Plasticity of a neural dictionary in prefrontal cortex

Abbreviated title: A neural dictionary in prefrontal cortex

Abhinav Singh[1], Adrien Peyrache[2], Mark D. Humphries[1*],

1 Faculty of Biology, Medicine and Health, University of Manchester, Manchester, M13 9PT, UK
2 Montreal Neurological Institute, McGill University, Montreal, QC H3A 1A1, Canada

***** Corresponding author: mark.humphries@manchester.ac.uk

Pages: 29
Figures: 9
Extended Data Items: 1
Abstract: 241
Introduction: 590
Discussion: 1440

Conflict of interest The authors declare no conflicts of interest.

Acknowledgments We thank the Humphries lab (Javier Caballero, Mat Evans, Silvia Maggi) for discussions; Rasmus Petersen for comments on the manuscript; and P. Berkes and M. Okun for respectively making their KL divergence and raster model code publicly available. A.S. and M.D.H were supported by a Medical Research Council Senior non-Clinical Fellowship award MR/J008648/1 to M.D.H. A.P. was supported by Human Frontier Science Program Fellowship LT000160/2011-l and National Institute of Health Award K99 NS086915-01.

Abstract

The dynamics of neural populations seem constrained to repeatedly visit a limited sub-set of all possible states. Within sensory populations, and especially in the retina, these repeated states take the form of millisecond-precise activity patterns. Enumerating a population's activity patterns thus defines a neural dictionary: the set of patterns that could potentially represent different things. Unknown is if such a dictionary is a general principle of cortical populations, and if learning changes the dictionary. To address these questions, we analysed population activity from the medial prefrontal cortex (mPfC) of rats learning new rules in a Y-maze. We found that patterns of co-active neurons on millisecond time-scales occurred far in excess of those predicted by firing rates alone. The set of activity patterns was strongly conserved between waking and sleep. Yet pattern frequencies detectably changed between the sleep epochs before and after a maze session. These changes were greatest for patterns that, during trials, were expressed at the maze's choice point and predicted the outcome of a trial. Successful learning of a rule systematically changed the dictionary of patterns, such that the probabilities of patterns after learning where maintained in post-learning sleep. By contrast, during stable behaviour there was no systematic change to the dictionary. Our data show that population activity in the mPfC contains a consistent yet plastic dictionary of task-encoding patterns at millisecond time-scales. We propose that these finding are a signature of the probabilistic representation of behavioural strategies in mPfC.

Significance statement

Cortex represents and computes information using the joint activity of many neurons. An open question is what experimentally observed features of this joint population activity are computationally relevant. We show here that the population activity from the prefrontal cortex of rats learning rules in a maze contains a specific dictionary of millisecond-precise activity patterns. This dictionary was altered during training, and encoded key parts of the task. But only during successful learning of a new rule was the dictionary seemingly permanently updated, because the changes could be detected during sleep after training. Our results thus further our understanding of the statistical structure of neural activity in cerebral cortex, and provides clues for the basis of cortical computation.

Introduction

Characterising the joint activity of cortical neurons is a step towards understanding how the cortex represents coding and computation (Cunningham and Yu, 2014; Yuste, 2015). A consistent observation is that the joint activity of cortical populations seems constrained to visit only a sub-set of all the possible states they could reach (Tsodyks et al., 1999; Luczak et al., 2009; Sadtler et al., 2014; Jazayeri and Afraz, 2017). Theoretical accounts propose that this manifold of possible states is specified by the connections into and within the network of cortical neurons (Galan, 2008; Marre et al., 2009; Buesing et al., 2011; Habenschuss et al., 2013; Kappel et al., 2015). An implicit prediction of these theories is that changing the network connections through learning would change the set of states visited by the joint activity of the population. If so, then these changes to the connections should constrain cortical activity across both evoked and spontaneous activity (Luczak et al., 2009; Ringach, 2009; Miller et al., 2014). Understanding how learning changes the set of states visited by a cortical population would provide strong constraints on theories for how cortex processes information.

It is convenient to characterise joint activity states as a set of binary activity patterns, or "words", each word identifying whether or not each neuron has fired in some small slice of time, typically on the order of a few milliseconds (Schneidman et al., 2006; Shlens et

al., 2006; Ohiorhenuan et al., 2010; Berkes et al., 2011). The set of unique words visited by a neural population defines a dictionary, and their rates of occurrence define the use of that dictionary to represent information (Schneidman et al., 2006; Tkacik et al., 2014; Marre et al., 2015; Ganmor et al., 2015; O'Donnell et al., 2017). Such approaches have been successfully applied in retina and V1, but not in higher cortices, and nor to learning or behaviour. So we ask here if such a dictionary of millisecond-precise activity patterns exist in higher cortices and, if so, what that dictionary encodes about behaviour and how it changes with learning.

The medial prefrontal cortex (mPFC) is a natural candidate for addressing these questions. It is necessary for learning new rules or strategies (Ragozzino et al., 1999; Rich and Shapiro, 2007). Changes in mPFC neuron firing times correlate with successful rule learning (Benchenane et al., 2010), suggesting that mPFC coding of task-related variables by the timing of spikes changes over learning. Further, mPFC population recording data from the outset of learning on a Y-maze task are available (Peyrache et al., 2009). We thus use that data here to test the hypothesis that mPFC population activity contains a dictionary of millisecond-precise activity patterns related to learning rules about the world.

We found that millisecond-precise patterns of joint activity occur far above chance levels defined by firing rates in the mPfC, across waking and sleeping, and irrespective of whether behaviour on the task was changing or stable. A set of these patterns changed their rate of occurrence between the sleep epochs before and after training sessions. During training, these changed patterns occurred around the choice point of the maze and predicted trial outcome. But only in learning sessions was the direction of change systematic, such that it brought the distribution of patterns in post-training sleep closer to the distribution sampled in training. We show how these findings are consistent with the mPfC representing and updating a sample-based internal model of the maze rules. Collectively, our results constrain models for mPfC dynamics, and suggest the existence of fine time-scale population codes.

Materials and Methods

Task and electrophysiological recordings

Four Long-Evans male rats with implanted tetrodes in prelimbic cortex were trained on a Y-maze task (Figure 1*A*). Each recording session consisted of a 20-30 minute sleep or rest epoch (pre-training epoch), in which the rat remained undisturbed in a padded flowerpot placed on the central platform of the maze, followed by a training epoch, in which the rat performed for 20-40 minutes, and then by a second 20-30 minute sleep or rest epoch (post-training epoch); see Figure 1*B*. Every trial started when the rat left the beginning of the departure arm and finished when the rat reached the end of one of the choice arms. Correct choice was rewarded with drops of flavoured milk. Each rat had to learn the current rule by trial-and-error, either: go to the right arm; go to the cued arm; go to the left arm; go to the uncued arm. To maintain consistent context across all sessions, the extra-maze light cues were lit in a pseudo-random sequence across trials, whether they were relevant to the rule or not.

The data analysed here were from a total set of 50 experimental sessions taken from the study of (Peyrache et al., 2009), representing a set of training sessions from naive until either the final training session, or until choice became habitual across multiple consecutive sessions (consistent selection of one arm that was not the correct arm). The four rats respectively had 13, 13, 10, and 14 sessions. From these we have used here ten learning sessions and up to 17 "stable" sessions (see below).

Tetrode recordings were spike-sorted only within each recording session for conservative identification of stable single units. In the sessions we analyse here, the populations ranged

in size from 15-55 units. Spikes were recorded with a resolution of 0.1 ms. For full details on training, spike-sorting, and histology see (Peyrache et al., 2009).

Session selection and strategy analysis

We primarily analysed here data from the ten learning sessions in which the previously-defined learning criteria (Peyrache et al., 2009) were met: the first trial of a block of at least three consecutive rewarded trials after which the performance until the end of the session was above 80%. In later sessions the rats reached the criterion for changing the rule: ten consecutive correct trials or one error out of 12 trials. Thus each rat learnt at least two rules, but none learnt the uncued-arm rule.

We also sought sessions in which the rats made stable choices of strategy. For each session, we computed the probability $P(rule)$ that the rat chose each of the three rules (left, right, cued arm) per trial. Whereas $P(left)$ and $P(right)$ are mutually exclusive, $P(cued-arm)$ is not, and has an expected value of 0.5 when it is not being explicitly chosen because of the random switching of the light cue. A session was deemed to be "stable" if $P(rule) > \theta$ for one of the rules, and the session contained at least 10 trials (this removed only two sessions from consideration). Here we tested both $\theta = 0.9$ and $\theta = 0.85$, giving $N = 13$ and $N = 17$ sessions respectively. These also respectively included 2 and 4 of the rule-change sessions. For the time-series in Figure 1C,E,F we estimated $P(rule)$ in windows of 7 trials, starting from the first trial, and sliding by one trial.

Activity pattern distributions

For a population of size N, we characterised population activity from time t to $t + \delta$ as an N-length binary vector with each element being 1 if at least one spike was fired by that neuron in that time-bin, and 0 otherwise. In the Results we predominantly report analyses using a bin size of $\delta = 2$ ms; key results were checked for bin sizes ranging over two orders of magnitude (Figure 2; Figure 7). We built patterns using the number of recorded neurons N, up to a maximum of 35 for computational tractability. The number of neurons used in each analysis is listed in Figure 2-1; where we needed to use less than the total number of recorded neurons, we ranked them according to their coefficient of variation of their firing rate between the three epochs, and choose the N least variable; in practice this sampled neurons from across the full range of firing rates.

To test the predicted proportion of co-activation patterns by independently firing neurons, we shuffled inter-spike intervals for each neuron independently, then reconstructed the activity patterns at the chosen bin size. This procedure kept the same inter-spike interval distribution for each neuron, but disrupted any correlation between neurons. As both the training and sleep epochs were broken up into chunks (of trials and slow-wave sleep bouts, respectively), we only shuffled inter-spike intervals within each chunk. We repeated the shuffling 20 times, and in Figure 2*B-E* we plot for the shuffled data the means and error bars of 99% confidence intervals (too small to see on the scales of the axes).

Comparing distributions

The probability distribution for the activity patterns in a given epoch of the task was compiled by counting the frequency of each pattern's occurrence and normalising by the total number of pattern occurrences. We quantified the distance $D(P|Q)$ between probability distributions P and Q using both the Kullback-Liebler divergence (KLD) and the Hellinger distance.

The KLD is an information theoretic measure to compare the similarity between two probability distributions. Let $P = (p_1, p_2, \ldots, p_n)$ and $Q = (q_1, q_2, \ldots, q_n)$ be two discrete probability distributions, for n distinct possibilities – for us, these are all possible individual activity patterns. The KLD is then defined as $D_{\mathrm{KLD}}(P|Q) = \sum_{i=1}^{n} p_i \log_2(\frac{p_i}{q_i})$. We

normalised this by unit time (2 ms bins except where noted) to obtain the information rate in bits/s.

There are 2^N distinct possible activity patterns in a recording with N neurons. The empirical frequency of these activity patterns is biased due to the limited length of the recordings (Panzeri et al., 2007). To counteract this bias, we used the Bayesian estimator and quadratic bias correction exactly as described in (Berkes et al., 2011). The Berkes estimator assumes a Dirichlet prior and multinomial likelihood to calculate the posterior estimate of the KLD; we used their code (github.com/pberkes/neuro-kl) to compute the estimator. We then computed a KLD estimate using all S activity patterns, and using $S/2$ and $S/4$ patterns randomly sampled without replacement. By fitting a quadratic polynomial to these three KLD estimates, we could then use the intercept term of the quadratic fit as an estimate of the KLD if we had access to recordings of infinite length (Strong et al., 1998; Panzeri et al., 2007). This final estimate varies according to the patterns sub-sampled in order to fit the quadratic; however, in our data the variation introduced by the sub-sampling was negligible on the scale of the distances measured (Figure 7).

We attempted here to characterise the population's joint activity as fully as possible, by making use of as many simultaneously recorded individual neurons as possible. We capped our activity patterns to a maximum of $N = 35$ neurons; but this still means that, for some populations, a full estimation of KLD using the above Bayesian estimator would mean enumerating all 2^{35} patterns every time we computed a KLD estimate. This is computationally intractable; moreover, in extensively checking the results and the raster model (see below) we produced thousands of KLD calculations for each recorded population. So we sought a practical solution, and set $P = 0$ for any activity pattern that was not in either distribution being compared (this was the vast majority of all potential patterns). Our data shows only a tiny fraction of activity patterns that appear in one distribution and do not appear in the other (Figure 2F), so we expected the disagreement between the KLD computed using the full enumeration of all 2^N patterns and using $P = 0$ to be small, and not to qualitatively affect results. We tested this explicitly for a full enumeration using patterns of $N = 15$ for all learning-session populations, and found that setting $P = 0$ did not qualitatively affect the results, nor showed a systematic bias in the distances measured by either approach (Figure 7). We note that this is not, in general, a safe assumption: we can only do this here because of the very low proportion of unique patterns in each compared distribution. Moreover, we checked the main results throughout with a different measure of inter-distribution distance - the Hellinger distance - that did not rely on any bias-correcting estimators or priors.

