Independent population coding of the present and the past in prefrontal cortex during learning

Silvia Maggi[1,2] and Mark D. Humphries[1,2*]

1. School of Psychology, University of Nottingham, Nottingham, UK.
2. Faculty of Biology, Medicine, and Health, University of Manchester, Manchester, UK.

* Contact: mark.humphries@nottingham.ac.uk

Abstract

Medial prefrontal cortex (mPfC) plays a role in both immediate behaviour and short-term memory. Unknown is whether the present and past are represented simultaneously or separately in mPfC populations. To address this, we analysed mPfC population activity of rats learning rules in a Y-maze, with self-initiated choice trials followed by a self-paced return during the inter-trial interval. Joint mPfC population activity encoded solely present events and actions during the trial, with decoding of the past at chance; conversely, population encoding of the same features in the immediately following inter-trial interval was solely of the past. Despite being contiguous in time, each population orthogonally encoded the present and past of the same events and actions. Consequently, only the population code of the present during the trials, and not the past coding of the inter-trials intervals, was re-activated in subsequent sleep. Our results suggest that representations of the past and present in the mPfC independently contribute to the learning of a new rule.

Keywords: decision making, mPFC, learning, neural ensembles, sleep, replay

Introduction

The medial prefrontal cortex (mPfC) plays key roles in adaptive behaviour, in the reshaping of behaviour in response to a changing or dynamic environment (Euston et al., 2012) or to errors in performance (Laubach et al., 2015; Narayanan and Laubach, 2008). Damage to mPfC prevents shifting behavioural strategies when the environment changes (Guise and Shapiro, 2017; Laskowski et al., 2016). Single neurons in mPfC shift the timing of spikes relative to hippocampal theta rhythms just before acquiring a new action-outcome rule (Benchenane et al., 2010). And multiple labs have reported that global shifts in mPfC population activity precede switching between behavioural strategies (Powell and Redish, 2016; Durstewitz et al., 2010; Rich and Shapiro, 2009; Karlsson et al., 2012; Schuck et al., 2015).

Adapting behaviour depends on knowledge of both the past and the present. Deep lines of research have established mPfC represents information about both. Memory of the immediate past is maintained in mPfC activity, both in tasks requiring explicit use of working memory (Miller and Cohen, 2001; Spellman et al., 2015; Baeg et al., 2003;

Averbeck and Lee, 2007; Miller et al., 2005; Machens et al., 2010; Fujisawa et al., 2008) and those that do not (Maggi et al., 2018; Jun et al., 2010). The use of such memory is seen in both the impairment arising from mPfC lesions (Rich and Shapiro, 2007; Young and Shapiro, 2009; Laskowski et al., 2016), and role of the mPfC in error monitoring (Laubach et al., 2015). Representations of stimuli and events happening in the present have been reported in a variety of decision-making tasks (Rigotti et al., 2013; Averbeck et al., 2006; Erlich et al., 2015; Hanks et al., 2015; **?**; Ito et al., 2015). But unknown is how mPfC represents the past and the present in the same population.

To address this question, we analyse here mPfC population activity from rats learning new rules on a Y-maze. Crucially, this task has distinct trial and inter-trial interval phases, in which we can respectively examine the encoding of the present (in trials) and the past (in the intervals) of the same task feature or event.

With these data, we first establish that population activity does encode multiple features of the present during a trial. We then show this encoding is orthogonal to the encoding of the same features in the past during the inter-trial interval. This independent encoding is functional: population activity encoding the present is reactivated in post-training sleep; but activity encoding the same features in the past is not reactivated. Finally, we show that the pattern of synchrony across the population is also independent between trials and inter-trial intervals. Thus converging evidence from encoding, sleep, and synchrony all suggest the same mPfC population can support independent coding of the past and present.

Results

We analyse here data from rats learning rules in a Y maze (Figure 1a). Animals were implanted with tetrodes in mPfC before the first session of training. Then, the rats were trained to learn one of four rules in order: go right, go to the lit arm, go left, go to the dark arm. We thus have two qualitatively different types of association learning to consider, direction rules and cue rules. The light was randomly presented at the end of one of the two arms, regardless of the rule in place. The animal self-initiated each trial by running along the central stem of the maze and choosing one of the arms. At the end of the arm, the trial ended, and correct choice was rewarded. During the following inter-trial interval, the rat made a self-paced return to the start of the central arm to initiate the next trial. The data-set contains 50 sessions, of between 7-51 trials.

We separate sessions into three groups by their qualitatively different behavioural events: learning, rule change, and other non-learning sessions. As in the original study (Peyrache et al., 2009), learning sessions (n=10) are identified post-hoc as those with three consecutive correct trials followed by a performance of at least 80% correct trials (Figure 1b). The first of the three correct trials is the learning trial, which identifies a clear step change in task performance (Figure 1b,d). In the eight rule-change sessions, the rule changed mid-session after ten consecutive correct trials or 11 correct out of 12 trials. The shift to a new rule produces a step change in the performance opposite to the learning sessions (Figure 1c,d). The remaining sessions, called here "Other", were characterised by a mixed behavioural performance (Maggi et al., 2018). Thus, the behaviour on this task lets us compare and contrast population encoding during step-like learning against encoding during externally-imposed changes in contingency, and during control periods of non-learning.

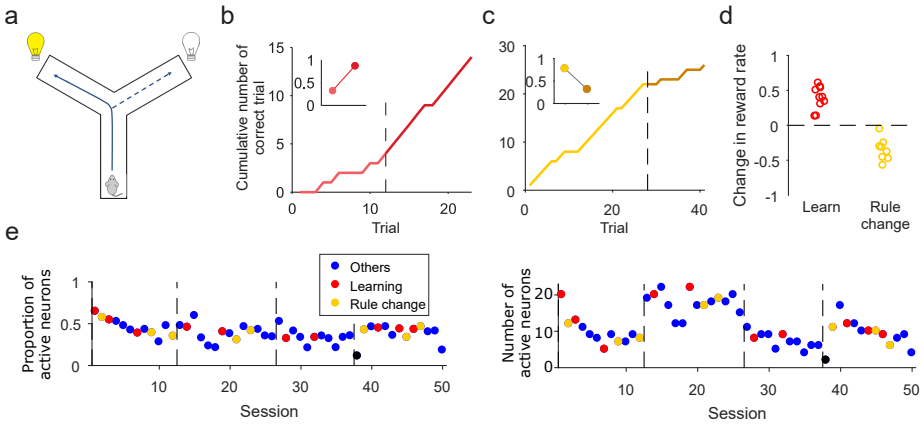

Figure 1: Task, behavioural changes and neural population size.
(**a**) A graphical representation of the Y maze. Each session is a series of trials followed by inter-trial intervals (ITIs). Animals were asked to learn one of 4 rules in sequence (go to the right arm, go to the lit arm, go to the left arm, go to the dark arm).
(**b**) Example learning curve from a learning session, plotting the cumulative number of correct trials. Black dashed line identifies the learning trial as the first of three consecutive correct trials followed by at least 80% correct trials. Inset: reward rates before (light red) and after (dark red) the learning trial. Reward rates were given by the slope of linear regressions fitted to the learning curve before and after the learning trial.
(**c**) Example learning curve from a rule change session. The black dashed line identifies the rule change trial. Inset: the reward rate before the rule change (light yellow) and after (dark yellow).
(**d**) Change in reward rate after learning trials (red) or rule change trials (yellow). Each symbol is a session.
(**e**) Size of retained neural populations. Left: the proportion of neurons active in every trial of a session (dots). No differences between learning, rule change and other non-learning sessions was observed here. Right: the actual number of neurons contributing to the retained population varied between 2 and 22 (right panel; mean ± SEM, 11.3 ± 0.7). All sessions with 4 or more neurons were included in our analyses, so excluding just one session (black dot). Sessions are plotted in time order; the vertical dashed lines separate sessions by animal.

Population activity encodes multiple features of the present

We first ask if we can decode current events and choices from population activity during the trials. In order to track population encoding across all trials, we select all the neurons active in every trial of a session. The resulting active populations range in size from 2 to 22 neurons across the sessions (Figure 1e). No differences in the proportion or number of retained neurons are observed between session types (learning, rule change and other non-learning). For each trial, we divide the firing activity of this active population according to five equally sized sections of the maze (Figure 2a,b). At each position, the population's firing rate vectors are used to train and test a cross-validated classifier. We train the same classifiers on shuffled events (see Methods and Supplementary Figure 1a), to define appropriate chance levels for each classifier given the unbalanced distribution of some task-features (such as outcome).

