact, ample evidence shows that in rat prefrontal cortex, particularly in its prelimbic subregion, neural population dynamics are reshaped before rule or reversal learning (Rich and Shapiro, 2009; Durstewitz et al., 2010; Karlsson et al., 2012; Powell and Redish, 2016; Malagon-Vina et al., 2018). Our results show that sudden population-wide reorganization is not exclusive to the acquisition of new behavioral strategies or to competition between conflicting strategies, but also occurs when reward seeking is suppressed per se. This finding offers circuit-level evidence in support of behavioral and cognitive theories of extinction learning, which view it as a form of new associative learning, rather than mere unlearning (Dunsmoor et al., 2015).

Our analysis further demonstrates that PL network reorganization during extinction learning is a global property, not anchored to a particular cognitive phase of the task. Indeed, we observed coordinated changes in neural activity during multiple trial stages, not only confined to specific windows within the trial, but also in resting periods between trials. Moreover, population firing rates predicted changes in animal behavior equally well regardless of the task window considered. These observations were therefore suggestive of the presence of two distinct states that globally characterized PL activity before and after the suppression of behavioral responses. In the presence of ambiguous sensory information, successful action selection is based on forming a reliable model of the environment as represented by the belief states of the animal (Babayan et al., 2018). The mPFC plays a fundamental role in the computation of belief states and

in cognitive control (Ridderinkhof et al., 2004; Gershman and Uchida, 2019; Sharpe et al., 2019). Our results suggest that updating belief regarding the availability of reward following extinction may correspond to a shift within the phase space of the resting state of the network.

Theoretical work (Katori et al., 2011) and experimental work (rat, PL: Durstewitz et al., 2010; primate, dorsolateral PFC: Wimmer et al., 2014) support the presence of attractor dynamics in PFC. Specifically, Redish et al. (2007) pushed forward the hypothesis that the prolonged absence of an expected reward would lead to the formation of a new attractor state within mPFC activity that represents the changed contingencies. Within this framework, a shift in phase space, as suggested by our data, may correspond either to a transition between two pre-existing attractor states led by external inputs [e.g., from the hippocampus (Euston et al., 2012; Sotres-Bayon et al., 2012) or from the amygdala (McGinty and Grace, 2008; Senn et al., 2014)] or to the formation of a new attractor state through plasticity (Toutounji and Pipa, 2014; Dunsmoor et al., 2015) or neuromodulatory processes (Harris and Thiele, 2011). Upon inspecting single-unit spike trains during extinction trials, we observed units within the same network with both slow and abrupt rate changes. This may suggest a third possible scenario in which learning slowly modulates the activity of a few neuronal ensembles, possibly on updates of reward expectancy, which internally lead network dynamics to undergo an abrupt transition between two pre-existing global attractor states.

The PL cortex in rodents has homologies with rostral ACC in humans (Brodmann's area 32) both at the anatomic and functional level (Laubach et al., 2018; van Heukelum et al., 2020). Like PL in rats, rostral ACC in humans is involved in cognitive control (di Pellegrino et al., 2007; Narayanan et al., 2013), behavioral flexibility (Kim et al., 2011), and drug seeking (Goldstein and Volkow, 2011). Beyond providing insights into motivational processes and learning, understanding reward extinction-learning mechanisms carries translational value for addiction research as well. Hence, while seeking reward is fundamental for survival, excessive drug seeking following cue exposure is a central component of addictive behavior. One behavioristic psychological approach that is used to treat individuals with alcoholism or drug addiction is cue exposure therapy (CET; i.e., extinction therapy). In CET, patients are exposed to relevant drug cues to extinguish conditioned responses. CET shows varying degrees of efficacy (Mellentin et al., 2017), and therefore it is of critical importance to understand its underlying neurobiological mechanisms. Our results indicate that extinction of alcohol-seeking behavior is not associated with a loss in mPFC responsiveness to conditioned stimuli. Instead, because of changes in the belief state of the subject, extinction manifests as a network-wide transition between two states corresponding to distinct behaviors: response (or consumption) and omission (or abstinence). This observation may thus suggest an alternative approach toward a pharmacologically driven CET that targets the relative strength and stability of the neural attractors representing the consumption and omission states. Such an intervention may go in two directions, either weakening the consumption state to facilitate extinction, thus avoiding maladaptive persistence of harmful behaviors, or strengthening the abstinence state to reduce the triggering effect of conditioned stimuli, thus reducing their valence and attenuating their ability to induce relapse.