The Hellinger distance for two discrete distributions P and Q is $D_{\mathrm{H}}(P|Q) = \frac{1}{2}\sum_{i=1}^{n}(\sqrt{p_i} - \sqrt{q_i})^2$. To a first approximation, this measures for each pair of probabilities (p_i, q_i) the distance between their square-roots. In this form, $D_{\mathrm{H}}(P|Q) = 0$ means the distributions are identical, and $D_{\mathrm{H}}(P|Q) = 1$ means the distributions are mutually singular: all positive probabilities in P are zero in Q, and vice-versa. The Hellinger distance is a lower bound for the KLD: $2D_{\mathrm{H}}(P|Q) \leq D_{\mathrm{KLD}}$.

To compare distances we computed a normalised measure of the relative "convergence" between the distributions. We computed the "convergence" score by computing the difference between a pair of distances between training and sleep epochs, and normalising by the the maximum distance between training and sleep epochs: $[D(Pre|X) - D(Post|X)]/\max\{D(Post|X), D(Pre|X)\}$. We express this here as a percentage, giving a range of $[-100, 100]\%$. Convergence greater than 0% indicates that the distance between the training epoch $P(X)$ and post-training sleep ($P(Post)$) distributions was smaller than that between the training and pre-training sleep (P(Pre)) distributions.

Estimating equivalence between distributions with finite samples

Even if two underlying probability distributions are exactly the same, empirical measurements of samples taken from them will not show exact equivalence $[D(P|Q) = 0]$ due to finite sampling effects. We estimated a baseline measure of equivalence for the activity distributions in the sleep epochs by bootstrapping the activity patterns within each epoch. To do this, we drew two sets of patterns with replacement from the set of empirically recorded patterns, and computed the distance between the two bootstrapped sets. This emulates the finite-sampling problem within the empirical data. We also tested a more severe version where the set of recorded activity patterns was split randomly in half and the distance computed between each half. However, as this procedure is itself halving the number of patterns, it induces more variation by further finite sampling; nonetheless, identical results to those in Figure 3 were obtained.

Statistics

Quoted measurement values are mean $\bar{x}$ and 95% confidence intervals for the mean $[\bar{x} - t_{\alpha/2,n}\mathrm{SE}, \bar{x} + t_{\alpha/2,n}\mathrm{SE}]$, where $t_{\alpha/2,n}$ is the value from the t-distribution at $\alpha = 0.05$ and given the number n of data-points used to obtain $\bar{x}$. All hypothesis tests used the non-parametric Wilcoxon signtest for a paired-sample test that the number of changes in sign (-,+) between each pair of samples exceeds that expected from the binomial distribution with $P = 0.5$. For testing the changes in convergence, we used the Wilcoxon signrank test. Throughout, we have $n = 10$ learning sessions and $n = 17$ stable sessions (using $\theta = 0.85$). All results were checked with the $\theta = 0.9$ criterion for identifying stable sessions, giving $n = 13$.

Relationship of location and change in pattern probability

We examined the spatial correlates of activity pattern occurrence for the learning and stable sessions. To rule out pure firing rate effects, we excluded all patterns with $K = 0$ and $K = 1$ spikes, considering only co-activation patterns $K \geq 2$; that is, those with two or more active neurons. The location of every occurrence of a co-activation pattern was expressed as a normalized position on the linearised maze (0: start of departure arm; 1: end of the chosen goal arm).

Within each session, we computed the absolute change $\delta_i = |p_i(pre) - p_i(post)|$ in each pattern's probability of occurrence between pre- and post-training slow-wave sleep. To combine data across sessions, for each session we normalised all changes by the maximum change in that session: $\delta_i^* = \delta_i / \max_i\{\delta\}$. Normalised change scores were pooled over all learning sessions.

Our main claim for this analysis was that activity patterns which changed probability between sleep epochs occur predominantly around the choice point of the maze, and so change and overlap (of the choice area) are dependent variables (Figure 4*A,C*). To test this claim, we compared this relationship against the null model of independent variables, by permuting the assignment of location centre-of-mass (median and interquartile range) to the activity patterns. For each permutation, we compute the proportion of patterns whose interquartile range overlaps the choice area, and bin as per the data. We permuted 5000 times to get the sampling distribution of the proportions predicted by the null model of independent variables: we plot the mean and 95% range of this sampling distribution as the grey region in Figure 4*B,D*.

Outcome prediction

We examined the correlates of co-activation pattern occurrence with behaviour for the learning sessions. To check whether individual activity patterns coded for the outcome on

each trial, we used standard receiver-operating characteristic (ROC) analysis. For each co-activation pattern, we computed the distribution of its occurrence frequencies separately for correct and error trials (as in the example of Figure 5A). We then used a threshold T to classify trials as "error" or "correct" based on whether the frequency on that trial exceeded the threshold or not. We found the fraction of correctly classified correct trials (true positive rate) and the fraction of error trials incorrectly classified as correct trials (false positive rate). Plotting the false positive rates against the true positive rates for all values of T gives the ROC curve. The area under the ROC curve gives the probability that a randomly chosen pattern frequency will be correctly classified as from a correct trial; we report this as P(predict outcome).

Relationship of sampling change and outcome prediction

Correlating P(predict outcome) against the change in pattern probability between sleep epochs δ_i^* showed that the better a pattern predicted trial outcome, the more it tended to change probability between pre- and post-training slow-wave sleep. But as most patterns had little change and little prediction of outcome, this correlation was skewed (Figure 5C).

Consequently, to better characterise the distributions of change between pre- and post-session sleep, we binned δ_i^* using variable-width bins of P(predict outcome): each consecutive bin-width was chosen in order to place the same number of data-points in every bin. We computed the empirical cumulative distribution in each bin, to visualise the distribution of changes in pattern probability between sleep epochs, and the change in that distribution with P(predict outcome). To quantify this change, we regressed P(predict outcome) against the median change in each bin; we used the mid-point of each variable-width bin as the value for P(predict outcome). Our main claim is that prediction and change are dependent variables (Figure 5C-G). To test this claim, we compared the data correlation against the null model of independent variables, by permuting the assignment of change scores to the activity patterns. For each permutation, we repeat the binning and regression. We permuted 5000 times to get the sampling distribution of the correlation coefficient R^* predicted by the null model of independent variables. To check robustness, all analyses were repeated for a range of fixed number of data-points per bin between 20 and 100.

Raster model

To control for the possibility that the systematic changes in activity pattern occurrence during learning were due solely to changes in the firing rates of individual neurons and of the total population between vigilance states, we used the raster model exactly as described in (Okun et al., 2012). For a given data-set of spike-trains N and bin size δ, the raster model constructs a synthetic set of spikes such that each synthetic spike-train has the same mean rate as its counterpart in the data, and the distribution of the total number of spikes per time-bin matches the data. In this way, it predicts the frequency of activity patterns that should occur given solely changes in individual and population rates.

For Figure 8 we generated 1000 raster models per session using the spike-trains from the post-training slow-wave sleep in that session. For each generated raster model, we computed the distance $D(Model|Data)$ between the distribution of patterns for that model $P(Model)$ and the corresponding data distribution $P(Data)$ of post-training slow-wave sleep patterns. For each generated raster model, we then computed the distance between its distribution of activity patterns and the data distribution for post-learning trials $D(Post-model|Learn)$. This comparison gives the expected distance between the training and post-training slow-wave sleep distributions due to firing rate changes alone. We plot the difference between the mean of $D(Post-model|Learn)$ over the 1000 raster models and the data $D(Post|Learn)$ in Figure 8.

Probabilistic reinforcement learning model

In the Results, we propose that our observed changes in pattern probability are consistent with the encoding and learning of a probabilistic internal model. To illustrate the expected behaviour of a probabilistic internal model during learning, we constructed a Bayesian reinforcement learning model of the Y-maze task. We modelled the trial-by-trial behaviour as a Bayesian multi-arm bandit problem (Ghavamzadeh et al., 2015), where the agent's task on each trial was to chose which strategy to adopt, based on a probabilistic estimate of the value of each strategy. We use this simplified representation as a proxy for more complex models with probability distributions over the uncertain values of individual actions and the transitions they cause between states in the maze, which collectively make a strategy.

Here we report results from modelling three strategies: go to the left arm; go to the right arm; and go to the cued arm. For each strategy x, the agent maintained a posterior probability distribution over the value of choosing that strategy $V_x \in [0, 1]$, given by a Beta distribution $P(V_x)$ with parameters (α_x, β_x). On each trial t, the winning strategy was chosen using Thompson sampling: a random value ζ_x was sampled from the probability distribution $P(V_x)$ for each strategy, and the strategy s with the highest sampled value was chosen. The corresponding action was then chosen: left, right, or cued arm (where, as per the experiment, the cued arm was randomly chosen on each trial). There was a small probability η of a mistake in choosing the corresponding action: if a mistake was made, then the opposite action was chosen (being the uncued arm for the cued-arm strategy). We used $\eta = 0.2$ for the simulations reported here. This was implemented to include noise into the decision process, providing a better replication of the rats' behaviour. Having taken the action, the agent received reward according to the current rule (left, right, or cued arm), with $R = 1$ if the action corresponded to the rule, and $R = 0$ otherwise. The reward was then used to update the probability distribution $P(V_s)$ of the chosen strategy s.

The full Bayesian update of the posterior should be proportional to $P(V_s|R = r) \propto P(R = r|V_s)P(V_s)$, where $P(R = r|V_s)$ is the likelihood function for the outcome r given the probability distribution over the strategy's value, and $P(V_s)$ is the prior distribution over that value. In simulation, we make use of the standard result that, assuming a binomial likelihood function $P(R = r|V_s)$ because each trial is a Bernoulli trial, then the Beta distribution $P(V_s)$ is the conjugate prior (Daw et al., 2005; Ghavamzadeh et al., 2015). Consequently, Bayesian updating is obtained by just updating the parameters of $P(R = r|V_s)$ by $(\alpha + r, \beta + (1 - r))$. Distributions $P(V_x)$ for trial 1 was set to the uniform distribution $(\alpha = 1, \beta = 1)$.

To make comparisons with the behavioural data, we made proxy estimates of learning trials, and then virtual "sessions" around those trials. For each simulation, the nominal "learning trial" was identified as the trial in the cumulative reward curve corresponding to the greatest inflection in reward rate. To do this, we fitted a piecewise linear slope around each trial t, with one line fitted to eleven trials before and including t, and one line fitted to eleven trials after and including t. The trial t_l with the greatest increase in slope of the lines before and after it was selected as the "learning" trial. A virtual session was given by the 14 trials before and after the chosen learning trial, giving a session length of 29 trials. The trials corresponding to the beginning (t_{pre}) and end (t_{post}) of this virtual session were deemed the pre- and post-training "sleep" epochs for the model.

In the Results, we claim that any such reinforcement-driven recursive updating of a set of probability distributions will stabilise those distributions over time. Estimating the probability distribution of some unknown value v_t (of, for example, a state or action) at time t, given all the rewards $(r_1, r_2, \ldots, r_t)$ up to time t, can be computed recursively using Bayes' theorem:

$$P(v_t|r_1, r_2, \ldots, r_t) \propto P(r_t|v_t)P(v_t|r_1, r_2, \ldots, r_{t-1}), \tag{1}$$

where the posterior distribution $P(v_t|r_1, r_2, \ldots, r_{t-1})$ for step $t-1$ becomes the prior distribution for step t. In general, given that r is stationary and given sufficient t, then the difference between the posterior and the prior $\delta = P(v_t|r_1, r_2, \ldots, r_t) - P(v_t|r_1, r_2, \ldots, r_{t-1})$ will become arbitrarily small. Thus, the posterior distribution will stabilise in any recursive Bayesian estimation.

This stabilisation of distributions is predicted to happen once our Bayesian reinforcement learning model has learnt the current rule. Once learnt, the agent will experience a long run of sustained rewards, with two consequences. First, for the Beta distribution $P_x(v)$ modelling the correct strategy x this will mean a continuously increasing α_x, with β_x approximately fixed. As a result, we expect $\alpha_x \gg \beta_x$. Second, the other Beta distributions, modelling the incorrect strategies, will be rarely updated (as they are only updated when selected). These distributions will thus be approximately stable.