We decode direction choice, light position and trial outcome to test encoding of anticipatory choice and reward prediction. We find that the current trial's direction choice, light position and outcome are consistently encoded in mPfC population activity from the choice point to the end of the maze (Figure 2c). Broken down by types of session and rule (direction or cued) the encoding of direction choice (Figure 2e, direction column) increases along the maze in almost all types of session and rule, consistent with a role

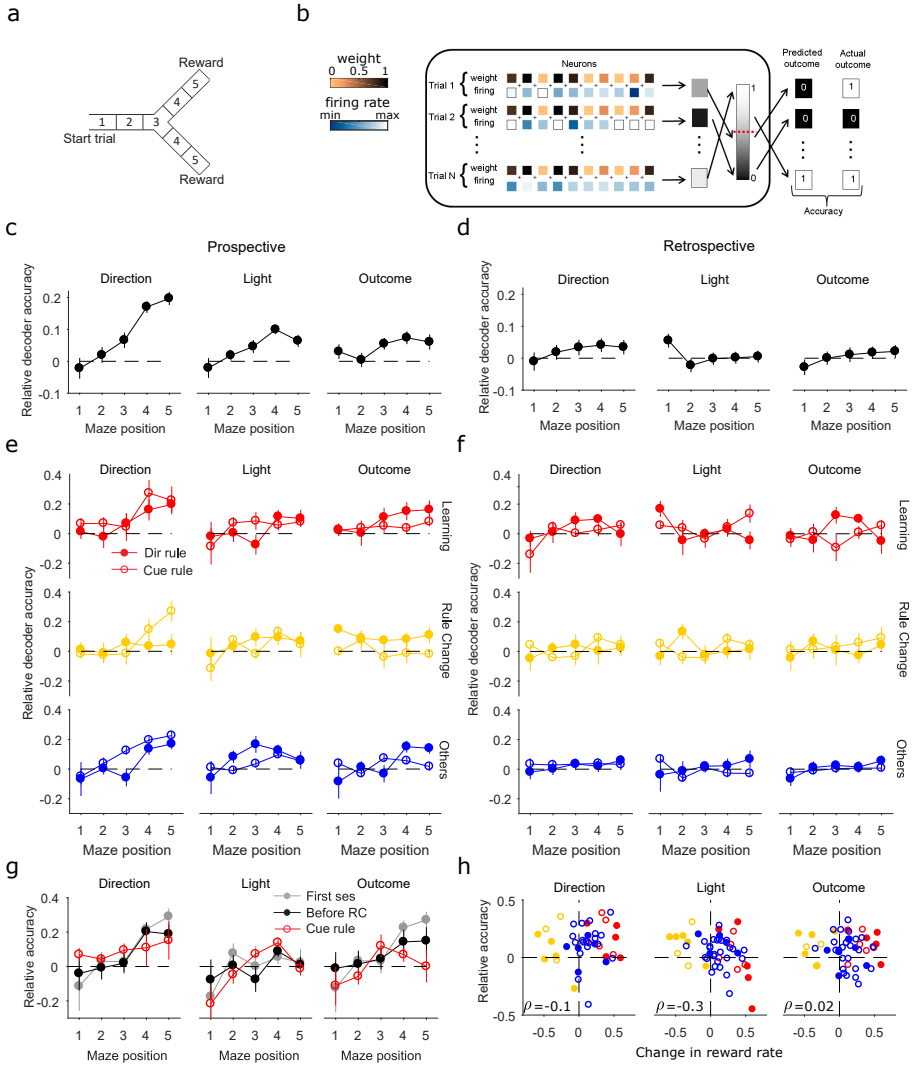

Figure 2: Population encoding of task features and events during trials.
(**a**) The maze is divided into 5 equally sized sections where position 1 represent the start of the first arm, position 3 is the choice point, and position 5 is the end of the two arms where the reward was delivered if correct choices were made. For each section and each trial, the population firing activity is decoded with a linear classifier.
(**b**) Graphical representation of a classifier. The population vector firing rate for each trial (shade of blue squares) is used as input for a linear classifier that fits the weight (shade of yellow squares) for each neuron across trials. A linear combination of weight and population firing rate for each trial is compared to a threshold (red dashed line) to predict the category to which the trial belongs. Classifier accuracy is the proportion of correctly predicted held-out trials.
(**c**) Decoding the animal's choice (of direction), the cue position, and the outcome of each trial as a function of maze position. We plot the relative decoder accuracy, the difference between the absolute decoder accuracy and the shuffled control for each feature (we plot example absolute accuracy results in SI Figure 1). Decoding accuracy above chance appear from the choice point (position 3) to the end of the maze. Each symbol is mean ± SEM over all sessions.
(**d**) Decoding of the previous trial's choice, cue position, and outcome as a function of maze position on the present trial. All values are mean ± SEM.
(**e**) Breakdown of the prospective decoding results in panel (**c**) by session type and rule type. Each line is mean ± SEM over sessions of the identified combination of session and rule types.
(**f**) Breakdown of the retrospective decoding results in panel (**d**) by session type and rule type. Conventions as per panel (e).
(**g**) Relative decoder accuracy for the first session of each animal (grey) is above chance for direction and outcome from the choice point, as well as all the sessions before the first rule change (all direction rule sessions, black lines). The learnt cue rule sessions for each animal (red lines) decode light at the choice point compared to the other sessions. All values are mean ± SEM.
(**h**) Relative decoder accuracy (on the whole maze) as a function of the change in reward rate during a session. Same colour scheme as per panel (**e**). ρ: Spearman's correlation coefficient.

in guiding choice. In contrast to direction choice, encoding of light position depends on both the types of session and rule. While outcome and direction choice are confounded on direction rules (solid symbols), Figure 2e shows we could decode the trial's outcome above chance for cue rules at maze locations from the choice point. These decodings of direction, light position, and outcome are strongly robust, being recapitulated by decoders applied to the whole trial (SI Fig 1b,d), and by a range of different classifiers (SI Fig 2b). Moreover, they are solely of the present: decoding of the previous trial is at chance levels for all features (Figure 2d,f and SI Fig 1c,e). Population activity in prefrontal cortex thus encodes multiple features and events of the present during trials.

Our decoder analysis established some specific links between population encoding and behaviour. Three are of particular note. First, we cannot decode direction choice during rule-change sessions that started with a direction rule (Figure 2e, yellow lines and solid symbols). In those sessions, each animal has near-perfect performance on the direction rule, suggesting that such habitual behaviour obviates the need for an explicit representation of direction choice in the mPfC. Second, direction choice can be decoded from the very first session of each animal, suggesting that this representation is not learnt (Figure 2g). Third, the encoding of light position is not present during the initial sessions that enforced a direction rule (Figure 2g).

While the population encoding could be linked to behaviour, we could not link it to learning per se. No encoded feature is unique to learning sessions (Figure 2e). Moreover, the strength of population encoding in a session does not predict the change in the animal's performance during the session (Figure 2h).

Orthogonal encoding of the past and the present

In the Y-maze task, the trial and following inter-trial interval are contiguous in time. We've just established that the mPfC population activity during the trial encodes features of the present. Previously we reported that during the inter-trial interval these populations encode the same features - choice, light, and outcome - of the preceding trial (Maggi et al., 2018). In Figure 3a,c we confirm and extend this result by showing that we can decode prior features from activity vectors constructed over the whole inter-trial interval.

An hypothesis to link the two observations is that the encoding during the inter-trial interval is a memory trace of encoding during the trial, a continuation of the neural representation in the trial into the immediately following inter-trial interval. Which brings us to the crux of the paper: this hypothesis predicts that the population encoding in the trial and following inter-trial interval should be the same. We show here they are not.

The active neurons during the trials are also active during the inter-trial interval (SI Fig 3c), so this shared common population could in principle carry the same encoding of the task. We use this common population to test the memory trace hypothesis: if the encoding is broadly the same, we should be able to use trial activity to train a decoder for the following inter-trial intervals; and vice-versa, using inter-trial activity to train a decoder for the preceding trials. In this cross-decoding test, we train a classifier for features of the present using the common population's activity during the trials, and then test the accuracy of the classifier when using the common population's activity during the inter-trial interval. The decoded features are the same - all that changes is the vector of population activity.