References

- Babayan BM, Uchida N, Gershman SJ (2018) Belief state representation in the dopamine system. *Nat Commun* 9:1891.
- Bartolo R, Averbach BB (2020) Prefrontal cortex predicts state switches during reversal learning. *Neuron* 106:1044–1054.e4.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 57:289–300.
- Brebner LS, Ziminski JJ, Margetts-Smith G, Sieburg MC, Reeve HM, Nowotny T, Hirrlinger J, Heintz TG, Lagnado L, Kato S, Kobayashi K, Ramsey LA, Hall CN, Crombag HS, Koya E (2020) The emergence of a stable neuronal ensemble from a wider pool of activated neurons in the dorsal medial prefrontal cortex during appetitive learning in mice. *J Neurosci* 40:395–410.
- Caballero JP, Scarpa GB, Ramage-Healey L, Moorman DE (2019) Differential effects of dorsal and ventral medial prefrontal cortex inactivation during natural reward seeking, extinction, and cue-induced reinstatement. *eNeuro* 6:ENEURO.0296-19.2019.
- Chen BT, Yau HJ, Hatch C, Kusumoto-Yoshida I, Cho SL, Hopf FW, Bonci A (2013) Rescuing cocaine-induced prefrontal cortex hypoactivity prevents compulsive cocaine seeking. *Nature* 496:359–362.
- Cho H, Fryzlewicz P (2015) Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *J R Stat Soc B* 77:475–507.
- Deliano M, Tabelow K, König R, Polzehl J (2016) Improving accuracy and temporal resolution of learning curve estimation for within- and across-session analysis. *PLoS One* 11:e0157355.
- di Pellegrino G, Ciaramelli E, Ladavas E (2007) The regulation of cognitive control following rostral anterior cingulate cortex lesion in humans. *J Cogn Neurosci* 19:275–286.
- Dunsmoor JE, Niv Y, Daw N, Phelps EA (2015) Rethinking extinction. *Neuron* 88:47–63.
- Durstewitz D, Vitoz NM, Floresco SB, Seamans JK (2010) Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron* 66:438–448.
- Enel P, Procyk E, Quilodran R, Dominey PF (2016) Reservoir computing properties of neural dynamics in prefrontal cortex. *PLoS Comput Biol* 12:e1004967.
- Euston DR, Gruber AJ, McNaughton BL (2012) The role of medial prefrontal cortex in memory and decision making. *Neuron* 76:1057–1070.
- Fragale JE, Khariv V, Gregor DM, Smith IM, Jiao X, Elkabes S, Servatius RJ, Pang KC, Beck KD (2016) Dysfunction in amygdala-prefrontal plasticity and extinction-resistant avoidance: a model for anxiety disorder vulnerability. *Exp Neurol* 275:59–68.
- Friedman JH, Silverman BW (1989) Flexible parsimonious smoothing and additive modeling. *Technometrics* 31:3–21.
- Gallistel CR, Fairhurst S, Balsam P (2004) The learning curve: implications of a quantitative analysis. *Proc Natl Acad Sci U S A* 101:13124–13131.
- Gershman SJ, Uchida N (2019) Believing in dopamine. *Nat Rev Neurosci* 20:703–714.
- Goldstein RZ, Volkow ND (2011) Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nat Rev Neurosci* 12:652–669.
- Harris KD, Thiele A (2011) Cortical state and attention. *Nat Rev Neurosci* 12:509–523.
- Heinz A, Kiefer F, Smolka MN, Endrass T, Beste C, Beck A, Liu S, Genauk A, Romund L, Banaschewski T, Birmphohl F, Deserno L, Dolan RJ, Durstewitz D, Ebner-Priemer U, Flor H, Hansson AC, Heim C, Hermann D, Kiebel S, et al. (2020) Addiction research consortium: losing and regaining control over drug intake (ReCoDe)-from trajectories to mechanisms and interventions. *Addict Biol* 25:e12866.
- Hirokawa J, Vaughan A, Masset P, Ott T, Kepecs A (2019) Frontal cortex neuron types categorically encode single decision variables. *Nature* 576:446–451.
- Jonkman S, Mar AC, Dickinson A, Robbins TW, Everitt BJ (2009) The rat prelimbic cortex mediates inhibitory response control but not the consolidation of instrumental learning. *Behav Neurosci* 123:875–885.
- Karlsson MP, Tervo DG, Karpova AY (2012) Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. *Science* 338:135–139.
- Katori Y, Sakamoto K, Saito N, Tanji J, Mushiaki H, Aihara K (2011) Representational switching by dynamical reorganization of attractor structure in a network model of the prefrontal cortex. *PLoS Comput Biol* 7:e1002266.