For our specific model using Beta distributions, we show here the explicit stabilisation of $P_x(v)$ by calculating the change in the distribution's mean and variance as a function of the number of rewards α. The mean of $P_x(v)$ is:

$$E(v) = \frac{\alpha}{\alpha + \beta}, \tag{2}$$

so the change in mean with increasing accumulated rewards is:

$$\frac{dE(v)}{d\alpha} = \frac{\beta}{(\alpha + \beta)^2}. \tag{3}$$

It is easy to see that as $\alpha \gg \beta$, so $dE(v) \to 0$: the mean stops changing over time with increasingly obtained reward.

The variance of $P_x(v)$ is

$$Var(v) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}, \tag{4}$$

so the change in variance with increasing accumulated rewards is:

$$\frac{dVar(v)}{d\alpha} = \frac{\beta(-2\alpha^2 - \alpha(\beta + 1) + \beta(\beta + 1))}{(\alpha + \beta)^3(\alpha + \beta + 1)^2} \tag{5}$$

Thus as $\alpha \gg \beta$, so $dVar(v) \approx (-2\alpha^2 - \alpha)/\alpha^5$; given the dominance of raising to the fifth power in the denominator, this also ensures $dVar(v) \to 0$: the variance stops changing over time with increasingly obtained reward. Thus, long runs of reward are expected to stabilise $P_x(v)$.

Results

Rats with implanted tetrodes in the mPfC learnt one of four rules on a Y-maze: go right, go to the randomly-cued arm, go left, or go to the uncued arm (Figure 1*A*). Rules were changed in this sequence, unsignalled, once the rat did 10 correct trials in a row, or 11 correct trials out of 12. Each rat experienced at least two of the rules, starting from a naive state. Each training session was a single day containing 3 epochs totalling typically 1.5 hours: pre-training sleep/rest, behavioural training on the task, and post-training sleep/rest (Figure 1*B*). Here we consider bouts of slow-wave sleep throughout, to unambiguously identify periods of sleep. Tetrode recordings were spike-sorted within each session, giving populations of single neuron recordings ranging between 12 and 55 per session (see Figure 2-1 for details of each session and each epoch within a session).

In order to test for the effects of learning on the states of joint population activity, we needed to compare sessions of learning with those containing no apparent learning as defined by the rats' behaviour. In the original study containing this data-set, Peyrache et

al. (2009) identified 10 learning sessions as those in which three consecutive correct trials were followed by at least 80% correct performance to the end of the session: the first of the initial three was considered the learning trial. By this criterion, the learning trial occurred before the mid-point of the session (mean 45%; range 28-55%). We first confirmed that this criterion indeed corresponded to clear learning: in each of the ten sessions there was an abrupt step change in reward accumulation around the identified learning trial (Figure 1*C,D*), corresponding with the switch to a consistent, correct strategy within that session (Figure 1*E*). We further identified a set of 17 sessions with a stable behavioural strategy throughout, defined as a session with the same strategy choice (left, right, cue) on more than 85% of trials (Figure 1*F*). This set included 4 sessions in which the rule changed. Setting this criterion to a more conservative 90% reduced the number of sessions to 13 (including two rule change sessions), but did not alter the results of any analysis; we thus show the 85% criterion results throughout.

mPfC population activity contains a dictionary of millisecond precise co-activation patterns

We first tested whether the joint activity of mPfC populations contains above-chance statistical structure on millisecond time-scales. Population-wide activity patterns were characterised as a binary vector (or "word") of active and inactive neurons within some small time window (Figure 2*A*). We used up to 35 neurons per session to construct the patterns for computational tractability; this meant using every neuron in all but 2 learning and 6 stable sessions (Figure 2-1). Our primary interest was in co-activation patterns of more than one neuron firing together, as the occurrences of each pattern with a single active neuron (a single "1") can correlate strongly with that neuron's firing rate. We thus first determined the time-scales at which co-activation patterns appear.

Figure 2*B,D* show that at low millisecond time-scales the proportion of activity patterns containing co-active neurons increases by an order of magnitude when doubling the bin size, for both learning and stable sessions. The smallest bin size with a non-negligible proportion of co-activation patterns was 2 ms, with $\sim$ 1% (89731/7452300) of all patterns in learning sessions. This was also true for each epoch considered separately, for both learning (Figure 2*C*) and stable (Figure 2*E*) sessions. We thus used a 2 ms bin size throughout, as this was the smallest time-scale with consistent co-activation patterns.

Such co-activation patterns could be due to persistent, precise correlations between spike-times in different neurons, or just due to coincident firing of otherwise independent neurons. We found that the proportion of co-activation patterns in the data exceeded those predicted for independent neurons by a factor of 3 (Figure 2*B*) at low millisecond time-scales. This was also true for each separate epoch (Figure 2*C*), extending up to a factor of at least 6 for the task trials. We found similar results for the stable sessions (Figure 2*E*); though we noted that the difference between the data and the predictions for independent neurons was not as consistent as it was for the learning sessions, with the greatest departure being at a bin size of around 20 ms. Nonetheless, these analyses rule out the possibility that the excess of precise correlations was due to differences in vigilance state.

While these results show there exist non-trivial fine time-scale patterns in mPfC population activity, they do not yet show they are the same structure in different states. We found that each recorded population of N neurons had the same sub-set of all 2^N possible activity patterns in all epochs (Figure 2*F*). This was true in both learning and stable sessions, with no apparent difference between them. Consequently, the overwhelming majority of words appeared in both the task-evoked activity of waking and the spontaneous activity of slow-wave sleep. This is consistent with the theoretical idea that a cortical population is constrained to a manifold of specific states - its dictionary - irrespective of the vigilance state of the animal.

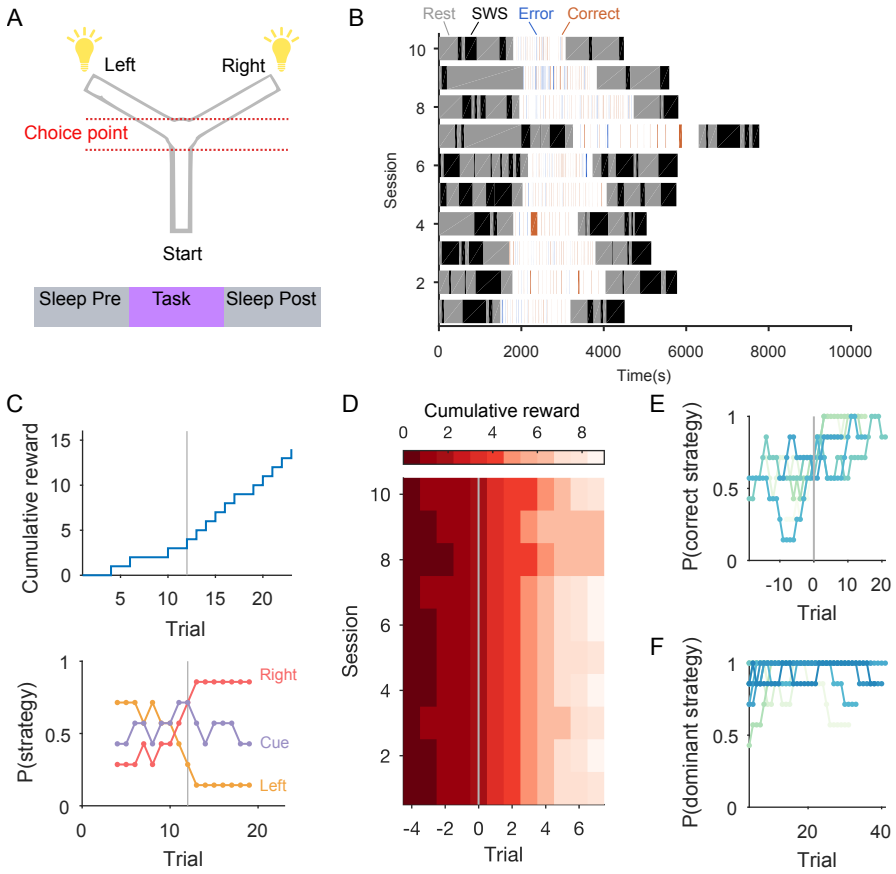

Figure 1. Task and behaviour. (A) Y-maze task set-up (top); each session included the epochs of pre-training sleep/rest, training trials, and post-training sleep/rest (bottom). One of four target rules for obtaining reward was enforced throughout a session: go right; go to the cued arm; go left; go to the uncued arm. No rat successfully learnt the uncued-arm rule. (B) Breakdown of each learning session into the duration of its state components. The training epoch is divided into correct (red) and error (blue) trials, and inter-trial intervals (white spaces). Trial durations were typically 2-4 seconds, so are thin lines on this scale. The pre- and post-training epochs contained quiet waking and light sleep states ("Rest" period) and identified bouts of slow-wave sleep ("SWS"). (C) Internally-driven behavioural changes in an example learning session: the identified learning trial (grey line) corresponds to a step increase in accumulated reward and a corresponding shift in the dominant behavioural strategy (bottom). The target rule was 'go right'. Strategy probability was computed in a 7-trial sliding window; we plot the mid-points of the windows. (D) Peri-learning cumulative reward for all ten identified learning sessions: in each session, the learning trial (grey line) corresponded to a step increase in accumulated reward. (E) Peri-learning strategy selection for the correct behavioural strategy. Each line plots the probability of selecting the correct strategy for a learning session, computed in a 7-trial sliding window. The learning trial (grey line) corresponds to the onset of the dominance of the correct behavioural strategy. (F) Strategy selection during stable behaviour. Each line plots the probability of selecting the overall dominant strategy ($P > 0.85$ for the session) computed in a 7-trial sliding window.

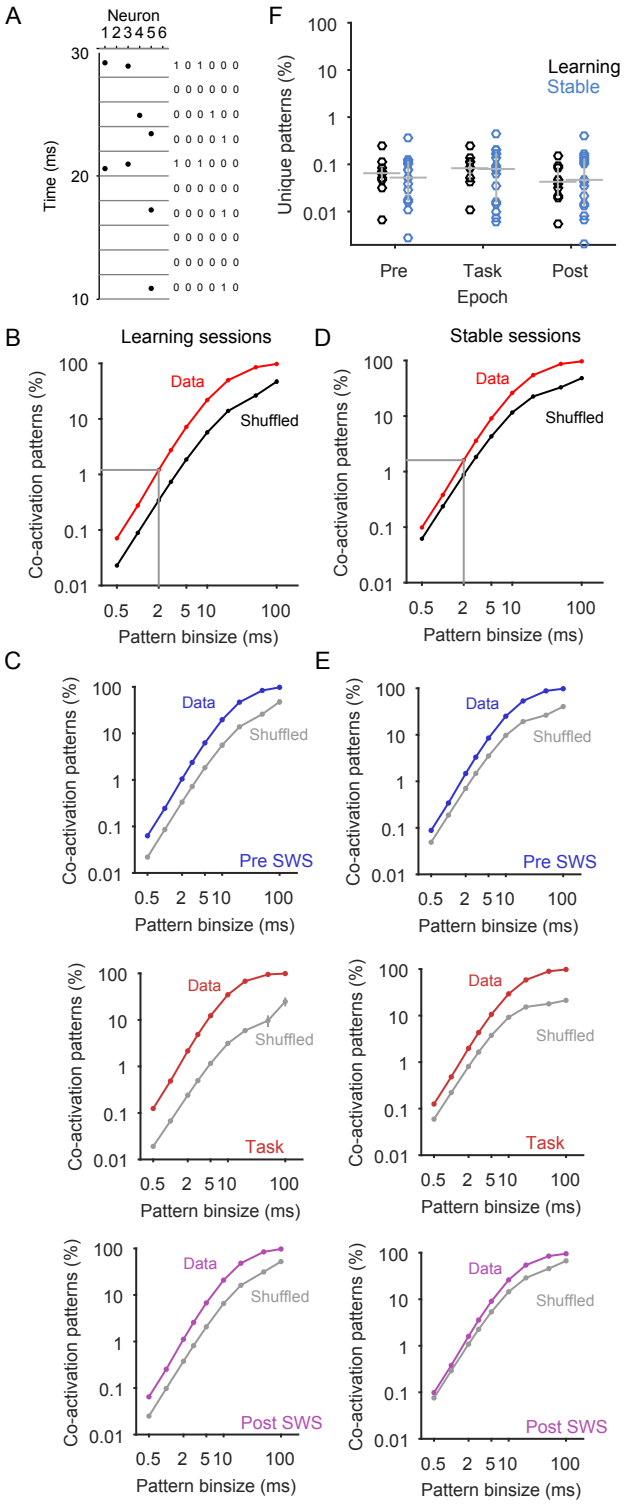

Figure 2. Activity patterns occur more than predicted by firing rates and are conserved between epochs. (*A*) The population activity of simultaneously recorded spike trains was represented as a binary activity pattern in some small time-bin (here 2 ms). (*B*) The proportion of co-activation patterns in the learning sessions, per bin size (red line). Here we count every occurrence of every pattern. The grey line indicates the proportion of $\sim 1\%$ at a bin size of 2 ms. In black we plot the corresponding proportion of co-activation patterns predicted if all neurons were firing independently; these are obtained by shuffling the inter-spike intervals of each neuron and recomputing the activity patterns. Error bars of 99% CI are too small to see on this scale. (*C*) Proportion of co-activation patterns per epoch of the learning sessions. Predicted proportions by independently-firing neurons are in grey. Error bars of 99% CI are too small to see on this scale. (*D-E*) As *B-C*, for stable sessions. (*F*) Consistent sampling of activity patterns across session epochs. Each circle is the proportion of unique activity patterns (2 ms time-bin) from the entire session that appeared only in that epoch. Unique activity patterns are defined as those that occurred at least once in the entire recording. Grey bar and line give the median and interquartile range across the 10 (learning) or 17 (stable) sessions. Note the log-scale, showing that the median proportion of patterns was less then 0.1% in all three epochs. Figure 2-1 gives the numbers of patterns in each epoch.