We find that cross-decoding is consistently poor in both directions, whether we train on trial activity and test on inter-trial intervals, or vice-versa (Figure 3b,d). Decoding of all features is at or close to chance, across different types of session and different types of rules, strikingly at odds with the within-trial (SI Fig 1d) or within-interval (Figure 3a,c)

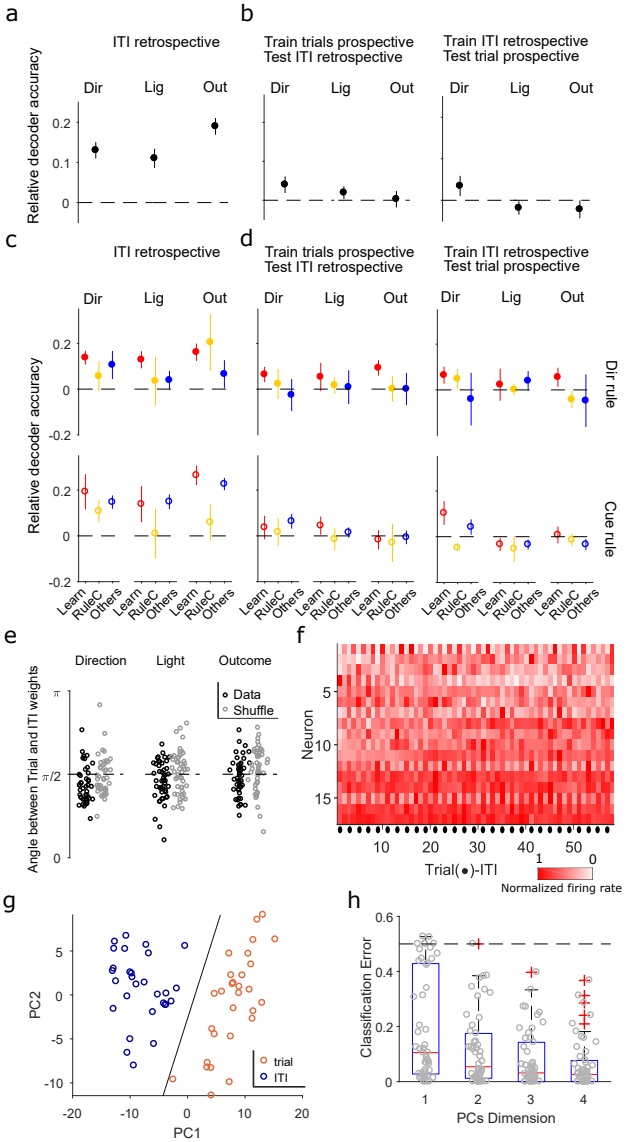

Figure 3: Orthogonal population encoding of past and present task events.
(**a**) Decoding accuracy for features of the preceding trial from inter-trial interval population activity over the whole maze. Each symbol gives mean ± SEM performance over all sessions.
(**b**) Cross-decoding performance. Left: mean ± SEM relative decoding accuracy of a logistic regression classifier trained on all trials prospectively and tested on all inter-trial intervals retrospectively. Black dashed line are the chance levels obtained training the classifier on shuffled labels for the trials and testing on inter-trial intervals given the same shuffled labels. Right: relative decoding accuracy for a classifier trained on all inter-trial intervals retrospectively and tested on all trials prospectively.
(**c**) Breakdown of the whole-maze decoding results in panel (a) by session and rule type. (**d**) Breakdowns of the whole-maze cross-decoding results in panel (b) by session and rule type. (**e**) Comparison of decoding vector weights between trials and inter-trial intervals. For each session we plot the angle between its trial and inter-trial interval decoding weight vectors, obtained from the trained classifiers in panel **b**. For reference, we also compute the angle between trial and inter-trial interval decoding vectors obtained by training on shuffled label data (grey circles).
(**f**) Population activity vectors for the trials (•) and following inter-trial intervals of one session. The heat-map shows the normalized firing rate for each neuron.
(**g**) Projection of that session's population activity vectors on to two dimensions shows a complete separation of trial and inter-trial interval activity. The black line is the linear separation found by the classifier. PC: principal component.
(**h**) Summary of classification error over all sessions, as a function of the number of dimensions. Each grey dot is the error for one session at that number of projecting dimensions. Dashed line gives chance performance. Boxplots show medians (red line), interquartile ranges (blue box), and outliers (red pluses).

decoding. Using leave-one-out cross-decoding instead (e.g. we leave out ith trial-ITI pair, train on $N-1$ trials, and predict the ith ITI), still results in decoding at or about chance levels (SI Fig 3a,b). These results suggest that population encoding of prior events in the inter-trial interval is not simply a memory trace of similar activity in the trial. Instead, they show that the same mPfC population is separately and independently encoding the present and past of the same features.

To quantify the coding independence, we turn to the vector of decoding weights for the trials and the equivalent vector for the inter-trial intervals of the same session. These weights, obtained from the decoder trained once on all trials and again on all inter-trial intervals, give the relative contribution of each neuron to the encoding of task features. We find the angles between the trial and inter-trial interval weight vectors cluster around $\pi/2$ (or, equivalently, their dot-product clusters around zero) (Fig 3e; see SI Fig 3e for a breakdown by session and rule types). Indeed, the distributions of angles between the trial and inter-trial interval vectors is well-captured by decoders trained on shuffled data (grey symbols in Fig 3e and SI Fig 3e). Moreover, we see no obvious sign here of the curse of dimensionality, with no clear dependence of the similarity between the decoding vectors on the dimensions of those vectors (SI Fig 3d). The population encoding in the inter-trial interval is then not a memory trace: to a good approximation, the past and present are orthogonally encoded in the same mPfC population

Independent encoding implies independent population activity between the trials and inter-trial intervals of a session (Fig 3f). If true, we should find that trial and inter-trial interval population activity vectors are easily separable. To test this, we project all population activity vectors of a session into the same low dimensional space (Fig 3g) and then quantify how easily we can separate them. Using just one dimension is sufficient for near-perfect separation in many of sessions; using two is sufficient for above chance performance in all sessions (Fig 3h; and see SI Fig 4 for a breakdown of each session's dependence on the number dimensions). Thus population activity in mPfC is independent between the trials and inter-trial intervals of this task.

Population representations of trial features re-activate in sleep

Strong evidence for true independent encoding of the past and present would be that the trial and inter-trial interval population codes are also functionally independent. One candidate for functional independence is that one population code is consolidated in memory, and the other is not. Prior reports showed patterns of mPfC population activity during training are preferentially repeated in post-training slow-wave sleep (Euston et al., 2007; Peyrache et al., 2009; Singh et al., 2019), consistent with a role in consolidation. However, what features these repeated patterns encode, and whether they encode their past or present, is unknown. Thus, we ask here which, if any, of the trial and inter-trial interval codes are re-activated in sleep. As a consequence we also examine here what is encoded in mPfC activity in sleep.

We first test whether population activity representations in trials re-activate more in post-training than pre-training sleep. For each feature of the task happening in the present (e.g choosing the left arm), we follow the decoding results by creating a population vector of the activity specific to that feature during a session's trials. To seek their appearance in slow-wave sleep, we compute population firing rate vectors in pre- and post-training slow-wave sleep in time bins of 1 second duration, and correlate each sleep vector with the feature-specific trial vector (Figure 4a). We thus obtain a distribution of correlations between the trial-vector and all pre-training sleep vectors, and a similar distribution between the trial-vector and all post-training sleep vectors. Greater correlation with post-training

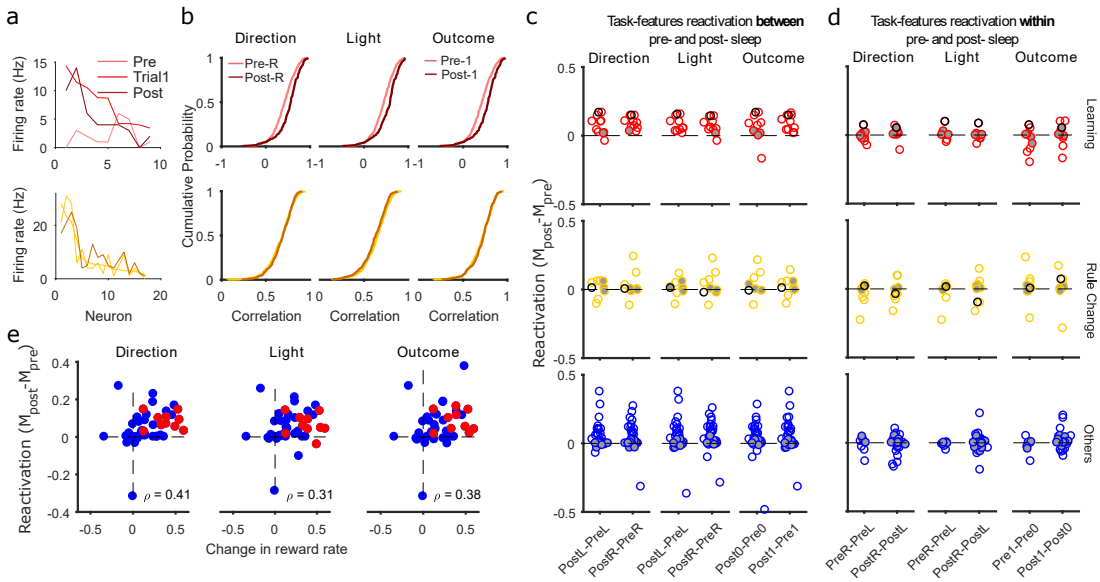

Figure 4: Reactivation of trial population coding in post-training sleep.
(**a**) Example population activity vectors. Upper panel: from one learning session, we plot the average firing rate vector for correct trials (Trial1). For comparison, we also plot examples of the one second time bin firing rate vectors from pre- and post-training slow-wave sleep. Neurons are ranked in order of their firing rates in the trial vector. Lower panel: as for the upper panel, for an example rule change session.
(**b**) Example distributions of Spearman's rank correlations between trial and sleep population activity. Upper panels: for the same learning session as panel (**a**), we plot the distributions of correlations between each vector of feature-specific trial activity and the population activity vectors in pre- and post-training slow-wave sleep. Lower panels: as for the upper panels, for the example rule change session in panel (**a**).
(**c**) Summary of reactivations across all sessions. For each feature, we plot the difference between the medians of the pre- and post-training correlation distributions (**b**). A difference greater than zero indicates greater correlation between trials and post-training sleep. Each symbol is a session. Empty symbols are sessions with significantly different correlation distributions at $p < 0.05$ (Kolmogorov-Smirnov test). Grey filled symbols are not significantly different. One black circle for learning and one for rule change sessions identify the two example sessions in panels (**a**) and (**b**).
(**d**) As for panel c, but plotting the median differences between distributions for paired features within the same sleep epoch. For example, in the left-most column, we plot the difference between the correlations with pre-session sleep activity for right-choice and left-choice specific trial vectors (PreR - PreL).
(**e**) Reactivation as a function of the change in reward rate in a session. One symbol per session: learning (red); "other" (blue). ρ: Spearman's correlation coefficient. We plot here reactivation of vectors corresponding to left (direction and light) or errors; correlations for other vectors are similar in magnitude: 0.41 (choose right); 0.38 (cue on right); 0.33 (correct trials).

sleep activity would then be evidence of preferential reactivation of feature-specific activity in post-training sleep.