- Kihlberg J, Herson J, Schotz W (1972) Square root transformation revisited. *Appl Statist* 21:76–81.
- Kim C, Johnson NF, Cilles SE, Gold BT (2011) Common and distinct mechanisms of cognitive flexibility in prefrontal cortex. *J Neurosci* 31:4771–4779.
- Kim CK, Ye L, Jennings JH, Pichamoorthy N, Tang DD, Yoo AW, Ramakrishnan C, Deisseroth K (2017) Molecular and circuit-dynamical identification of top-down neural mechanisms for restraint of reward seeking. *Cell* 170:1013–1027.e14.
- Laubach M, Amarante LM, Swanson K, White SR (2018) What, if anything, is rodent prefrontal cortex? *eNeuro* 5:ENEURO.0315-18.2018.
- Leon MI, Shadlen MN (1999) Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. *Neuron* 24:415–425.
- Ma L, Hyman JM, Durstewitz D, Phillips AG, Seamans JK (2016) A quantitative analysis of context-dependent remapping of medial frontal cortex neurons and ensembles. *J Neurosci* 36:8258–8272.
- Malagon-Vina H, Ciocchi S, Passecker J, Dorffner G, Klausberger T (2018) Fluid network dynamics in the prefrontal cortex during multiple strategy switching. *Nat Commun* 9:309.
- Mante V, Sussillo D, Shenoy KV, Newsome WT (2013) Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503:78–84.
- Marek R, Xu L, Sullivan RKP, Sah P (2018) Excitatory connections between the prelimbic and infralimbic medial prefrontal cortex show a role for the prelimbic cortex in fear extinction. *Nat Neurosci* 21:654–658.
- McGinty VB, Grace AA (2008) Selective activation of medial prefrontal-to-accumbens projection neurons by amygdala stimulation and Pavlovian conditioned stimuli. *Cereb Cortex* 18:1961–1972.
- Mellentin AI, Skøt L, Nielsen B, Schippers GM, Nielsen AS, Stenager E, Juhl C (2017) Cue exposure therapy for the treatment of alcohol use disorders: a meta-analytic review. *Clin Psychol Rev* 57:195–207.
- Moorman DE, Aston-Jones G (2015) Prefrontal neurons encode context-based response execution and inhibition in reward seeking and extinction. *Proc Natl Acad Sci U S A* 112:9472–9477.
- Moorman DE, James MH, McGlinchey EM, Aston-Jones G (2015) Differential roles of medial prefrontal subregions in the regulation of drug seeking. *Brain Res* 1628:130–146.
- Narayanan NS, Kimchi EY, Laubach M (2005) Redundancy and synergy of neuronal ensembles in motor cortex. *J Neurosci* 25:4207–4216.
- Narayanan NS, Cavanagh JF, Frank MJ, Laubach M (2013) Common medial frontal mechanisms of adaptive control in humans and rodents. *Nat Neurosci* 16:1888–1895.
- Olshen AB, Venkatraman ES, Lucito R, Wigler M (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5:557–572.
- Pelánek R (2017) Bayesian knowledge tracing, logistic models, and beyond: an overview of learner modeling techniques. *User Model User Adap Inter* 27:313–350.
- Pfarr S, Schaafl L, Reinert JK, Paul E, Herrmannsdörfer F, Roßmanith M, Kuner T, Hansson AC, Spanagel R, Körber C, Sommer WH (2018) Choice for drug or natural reward engages largely overlapping neuronal ensembles in the infralimbic prefrontal cortex. *J Neurosci* 38:3507–3519.
- Powell NJ, Redish AD (2014) Complex neural codes in rat prelimbic cortex are stable across days on a spatial decision task. *Front Behav Neurosci* 8:120.
- Powell NJ, Redish AD (2016) Representational changes of latent strategies in rat medial prefrontal cortex precede changes in behaviour. *Nat Commun* 7:12830.
- Puchalla JL, Schneidman E, Harris RA, Berry MJ (2005) Redundancy in the population code of the retina. *Neuron* 46:493–504.
- Quirk GJ, Mueller D (2008) Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology* 33:56–72.
- Ramanathan KR, Jin J, Giustino TF, Payne MR, Maren S (2018) Prefrontal projections to the thalamic nucleus reuniens mediate fear extinction. *Nat Commun* 9:4527.
- Redish AD, Jensen S, Johnson A, Kurth-Nelson Z (2007) Reconciling reinforcement learning models with behavioral extinction and renewal: implications for addiction, relapse, and problem gambling. *Psychol Rev* 114:784–805.

- Riaz S, Puveendrakumaran P, Khan D, Yoon S, Hamel L, Ito R (2019) Prelimbic and infralimbic cortical inactivations attenuate contextually driven discriminative responding for reward. *Sci Rep* 9:3982.
- Rich EL, Shapiro M (2009) Rat prefrontal cortical neurons selectively code strategy switches. *J Neurosci* 29:7208–7219.
- Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S (2004) The role of the medial frontal cortex in cognitive control. *Science* 306:443–447.
- Schuck NW, Gaschler R, Wenke D, Heinze J, Frensch PA, Haynes JD, Reverber C (2015) Medial prefrontal cortex predicts internally driven strategy shifts. *Neuron* 86:331–340.
- Senn V, Wolff SB, Herry C, Grenier F, Ehrlich I, Gründemann J, Fadok JP, Müller C, Letzkus JJ, Lüthi A (2014) Long-range connectivity defines behavioral specificity of amygdala neurons. *Neuron* 81:428–437.
- Sharpe MJ, Stalnaker T, Schuck NW, Killcross S, Schoenbaum G, Niv Y (2019) An integrated model of action selection: distinct modes of cortical control of striatal decision making. *Annu Rev Psychol* 70:53–76.
- Singh A, Peyrache A, Humphries MD (2019) Medial prefrontal cortex population activity is plastic irrespective of learning. *J Neurosci* 39:3470–3483.
- Smith AC, Frank LM, Wirth S, Yanike M, Hu D, Kubota Y, Graybiel AM, Suzuki WA, Brown EN (2004) Dynamic analysis of learning in behavioral experiments. *J Neurosci* 24:447–461.
- Sotres-Bayon F, Sierra-Mercado D, Pardiña-Delgado E, Quirk GJ (2012) Gating of fear in prefrontal cortex by hippocampal and amygdala inputs. *Neuron* 76:804–812.
- Sparta DR, Hovelsø N, Mason AO, Katak PA, Ung RL, Decot HK, Stuber GD (2014) Activation of prefrontal cortical parvalbumin interneurons facilitates extinction of reward-seeking behavior. *J Neurosci* 34:3699–3705.
- Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. Cambridge, MA: MIT.
- Toutounji H, Durstewitz D (2018) Detecting multiple change points using adaptive regression splines with application to neural recordings. *Front Neuroinform* 12:67.
- Toutounji H, Pipa G (2014) Spatiotemporal computations of an excitable and plastic brain: neuronal plasticity leads to noise-robust and noise-constructive computations. *PLoS Comput Biol* 10:e1003512.
- van Heukelum S, Mars RB, Guthrie M, Buitelaar JK, Beckmann CF, Tiesinga PHE, Vogt BA, Glennon JC, Havenith MN (2020) Where is cingulate cortex? A cross-species view. *Trends Neurosci* 43:285–299.
- Voigts J, Siegle JH, Pritchett DL, Moore CI (2013) The flexDrive: an ultra-light implant for optical control and highly parallel chronic recording of neuronal ensembles in freely moving mice. *Front Syst Neurosci* 7:8.
- Wang XJ (2002) Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* 36:955–968.
- Watanabe M (1996) Reward expectancy in primate prefrontal neurons. *Nature* 382:629–632.
- Wimmer K, Nykamp DQ, Constantinidis C, Compte A (2014) Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat Neurosci* 17:431–439.