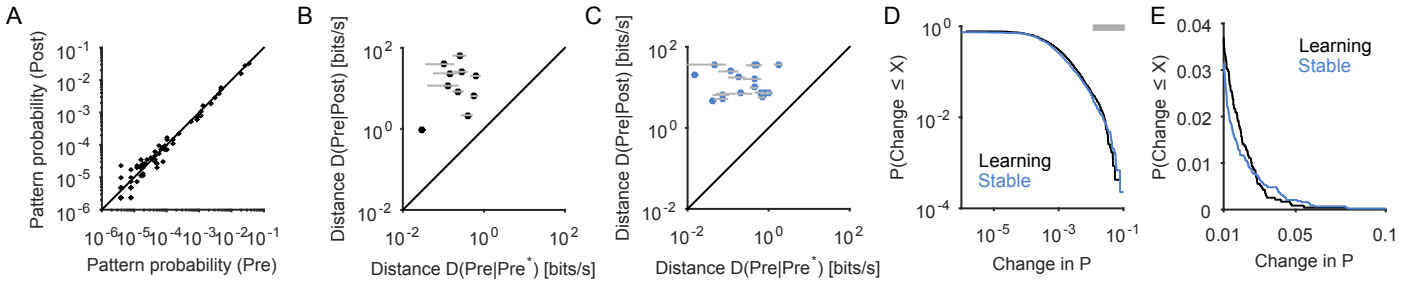

Figure 3. Distributions of joint activity patterns change between pre- and post-learning sleep. (A) The joint frequency of every occurring pattern (dots) in pre-training sleep (distribution $P(Pre)$) and post-training sleep (distribution $P(Post)$) for one learning session. (B) Distances between pre- and post-training sleep distributions (y-axis) for every learning session, compared to a per-session estimate of baseline differences (x-axis), obtained by bootstrap sampling of patterns within the pre-training sleep epoch. Distances are computed using the Kulback-Liebler divergence (see Materials and Methods). Error bars give the mean and 95% confidence interval on the bootstrapped within-epoch distance $D(Pre|Pre^*)$; identical results were obtained when using $D(Post|Post^*)$. (C) As for B, for stable sessions (85% criterion). (D) Cumulative distribution of the change in pattern probability between sleep epochs, over all sessions. Grey line indicates the tail plotted in panel E. (E) Tail of the cumulative distribution in panel D, on a linear scale.

The dictionary's activity patterns change in probability between sleep epochs

These results demonstrate there is a dictionary of "words" in the mPfC population activity, and that dictionary is conserved across vigilance states. Because it is conserved, we expect that changes to the dictionary due to learning, if any, will thus be mostly expressed as the changes in the probability of particular patterns appearing, rather than the appearance of new patterns or the suppression of existing patterns.

To test this idea, we asked whether training changed the probability of activity patterns appearing. We did this by comparing the probability distributions of patterns in sleep before $P(Pre)$ and after $P(Post)$ training (Figure 3A), and measuring the distance between them $D(Pre|Post)$ using the Kulback-Liebler divergence (see Material and Methods). Due to the finite duration of the two sleep epochs, and so the limited sampling of each activity pattern, identical underlying probability distributions will give rise to similar but not identical distributions of activity patterns. We thus estimated the expected distances for identical distributions by bootstrap sampling within each epoch, giving estimates of $D(Pre|Pre^*)$ and $D(Post|Post^*)$ for the distances between sets of patterns drawn from identical underlying distributions.

In every learning (Figure 3B) and stable (Figure 3C) session, we found the distance between sleep-epoch distributions $D(Pre|Post)$ was greater than their estimate of equivalence. We found similar results when we estimated $D(Pre|Pre^*)$ by randomly dividing the sleep epochs into two sets of samples and computing the distance between the two (results not shown). Pooling over all sessions, we found that the changes in probability followed a long-tailed distribution (Figure 3D), with the vast majority of changes close to zero and the largest changes only occurring for a small proportion ($\approx 5\%$) of activity patterns. The distributions were not identical (Kolmogorov-Smirnov test, $P = 5.4 \times 10^{-4}$; $D_n = 0.0517$; $N(\text{learning}) = 2353$; $N(\text{stable}) = 4374$), as the learning sessions showed a consistently higher probability of large changes except at the extreme end of the distribution (with less than 1% of patterns) (Figure 3E). Together, these results show there were detectable changes to the dictionary between sleep epochs either side of training in both learning and stable sessions, with a suggestion of a higher probability of a larger change in pattern frequency during learning.

The dictionary encodes task elements necessary for learning the maze rules

What is it that is changing in the dictionary between sleep states? If the mPfC dictionary is encoding some rules or regularity of the world, then the patterns which change between sleep states, putatively reflecting some update during training, should correlate with aspects of the task critical for understanding those rules. To test this, we examined the spatial correlates of the co-activation patterns.

In Figure 4*A,C*, we plot the positions at which each co-activation pattern occurred during training as a function of its change in probability between the sleep epochs. We found that the most-changed activity patterns almost exclusively occurred around the choice point of the maze (Figure 4*A,C*). Particularly striking was that the most-changed patterns rarely occurred in the departure arm. Both these properties were true in both learning and stable sessions.

To test the visual impression of the strong correlation between occurrence at the choice point and change in probability, we tested whether this correlation could have arisen by chance. This may have occurred if, for example, the overall spatial distribution of patterns was strongly peaked at the choice point: then sampling at random a small proportion of patterns (the small proportion with the largest changes) would most likely cause this random set of patterns to fall across the choice point. To rule out this possibility, we randomly permuted assignment of positions to patterns, and computed the percentage of permuted patterns that occurred around the choice point. We found that the percentage of data patterns occurring around the choice point well-exceeded the upper-bound predicted by random sampling (Figure 4*B,D*). Again this was true for both learning and stable sessions. Consequently, the spatial correlates of the updated activity patterns are consistent with the dictionary encoding rule-related aspects of the task.

For the learning sessions we could also check if the activity patterns correlated with the decision on each trial, and thus reflected knowledge of the current rule (by definition, we could not do this for the stable sessions, as their behavioural choice was inflexible). To test this, for each co-activation pattern we found its ability to predict a trial's outcome by its rate of occurrence on that trial (Fig 5*A*). We then compared this outcome prediction P(predict outcome) during trials to the change in pattern probability between pre- and post-training sleep (Figure 5*B*). Figure 5C shows there was a notable positive correlation between P(predict outcome) and the change in pattern probability. Nonetheless, the majority of patterns did not markedly change their probability (Figure 2D), nor were they predictive of outcome (72% (1699/2353) have P(predict outcome) ≤ 0.6), so fitting a linear regression is not robust as it is dominated by fitting to this majority that do not change. Rather it is clear from Figure 5C that the *distribution* of change in probability depends on P(predict outcome).

To better quantify this dependence, we constructed the distributions explicitly: we discretised P(predict outcome) into bins containing a fixed number of patterns, and then constructed the distribution of probability change per bin (Figure 5*D*). We then quantified the relationship between P(predict outcome) and the likelihood of a pattern changing its sampling between the pre- and post-training sleep, and found a strong correlation (Figure 5*D-E*). This correlation was highly robust to how we constructed the distributions of change between sleep epochs (Figure 5*E-H*). Consistent with this correlation between change and prediction, highly predictive patterns also preferentially occurred around the choice point of the maze (Figure 5*I*). Again, the percentage of patterns around the choice point well-exceeded the upper bound from a permutation test (results not shown). Thus, not only did patterns that updated between sleep epochs appear at the decision point for each rule during training, but they also predicted the outcome of the trial, consistent with an internal representation of a strategy.

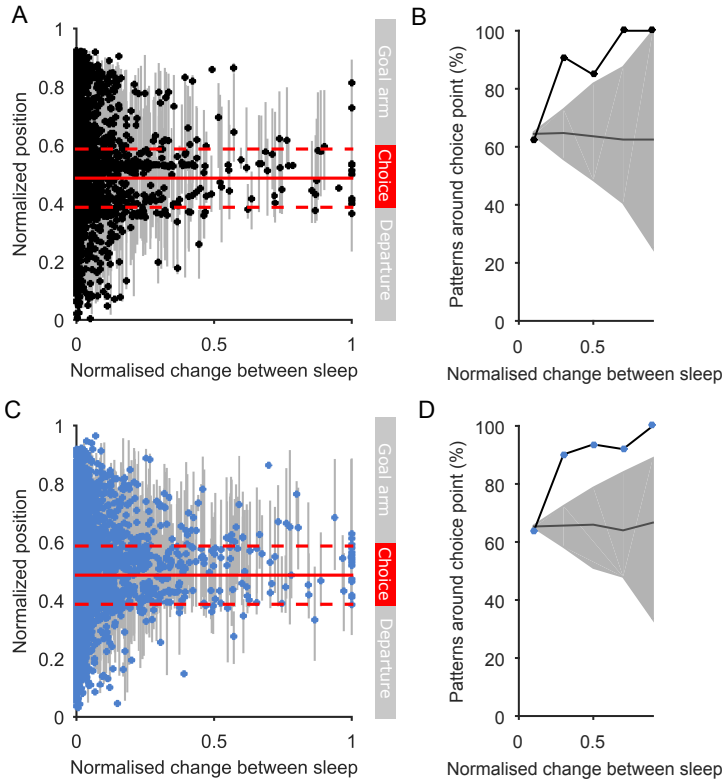

Figure 4. Activity patterns that change probability between sleep epochs are sampled in the choice area during behaviour. (*A*) For learning sessions, the scatter plot of each pattern's change in probability between sleep epochs against the positions of its occurrence in the maze during trials. Change between sleep epochs is the absolute change in probability for that pattern, normalised to the maximum absolute change in that pattern's session. All positions are given as a proportion of the linearised maze from the start of the departure arm. Each dot is the median position; grey line is interquartile range. Red lines indicate the approximate centre (solid) and boundaries (dashed) of the maze's choice area (cf Figure 1*A*). (*B*) For learning sessions, the proportion of activity patterns whose interquartile range of positions enters the choice area (black dots and line). Patterns are binned in change intervals of 0.2. The grey region shows the median (line) and 95% range (shading) of proportions from a permutation test. The data exceed the upper limit of the expected proportions for all patterns that change their frequency between sleep epochs. (*C*-*D*) As A-B, for sessions of stable behaviour (85% criterion).