We plot in Figure 4b (upper panels) a clear example of a learning session with preferential reactivation. For all trial features the distribution of correlations between the trial and post-training sleep population activity is right-shifted from the distribution for pre-training sleep. For example, the population activity vector for choosing the left arm is more correlated with activity vectors in post-training (Post-L) than pre-training (Pre-L) sleep.

Such post-training reactivation is not inevitable. In Figure 4b (lower panels) we plot an example rule-change session in which the trial-activity vector equally correlates with population activity in pre- and post-training sleep. Even though specific pairs of features (such as the left and right light positions) differ in their overall correlation between sleep and trial activity, no feature shows preferential reactivation in post-training sleep.

These examples are recapitulated across the data (Figure 4c). In learning sessions, feature-specific activity vectors are consistently more correlated with activity in post- than pre-training sleep. By contrast, sessions with rule changes show no consistent preferential reactivation of any feature vector in post-training sleep. As a control here for statistical artefacts in our re-activation analysis, we look for differences in reactivation between paired features (e.g. left versus right arm choice) within the same sleep epoch, and see these all centre on zero (Figure 4d). Thus population representations of task features in the present are re-activated in sleep, and this occurs following learning but not rule-change sessions.

To check whether reactivation is unique to step-like learning, we turn to the "other" sessions: there we find a wide distribution of preferential re-activation, from many about zero to a few reactivated nearly as strongly as in the learning sessions (Figure 4c, blue symbols). Indeed, when pooled with the learning sessions, we find re-activation of a feature vector in post-training sleep is correlated with the increase in accumulated reward during the session's trials (Fig 4e). Consequently, re-activation of population encoding during sleep may be directly linked to the preceding improvement in performance.

Prior reports suggest that the replay of activity patterns in sleep can be faster or slower during sleep than they were during waking activity. We test the time-scale dependence of feature-vector re-activation by varying the size of the bins used to create population vectors in sleep, with larger bins corresponding to slower re-activation. We find preferential reactivation in post-training sleep in learning and (some) "other" sessions is robust over orders of magnitude of vector widths (Figure 5), as is the absence of reactivation in rule-change sessions. This consistency across broad time-scales suggests that it is the changes during trials to the relative excitability of neurons within the mPfC population that are carried forward into sleep (Singh et al., 2019).

No re-activation in sleep of inter-trial interval feature representations

To ask if this re-activation is unique to encoding of the present, we repeat the same re-activation analysis for population vectors from the inter-trial interval. Again following our decoding results, each population feature vector is created from the average activity during inter-trial intervals after that feature (e.g. choose left) has occurred. We then check for reactivation of this feature vector in pre- and post-training slow-wave sleep.

We find no preferential reactivation of population encoding in post-training sleep, for any feature in any type of session. Figure 6a shows that the lower bounds of approximate 95% confidence intervals contained zero difference between pre- and post-training correlation distributions over two orders of magnitude for the time-scales of sleep vectors. Consistent with this, we find no correlation between the change in performance during a

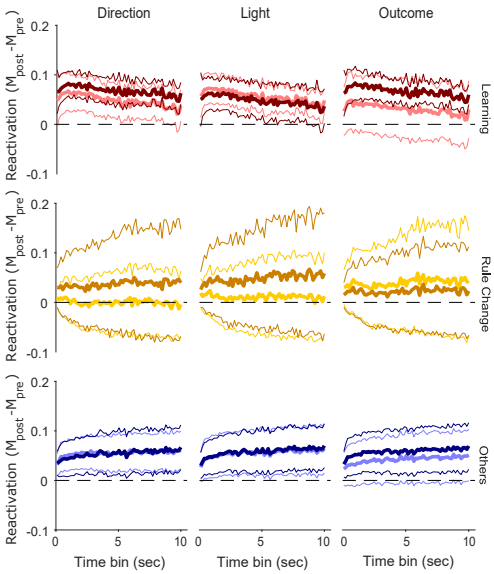

Figure 5: Robust reactivation of trial population coding across time-scales of sleep activity.
At each time bin used to construct population activity vectors in sleep, we plot the distribution over sessions of the median differences between pre- and post-training correlation distributions, for learning (top), rule-change (middle), and other (bottom) sessions. Distributions are plotted as the mean (thick lines) ± 2 SEM (thin lines); at the 1s bin, these summarise the distributions shown in full in Figure 4c. Each panel plots two distributions, one per pair of features: lighter colours indicate left or error trials; while darker colours indicate right or correct trials. Time bins range from 100 ms to 10 s, tested every 150 ms.

session and the re-activation of feature vectors after a session (Figure 6b). The orthogonal population encoding during sessions (Figure 3) thus appears functional: population encoding of features in the present is reactivated in sleep, but encoding of the same features until the start of the next trial is not.

Population synchrony is recalled in correct trials around learning

Our analyses thus far have shown evidence of independent coding of the past and the present in the same mPfC population. We now turn to consider what this independence means for the synchrony of activity in the population. We previously showed that these mPfC populations repeated similar patterns of population synchrony across inter-trial intervals, most reliably in the trials around the point of learning (Maggi et al., 2018). As this synchrony was triggered by reward, we suggested that reward acts as a trigger to synchronise the features of the past trial encoded within the population; in principle, this would be a powerful tool for learning, as the joint presence of those features is predictive of reward. We thus ask here if these populations have such a synchrony code in the trials too. We show that trials also have evidence of outcome-dependent synchrony in the population; in the next section we ask if these synchrony patterns are also re-activated in sleep.

We characterise the synchrony code by first computing the matrix of pairwise similarities between neurons on each trial (Figure 7a). We then compare matrices across trials to ask if the same pattern of synchrony is recalled on different trials (Figure 7b). In learning sessions, population synchrony is consistently more similar between trials with upcoming rewards than errors (Figure 7c), suggesting a recalled pattern of synchrony across the

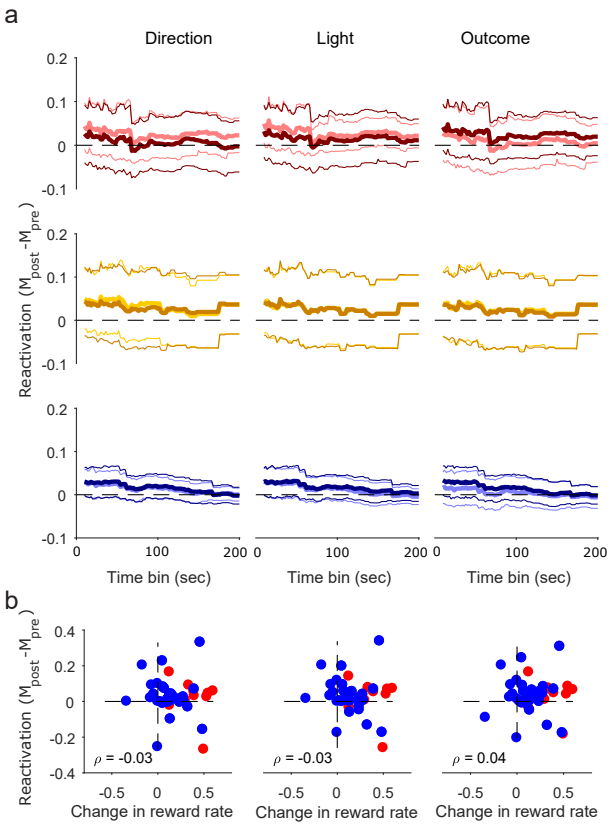

Figure 6: Absence of preferential reactivation of population encoding of the past.
(**a**) Similar to Figure 5, for reactivation of population feature-vectors constructed from inter-trial interval activity. We plot the distribution over sessions of the median differences between pre- and post-training correlation distributions, for learning (top), rule-change (middle), and other (bottom) sessions. Note that the range of sleep vector time-bins is an order of magnitude larger than for trials, as the inter-trial intervals themselves are an order of magnitude longer than trials (Maggi et al., 2018).
(**b**) Similar to Figure 4e, reactivation of the inter-trial interval population vector as a function of the change in reward rate in a session. Reactivation is computed for 22 s bins. One symbol per session: learning (red); "other" (blue). ρ: Spearman's correlation coefficient. We plot here reactivation of vectors corresponding to left (direction and light) or errors; correlations for other vectors are similar in magnitude: 0.03 (choose right); 0.01 (cue on right); -0.11 (correct trials).

population precedes correct choice on a trial.

This consistent recall of synchrony patterns across trials was specific to the combination of learning sessions and future outcomes. We find no consistent outcome-dependent recall of synchrony patterns in either rule-change or other sessions (Figure 7c). We observe no recall of synchrony patterns across trials that followed a correct trial (Figure 7d), ruling out a history-dependent effect of reward. We observe no consistently biased recall of synchrony patterns across trials conditioned on other task-related features, whether they were upcoming (Figure 7e,f) or in previous trials (SI Fig 5d,e). Finally, to ensure that the recall of synchrony patterns is not affected by the temporal precision of the spike-train correlation used, we repeat all these comparisons for different resolutions of the Gaussian width used to convolve spike-trains before computing the pair-wise similarity. Across an order of magnitude for the temporal resolution, we still observe recall effects only for learning sessions and only when preceding the outcome (SI Fig 5).

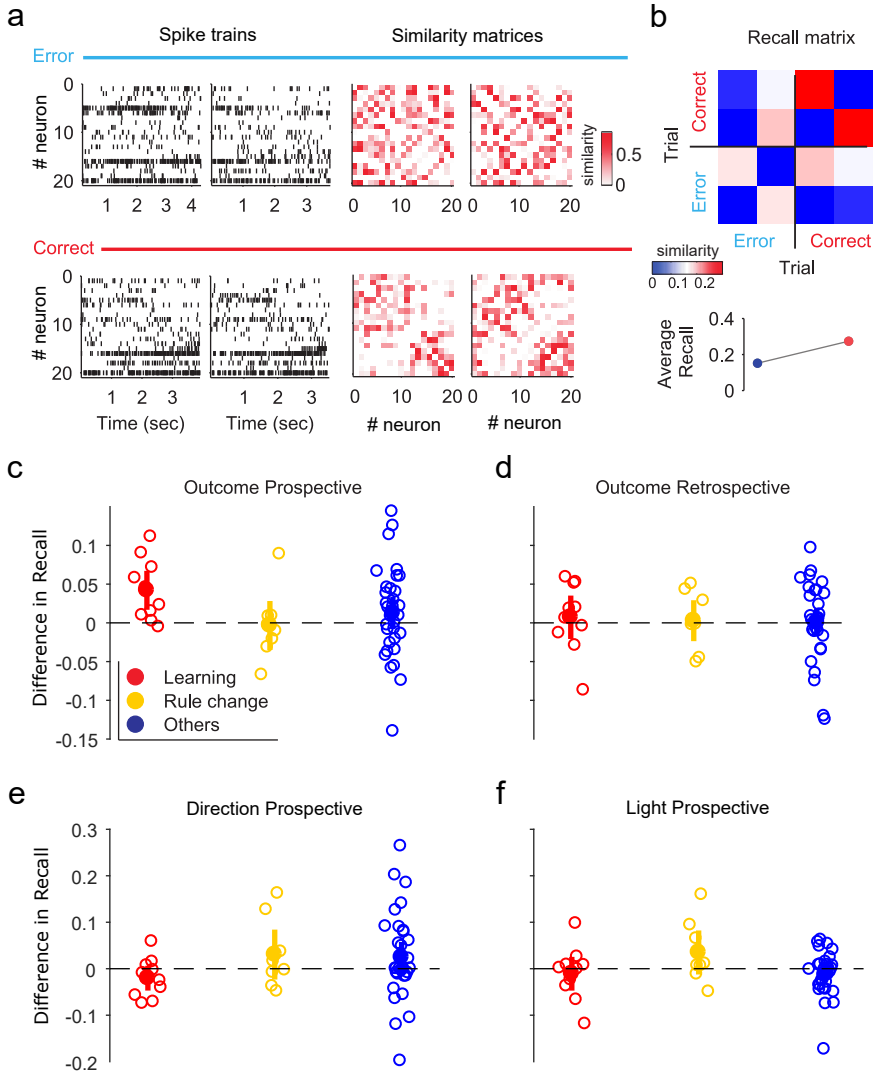

Figure 7: Population synchrony is repeated before rewards, but only in learning sessions.
(**a**) Example raster plots of spike trains during two error and two correct trials from the same session (left panels). Right panels: corresponding pairwise similarity matrices.
(**b**) Top, matrix of similarities between the example similarity matrices in panel (a). This "Recall" matrix shows higher values within correct trials compared to error trials, indicating that population synchrony was more similar between correct trials. Bottom: we summarise this by comparing the average recall values between error (blue) and correct (red) trials. (See SI Fig5a-c for further details of computing the Recall matrix)
(**c**) The difference in average recall between correct and error trials, across session types. Empty circles are single session values, while bold symbols are means ± 2SEM.
(**d**) As for (c), with the difference in recall computed with respect to the outcome of the preceding trial.
(**e**) As for (c), with the difference in recall computed with respect to the upcoming direction choice on each trial.
(**f**) As for (c), with the difference in recall computed with respect to the cue location on each trial.

Population synchrony is also independent between trials and inter-trial intervals

Paralleling the population's coding of features, both trials (present paper) and inter-trial intervals (Maggi et al., 2018) thus show outcome-dependent recall of population synchrony.

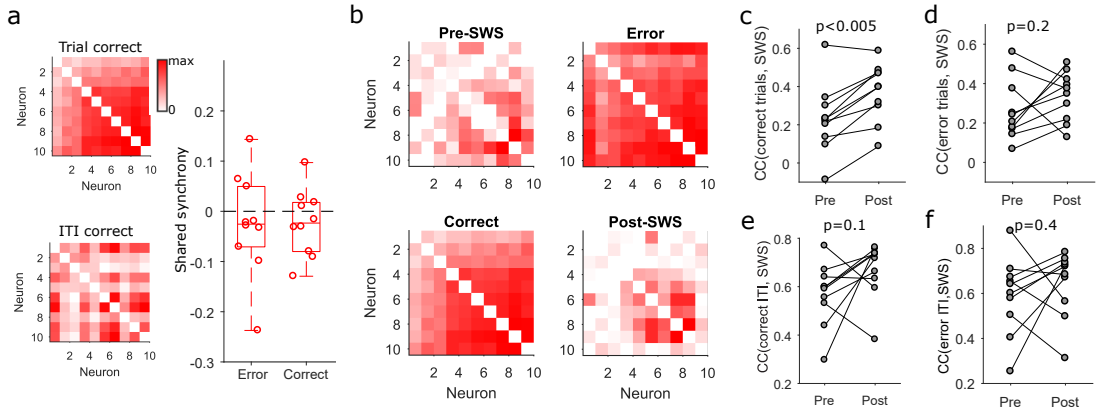

Figure 8: Population synchrony is also orthogonal between trials and inter-trial intervals.
(**a**) Example similarity matrices for the concatenated correct trials (upper panel) and concatenated correct ITIs (lower panel) on the left. On the right, distributions of relative similarity between ITIs and preceding trials compared to a shuffle control model. Distributions around zero indicate that trials and ITIs are independent (redrawn from Maggi et al. (2018)).
(**b**) Similarity matrices of concatenated pre-session SWS episodes, error trials, correct trials and post-session SWS episodes for an example learning session.
(**c**) Correlation between concatenated correct trials and pre-session SWS episodes compared to the correlation between concatenated correct trials and post-session SWS episodes. Each dot is a learning session. p-value from a paired T-test is shown on top.
(**d**) Correlation between concatenated error trials and pre-session SWS episodes compared to the correlation between concatenated error trials and post-session SWS episodes.
(**e**) As in panel (**c**) Correlation between concatenated correct ITIs and pre-session SWS episodes compared to the correlation between concatenated correct ITIs and post-session SWS episodes.
(**f**) As in panel (**d**) Correlation between concatenated error ITIs and pre-session SWS episodes compared to the correlation between concatenated error ITIs and post-session SWS episodes.

And both most consistently in sessions of learning. We thus ask if the trial and inter-trial interval synchrony patterns in learning sessions are also orthogonal: if they differ between a trial and the following inter-trial interval; and if they differ in their re-activation during sleep.

Within a learning session, the patterns of population synchrony in a trial and its following inter-trial interval are as different from each other as shuffled data (Fig 8a, redrawn from Maggi et al. (2018)). The synchrony between neurons in the population thus seems as independent between the task periods as the population encoding of past and present.