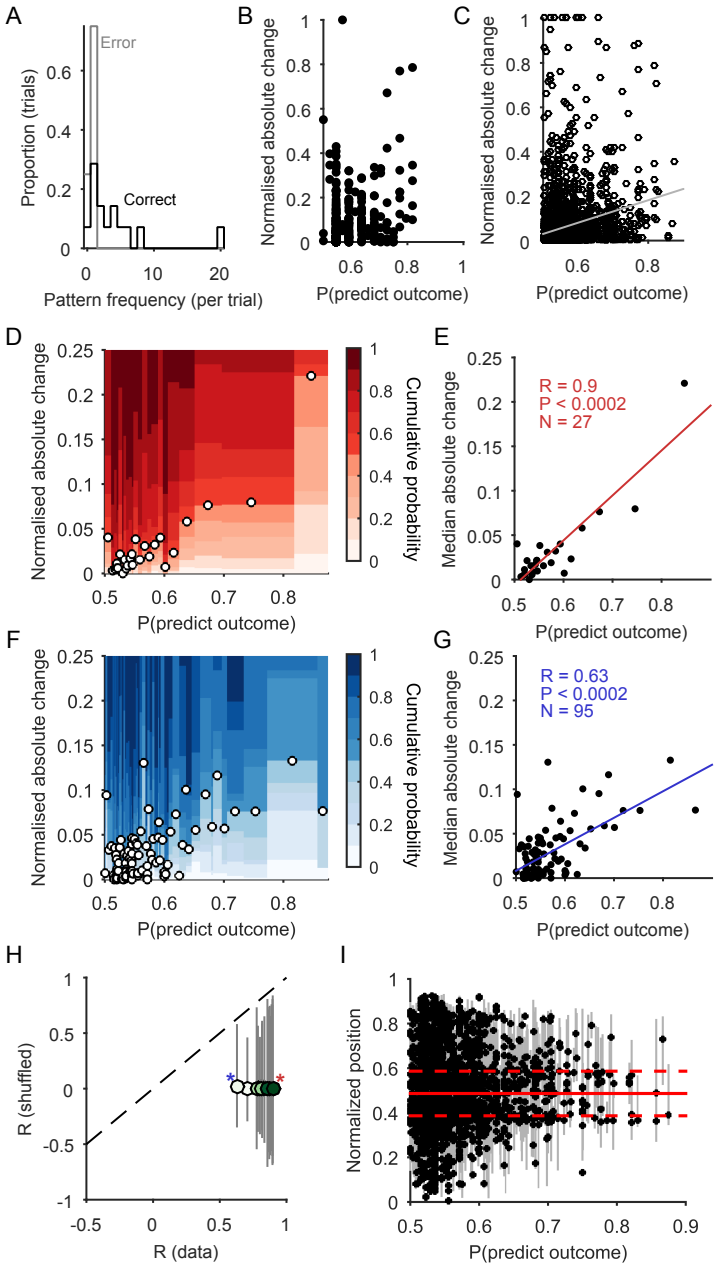

Figure 5. Activity patterns that change frequency between sleep epochs also encode choice behaviour during learning. (*A*) Example distributions for one pattern of its frequency on correct and error trials during one learning session. We define the ability to predict trial outcome from a pattern's frequency as P(predict outcome). (*B*) Scatter plot of outcome prediction and the (absolute) change in pattern frequency between pre- and post-training sleep for all co-activation patterns in one session. Change is normalised to the maximum change in the session. (*C*) Scatter plot of outcome prediction and the (absolute) change in pattern frequency between pre- and post-training sleep for all co-activation patterns in all learning sessions. Grey line: linear regression ($R = 0.22$, $P < 10^{-27}$, $N = 2353$). (*D*) Distributions of the change in pattern frequency as a function of the patterns' outcome prediction probability. Each column is the cumulative probability density for the change in pattern frequency between pre- and post-training sleep, over all patterns in that bin. Circles give the median absolute change for each distribution. Co-activation patterns from all ten learning sessions were binned by outcome prediction into variable size bins containing the same number of patterns. In this example, distributions were built using bins with 90 patterns each. (*E*) Correlation between the outcome prediction and the median change in pattern frequency between pre- and post-training sleep from *D*. Red line: linear regression (P from permutation test). (*F*)-(*G*) As D-E, for the worst-case correlation observed, using 25 patterns per bin. (*H*) Robustness of correlation between the outcome prediction and the median change in pattern frequency between sleep epochs. Circles are the correlation coefficient R between outcome prediction and median change in pattern frequency obtained for different binnings of the data; green colour-scale is proportional to the number of patterns per bin (light to dark: 20-100 per bin). Asterisks indicate data points correspond to panels D-E and F-G. Lines each give the entire range of R obtained from a 5000-repeat permutation test; none reach the equivalent data point (dashed line shows equality), indicating all data correlations had $P < 0.0002$. (*I*) Scatter plot of each pattern's outcome prediction against the positions of its occurrence in the maze during trials. Strongly predictive patterns appeared at the choice point. Compare Figure 4A.

Learning systematically updates the dictionary

The above results showed that the dictionary in mPfC changes between sleep during training, and the changes in pattern probability correlate with task features. This leaves open the question of whether or not the changes that occur during training have a consistent direction. There are two possibilities. One is that there is no direction: the changes in pattern probability during training are random with respect to the pre- and post-training sleep, so the distribution in training is equidistant on average from that in pre- and post-training sleep. The other is that there is a directional change: the changes in pattern probability during training move its pattern distribution systematically closer to one of the sleep distributions. If synaptic changes underlie successful behavioural learning, then we would expect those synaptic modifications to be detectable in the altered dictionary of post-training sleep in learning sessions. We thus hypothesised that learning would move the distribution of patterns in training closer to that in post-training sleep.

To test this, we computed the distances between the distributions of patterns in the sleeping and task epochs in the learning sessions. In order to reasonably compare the learning and stable sessions, for the learning sessions we computed the activity pattern distribution for the all trials after the learning trial $P(Learn)$ so that we only used the trials with stable behaviour during the learning sessions (Figure 1E). We compared the distance $D(Pre|Learn)$ between the distributions in pre-training sleep and learning trials to the equivalent distance $D(Post|Learn)$ between the distributions in post-training sleep and learning trials. If the dictionary of patterns in post-training sleep more closely resembled the dictionary in training than did the dictionary in pre-training sleep then $D(Pre|Learn) > D(Post|Learn)$ (Figure 6A). Remarkably, this is exactly what we found: Figure 6C shows that $D(Pre|Learn)$ was consistently larger than $D(Post|Learn)$, consistent with a convergence of the dictionaries in post-learning trial and post-training sleep activity.

If these changes are wrought by learning, then it follows that we should not see any systematic change to the dictionary when no learning is observed in the stable sessions. To test this prediction, we compared the distance $D(Pre|Stable)$ between the distributions in pre-training sleep and stable session trials to the equivalent distance $D(Post|Stable)$ between the distributions in post-training sleep and stable session trials. If there was no systematic change in the dictionary of patterns then $D(Pre|Stable) \approx D(Post|Stable)$ (Figure 6B). Again, this is exactly what we found: Figure 6C shows that $D(Pre|Stable)$ was not systematically different to $D(Post|Stable)$, consistent with a lack of direction in the change of dictionaries in training trials and post-training sleep activity.

It is also useful to consider not just which sleep distribution is closer to the training distribution, but how much closer. We express this as a convergence ratio $C = [D(Pre|X) - D(Post|X)]/\max\{D(Pre|X), D(Post|X\}$, for the training distribution $X = \{Learn, Stable\}$ in each session. Expressed as a percentage, C falls in the range $[-100, 100]\%$, and a value greater than zero means that the training-epoch distribution X of activity patterns is closer to the distribution in post-training sleep than the distribution in pre-training sleep. We found the post-learning distribution of patterns was on average 20.5% (95% CI=[7.4,33.7]%]) closer to the post-training than the pre-training sleep distribution, whereas the mean convergence for stable sessions was 5.8% (95% CI [-13.6,25.2]%). This further supports the finding that there was no systematic change for the stable sessions, as the magnitude of change also did not show a convergence.

Using all trials for the learning sessions (giving distribution $P(All)$) produced intermediate results. There was weaker evidence of a systematic change in direction between $P(Post)$ and $P(All)$ (8/10 sessions agreed, $p = 0.11$, Wilcoxon signtest), though the mean convergence between $P(Post)$ and $P(All)$ was still greater than zero (mean 18.1%; 95% CI=[2.7,33.7]%]). Correspondingly, the average difference between convergence when using all trials $[P(All)]$ or just the post-learning trials $[P(Learn)]$ was approximately zero (mean difference 2.3%; $p = 0.92$, Wilcoxon signrank). This partial agreement is consistent

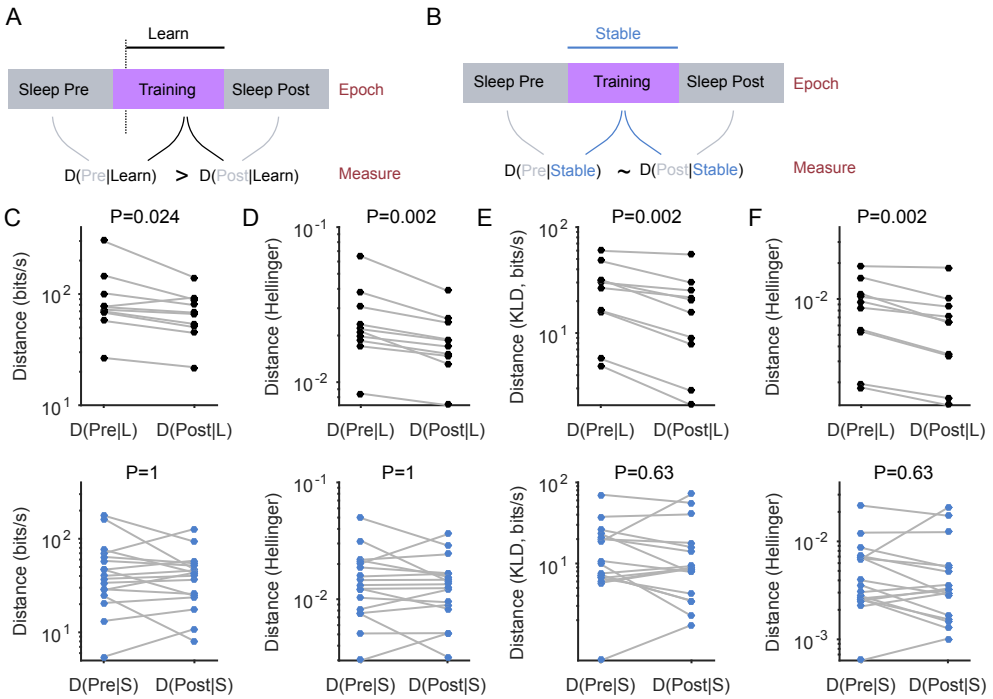

Figure 6. Differing convergence of activity pattern distributions between training and post-training sleep for learning and stable sessions. (A) Schematic of measuring distances between the activity distributions in sleep and training epochs for learning sessions. To detect changes due to learning, we construct the distribution $P(Learn)$ from all trials after the learning trial. $D(X|Y)$: distance between pattern distributions $P(X)$ and $P(Y)$ in epochs X and Y. (B) As A, for stable sessions. The distribution $P(Stable)$ is constructed over all training trials. (C) Distances between the distributions of pattern frequencies in sleep and training epochs for learning (top) and stable (bottom) sessions. One dot per session. L: post-learning trials. S: stable training trials. (D) As C, using the Hellinger distance to compare distributions. (E) As C, but using patterns up to a maximum of 15 neurons per session for more accurate estimation of the Kullback-Liebler divergence. (F) As panel E, using the Hellinger distance. All P-values from a signtest with $N = 10$ (learning) or $N = 17$ (stable) sessions.

with the majority of trials being post-learning in the learning sessions.

Robustness of systematic changes in learning

These results required careful checking. The available number of activity patterns in the trials was an order of magnitude smaller than for the sleep epochs, due to the short duration of the trials (Figure 1B; Figure 2-1). This raises the question of whether the convergence of dictionaries only in the learning and not the stable sessions is a fluke, due to a sampling issue in reliably estimating both the distributions themselves, and the Kulback-Liebler divergence between them. To counteract this possibility, we did three robustness tests. First, we recomputed all distances between probability distributions using a different measure (Figure 6D), the Hellinger distance, which does not suffer from the same sampling issues as the Kulback-Liebler divergence; moreover, it is an asymptotic lower bound for the Kulback-Liebler divergence. Indeed, there was a strong correlation between Hellinger distances and corresponding Kulback-Liebler divergences across the $D(Pre|Learn)$ and $D(Post|Learn)$ measurements (Spearman's $\rho = 0.86$, $N = 20$). Second, we estimated the Kulback-Liebler divergence more accurately by restricting patterns to 15 neurons (Figure 6F), chosen in the same way as the 35 neuron patterns. Third, we checked the 15 neuron patterns using the Hellinger distance (Figure 6F). In all three tests, we found $D(Pre|Learn)$ was systematically larger than $D(Post|Learn)$; and $D(Pre|Stable)$ was not systematically different to $D(Post|Stable)$. Consequently, these results suggest learning systematically changes the mPfC dictionary, and this new dictionary is sampled in post-training sleep.

We also checked if observing systematic changes only in learning sessions could be a

consequence of choices made in our calculations. One possibility was that we happened to chose a binsize for the activity patterns that uniquely obtained a convergence for the learning sessions. As we show in Figure 7A-B, we found a systematic convergence across a range of bin sizes for the patterns, and for patterns of either 35 or 15 neurons. Irrespective of the number of neurons used, the maximum convergence was obtained at low millisecond bin sizes, suggesting this convergence of dictionaries is specific to fine time-scale patterns of synchrony between neurons.