To address the re-activation of synchrony patterns in sleep, we first compute two session-wide synchrony patterns for each session, one from concatenating all correct trials; the other from concatenating all error trials (Figure 8b). We find the population synchrony in correct trials is a closer match to the synchrony in post- than pre-training slow-wave sleep (Figure 8c). Error trials show no preferential match of synchrony to either sleep epoch (Figure 8d). Repeating the same analysis for the inter-trial intervals, we find no preferential match of synchrony for either sleep epoch, neither for intervals following correct outcomes (Figure 8e), or those following error (Figure 8f). Outcome dependent population synchrony in trials is thus recapitulated in sleep independently of the population synchrony in inter-trial intervals.

Discussion

We have shown that the medial PfC population activity independently represents the past and present of the same task features. Three lines of evidence support this conclusion. First, the same task feature, such as the choice of arm, is orthogonally encoded in the trials and the inter-trial intervals. Second, these orthogonal encodings are functionally independent: population activity representations of features during the trials are re-activated in post-training sleep, but inter-trial interval representations are not. Third, the reward-evoked pattern of population synchrony is also independent between trials and inter-trial intervals, and only the trials' synchrony pattern is recapitulated in post-training sleep.

Mixed population coding in mPfC

Consistent with prior reports of mixed or multiplexed coding by single mPfC neurons (Rigotti et al., 2013; Fusi et al., 2016; Jung et al., 1998; Horst and Laubach, 2012), we find here that mPfC populations can sustain mixed encoding of two or more of the current direction choice, light position, and anticipated outcome. The extent of mixed encoding depends on the enforced rule and the learning status of the animal. These encodings are also position-dependent. Encoding of direction choice reliably occurs from the mazes choice point onwards, but it is unclear whether this represents a causal role in the choice itself, or an ongoing representation of a choice being made.

We also report that these mixed encodings of the present within each population reactivate in post-training sleep. This finding goes beyond prior reports that specific patterns of trial activity replay in sleep (Euston et al., 2007; Peyrache et al., 2009; Singh et al., 2019) to indicate the semantic content of sleep activity. It seems mixed encoding is a feature of sleep too.

As we showed in (Maggi et al., 2018), population activity during the inter-trial interval also has mixed encoding of features of the past. This encoding decays in strength as the rat gets closer to the start of the maze: we show here this mixed encoding is still evident even if we disregard maze positions. Collectively, our results show that population activity switches from mixed encoding of the present in a trial to mixed encoding of the past in the following inter-trial interval.

Independent mixed population codes

Such mixed encoding implies a distributed representation across a network of neurons (Mante et al., 2013). As we noted in Results, switching from a mixed encoding of the present to the past is most simply explained by the hypothesis that network-wide activity in the trials is simply carried forward into the inter-trial interval. The encoding of the past is then an activity trace, that functions as a memory. But our demonstration of orthogonal encoding in the population between trials and the following intervals rules out this hypothesis.

Finding orthogonal decoding vectors in a single group of neurons implies there are at least two distinct population codes within the same network. However, orthogonal decoding vectors do not mean there are distinct groups of active neurons. We ruled out this interpretation by only decoding from neurons active in every trial or every inter-trial interval. Rather, large magnitude decoding weights identify neurons within a population who vary their rates most consistently with the change in a feature (e.g. choice of direction). So finding these orthogonal decoding vectors implies one of two encoding mechanisms. Either the group of neurons with feature-dependent variations in rate during trials overlap little,

if at all, with the group of neurons doing the same in the inter-trial intervals. Or, perhaps less likely, the same group of neurons inverts its encoding between trials and inter-trial intervals (for example, neurons encoding "left" in trials encode "right" in the inter-trial intervals). Either way, there are at least two independent population codes within the same network.

An open question here is the extent to which this clean independence between the encoding of the past and present is dependent on the behavioural task. In the Y-maze task design, there is a qualitative distinction between trials (with a forced choice) and inter-trial intervals (with a self-paced return to the start arm), which may support clearly separate roles for encoding of the present and the past. Such independent coding may be harder to uncover in tasks without a distinct separation of decision and non-decision phases. For example, tasks where the future choice of arm depends on recent history, such as double-ended T-mazes (Jones and Wilson, 2005), multi-arm sequence mazes (Poucet et al., 1991), or delayed non-match to place (Spellman et al., 2015), blur the separation of the present and the past. Comparing population-level decoding of the past and present in such tasks would give useful insights into when the two are, and are not, independently coded

Mechanisms for rapid switching of population codes

The trial and inter-trial interval are contiguous in time. So the orthogonal encoding in the population implies a rapid switch from one encoding to another. How might such a rapid switch of network-wide encoding be achieved?

Such rapid switching in the state of a network suggests a switch in the driver inputs to the network. In this model, drive from one source input creates the network states for population encoding A; a change of drive – from another source, or a qualitative change from the same source - creates the network states for population encoding B (either set of states may of course arise solely from internal dynamics). One option for a switching drive is the hippocampal-prefrontal pathway.

Previous work (Benchenane et al., 2010; Peyrache et al., 2009) showed that learning correlated with increased cortico-hippocampal coherence at the choice point of this Y maze. This coherence recurred during slow-wave ripples in post-training sleep. These data and our analyses here are consistent with a picture where the population encoding of the trials is (partly) driven by hippocampal input; and the re-activation of only the trial representations in sleep are explained as the recruitment of those states by hippocampal input during slow-wave sleep. Moreover, the increased coherence between hippocampus and mPfC activity may act as a window for synaptic plasticity of that pathway (Benchenane et al., 2010, 2011). Consistent with this, we see a correlation between performance improvement in trials and reactivation in sleep (Maingret et al., 2016).

All of which suggests the encoding of the past during the inter-trial interval is not driven by the hippocampal input to mPfC, as its representation is not re-activated in sleep. (Spellman et al. (2015) report hippocampal input to mPfC is necessary for the maintenance of a cue location; though unlike in our task, actively maintaining the location of this cue was necessary for a later direction decision). Rather, the population coding during the inter-trial interval could reflect the internal dynamics of the mPfC circuit. Indeed, network models of working memory in the prefrontal cortex focus on attractor states created by its local network (Durstewitz et al., 2000; Miller et al., 2005; A. Compte, 2000; Wimmer et al., 2014). If somewhere close to the truth, this account of rapid switching suggests that the hippocampal input to mPfC drives population activity in the trial, and a change or reduction in that input allows the mPfC local circuits to create a different internal state:

and thus a switch from population coding A to coding B. A clear prediction of this account is that perturbation of the hippocampal input to the PFC could disrupt its encoding of the past and present in different ways.

Reconciling mPfC roles in memory and choices

We previously showed that the reward-driven pattern of population synchrony in the inter-trial interval appears immediately before the learning trial of a session (Maggi et al., 2018). But we show here that neither the encoding nor synchrony pattern in the inter-trial interval are carried forward long-term into sleep. These new results add further support to our hypothesis in (Maggi et al., 2018) that the role of memory encoding in the inter-trial interval is to guide learning online: reward tags past features whose conjunction led to successful outcomes (e.g. to turn left when the light is on in the left arm).

By contrast, representations of the present in the trial are carried forward and reactivated in sleep. This reactivation may be consolidating the conjunction of present features and choice that is going to be successful when re-used in future.

Further insight into these and other ideas here would come from stable recordings of the same population across multiple sessions, to track how encoding of the past and present evolves and is or is not reused. In particular, it would be insightful to establish if re-activated trial representations in sleep reappear in subsequent sessions.

The medial prefrontal cortex plays a key role in both short-term memory (Maggi et al., 2018; Fujisawa et al., 2008; Jun et al., 2010) and choice behaviour (Rigotti et al., 2013; Averbeck et al., 2006; Erlich et al., 2015; Hanks et al., 2015). Our finding here of independent coding of the past and the present suggest these roles can be carried out sequentially within the same mPfC neural population.

Acknowledgments

We thank Adrien Peyrache for the data, discussions, and comments on the manuscript, and the Humphries' lab (Abhinav Singh, Javier Caballero, Mat Evans) for discussion. M.D.H and S.M. were supported by a Medical Research Council Senior non-Clinical Fellowship award (MR/J008648/1) and MRC Project Grant MR/P005659/1. The original data collection was supported by the EU Framework (FP6) "ICEA" grant.

Author Contributions

M.D.H and S.M. designed the analyses. S.M. analysed the data. M.D.H and S.M. wrote the manuscript.

Declaration of Interest

The authors declare no conflicts of interest.

References

A. Compte, N. Brunel, P. G.-R. . X. W. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb Cortex*, 10:910–923.

Averbeck, B. B. and Lee, D. (2007). Prefrontal neural correlates of memory for sequences. *J Neurosci*, 27:2204–2211.

Averbeck, B. B., Sohn, J.-W., and Lee, D. (2006). Activity in prefrontal cortex during dynamic selection of action sequences. *Nat Neurosci*, 9:276–282.

Baeg, E. H., Kim, Y. B., Huh, K., Mook-Jung, I., Kim, H. T., and Jung, M. W. (2003). Dynamics of population code for working memory in the prefrontal cortex. *Neuron*, 40(1):177–188.

Benchenane, K., Peyrache, A., Khamassi, M., Tierney, P. L., Gioanni, Y., Battaglia, F. P., and Wiener, S. I. (2010). Coherent theta oscillations and reorganization of spike timing in the hippocampal- prefrontal network upon learning. *Neuron*, 66(6):921–936.

Benchenane, K., Tiesinga, P. H., and Battaglia, F. P. (2011). Oscillations in the prefrontal cortex: a gateway to memory and attention. *Curr Opin Neurobiol*, 21(3):475–485.

Durstewitz, D., Seamans, J. K., and Sejnowski, T. J. (2000). Neurocomputational models of working memory. *Nature Neurosience*, 3:1184–1191.

Durstewitz, D., Vittoz, N. M., Floresco, S. B., and Seamans, J. K. (2010). Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron*, 66(3):438–448.

Erlich, J. C., Brunton, B. W., Duan, C. A., Hanks, T. D., and Brody, C. D. (2015). Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat. *eLife*, 4.

Euston, D. R., Gruber, A. J., and McNaughton, B. L. (2012). The role of medial prefrontal cortex in memory and decision making. *Neuron*, 76(6):1057–1070.

Euston, D. R., Tatsuno, M., and McNaughton, B. L. (2007). Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. *Science*, 318(5853):1147–1150.

Fujisawa, S., Amarasingham, A., Harrison, M. T., and Buzsaki, G. (2008). Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nat Neurosci*, 11(7):823–833.

Fusi, S., Miller, E. K., and Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. *Curr Opin Neurobiol*, 37:66–74.

Guise, K. G. and Shapiro, M. L. (2017). Medial prefrontal cortex reduces memory interference by modifying hippocampal encoding. *Neuron*, 94:183–192.e8.

Hanks, T. D., Kopec, C. D., Brunton, B. W., Duan, C. A., Erlich, J. C., and Brody, C. D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature*, 520:220–223.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer, Berlin.

Horst, N. K. and Laubach, M. (2012). Working with memory: evidence for a role for the medial prefrontal cortex in performance monitoring during spatial delayed alternation. *Journal of neurophysiology*, 108(12):3276–3288.

Ito, H. T., Zhang, S.-J., Witter, M. P., Moser, E. I., and Moser, M.-B. (2015). A prefrontal–thalamo–hippocampal circuit for goal-directed spatial navigation. *Nature*, 522(7554):50.

Jones, M. W. and Wilson, M. A. (2005). Theta rhythms coordinate hippocampal-prefrontal interactions in a spatial memory task. *PLoS Biol*, 3(12):e402.

Jun, J. K., Miller, P., Hernndez, A., Zainos, A., Lemus, L., Brody, C. D., and Romo, R. (2010). Heterogenous population coding of a short-term memory and decision task. *J Neurosci*, 30(3):916–929.

Jung, M. W., Qin, Y., McNaughton, B. L., and Barnes, C. A. (1998). Firing characteristics of deep layer neurons in prefrontal cortex in rats performing spatial working memory tasks. *Cereb Cortex*, 8(5):437–450.

Karlsson, M. P., Tervo, D. G., and Karpova, A. Y. (2012). Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. *Science*, 338(6103):135–139.

Laskowski, C. S., Williams, R. J., Martens, K. M., Gruber, A. J., Fisher, K. G., and Euston, D. R. (2016). The role of the medial prefrontal cortex in updating reward value and avoiding perseveration. *Behavioural Brain Research*, 306:52–63.

Laubach, M., Caetano, M. S., and Narayanan, N. S. (2015). Mistakes were made: neural mechanisms for the adaptive control of action initiation by the medial prefrontal cortex. *Journal of Physiology-Paris*, 109(1):104–117.

Machens, C. K., Romo, R., and Brody, C. D. (2010). Functional, but not anatomical, separation of "what" and "when" in prefrontal cortex. *J Neurosci*, 30:350–360.

Maggi, S., Peyrache, A., and Humphries, M. D. (2018). An ensemble code in medial prefrontal cortex links prior events to outcomes during learning. *Nature Communications*, 9.

Maingret, N., Girardeau, G., Todorova, R., Goutierre, M., and Zugaro, M. (2016). Hippocampo-cortical coupling mediates memory consolidation during sleep. *Nature Neuroscience*, 19:959–964.

Mante, V., Sussillo, D., Shenoy, K. V., and Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503:78–84.

Miller, E. K. and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, 24(1):167–202.

Miller, P., Brody, C. D., Romo, R., and Wang, X.-J. (2005). A recurrent network model of somatosensory parametric working memory in the prefrontal cortex. *Cereb Cortex*, 15:679–679.

Narayanan, N. S. and Laubach, M. (2008). Neuronal correlates of post-error slowing in the rat dorsomedial prefrontal cortex. *Journal of neurophysiology*, 100(1):520–525.

Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S. I., and Battaglia, F. P. (2009). Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nature Neuroscience*, 12:919–926.

Poucet, B., Lucchessi, H., and Thinus-Blanc, C. (1991). What information is used by rats to update choices in the radial-arm maze? *Behav Processes*, 25:15–26.

Powell, N. J. and Redish, A. D. (2016). Representational changes of latent strategies in rat medial prefrontal cortex precede changes in behaviour. *Nature Communications*, 7.

Rich, E. L. and Shapiro, M. (2009). Rat prefrontal cortical neurons selectively code strategy switches. *Journal of Neuroscience*, 29(22):7208–7219.

Rich, E. L. and Shapiro, M. L. (2007). Prelimbic/infralimbic inactivation impairs memory for multiple task switches, but not flexible selection of familiar tasks. *J Neurosci*, 27(17):4747–4755.

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497(7451):585–590.

Schuck, N., Gaschler, R., Wenke, D., Heinzle, J., Frensch, P., Haynes, J.-D., and Reverberi, C. (2015). Medial prefrontal cortex predicts internally driven strategy shifts. *Neuron*, 86:331–340.

Singh, A., Peyrache, A., and Humphries, M. D. (2019). Medial prefrontal cortex population activity is plastic irrespective of learning. *Journal of Neuroscience*, pages 1370–17.

Spellman, T., Rigotti, M., Ahmari, S. E., Fusi, S., Gogos, J. A., and Gordon, J. A. (2015). Hippocampal-prefrontal input supports spatial encoding in working memory. *Nature*, 522(7556):309–314.

Wimmer, K., Nykamp, D. Q., Constantinidis, C., and Compte, A. (2014). Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature Neuroscience*, 17:431–439.

Young, J. J. and Shapiro, M. L. (2009). Double dissociation and hierarchical organization of strategy switches and reversals in the rat pfc. *Behavioral Neuroscience*, 123:1028–1035.

Methods

Task description and electrophysiological data

All the data in this study comes from previously published data (Peyrache et al., 2009). The full details of training, spike-sorting and histology can be found in (Peyrache et al., 2009). The experiments were carried out in accordance with institutional (CNRS Comité Opérationnel pour l'Ethique dans les Sciences de la Vie) and international (US National Institute of Health guidelines) standards and legal regulations (Certificate no. 7186, French Ministère de l'Agriculture et de la Pêche) regarding the use and care of animals.

Four Long-Evans male rats were implanted with tetrodes in prelimbic cortex and trained on a Y-maze task (Figure 1a). During each session, prelimbic activity was recorded for 20-30 minutes of sleep or rest epoch before the training phase, in which rats worked at the task for 20-40 minutes. After that, another 20-30 minutes of sleep or rest epoch recording followed. During the sleep epochs, intervals of slow-wave sleep were identified offline from local field potential (details in ref (Peyrache et al., 2009)).

The Y-maze had symmetrical arms, 85 cm long, 8 cm wide, and separated by 120 degrees, connected to a central circular platform (denoted as the choice point throughout). Each rat worked at the task phase by self-initiating the trial, leaving the beginning of the start arm. Trial finished when the rat reached the end of the chosen goal arm. If the chosen arm was correct according to the current rule, the rat was rewarded with drops of

flavoured milk. As soon as the animal reached the end of the chosen arm an inter-trial interval started and lasted until the rat completed its self-paced return to the beginning of the start arm.

Each rat was exposed to the task completely naïve and had to learn the rule by trial-and-error. The rules were presented in sequence: go to the right arm; go to the cued arm; go to the left arm; go to the uncued arm. The light cues at the end of the two arms were lit in a pseudo-random sequence across trials, regardless the rule in place.

The recording sessions taken from the study of Peyrache and colleagues (Peyrache et al., 2009) were 53 in total. Each of the four rats learnt at least two rules, and they contributed with 14, 14, 11, and 14 sessions. The learning, rule change, and "other" sessions for each rat were intermingled. We used 50 sessions for most of the analysis. One session was omitted for missing position data, one for consistent choice of the right arm (in a dark arm rule) preventing decoder analyses (see below), and one for missing spike data in a few trials. An additional session has been excluded from Figure 2 ahead for having only two neurons firing in all trials. Tetrode recordings were spike-sorted within each recording session. In the sessions we analysed here, the populations ranged in size from 4-25 units. Spikes were recorded with a resolution of 0.1 ms. Simultaneous tracking of the rat's position was recorded at 30 Hz.