Two other choices with a potential source of error were in estimating the Kulback-Liebler divergence itself. The first is that we used a quadratic estimator to correct the inherent bias in estimating the Kulback-Liebler divergence when using finite samples (see Materials and Methods). However, this estimator means the Kulback-Liebler divergence will vary every time it is calculated. As we show in Figure 7C, this variation was tiny on the scales of the distances between distributions, so did not affect the main result of convergence for the learning sessions. The second is that we computed Kulback-Liebler divergences by setting $P = 0$ for any pattern that does not appear in either epoch we were comparing, to make the computations tractable (see Materials and Methods); while this is the empirical estimate, it could underestimate the actual P due to finite sampling. Our choice of setting $P = 0$ thus rests on the assumption that such infrequent patterns will not systematically alter the main distance results. To check this, we estimated the Kulback-Liebler divergence using a full prior distribution for the probabilities for all possible patterns (see Materials and Methods). Figure 7D shows that the choice of setting $P = 0$ or using the full estimator did not systematically change the results: all convergence scores were still positive, so all sessions showed a convergence between the distributions in training and post-training sleep activity. Thus, the learning sessions had a systematic convergence of dictionaries between training and post-training sleep irrespective of how we calculated our distributions or calculated the distances between them.

Convergence of dictionaries is a consequence of changes to correlation not firing rate

The convergence of distributions in the learning sessions was measured across a change in brain state between waking and sleeping. While within each vigilance state the occurrence of co-activation patterns exceeds chance by up to a factor of six (Figure 2), this still leaves open the possibility that the global change in population dynamics between waking and sleeping could artificially cause their activity pattern distributions to increase in similarity (Okun et al., 2012; Fiser et al., 2013). Indeed, within each learning session, it was clear that neurons had different median firing rates between training and sleep (Figure 8A), though there was no consistent difference between the rates in pre and post-training sleep (Figure 8B).

To test whether the change in vigilance state could account for the convergence of distributions, we used the "raster" model (Okun et al., 2012) to generate predictions for how the change in population firing statistics would alter the occurrence of activity patterns. The raster model generates surrogate sets of spike-trains from the data by permuting spikes within the population's activity patterns, constrained such that each neuron has the same number of spikes and the population has the same distribution of total spikes per time-bin (Figure 8C). Consequently, by counting the probability of each activity pattern in the raster model, we obtain an estimate for the expected occurrence of each pattern due just to the change in vigilance state.

We fitted the raster model to the post-training sleep population activity for each learning session, obtaining the model-derived probability distribution for activity patterns $P(Post-model)$. Our activity patterns were built from single units, unlike previous work using multi-unit activity (Schneidman et al., 2006; Berkes et al., 2011; Okun et al., 2012; Tkacik et al., 2014; Ganmor et al., 2015), so we expected our patterns to be sparse with rare synchronous activity. Indeed our data are dominated by activity patterns with

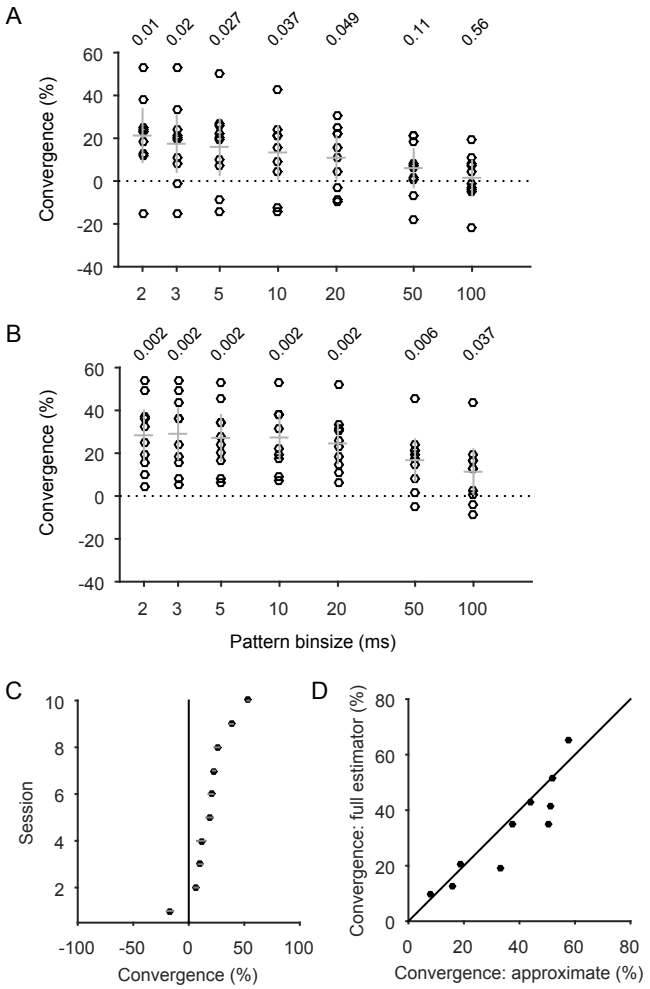

Figure 7. Convergence of activity pattern distributions during learning is robust. (A) Dependence of convergence on the bin size used to construct activity patterns for the full recorded population, up to a maximum of 35 neurons per pattern. Circles are individual learning sessions ($N = 10$); lines give means and 95% confidence intervals. All P-values above the strip-plots are from a Wilcoxon signrank test. (B) As panel A, but using a maximum of 15 neurons per population. (C) Variation in estimating the Kullback-Liebler distance is small. Here we plot the variation in the convergence score for each of the learning sessions over 100 repeated calculations of the Kullback-Liebler distances; symbols give mean distances; error bars plot two standard deviations - on this scale, they are approximately the width of the symbols. (D) Comparison of the convergence estimates for the learning sessions when using the full prior estimator of the unobserved portion of the activity pattern probability distribution (y-axis), and when using our approximation (x-axis). Here we use a maximum of 15 neurons per session, to allow tractable calculation of the full estimator.

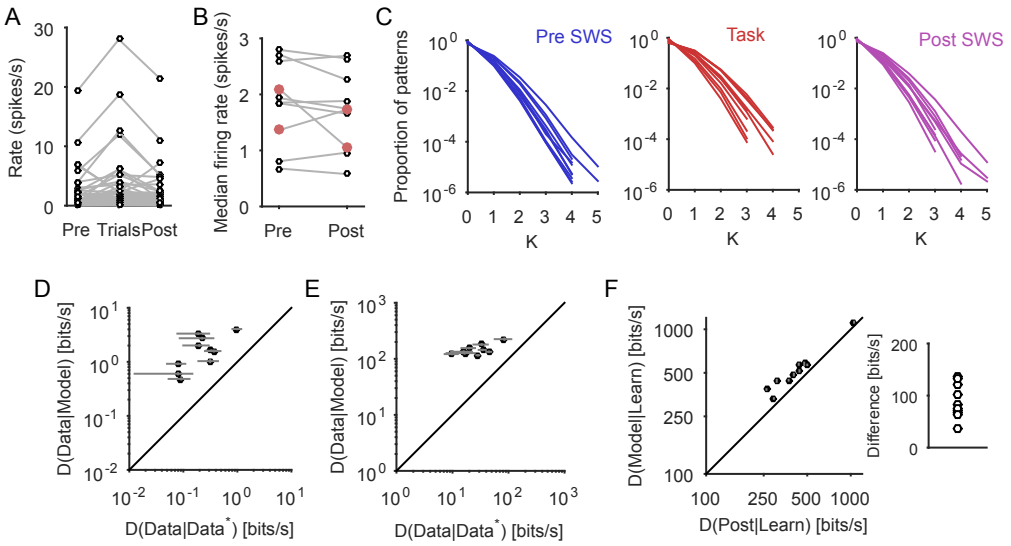

Figure 8. Convergence in learning sessions is caused by changes in correlation, not population firing rate. (A) The distributions of firing rates in the three epochs of one learning session. Firing rates within epochs have a long-tailed distribution, with low firing rates dominating. (B) The median firing rate in each sleep epoch, by session. The red symbols indicate the only two sessions with a detectable shift in firing rates between the sleep epochs at $\alpha = 0.05$ (Wilcoxon signrank test; see Figure 2-1 for numbers of neurons per session). (C) Distributions of the proportion of activity patterns containing exactly K spikes for each epoch of the learning sessions. Each line is the distribution for one session. (D) Distances between model and data distributions for post-training sleep epochs (y-axis) for every learning session, compared to a per-session estimate of baseline differences (x-axis), obtained by bootstrap sampling of patterns within the post-training sleep epoch. Error bars give the mean and 95% confidence intervals on the bootstrapped within-epoch distance $D(Post|Post^*)$ [x-axis], and the 100 repeats of the raster model (y-axis). (E) As panel D, using only co-activation patterns from data and model. (F) The distance between the task and post-task sleep distributions $D(Post|Learn)$ is always smaller than the distance $D(Model|Learn)$ predicted by population firing rate changes during sleep alone, as given by the raster model. Error bars give the mean and 95% confidence intervals over the 100 repeats of the raster model, too small to see on this scale. Inset: plot of the difference between model mean and the data for each session: $D(Model|Learn)$ - $D(Post|Learn)$

no spikes or one spike (Figure 8C). If all patterns had only no spikes or one spike, then the raster model spike trains would be exactly equivalent to the data. Yet despite the relative sparsity ($\sim 1\%$) of co-activation patterns in our data, we found that the distance $D(Post|Post-model)$ between data and model-derived distributions in post-training sleep was always greater than baseline estimates of equivalence between distributions (Figure 8D).

It follows that the true difference between data and model is in the relative occurrence of co-activation patterns. To check this, we applied the same analysis to distributions built only from these co-activation patterns, drawn from data ($P(Post)$) and from the raster model ($P(Post - model)$) fitted to the complete data. With the co-activation patterns, we found that the distance between data and model-derived distributions in post-training sleep was up to an order of magnitude greater than estimates of equivalence (Figure 8E). Consequently, when we then checked the distances between the training and sleep distributions, we found that the data-derived distance $D(Post|Learn)$ was always smaller than the distance $D(Post - model|Learn)$ predicted by the raster model (Figure 8F). These results show that the convergence between training and post-training sleep distributions could not be accounted for by the change in global brain state; rather, the convergence is due to selective changes in when two or more neurons are co-active.

The mPfC dictionary as a probabilistic representation of strategy

Having described the existence of and changes to a dictionary of activity patterns in mPfC, we now propose an explanation for what computations are represented by this dictionary. The recent inference-by-sampling hypothesis (Fiser et al., 2010; Buesing et

al., 2011; Berkes et al., 2011; Habenschuss et al., 2013; Haefner et al., 2016) proposes that the joint activity of a population of neurons represents samples from an encoded probability distribution – or internal model – for some features of the world. We illustrate this idea in Figure 9A for the simple case of two neurons representing a two-dimensional probability distribution. This joint activity, defined by the synaptic connections within and into the population, will specify a particular set of activity patterns: a particular dictionary (Figure 9A). In the sampling theory, neural activity evoked by external input represents samples from a "posterior" probability distribution for the world being in a particular state. A strong prediction of this theory is that if the internal model is encoded by synaptic weights, then spontaneous activity of the same neurons must still represent samples from the internal model (Fiser et al., 2010; Berkes et al., 2011). In the absence of external input, these are then samples from the "prior" probability distribution over the expected properties of the world. As we have observed a consistent dictionary between waking and sleeping states, and which encodes task features, our data are consistent with mPfC activity being samples from the same internal model of the task in both evoked (training) and spontaneous (sleeping) states. Thus we propose the hypothesis that the dictionary in the mPfC is a signature of representing and learning a probabilistic internal model.

To test the plausibility of this hypothesis, we constructed a model for how a probabilistic internal model changes over learning. A candidate for the mPfC internal model is the probabilistic representation of the behavioural strategies that correspond to the rules or regularities of the world (Ragozzino et al., 1999; Rich and Shapiro, 2007, 2009; Durstewitz et al., 2010; Karlsson et al., 2012; Powell and Redish, 2016). Indeed, we have already seen that the mPfC dictionaries selectively change according to relevant features of the world. We thus used a model of probabilistic reinforcement learning of strategies on a simulated Y-maze task.

Our model maintains a probability distribution over the expected reward obtained by choosing each strategy (Figure 9B). As the actual distributions encoded by mPFC are unknown, we use this simplified representation as a proxy for more complex models with distributions over the uncertain values of individual actions and the transitions they cause between states in the maze, which collectively make a strategy. On each simulated trial, the model stochastically chooses a strategy, takes the corresponding action, and observes the resultant feedback. The probability distribution of the selected strategy is then updated to increase or decrease the expected value and the variance around it, according to the feedback. The model is thus an example of general algorithms for updating probabilistic internal models from feedback.

Simulating the model shows how learning the correct strategy corresponds to the probability distributions stabilising (Figure 9C-E). Like the rat, the model shows a marked increase in reward accumulation (Figure 9C) when it learns to consistently select the correct strategy. Consistent reward accumulation will cause the probability distributions to stabilise (Figure 9D), as their changes asymptotically decrease with continual successful outcomes. We illustrate this asymptotic stabilisation in the mean of the distributions in Figure 9E. As we show in the Materials and Methods, irrespective of the details of the algorithm, a basic prediction of any probabilistic reinforcement learning model is that the probability distributions stabilise in this way after sufficient reinforcement.

In this model, we assume that the probability distributions over strategies obtained by the end of a training session are then sampled in post-training sleep. This is consistent with our observation that the dictionary of patterns is conserved between sleep and waking, and with the changes in the probabilities of specific patterns between pre- and post-training sleep (Figure 9G). For comparison with the rats' behaviour, we consider the set of trials around the model's learning trial as a learning session (Figure 9C). The internal model will change within a learning session less after the learning trial than before it (Fig 9E-F), because of the increased stability of the probability distributions

with learning (Figure 9D). Consequently, the probability distributions over strategies in post-learning trials will be closer to those in post-training sleep then pre-training sleep; thus, so will be the distributions over activity patterns that represent samples from these probability distributions, giving $D(Pre|Learn) > D(Post|Learn)$, just as we observed. We thus suggest that the convergence of activity pattern dictionaries in mPfC specifically during learning is a signature of stabilised probability distributions encoded by the mPfC population activity.

Discussion

Here we sought to address whether mPFC population activity contains a dictionary of millisecond-precise activity patterns, and if that dictionary related to learning rules about the world. We found that the set of patterns describing millisecond-scale co-activations of neurons occurred well in excess of the levels predicted by neuron firing rates alone. The same set of patterns was conserved between sleeping and waking. Yet during training the probability of pattern occurrence changed selectively for patterns that occurred at the maze's choice point and predicted the outcome of trials. The direction of change was systematic only during sessions of clear learning, and not during sessions of stable behaviour. Thus, we have described a dictionary of words in mPfC population activity that encodes task features and is updated during training.

Our finding of a highly similar set of precisely-timed activity patterns across sleeping and task performance suggests that mPfC population activity is underpinned by similar constraints in both vigilance states. These results extend to fine time-scale activity patterns the observations in previous studies of strong similarities between spontaneous and evoked firing rates (Tsodyks et al., 1999; Hromádka et al., 2008; O'Connor et al., 2010; Wohrer et al., 2013), firing sequences (Luczak et al., 2009) or rate ensembles (Miller et al., 2014; Carrillo-Reid et al., 2015) in cortex. These findings imply that the underlying cortical circuit has similarly constrained dynamics in both spontaneous and evoked states (Galan, 2008; Marre et al., 2009). Maass and colleagues (Buesing et al., 2011; Habenschuss et al., 2013) have shown that a range of cortical network models can produce specific distributions of such precise activity patterns, provided they have a source of noise (such as synaptic release failure) to produce stochastic wandering of the global activity level. Our data support these models, and suggest that global activity oscillations during slow-wave sleep (Destexhe et al., 1999; Steriade et al., 2001) do not prevent the stochastic sampling of activity patterns, providing a target for future modelling studies.

Studies of prefrontal cortex coding generally assume that information is encoded by firing rates (Ito et al., 2015; Pinto and Dan, 2015; Siegel et al., 2015; Spellman et al., 2015) or ensemble rate correlations (Baeg et al., 2003; Averbeck et al., 2006; Baeg et al., 2007). By contrast, here we show evidence of population coding at highly precise time scales of both position dependence and outcome. That we could extract anything of interest at this resolution was unexpected, and we checked these results extensively, including the use of large-repeat permutation tests. Previously, such fine-scale structure of stimulus-evoked population activity patterns has only been observed in the retina and V1 during passive observation of stimuli (Schneidman et al., 2006; Berkes et al., 2011; Tkacik et al., 2014), and only recently have attempts been made to decode information from these patterns in the retina (Ganmor et al., 2015; Marre et al., 2015). We extend these results to show that such fine time-scale correlation structure can be observed in mPfC, across sleep and behaviour.

We found that the patterns selectively changed during training were just those which occurred at the maze's choice point and, in the learning sessions, predicted trial outcome. This is consistent with plastic changes to the connections within and into the recorded population during training, which in turn changes the frequency of visiting the possible states of the network. But these changes in pattern probability were only systematic in

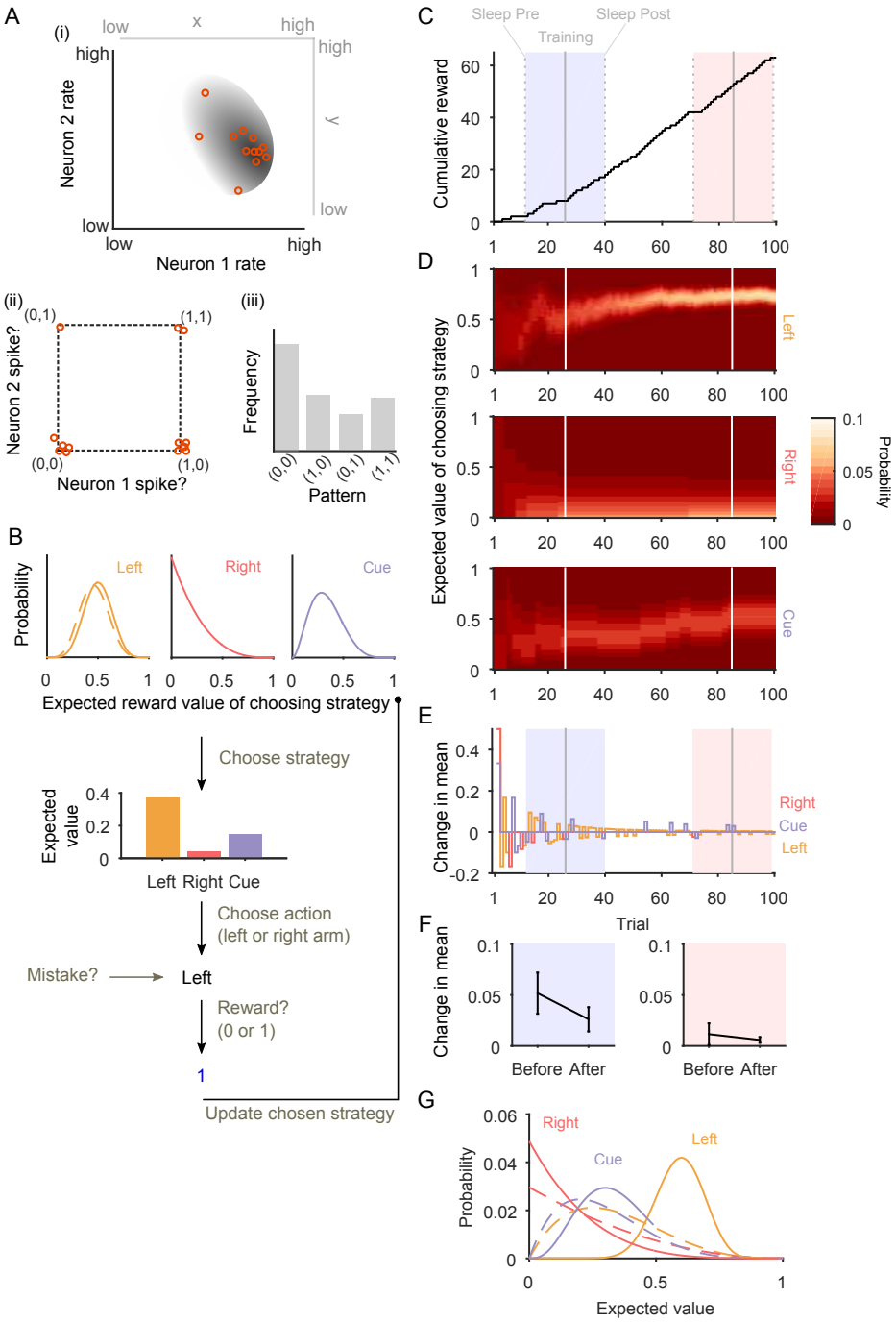

Figure 9. Neural sampling and probabilistic reinforcement learning models. (*A*) Schematic of inference-by-sampling, showing how an arbitrary joint probability distribution can be represented by the joint firing of a neuron population. (i) A two dimensional probability distribution (grey shading) for the values of two variables (x, y: grey axes) can be encoded by the joint firing rate of two neurons (orange circles): high probability regions correspond to frequent pairs of rates. (ii) By finely discretising time, these joint rates correspond to the four possible joint spiking patterns (orange circles are occurrences of each pattern, jittered). (iii) The frequency of these activity patterns is thus a direct function of the underlying encoded probability distribution, and represent samples from that distribution. (*B*) Schematic of the probabilistic reinforcement learning model. The model maintains probability distributions over the expected value of choosing each strategy (dashed lines). On each trial, a strategy is selected according to the highest sample drawn from each distribution. The corresponding action, of selecting the left or right arm on the maze, is executed. Noise is introduced here as a small probability of executing the opposite action (labelled 'Mistake?'). Reward is obtained, and the probability distribution for the chosen strategy is correspondingly updated (solid lines). (*C*) Cumulative reward curve from an example simulation with reward for "go left". The blue shading identifies a virtual "learning" session, a group of trials around the identified learning trial (solid grey line; see Methods). The dotted grey lines identify the trials whose distributions are then sampled in sleep. Red shading identifies an arbitrary later virtual session of stable behaviour, with consistent accumulation of rewards – see Discussion; the grey lines here identify the mid-session, and putative pre- and post-session sleep. (*D*) Corresponding trial-by-trial probability distributions for the expected value of each strategy. Colour-scale gives probability; white lines indicate the learning and stable session mid-points. (*E*) Corresponding trial-by-trial change in the mean of the probability distribution updated on each trial. Shading conventions as per panel C. (*F*) The distribution of changes to the mean before and after the session mid-point, for the learning session (blue) and stable session (red). Error bars plot means and standard deviations. (*G*) Probability distributions for each strategy in pre- and post-training sleep for the learning session (dashed: pre; solid:post).

learning sessions and not during stable behaviour. These results come with the technical issue that the theoretically best distance estimator - the Kullback-Liebler divergence - is also the most difficult to measure accurately with finite samples (Panzeri et al., 2007). To counteract this, we have extensively checked its behaviour, and re-checked our key results with a different, non-parametric distance measure. All showed that the sampling of the dictionary converged between training and post-training sleep in learning but not stable sessions. Clinching confirmatory data for this difference between learning and stable sessions would need recordings from the same set of neurons across multiple sessions, for which we await stable long-term population recordings at millisecond resolution (Jun et al., 2017).

This difference in the changes during learning and stable sessions could be underpinned by two different forms of plasticity. During successful learning of the current rule in the Y-maze, it is plausible that mPfC populations undergo reinforcement-driven plasticity, changing synaptic weights into and between neurons whose co-activity tends to lead to reward (Izhikevich, 2007; Benchenane et al., 2011). During stable behaviour, in which behavioural choice is decoupled from reinforcing feedback, it is plausible that the dictionary changes are driven by synaptic turnover (Wolff et al., 1995), allowing exploration of the possible network states (Kappel et al., 2015; Maass, 2016). Testing such ideas would again require stable long-term recordings of the same population across learning and asymptotic behaviour.

Replay and dictionary sampling

That the spontaneous activity of sleep and task-evoked activity during waking are sampling from the same, highly conserved dictionary suggests an alternative interpretation of "replay" phenomena (Euston et al., 2007; Peyrache et al., 2009). Replay of neural activity during waking in a subsequent episode of sleep has been inferred by searching for matches of patterns of awake activity in sleep activity, albeit at much coarser time-scales than used here. The better match of waking activity with subsequent sleep than preceding sleep has been taken as evidence that replay is encoding recent experience, perhaps to enable memory consolidation. However, our observation that the distributions of patterns in stable sessions' trials are not specifically sampled in post-training sleep (Figure 6) is incompatible with the simple replay of experience-related activity in sleep.

Rather, our results suggest that the similarity between waking and sleep activity is due to the constraints placed on them by the cortical network, and how that network is changed by learning, not recent experience per se. The similarity of the patterns in waking and subsequent sleep is then caused by sampling from the same dictionary, not by explicitly recalling specific patterns that occurred in waking activity. Our data thus suggest that replay may be a signature of resampling.

Dictionary sampling as a probabilistic internal model

While the above discussed results give strong constraints on mPfC dynamics and coding, they do not in themselves make an obvious connection to neural computation. To address this, we have proposed a computational hypothesis that the dictionary itself represents a probabilistic internal model in mPfC. While the hypothesis that brains compute using probabilities is widely-discussed, most evidence for it has been from observations of behaviour that is consistent with probabilistic inference (Wolpert et al., 1995; Körding and Wolpert, 2004; Pouget et al., 2013). Strong evidence for probabilistic brains requires detecting the representation and use of probability distributions in circuit-level neural activity (Knill and Pouget, 2004). Theoretical work has elucidated potential mechanisms for how cortical populations represent and compute with probabilities (Zemel et al., 1998; Ma et al., 2006; Beck et al., 2008; Fiser et al., 2010; Buesing et al., 2011; Haefner et al., 2016). One popular account is probabilistic population codes (Ma et al., 2006; Beck et al., 2008;

Pouget et al., 2013), but these have been elucidated in the context of sensory variables and rely on neurons encoding a set of basis functions to represent the range of each sensory variable; it is not immediately clear what such a basis function set would be in mPfC or rule learning. Rather, our finding of a consistent dictionary of activity patterns is more easily interpreted in the context of the inference-by-sampling theory (Figure 9A), which allows for the representation of arbitrary probability distributions (Fiser et al., 2010).

We demonstrated how a probabilistic reinforcement learning model could learn an internal model for the Y-maze task using the probabilistic representation of strategies. Our model predicts that the internal model stabilises during successful learning of the correct strategy. Given the sampling theory's correspondence between the probability distributions of the internal model and the sampled activity patterns, this stabilisation predicts that the dictionary of activity patterns in sleep should be closer to the distribution in the trials after the behavioural strategy changes, just as we observed in the data. The model also makes the prediction that, should stable recordings of the same neurons across sessions be obtained, then the dictionaries within each training epoch should become progressively more similar in consecutive sessions (compare blue and red shading in Figure 9C-F). Thus here we suggest that the existence and changes to the neural dictionary in mPfC are a signature of probabilistic internal models and their updating through learning. Such an account extends the applications of probabilistic internal models to a candidate general computational principle of cortex.

Extended Data Legends

Figure 2-1 Numbers of neurons in each session, and numbers of activity patterns in each epoch for each session.

References

Averbeck BB, Sohn JW, Lee D (2006) Activity in prefrontal cortex during dynamic selection of action sequences. *Nat Neurosci* 9:276–282.

Baeg EH, Kim YB, Huh K, Mook-Jung I, Kim HT, Jung MW (2003) Dynamics of population code for working memory in the prefrontal cortex. *Neuron* 40:177–188.

Baeg EH, Kim YB, Kim J, Ghim JW, Kim JJ, Jung MW (2007) Learning-induced enduring changes in functional connectivity among prefrontal cortical neurons. *J Neurosci* 27:909–918.

Beck JM, Ma WJ, Kiani R, Hanks T, Churchland AK, Roitman J, Shadlen MN, Latham PE, Pouget A (2008) Probabilistic population codes for Bayesian decision making. *Neuron* 60:1142–1152.

Benchenane K, Peyrache A, Khamassi M, Tierney PL, Gioanni Y, Battaglia FP, Wiener SI (2010) Coherent theta oscillations and reorganization of spike timing in the hippocampal- prefrontal network upon learning. *Neuron* 66:921–936.

Benchenane K, Tiesinga PH, Battaglia FP (2011) Oscillations in the prefrontal cortex: a gateway to memory and attention. *Curr Opin Neurobiol* 21:475–485.

Berkes P, Orbán G, Lengyel M, Fiser J (2011) Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* 331:83–87.

Buesing L, Bill J, Nessler B, Maass W (2011) Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons. *PLoS Comput Biol* 7:e1002211.

Carrillo-Reid L, Miller JEK, Hamm JP, Jackson J, Yuste R (2015) Endogenous sequential cortical activity evoked by visual stimuli. *J Neurosci* 35:8813–8828.

Cunningham JP, Yu BM (2014) Dimensionality reduction for large-scale neural recordings. *Nat Neurosci* 17:1500–1509.

Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci* 8:1704–1711.

Destexhe A, Contreras D, Steriade M (1999) Spatiotemporal analysis of local field potentials and unit discharges in cat cerebral cortex during natural wake and sleep states. *J Neurosci* 19:4595–4608.

Durstewitz D, Vittoz NM, Floresco SB, Seamans JK (2010) Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron* 66:438–448.

Euston DR, Tatsuno M, McNaughton BL (2007) Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. *Science* 318:1147–1150.

Fiser J, Berkes P, Orbán G, Lengyel M (2010) Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn Sci* 14:119–130.

Fiser J, Lengyel M, Savin C, Orbán G, Berkes P (2013) How (not) to assess the importance of correlations for the matching of spontaneous and evoked activity. p. arXiv:1301.6554.

Galan RF (2008) On how network architecture determines the dominant patterns of spontaneous neural activity. *PloS One* 3:e2148.

Ganmor E, Segev R, Schneidman E (2015) A thesaurus for a neural population code. *Elife* 4:e06134.

Ghavamzadeh M, Mannor S, Pineau J, Tamar A (2015) Bayesian reinforcement learning: A survey. *Foundations and Trends in Machine Learning* 8:359–492.

Habenschuss S, Jonke Z, Maass W (2013) Stochastic computations in cortical microcircuit models. *PLoS Comput Biol* 9:e1003311.

Haefner RM, Berkes P, Fiser J (2016) Perceptual decision-making as probabilistic inference by neural sampling. *Neuron* 90:649–660.

Hromádka T, Deweese MR, Zador AM (2008) Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol* 6:e16.

Ito HT, Zhang SJ, Witter MP, Moser EI, Moser MB (2015) A prefrontal-thalamo-hippocampal circuit for goal-directed spatial navigation. *Nature* 522:50–55.

Izhikevich EM (2007) Solving the distal reward problem through linkage of stdp and dopamine signaling. *Cereb Cortex* 17:2443–2452.

Jazayeri M, Afraz A (2017) Navigating the neural space in search of the neural code. *Neuron* 93:1003–1014.

Jun JJ, Mitelut C, Lai C, Gratiy S, Anastassiou C, Harris TD (2017) Real-time spike sorting platform for high-density extracellular probes with ground-truth validation and drift correction. *bioRxiv* .

Kappel D, Habenschuss S, Legenstein R, Maass W (2015) Network plasticity as bayesian inference. *PLoS Comput Biol* 11:e1004485.

Karlsson MP, Tervo DGR, Karpova AY (2012) Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. *Science* 338:135–139.

Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci* 27:712–719.

Körding KP, Wolpert DM (2004) Bayesian integration in sensorimotor learning. *Nature* 427:244–247.

Luczak A, Barthó P, Harris KD (2009) Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron* 62:413–425.

Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. *Nat Neurosci* 9:1432–1438.

Maass W (2016) Searching for principles of brain computation. *Curr Opin Behav Sci* 11:81–92.

Marre O, Botella-Soler V, Simmons KD, Mora T, Tkacik G, Berry II MJ (2015) High accuracy decoding of dynamical motion from a large retinal population. *PLoS Comput Biol* 11:e1004304.

Marre O, Yger P, Davison AP, Frégnac Y (2009) Reliable recall of spontaneous activity patterns in cortical networks. *J Neurosci* 29:14596–14606.

Miller JeK, Ayzenshtat I, Carrillo-Reid L, Yuste R (2014) Visual stimuli recruit intrinsically generated cortical ensembles. *PNAS* 111:E4053–E4061.

O'Connor DH, Peron SP, Huber D, Svoboda K (2010) Neural activity in barrel cortex underlying vibrissa-based object localization in mice. *Neuron* 67:1048–1061.

O'Donnell C, Gonçalves JT, Whiteley N, Portera-Cailliau C, Sejnowski TJ (2017) The population tracking model: A simple, scalable statistical model for neural population data. *Neural Computation* 29:50–93.

Ohiorhenuan IE, Mechler F, Purpura KP, Schmid AM, Hu Q, Victor JD (2010) Sparse coding and high-order correlations in fine-scale cortical networks. *Nature* 466:617–621.

Okun M, Yger P, Marguet SL, Gerard-Mercier F, Benucci A, Katzner S, Busse L, Carandini M, Harris KD (2012) Population rate dynamics and multineuron firing patterns in sensory cortex. *J Neurosci* 32:17108–17119.

Panzeri S, Senatore R, Montemurro MA, Petersen RS (2007) Correcting for the sampling bias problem in spike train information measures. *J Neurophysiol* 98:1064–1072.

Peyrache A, Khamassi M, Benchenane K, Wiener SI, Battaglia FP (2009) Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nat Neurosci* 12:916–926.

Pinto L, Dan Y (2015) Cell-type-specific activity in prefrontal cortex during goal-directed behavior. *Neuron* 87:437–450.

Pouget A, Beck JM, Ma WJ, Latham PE (2013) Probabilistic brains: knowns and unknowns. *Nat Neurosci* 16:1170–1178.

Powell NJ, Redish AD (2016) Representational changes of latent strategies in rat medial prefrontal cortex precede changes in behaviour. *Nat Commun* 7:12830.

Ragozzino ME, Detrick S, Kesner RP (1999) Involvement of the prelimbic-infralimbic areas of the rodent prefrontal cortex in behavioral flexibility for place and response learning. *J Neurosci* 19:4585–4594.

Rich EL, Shapiro M (2009) Rat prefrontal cortical neurons selectively code strategy switches. *J Neurosci* 29:7208–7219.

Rich EL, Shapiro ML (2007) Prelimbic/infralimbic inactivation impairs memory for multiple task switches, but not flexible selection of familiar tasks. *J Neurosci* 27:4747–4755.

Ringach DL (2009) Spontaneous and driven cortical activity: implications for computation. *Curr Opin Neurobiol* 19:439–444.

Sadtler PT, Quick KM, Golub MD, Chase SM, Ryu SI, Tyler-Kabara EC, Yu BM, Batista AP (2014) Neural constraints on learning. *Nature* 512:423–426.

Schneidman E, Berry MJ, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440:1007–1012.

Shlens J, Field GD, Gauthier JL, Grivich MI, Petrusca D, Sher A, Litke AM, Chichilnisky EJ (2006) The structure of multi-neuron firing patterns in primate retina. *J Neurosci* 26:8254–8266.

Siegel M, Buschman TJ, Miller EK (2015) Cortical information flow during flexible sensorimotor decisions. *Science* 348:1352–1355.

Spellman T, Rigotti M, Ahmari SE, Fusi S, Gogos JA, Gordon JA (2015) Hippocampal-prefrontal input supports spatial encoding in working memory. *Nature* 522:309–314.

Steriade M, Timofeev I, Grenier F (2001) Natural waking and sleep states: a view from inside neocortical neurons. *J Neurophysiol* 85:1969–1985.

Strong SP, Koberle R, de Ruyter van Steveninck RR, Bialek W (1998) Entropy and information in neural spike trains. *Phys Rev Lett* 80:197–200.

Tkacik G, Marre O, Amodei D, Schneidman E, Bialek W, Berry n MJ (2014) Searching for collective behavior in a large network of sensory neurons. *PLoS Comput Biol* 10:e1003408.

Tsodyks M, Kenet T, Grinvald A, Arieli A (1999) Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science* 286:1943–1946.

Wohrer A, Humphries MD, Machens C (2013) Population-wide distributions of neural activity during perceptual decision-making. *Prog Neurobiol* 103:156–193.

Wolff JR, Laskawi R, Spatz WB, Missler M (1995) Structural dynamics of synapses and synaptic components. *Behav Brain Res* 66:13–20.

Wolpert DM, Ghahramani Z, Jordan MI (1995) An internal model for sensorimotor integration. *Science* 269:1880–1882.

Yuste R (2015) From the neuron doctrine to neural networks. *Nat Rev Neurosci* 16:487–497.

Zemel RS, Dayan P, Pouget A (1998) Probabilistic interpretation of population codes. *Neural Comput* 10:403–430.

Extended Data Tables

Neurons	Pre-training SWS	Post-learning trials	Post-training SWS	Rest
23	281001	57419	193500	315508
20	65007	49029	165519	350335
20	270012	34910	99488	282512
35	240992	20417	461972	92504
35	558510	43682	322499	131011
31	362007	26713	330485	206006
23	351996	50058	414982	205510
12	433009	29612	266493	204506
25	388006	50995	568512	105997
27	371013	64785	453993	90008

Table S1: Learning sessions: neurons and patterns. The Neurons column give the number of neurons used from each of the ten learning sessions to build the activity patterns; eight used all recorded neurons, two were capped at 35. The other columns give the total number of activity patterns in each epoch.

Neurons	Pre-training SWS	All trials	Post-training SWS
21	433009	42006	266493
19	377028	70435	468512
35	262999	76452	262511
35	341040	40062	250509
35	166511	70159	389510
35	104998	66319	16998
35	286491	66880	260521
35	109992	46539	209005
21	127997	71266	302997
19	346530	449624	448510
22	238523	30048	139999
17	521982	66071	330505
29	154498	144571	214992
12	107994	111723	204010
19	441977	108721	168996
21	90498	86011	112500
22	99508	97662	81003

Table S2: Stable sessions: neurons and patterns. The Neurons column give the number of neurons used from each of the 17 stable sessions (using the threshold of 85%) to build the activity patterns; nine used all recorded neurons, six were capped at 35. The other columns give the total number of activity patterns in each epoch.