Behavioural analysis

Each session was classified according to its behavioural features. The learning sessions were identified according to the original study (Peyrache et al., 2009) as the ones with three consecutive correct trials followed by a performance of at least 80% correct. The first of the three correct trials was the learning trial. Only ten sessions satisfied this criterion. We quantified this learning as a step-like change in performance by fitting a robust regression line to the cumulative reward curve before and after the learning trial. The slopes of the two lines gave us the rate of reward accumulation before (r_{before}) and after (r_{after}) the learning trial.

Eight sessions were characterised by 10 consecutive correct trials or eleven correct out of twelve trials followed by a change in the rule. The first trial with the new rule was identified as the rule change trial. The change in performance in these sessions, labelled "Rule change" sessions, was quantified with the same method above. A robust regression line was fitted to the cumulative reward curve before and after the rule change trial.

To identify other possible learning session, we fitted this piece-wise linear regression model to each trial in turn (allowing a minimum of 5 trials before and after each tested trial). We then found the trial at which the increase in slope ($r_{after} - r_{before}$) was maximised, indicating the point of steepest inflection in the cumulative reward curve. In the learning sessions, the learning trial largely agreed with this method. Amongst the remaining sessions, labelled "Other", we searched for signs of incremental learning using this method. We found 22 sessions falling in this category in addition to the 10 learning sessions. We called these new sessions "minor-learning".

Decoder analysis and independence of ensembles encoding

To predict which task feature was encoded into the ensemble we trained and tested a range of linear decoders (Hastie et al., 2009; Maggi et al., 2018). In the main text we report the results obtained using a logistic regression classifier, but for robustness we also tested three other decoders: linear discriminant analysis; (linear) support vector machines; and

a nearest neighbours classifier. The full details of the decoding analysis can be found in Maggi et al. (2018).

Briefly, we linearised the maze in five equally-sized sections and we computed the firing rate vector of the core population of length N for each position p, $\mathbf{r^p}$. For each trial $t = 1, \ldots, T$ and each section of the maze $p = 1, \ldots, 5$, the set of population firing rate vectors $\{\mathbf{r^p}(1), \ldots, \mathbf{r^p}(T)\}$ was used to train the decoder. Each relevant trials' task information was binary labelled: outcome (labels: 0, 1), the chosen arm (labels: left, right) and the position of the light cue (labels: left, right). The same classifier was also trained to decode previous trials' task information. We used leave-one-out cross-validation and we quantified the accuracy of the decoder as the proportion of correctly predicted labels over the T held out trial intervals.

For each rat and each session, the distribution of outcomes and arm choices depended on the rats' performance, which could differ from 50%. Therefore, we trained and cross-validated the same classifier on the same data-sets, but shuffling the labels of the task features. In this way we obtained the accuracy of detecting the right labels by chance. We repeated the shuffling and fitting 50 times and we averaged the accuracy across the 50 repetitions. The results displayed in the figures showed the difference between the decoder accuracy on the original data with the accuracy of the shuffled label.

To compare the decoding accuracy between trials and ITIs, we trained again the classifier using the population firing rate vectors computed on the entire maze $\{\mathbf{r}(1), \ldots, \mathbf{r}(T)\}$. We then trained the classifier on all the trials. We saved the population vector of weights and we tested the model, optimised to decode trial activity, on every ITI to evaluate the accuracy in decoding retrospective ITI labels. The same procedure was used to train the linear classifier on all the ITI to test its accuracy in decoding trials activity. The population vector of weight was also saved for this model.

The angle, θ, between the population vector of trials', w_t, and ITIs', w_I, weights was computed as $\theta = \cos^{-1}\left(\frac{w_t \cdot w_I}{\|w_t\|\|w_I\|}\right)$.

We further evaluated the independence of trial and ITI population vectors by quantifying their separability in a low dimensional space. We used principal components analysis (PCA) to project the population vectors of a session onto a common set of dimensions. To do so, we constructed the data matrix $\mathbf{X}$ from the firing rate vectors of the core population, by concatenating trials and ITIs in their temporal order $\{\mathbf{r}_t(1), \mathbf{r}_I(1), \ldots, \mathbf{r}_t(T), \mathbf{r}_I(T)\}^{\mathrm{T}}$; the resulting matrix thus had dimensions of $2T$ rows and N (neurons) columns. Applying PCA to X, we projected the firing rate vectors on to the top d principal axes (eigenvectors of $\mathbf{X^T X}$) to create the top d principal components. For each set of d components, we quantified the separation between the projected trial and ITI population vectors using a linear classifier (Support Vector Machine, SVM), and report the proportion of misclassified vectors. We repeated this for between $d = 1$ and $d = 4$ axes for each session.

Replay of task-features in sleep

In order to quantify the replay activity in pre- and post-session sleep, we used the population firing rate vectors computed for the decoder. We considered here the average population vector computed across all the trials in the all maze for each feature. In details, we quantified separately the average population firing rate vector for all the right choice trials and all the left choice trials. Similarly we did for right and left cue location and for correct and error trials. We then compare the ranked average population firing rate vector for each feature with the firing rate vector of each 1 second time bin of SWS pre- and post-session. We used Spearman's correlation coefficient to compare them and

to quantify the difference between the distributions of each feature and the SWS pre- and post-session. Spearman's coefficient was chosen specifically to remove any effects of global rate variations across the vectors within or between epochs. In order to have a replay of activity in post-session sleep, we expected the distribution of Spearman CC between a feature and pre-session SWS to be leftward shifted compare to the distribution of Spearman CC between the same feature and post-session SWS. We quantified this shift by measuring the difference in the medians distribution between the two Spearman distributions. If the delta medians was positive then we had a higher correlation of the population firing vector with the post-session SWS compared to the pre-session SWS. If the delta median (called Reactivation ($M_{post} - M_{pre}$) in the text) was negative, then the population firing rate vector was more similar to the pre-session SWS population vector. To then control for different time scales of replaying in sleep we repeated the same procedure changing the time bin in the SWS pre- and post-session. We used time bins from 100 ms to 10 sec every 150 ms for trials and from 10 sec to 200 sec every 2 sec for ITIs.

Testing for reinforcement-driven ensembles

The reinforcement-dependent recall of ensemble activity was identified as per Maggi et al. (2018). We firstly selected the spike-trains of the N neurons active in every trials. We convolved these spike-trains with a Gaussian ($\sigma = 100$ ms) to obtain a spike-density function f_k for the kth spike-train. All the recall analyses were repeated for different Gaussian widths ranging from 20 ms to 240 ms (SI Fig. 6). Each spike-train was then z-scored to obtain a normalised spike-density function f^* of unit variance: $f_k^* = (f_k - \langle f_k \rangle)/\sigma_k$, where $\langle f_k \rangle$ is the mean of f_k, and σ_k its standard deviation, taken over all the trials of a session.

We then wanted to track the changes in the co-activity pattern of the core population along the sessions. We first computed the pairwise similarity matrix between the spike density functions of each neuron in trial t, $\mathbf{S_t}$. This matrix was rectified in order to keep track only of those pair with positive co-activity pattern (Figure 7a). To compare the co-activity patterns across trials along the sessions, we then computed the recall matrix (Figure 7b, $\mathbf{R}$, where each entry $R(t, u)$) is the rectify correlation coefficient between the similarity matrices $\mathbf{S_t}$ and $\mathbf{S_u}$.

We grouped the entries of $\mathbf{R}$ into two groups according to the outcome of the trials. In such way we obtained two block diagonals $\mathbf{R_1}$ and $\mathbf{R_2}$ (such as $\mathbf{R_{error}}$ and $\mathbf{R_{correct}}$, as illustrated in Figure 7b). We summarised the recall between groups by computing the mean of each block.

In the main text, we report that there is higher average similarity in $\mathbf{R_{correct}}$ than $\mathbf{R_{error}}$ in the learning sessions. However, as a control and to keep comparison with the study of inter-trial intervals (Maggi et al., 2018) we defined a null model to dissect the contribution in the correlation of the different trials durations. For each session we created a predicted recall matrix $\hat{\mathbf{R}}$, by averaging 1000 random recall matrices, each computed from shuffled spike trains. Each spike-train was shuffled by randomly re-ordering its inter-spike intervals. This shuffling was meant to destroy any specific temporal pattern of the spike train, allowing to quantify the pairwise similarity contribution due exclusively to the duration of trial. Our final residual recall matrix $\tilde{\mathbf{R}} = \mathbf{R} - \hat{\mathbf{R}}$ was obtained as the difference between the Recall matrix and the average shuffled recall matrix (SI Fig 5a,b). The results shown in the main text summarise the differences between the two groups ($\mathbf{R_{correct}}$ and $\mathbf{R_{error}}$) as detailed above.

Data Availability

The spike-train and behavioural data that support the findings of this study are available in CRCNS.org (DOI: 10.6080/K0KH0KH5) (ref. (Peyrache et al., 2009)). Code to reproduce the main results of the paper is available at: