Complementary task representations in hippocampus and prefrontal cortex for generalising the structure of problems

Veronika Samborska[1*], James L Butler[2], Mark E Walton[1,3] & Timothy EJ Behrens[1,4,5†*], Thomas Akam[1,3†]

[1] Wellcome Centre for Integrative Neuroimaging, University of Oxford, UK

[2] Sobell Department of Motor Neuroscience, University College London, UK

[3] Department of Experimental Psychology, University of Oxford, UK

[4] Wellcome Centre for Human Neuroimaging, University College London, UK

[5] Sainsbury Wellcome Centre for Neural Circuits and Behaviour, University College London, UK

[*] Correspondence to : veronika.samborska@ndcn.ox.ac.uk, behrens@fmrib.ox.ac.uk

[†] these authors contributed equally to this work

Abstract

Few situations in life are completely novel. We effortlessly generalise prior knowledge to solve novel problems, abstracting common structure and mapping it onto new sensorimotor specifics. Here we trained mice on a series of reversal learning tasks that shared the same structure but had different physical implementations. Performance improved across tasks, demonstrating transfer of knowledge. Neurons in medial prefrontal cortex (mPFC) maintained similar representations across multiple tasks, despite their different sensorimotor correlates, whereas hippocampal (dCA1) representations were more strongly influenced by the specifics of each task. Critically, this was true both for representations of the events that comprised each trial, and those that integrated choices and outcomes over multiple trials to guide subjects' decisions. These data suggest that PFC and hippocampus play complementary roles in generalisation of knowledge, with the former abstracting the common structure among related tasks, and the latter mapping this structure onto the specifics of the current situation.

1 Introduction

When we walk into a new restaurant, we know what to do. We might find a table and wait to be served. We know that the starter will come before the main, and when the bill arrives, we know it is the food we are paying for. This is possible because we already know a lot about how restaurants work, and only have to map this knowledge onto the specifics on the new situation. This requires that the common structure is abstracted away from the sensorimotor specifics of experience, so it can be applied seamlessly to new but related situations.

Such abstraction has been variously described as a schema (in the context of human behaviour[1] and memory research[2,3]), learning set[4] (in the context of animal reward-guided behaviour), transfer learning[5] and meta-learning[6] (in the context of machine learning).

We have little understanding of how the necessary abstraction is achieved in the brain, or how abstract representations are tied to the sensorimotor specifics of each new situation. However, recent data suggest that interactions between frontal cortex and the hippocampal formation play an important role. Neurons[7,8] and fMRI voxels[9,10] in these brain regions form representations that generalise over different sensorimotor examples of tasks with the same structure, and track different task rules embedded in otherwise similar sensory experience[11,12].

The involvement of these regions in abstraction is also of interest from a theoretical perspective. Both frontal cortex[13–16] and hippocampus[17–25] have been hypothesized to represent task states and the relationships between them. It has not been clear what distinguishes the representations in these regions, but some insight might be gained by considering hippocampal representations underlying spatial cognition. In rodent hippocampus, place cells are specific to each particular environment[26–28], but firing patterns in neighbouring entorhinal cortex (including grid cells) generalise across different environments – that is, they are abstracted from sensorimotor particularities[29–33]. Similarly, there is evidence that mPFC representations of spatial tasks generalise across different paths[34–36].

One possibility is that, as in space, abstracted or schematic representations of tasks in cortex might be flexibly linked with the sensorimotor characteristics of a particular environment to rapidly construct concrete task representations in hippocampus, affording immediate inferences[37,38]. Indeed, hippocampal manipulations appear particularly disruptive when new task rules must be inferred, either at the beginning of training[39] or when task contingencies change[40,41].

To probe cortical and hippocampal contributions to generalisation, we developed a novel behavioural paradigm where we presented mice with a series of tasks with the same abstract structure (probabilistic reversal learning), but different physical instantiations, and hence different sensorimotor correlates. We recorded single units in medial prefrontal cortex (mPFC) and hippocampus (dCA1) across multiple physical task layouts in each recording session. We examined neuronal representations both of the individual elements of each trial, and of the cross-trial learning that controlled animal's choices. Prefrontal representations generalised across tasks, with neurons coding for a given task event, irrespective of the sensorimotor particulars of the current task. In contrast, hippocampal neurons were more task specific - different neuronal populations participated in each task representation. Both hippocampus and prefrontal cortex also contained representations of animals' current policy that integrated events over multiple trials. These policy representations were again abstract in prefrontal cortex but tied to sensorimotor specifics in hippocampus.

2 Results

2.1 Mice generalise knowledge between structurally equivalent tasks

Subjects serially performed a set of reversal learning tasks which shared the same structure but had different physical layouts. In each task, every trial started with an 'initiation' nose-poke port lighting up. Poking this port illuminated two 'choice' ports, which the subject chose between for a probabilistic reward (Figure 1A). Once the subject consistently (75 % of trials) chose the high reward probability port, reward contingencies reversed (Figure 1B). Once subjects completed ten reversals on a given port layout (termed a 'task'), they were moved onto a new task where the initiation and choice ports were in different physical locations (Figure 1C). All tasks therefore shared the same trial structure (initiate in the illuminated poke, then choose between the two illuminated pokes) and a common abstract rule (one port has high and one low reward probability, with occasional reversals), but required different motor actions due to the different port locations. In this phase of the experiment, task switches occurred between sessions, and subjects completed ten different tasks.

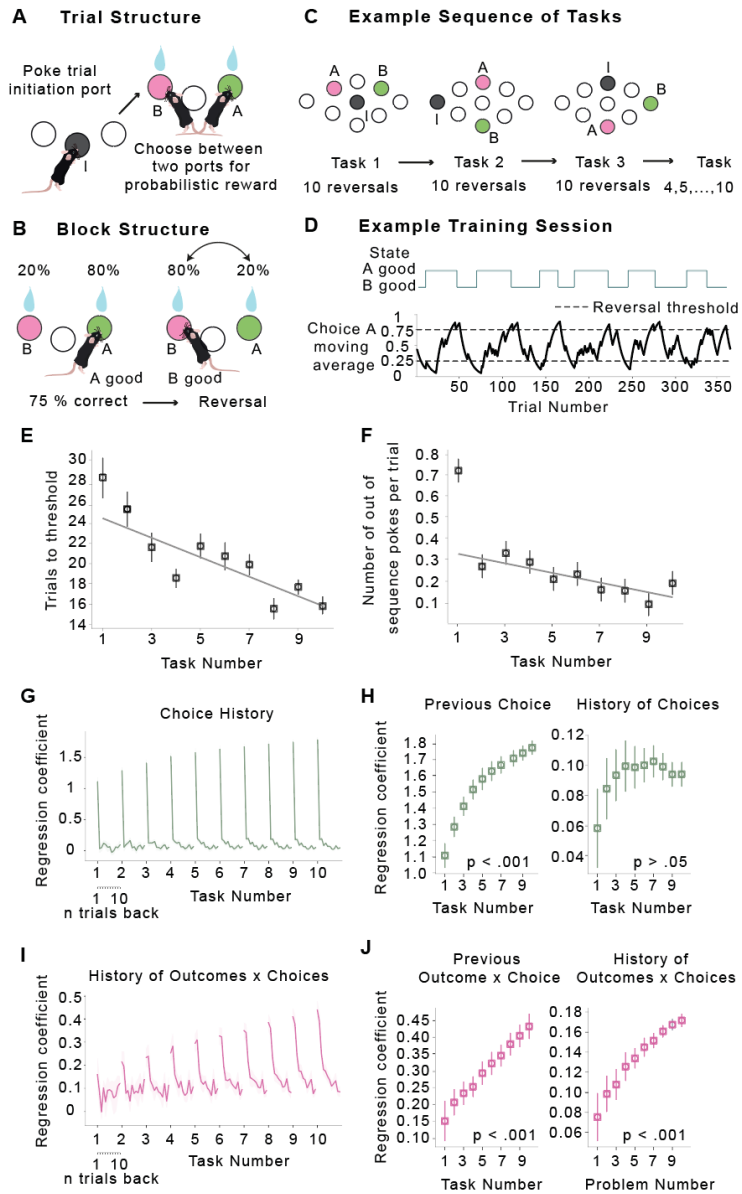

Fig: 1. Transfer learning in mice. **A)** Trial structure of the probabilistic reversal-learning task. Mice poked in an initiation port (grey), then chose between two choice ports (green and pink) for a probabilistic reward. **B)** Block structure of the probabilistic reversal-learning task. Reward contingencies reversed after the animal consistently chose the high reward probability port. **C)** Example sequence of tasks used for training, showing different locations of the initiation (I) and two choice ports (A & B) in each task. **D)** Example behavioural session late in training in which the animal completed 12 reversals. Top panel shows which side has high reward probability; bottom panel shows exponential moving average of subjects' choices (tau=8 trials). **E)** Number of trials following a reversal taken to reach the threshold to trigger the next reversal, as a function of task number. **F)** Number of pokes per trial to a choice port that was no longer available because the subject had already chosen the other port, as a function of task number. **G, I)** Coefficients from a logistic regression predicting current choices using the history of previous choices **(G)**, outcomes (not shown) and choice-outcome interactions **(I)**. For each task and predictor the coefficients at lag 1-11 trials are plotted. **H, J)** Coefficients for the previous trial (lag 1, left) and average coefficients across lags 2-11 (right), as a function of task number. Error bars on all plots show mean ± SEM across mice.

We first asked whether subjects showed evidence of generalising the abstract task structure (one port is good a time, with reversals) to new tasks (Figure 1B). Mice took fewer trials to reach the 75 % correct threshold for triggering a reversal within each task ($F_{(9,\ 72)} = 3.23$, $p = .002$; Supplementary Figure 2A), and crucially also across tasks ($F_{(9,\ 71)} = 3.88$, $p < .001$; Figure 1E), consistent with generalising knowledge of this abstract structure. Improvement across tasks in a subject's ability to track the good port might reflect an increased ability to integrate the recent history of outcomes and choices across trials. To assess this, we fit a logistic regression model predicting subjects' choices using the choices, outcomes and choice-outcome interactions over the past history of trials. Across tasks, the influence of both the most recent ($F_{(9,\ 71)} = 5.50$, $p < .001$; Figure 1I, J) and earlier ($F_{(9,\ 71)} = 4.33$, $p < .001$; Figure 1I, J) choice-outcome interactions increased. Subjects' choices were also increasingly strongly influenced by their previous choices ($F_{(9,\ 71)} = 11.18$, $p < .001$; Figure 1G, H), suggesting a decrease in spontaneous exploration with learning.

We also looked at whether subjects showed evidence of generalising the trial structure (initiate then choose; Figure 1A) across tasks, by assessing how often they made nose pokes that were inconsistent with this sequence (i.e., pokes to the alternative choice port after having made a choice, instead of going straight back to initiation). Mice made fewer such out-of-sequences pokes across reversals within each task ($F_{(9,\ 72)} = 5.43$, $p < .001$; Supplementary Figure 2B), but importantly also across tasks ($F_{(9,\ 71)} = 18.40$, $p < .001$; Figure 1F).

These data suggest that mice learned to generalise both the block and trial structure across tasks. We next searched for evidence of neural representations that abstracted the task structure away from its physical details, allowing generalisation of knowledge.

Abstract and task-specific representations of trial events by PFC and CA1 units

We recorded single units from dorsal CA1 (345 neurons, n = 3 mice, 91 to 162 neurons per mouse) and medial prefrontal cortex (mPFC, 556 neurons, n = 4 mice, 117 to 175 neurons per mouse; Supplementary Figure 1, Figure 2) using electrophysiology. For recording sessions, we modified the behavioural task such that changes from one task to the next occurred within a session, with the transition to the next task triggered once subjects had completed four reversals on the current task, up to a maximum of three tasks in one session. Subjects adapted well to this change and in most recording sessions performed at least four reversals in three different task layouts, allowing us to track the activity of individual units across tasks (Figure 2B). Cross-task learning reached asymptote prior to starting recordings, i.e., during recording sessions mice no longer showed improvement across tasks (Supplementary Figure 3).

During recording sessions, we used ten different port layouts, but to simplify the analysis they were all reflections of three basic layout types (Figure 2B), each of which occurred once in every session. In the first layout type, the initiation port (I1) was the top or bottom port, and the choice ports were the far left and far right ports. One of these choice ports remained in the same location in all three layouts used in a session, and will be referred to as the A choice. This acted as a control for physical location, allowing us to assess how the changing context of the different tasks affected the representation of choosing the same physical port. Both the other choice port (B choice), and the initiation port, moved physical locations between tasks. In the second layout type, both the initiation port (I2) and B choice port (B2) were in locations that were not used in layout type 1. In the third layout type, the initiation port was the same as the initiation port in layout type 1 (I3 = I1), and the B choice port was the same as the initiation port from layout type 2 (B3 = I2). Hence, in every recording session, we had examples of (1) the same port playing the same role across tasks, (2) different ports playing the same role across tasks and (3) the same port playing different roles across tasks (I3 and B2). The order of the layout types

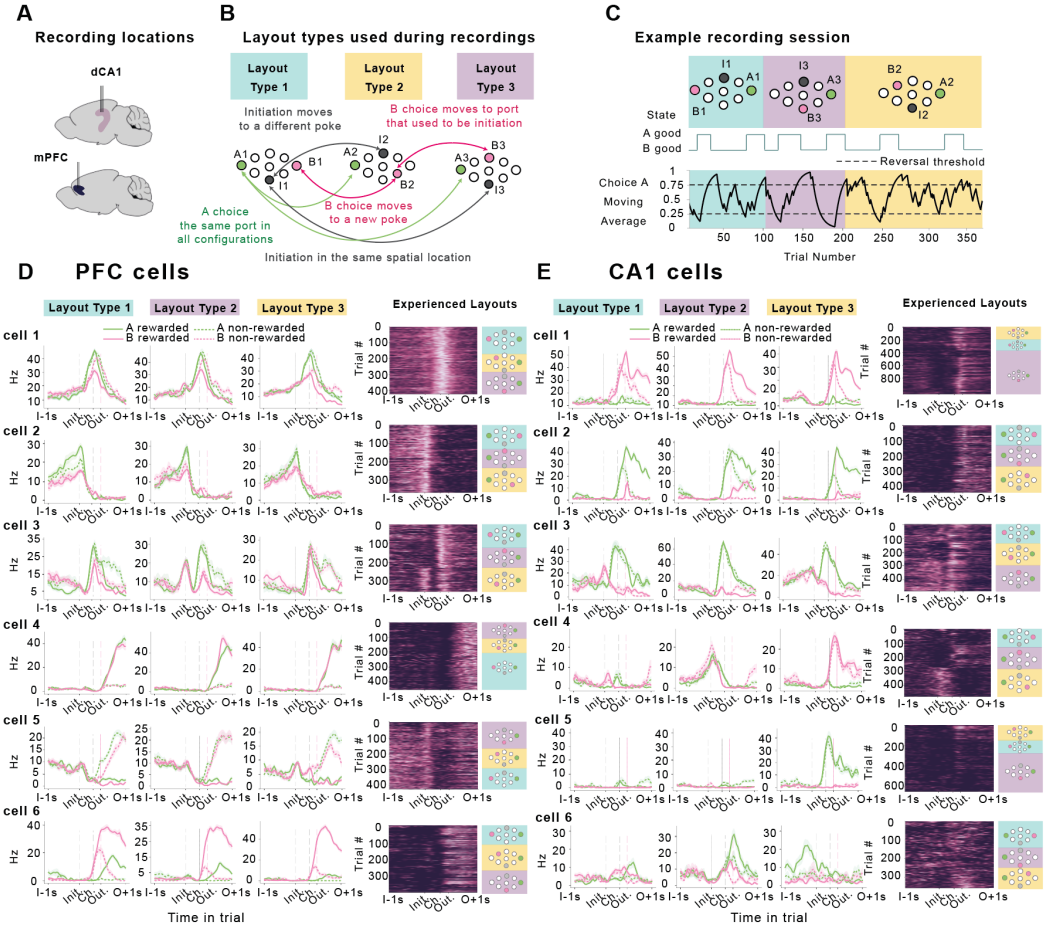

Fig: 2. Recording units across multiple tasks in a single session. **A)** Silicon probes targeting hippocampal dorsal CA1 and medial PFC were implanted in separate groups of mice. **B)** Diagram of task layouts types used during recording sessions. **C)** Example recording session in which a subject completed four reversals in each of three tasks. Top panel shows the ports participating in each task colour coded by layout type. Bottom panel shows the exponential moving average of choices, with the blocks shown above. **D) Example PFC neurons**. Cell 1 in PFC fired selectively to both choice ports (but not initiation) in each task, even though the physical location of the choice ports was different both within and across tasks. Cell 2 fired at the initiation port in every task, even when its physical location changed. Cell 3 fired at B choice ports in all tasks, but also gained a firing field when initiation port moved to the previous B choice port (showing PFC does have some port-specific activity). Cell 4 responded to reward at every choice port in every task. Cell 5 responded to reward omission, and had high firing during the ITI. Cell 6 responded to reward at B choice port (that switched location) in each task. **E) Example CA1 neurons**. Some CA1 cells also had task general firing properties (cell 1 and 2). Cell 1 fired at B choice that switched physical location between tasks. Cell 2 responded to the same port in all tasks and modulated its firing rate depending on whether it was rewarded or not. Cell 3 fired at the same port in all task layouts. Cell 4 switched its firing preference from initiation to B choice that shared physical locations, analogous to 'place cells' firing at a particular physical location. This port selectivity was more pronounced in CA1 than PFC (Supplementary Figure 4). Cell 5 and 6 'remapped' - showing interactions between task and space. Cell 5 fired at a given port in one layout but not when the same port was visited in a different layout. Cell 6 fired at choice time at a given port in one layout and changed its preferred firing time to pre-initiation in a different layout.

was randomised in each recording session.

As animals transferred knowledge of the trial structure across tasks, we reasoned that neurons may exhibit 'task general' representations of the abstract stages of the trial (initiate, choose, outcome) divorced from the sensorimotor specifics of each task. On inspection, such cells were common in PFC (Figure 2D). To respond flexibly when a novel task with the same trial structure is encountered, abstract knowledge should be mapped onto the sensorimotor specifics of the new experience. In line with this, although we observed some task-general firing in CA1, hippocampal cells were more likely to respond to the specifics of each task (Figure 2E). These single unit examples suggest that although task general representations might exist in both regions, PFC activity appears to generalise more across tasks, while CA1 represents physical location more strongly, and additionally exhibits 'remapping' between tasks in which neurons change their tuning to both physical location and task events.

PFC population activity generalises more strongly across tasks than CA1

To assess whether our single unit observations hold up at the population level, we sought to characterise how neural activity in each region represented task events, and how these representations generalised across tasks.

We first assessed the influence of different task variables in each region using linear regression to predict spiking activity of each neuron, at each time point across the trial, as a function of the choice, outcome, and outcome x choice interaction on that trial (Figure 3A). We quantified how strongly each variable affected population activity as the population coefficient of partial determination (i.e., the fraction of variance uniquely explained by each regressor) at every time point across the trial (Figure 3B). This analysis was run separately for each task in the session and the results were averaged across tasks and sessions. Both regions represented current choice, outcome, and choice x outcome interaction, but there was regional specificity in how strongly each variable was represented. Choice (A vs B) representation was more pronounced in CA1 than PFC (peak variance explained - CA1: 8.4 %, PFC: 4.8 %, $p < .001$), whereas outcome (reward vs no reward) coding was stronger in PFC (peak variance explained – CA1: 7.1 %, PFC: 12.9 %, $p < .001$). Furthermore, choice x outcome interaction explained more variance in CA1 than PFC (peak variance explained – CA1: 3.7 %, PFC: 2.4 %, $p <.001$).

Though highlighting some differences in population coding between regions, this approach cannot assess the relative contribution of abstract representations that generalise across tasks versus task specific features such as the physical port location. This requires comparing activity both across time points in the trial and across tasks, which we did using representational similarity analysis (RSA)[42]. We extracted firing rates around initiation and choice port entries (40ms window) and categorised these windows by which task they came from, whether they were initiation or choice, and - for choice port entries whether the choice was A or B and whether it was rewarded - yielding a total of 15 categories (Figure 3C). For each session we computed the average activity vector for each category, then quantified the similarity between categories as the correlation between the corresponding activity vectors. We show RSA matrices for this 'choice time' analysis (Figure 3C, left panels), and also an 'outcome time' analysis (Figure 3C, right panels) where the windows for choice events were moved 240ms after port entry, holding the time window around trial initiations constant.

To quantify the factors influencing representation similarity, we created representational similarity design matrices (RDMs) which each encapsulated the predicted pattern of similarities under the assumption that activity was influenced by a single task feature (Figure 3D). For example, if the population activity represented only which physical port the animal was at, its correlation matrix would look like Figure 3D, Port. We included design matrices for a set of task-general features; the trial stage ('Initiation vs Choice'), choice (A vs B), trial outcome (both on its own as 'Outcome', and in conjunction with

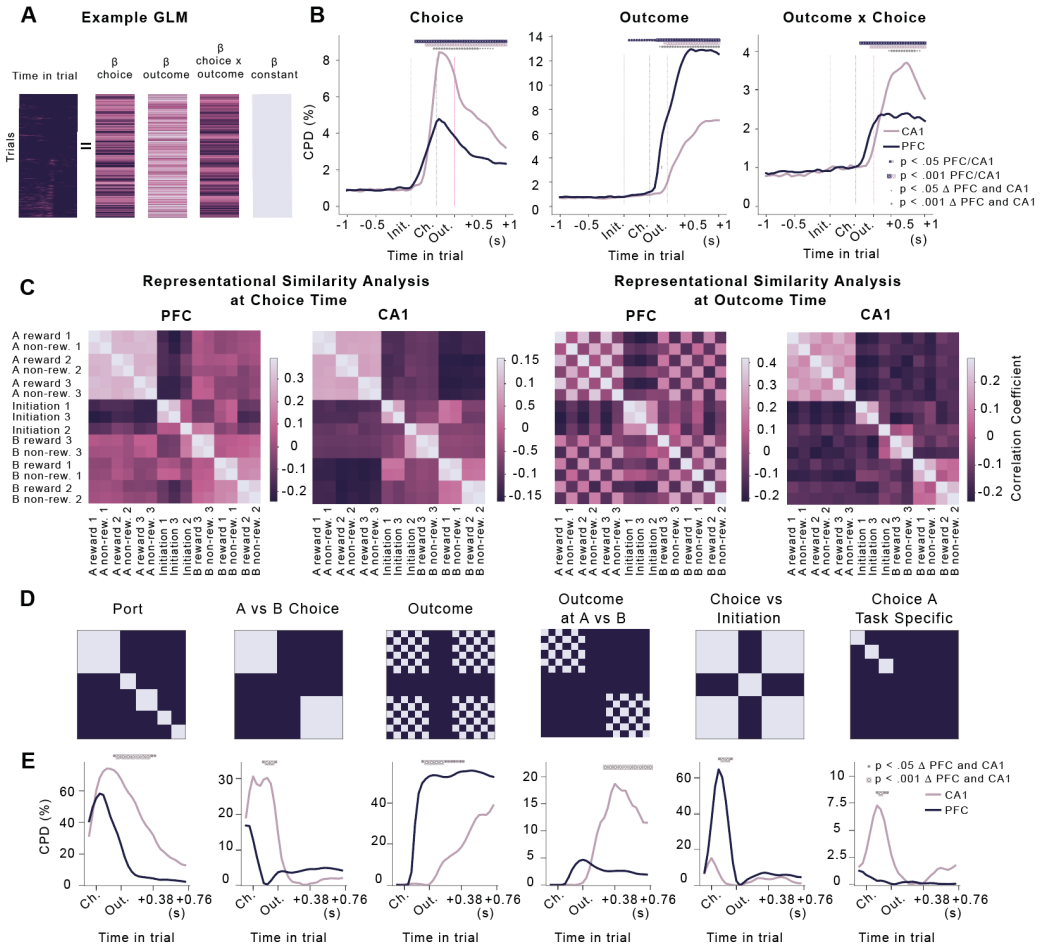

Fig: 3. Task-general and task-specific representations in PFC and CA1 population activity. **A)** Linear regression predicting activity of each neuron at each time point across the trial, as a function of the choice, outcome and outcome x choice interaction. **B)** Coefficients of partial determination from the linear model shown in **A** for choice, outcome and outcome x choice regressors in PFC and CA1. **C)** Representation similarity at 'choice time' (left) and 'outcome time' (right), quantified as the Pearson correlation between the demeaned neural activity vectors for each pair of conditions. **D)** Representational Similarity Design Matrices (RDMs) used to model the patterns of representation similarity observed in the data. Each RDM codes the expected pattern of similarities among categories in **C** under the assumption that the population represents a given variable. The *Port* RDM models a representation of the physical port poked (e.g., far left) irrespective of its meaning in the task. *A vs B Choice* models a representation of A/B choices irrespective of physical port. The *Outcome* RDM models representation of reward vs reward omission. The *Outcome at A vs B* RDM models separate representations of reward vs omission following A and B choices. *Choice vs Initiation* models representation of the stage in the trial. *Choice A Task Specific* models separate representation of the A choice in different tasks. **E)** Coefficients of partial determination in a regression analysis modelling the pattern of representation similarities using the RDMs shown in **D**. The time-course is given by sliding the windows associated with choices from being centered on choice port entry to 0.76 s after choice port entry, while holding time windows centered on trial initiations fixed. Stars indicated time points where regression weight for each RDM was significantly different between the two regions ($p < .05$ (small stars) and $p < .001$ (big stars), permutation test across sessions corrected for multiple comparison over time points. For more details on permutation tests see Methods.

choice 'Outcome at A vs B'). Changes in activity across tasks might occur simply due to neurons being tuned for particular physical locations, which will be captured by the 'Port' RDM. However, it is also possible that the changing context provided by different tasks modifies the representation of choosing the same physical port at the same trial stage. To assess such 'remapping', we included an RDM 'Choice A task specific' which modelled task specific representations of the A choice, which shares the same physical location and meaning across tasks. We modelled the observed pattern of similarities in the data as a linear combination of these RDMs, quantifying the influence of each by its corresponding weight in the linear fit. To be able to examine the temporal evolution of these effects we run a series of regressions onto the data. In each, the data around initiation port entry was the same but the data around the choice port entry progressed serially through time from choice point until after the reward was delivered (Figure 3E).

Consistent with our single unit observations, both PFC and CA1 represented both task specific and task general features to some extent. However, there was a marked regional specificity in how strongly different features were encoded (Figure 3E). PFC had stronger, abstract, sensorimotor-invariant representation of trial stage (initiation vs choice) and trial outcome ($p < .001$). In contrast, CA1 had stronger representation of the physical port the subjects was poking, and whether it was an A vs B choice ($p < .001$). Additionally, CA1 but not PFC showed a task specific representation of A choices ($p < .001$). This is striking because during A choices both the physical port and its meaning are identical across tasks, indicating that the changing task context alone induced some 'remapping' in CA1 but not PFC. Finally, there was a regional difference in the representation of trial outcome. PFC outcome representations were more general (the same neurons responded to reward or reward omission across ports and tasks - $p < .001$). CA1 also maintained an outcome representation, but this was more likely to be conjunctive than in PFC – different neurons would respond to reward on A and B choices ($p < .001$). To exclude the possibility that task specificity in CA1 might be driven by CA1 representations drifting slowly over time we confirmed that task representation changed abruptly at transitions between tasks (Supplementary Figure 5).

Low dimensional temporal structure of activity is invariant across tasks and regions, but cell assemblies generalise more strongly in PFC than CA1

To further explore how the structure of population activity generalised between tasks, we used singular value decomposition to compare the principal temporal and cellular modes across the different tasks. We decomposed activity in each task into a set of cellular (across neurons) and temporal modes (across trial and time). For each cell in each task, we computed the average firing rate at each time point across the trial, for four types of trials – rewarded A choices, A non-rewarded, B rewarded, and B non-rewarded. We concatenated these four time series for each cell to create an activity matrix D where each row contained the average activity of one neuron in one task across each time point of the four trial types (Figure 4A). Using SVD, we decomposed each activity matrix into cellular and temporal modes U and V, linked by a diagonal weight matrix Σ.

$$D = U\Sigma V^T$$

Each cellular mode in U is a vector with a weight for each cell. They can be thought of as cell assemblies, as they correspond to sets of neurons whose activity covaries over time. Cellular and temporal modes come in pairs, such that each cellular mode has an associated temporal mode in V, which is a vector of weights indicating how strongly the cellular mode contributes to population activity at each time point across the four trial types. The cellular and temporal modes are both unit vectors, so the contribution of each pair to the total data variance is determined by the corresponding element of the diagonal matrix Σ. The first cellular and temporal mode of PFC activity in three different tasks is shown in Figure 4B, C. It is high throughout the ITI and trial with a

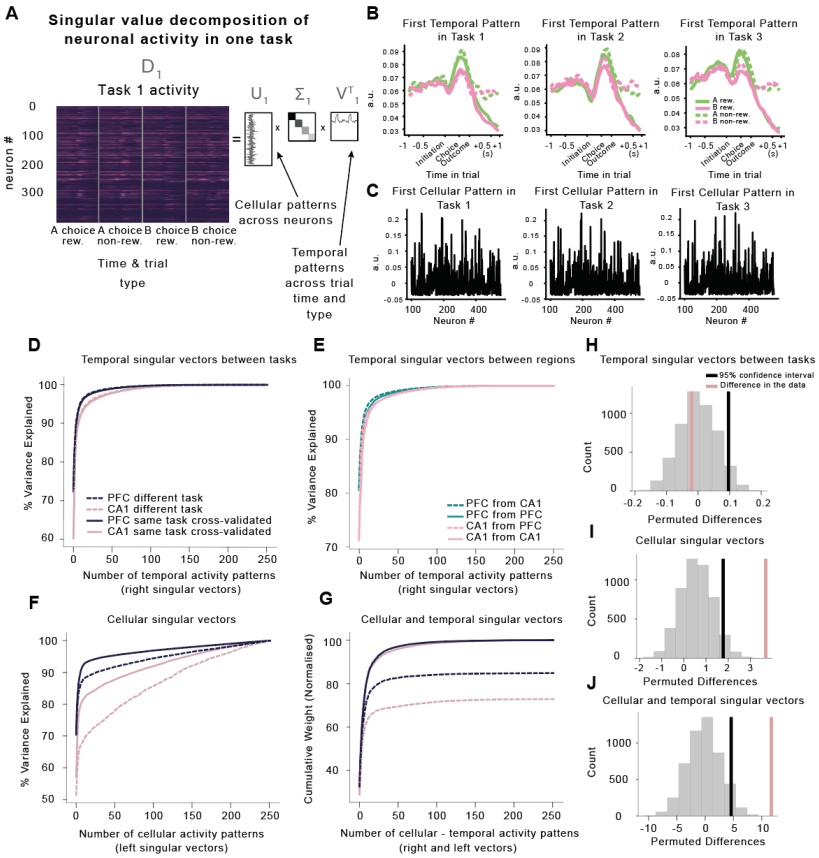

Fig: 4. Generalisation of low dimensional representations of trial events. **A)** Diagram of singular value decomposition (SVD) analysis. A data matrix comprising the average activity of each neuron across time points and trial types was decomposed into the product of three matrices, where diagonal matrix Σ linked a set of temporal patterns across trial type and time (rows of V^T) to a set of cellular patterns across cells (columns of U). **B)** First temporal mode in V^T from SVD decomposition of data matrix from PFC plotted in each task separately for clarity and separated by A (green) and B (pink) rewarded (solid) non-rewarded (dashed) choices. **C)** First cellular mode from SVD decomposition of data matrix from PFC in each task showing similar pattern of cells participate in all tasks. **D)** Variance explained when using temporal activity patterns V_1^T from one task to predict either held out activity from the same task (solid lines) or activity from a different task (dash lines). **E)** Variance explained when using temporal activity patterns V_1^T to predict either activity from the same task and brain region (solid lines) or a different brain region and the same task (dash lines) D_2. **F)** Variance explained when using cellular activity patterns U_1 from one task to predict either held out activity from the same task (solid lines) or activity from a different task (dash lines). **G)** Cumulative weights along the diagonal Σ using pairs of temporal V_1^T and cellular U_1 activity patterns from one task to predict either held out activity from the same task (solid lines) or activity from a different task (dash lines). Weights were normalised by peak cross-validated cumulative weight computed on the activity from the same task. **H)** To assess whether the temporal singular vectors generalised significantly better between tasks in PFC than CA1, we evaluated the area between the dash and solid lines in **D** for CA1 and for PFC separately, giving a measure for each region of how well the singular vectors generalised. We computed the difference in this measure between CA1 and PFC (pink line in **H**), and compared this difference to the null distribution obtained by permuting sessions between brain regions (grey histogram, black line shows 95th percentile of distribution). For more details on permutation tests see Methods.Temporal singular vectors generalised equally well between tasks in the two regions. **I)** Cellular singular vectors generalised significantly better between tasks in PFC than CA1. Computed as in **H** but using the solid / dash lines from **F**. **G)** Pairs of cellular and temporal singular vectors generalised significantly better between tasks in PFC than CA1. Computed as in **H** but using the solid / dash lines from **G**.

peak at choice time, but strongly suppressed following reward (similar to cell 5 in Figure 2D).

We reasoned that: (i) if the same events were represented across tasks (e.g. initiation, A/B choice, reward), then the temporal modes would be exchangeable between tasks, no matter whether these representations were found in the same cells; (ii) if the same cell assemblies were used across tasks, then the cellular modes would be exchangeable across tasks, no matter whether the cell assemblies played the same role in each task; and (iii) if the same cell assemblies performed the same roles in each task, then pairs of cellular and temporal modes would be exchangeable across tasks.

To see whether the same representations existed in each task, we first asked how well the temporal modes from one task could be used to explain recordings from other tasks. Since V is an orthonormal basis, any data of the same rank or less can be perfectly explained when using all the temporal modes. However, population activity in each task is low dimensional so a small number of modes explain a great majority of the variance. Modes that explain a lot of variance in one task will only explain a lot of variance in the other task if the structure captured by the mode is prominent in both tasks. The question is therefore how quickly variance is explained in data set B, when ordering the modes according to variance explained in data set A. To assess this, we regressed the temporal modes from one task onto the data matrix from the other, and plotted cumulative variance explained (Figure 4D). To control for drift in neuronal representations across time, we computed the data matrices separately for the first and second halves of each task. We compared the amount of variance explained using modes from the first half of one task to model activity in the second half of the same task, with the variance explained using modes from the second half of one task to model activity from the first half of the next task.

In both PFC and CA1, the cumulative variance explained as a function of the number of temporal modes used, did not depend on whether the two data sets were from the same task (solid) or different tasks (dashed) (Figure 4D, H, $p > .05$). This indicates that the temporal patterns of activity, and therefore the trial events represented, did not differ across tasks in either brain area. However, as this analysis used only the temporal modes, it says nothing about whether the same or different neurons represented a given event across tasks. In fact, we can even explain activity in one brain region using temporal modes from another region and mouse. (Figure 4E).

The pattern was very different when we used cellular modes (i.e., assemblies of co-activating neurons) from one task to explain activity in another. In both PFC and CA1, cellular modes in U that explained a lot of variance in one task, explained more variance in the other half of the same task than they did in an adjacent task (Figure 4F - differences between solid and dashed lines). However, the within task vs cross task difference was larger in CA1 than PFC (Figure 4I, $p < .05$). This indicates that PFC neurons whose activity covaried in one task were more likely to also covary in another task, when compared to CA1 neurons. As this analysis considered only the cellular modes it does not indicate whether a given cell assembly carried the same task information across tasks.

To assess how well the cellular-temporal activity patterns from one task explained activity in another, we projected one data set D_2 onto the cellular and temporal mode pairs of the other (U_1^T, V_1).

$$\Sigma_2 = U_1^T D_2 V_1$$

If the same cell assemblies perform the same roles in two different tasks, the temporal and cellular modes will align, and Σ_2 will have high weights on the diagonal. We therefore plotted the cumulative weight of the diagonal elements of Σ within and between tasks (Figure 4G). In both PFC and CA1 cellular and temporal modes aligned better in different data sets from

the same task (solid lines), than for different tasks (dashed lines). However, this difference was substantially larger for CA1 than PFC (Figure 4J, $p < .05$).

These data show that although the temporal structure of activity in both regions generalises perfectly across tasks, brain regions and subjects – a consequence of the same set of trial events being represented in each, the cell assemblies used to represent them generalised more strongly in PFC than CA1.

Policy representations are abstract in PFC, but linked to sensorimotor experience in CA1

So far, we have focused on the neuronal representations of events on individual trials, and how they generalise across tasks. But to maximise reward, the subject must also track which option is currently best by integrating the history of choices and outcomes across trials. To be useful for generalisation, this policy representation should also be divorced from the current sensorimotor experience of any specific task.

To obtain an estimate of subjects' beliefs about which option was best, we used a logistic regression predicting current choices as a function of the history of previous rewards, choices and their interactions (Figure 5A). This allowed us to compute, trial-by-trial, the probability that the animal would choose A vs B – i.e., the animal's policy. When we used this policy as a predictor of neural activity, it explained variance that was not captured by within-trial regressors such as choice, reward and choice x reward interaction. Specifically, the subjects' policy interacted with the current choice explained variance (Figure 5B, $p < .001$). Notably, this signal became prominent around the time of trial initiation, when it would be particularly useful for guiding the decision.

To examine whether policy representations generalised across tasks, we evaluated the correlation across tasks between the policy weights in the neural regression. Because the A port was the same on each task, but the B port varied between tasks, we computed policy regression weights at each time point separately for A and B choices (controlling for reward). We then computed the average across-task correlation of these weights between every pair of timepoints (Figure 5C). The diagonal elements of these matrices show the average correlation across tasks at the same time point in each task. Visual inspection (Figure 5C), and permutations tests of differences between sums of the diagonals of Policy on A and B choices correlation matrices ($p < .05$), revealed that these correlations were larger in PFC than CA1 (Figure 5D). On average, therefore, cellular representations of policy generalised across tasks better in PFC than CA1 on both A and B choices.

One possible explanation is that PFC simply represented action values in a task-general way. A more interesting possibility is that current policy shapes the representation of each trial stage differently, but in CA1 these representations are more tied to the sensorimotor specifics of the current task. To test this, we examined time-slices through the correlation matrices at initiation, choice, and outcome times (Figure 5E). In PFC, all three correlation profiles on both A and B trials peaked at the correct time point (the equivalent to the diagonal elements of the matrix) – i.e., the policy representations generalised across problems, but were specific to the different parts of the trial (initiate, choose, outcome). A similar pattern was present in CA1, but only on A choices (which are the same physical port across tasks). No CA1 correlation was significantly above zero on B choices. Indeed, whilst PFC policy correlations were greater than CA1 correlations for all representations (all $p < .05$) on both A and B choices, CA1 correlations showed a greater difference between A and B trials at outcome time (Figure 5E, all $p < .05$).

Overall, therefore, both PFC and CA1 maintained representations of the subject's current policy that were not simple value representations – as they differed depending on the trial stage. These representations were abstracted across tasks in PFC, but tied to the sensori-

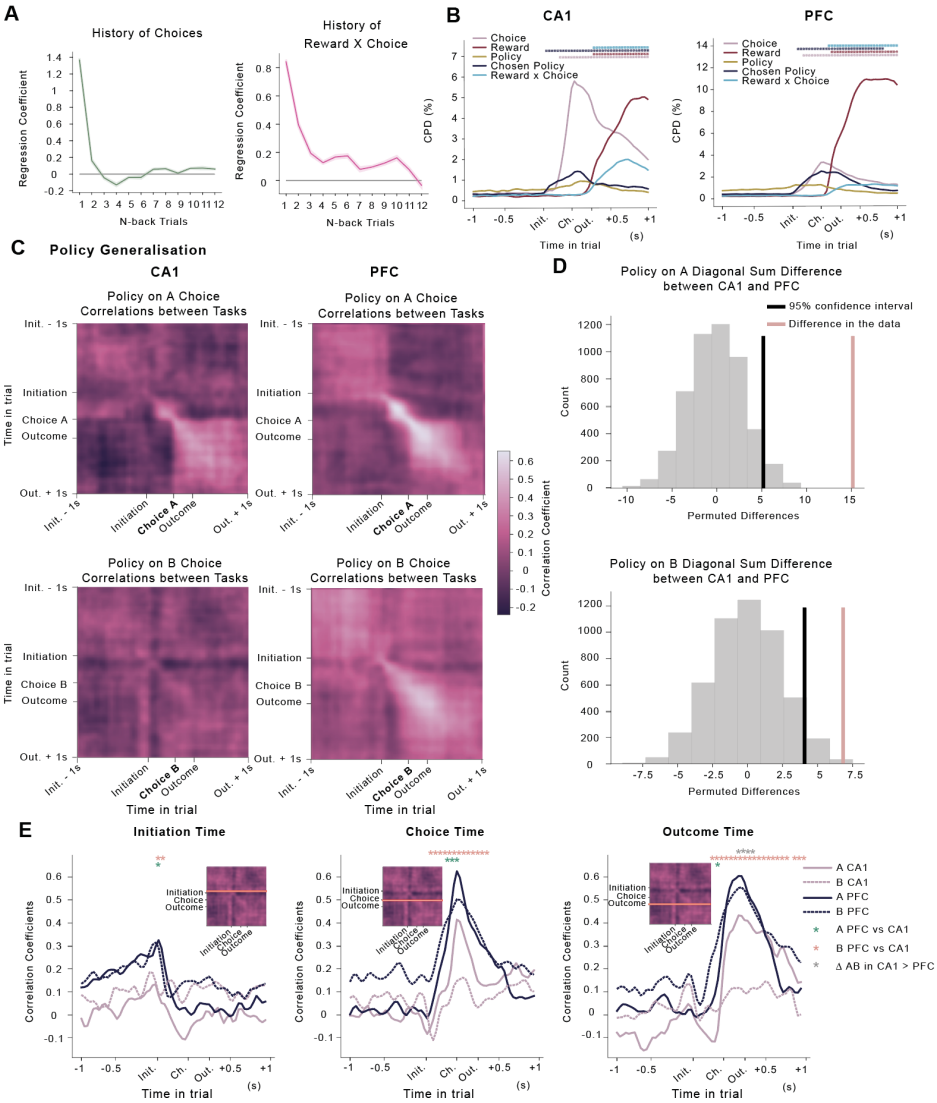

Fig: 5. Policy Generalisation in PFC and CA1. **A)** Weights from logistic regression predicting choices in recording sessions using choices, rewards and choice x reward interactions over the previous twelve trials as predictors. The effect of choice x outcome interaction history was above zero on up to eleven trials back. Error bars report the mean ± SEM across mice. **B)** CPDs from regression models predicting neural activity using current trial events, subjects' policy (estimated using the behavioural regression in **A**), and policy interacted with current choice. Stars denote the time points at which each regressor explained significantly more variance than expected by chance (permutation test across sessions - $p < .001$, corrected for multiple comparisons; for more details on permutation tests see Methods.) **C)** Correlations across tasks between policy weights in regressions predicting neural activity. Regressions were run separately for A (top panels) and B (bottom panels) choices in each task, and at each time point across the trial. Correlations of policy representations between all task pairs were evaluated for each pair of time points, values on the diagonal show how correlated policy representation was at the same time point in both tasks. Positive correlation indicates that the same neurons coded policy with the same sign in both tasks. **D)** To quantify whether policy generalised more strongly between tasks in PFC than CA1, we computed the between region difference in the sum along the diagonal of the correlation matrices in **C)**, separately for A and B choices, and compared it against the null distribution obtained by permuting sessions between brain regions. Policy representation on both A and B choices generalised more strongly in PFC than CA1. **E)** Slices through the correlation matrices at initiation (left), choice (center) and outcome (right) times for A (solid) and B (dash line) choices. Significant differences between conditions are indicated by stars as shown in legend.

motor specifics in CA1. A portion, but not all, of this task specificity in CA1 was accounted for by the port identity.

3 Discussion

Humans and other animals effortlessly generalise prior experience to novel situations that are only partially related. This ability relies on our understanding of the abstract structure in the regularities we experience in the world. Here we developed a novel behavioural paradigm to measure this generalisation of abstract knowledge between reinforcement learning problems with the same structure - probabilistic reversal learning - but different sensorimotor particularities. Mice generalised knowledge about two elements of the task structure between different but related problems - the sequence of responses required within a trial, and the between-trial policy required to obtain rewards. Recordings from hippocampal CA1 and mPFC revealed that both abstract and task-specific representations existed in both brain areas but in markedly different proportions, such that population responses in mPFC but not CA1 were dominated by task invariant, abstract representations. By contrast, the CA1 responses contained major sources of variance that were either invariant to the sensorimotor particularities (port selective), or intriguingly, the interaction of these with the task (demonstrating 'remapping' between tasks). Notably, this was true both for correlates of the elements of an individual trial, and for correlates of the long-term behavioural policy that guided between-trial behaviour.

Recent data have highlighted the low dimensional structure of task representations in rodent OFC[8]. We show that these low dimensional temporal modes are also consistent across tasks in both mPFC and CA1. We also confirm that they are consistent between animals and further demonstrate they are consistent between different brain areas (mPFC and CA1), suggesting this low dimensional structure does not reflect the unique representational properties of a particular brain area. Our manuscript makes further unique contributions. Because we record across the same neurons in different tasks, we are able to ask not only whether the temporal dimensions are preserved across tasks, but whether these temporal modes align to the same neurons in each task, i.e., whether the same neurons represent the same trial events across tasks. They do so significantly more in PFC than CA1. Whilst transfer learning relies on building abstractions, it must also tie these abstractions to the sensorimotor properties of each new task. In this context it is intriguing that CA1 representations contained distinct portions of variance aligned to abstract task coordinates, to sensorimotor coordinates and to the interaction of the two coordinate sets. Lastly, our paper extends these ideas to variables that must integrate information over many different experiences (such as the animals' choice policy) and shows a similar distinction between mPFC and CA1 in performing such computations.

A second recent line of work has examined related ideas in primate PFC and hippocampus[7,43]. Where data are available from both structures in the same task, representations are found to be geometrically arranged in line with task coordinate space and no clear differences are observed between structures. Whilst it is tempting to postulate a species difference, careful examination reveals another possibility. Because these data are acquired in tasks that share no sensory elements (no overlapping images between tasks), and because motor coordinates are aligned to task coordinates, it is not possible in these data to discern whether there is also a sensorimotor component to the hippocampal representation, as we observe in our data.

We found that prefrontal neurons encoded abstract meanings of different stages of the task (initiation, choice, and outcome), which might underlie animals' ability to quickly know how to do a trial on any set of physical ports. The identification of a common representation of the sequential structure of different states/actions aligns with theoretical arguments about abstracting the structure of behaviour[38,44]. Such theories suggest that these abstractions need not be limited to representing exact sequences, but can also abstract the rules and regularities that constrain possible sequences. These ideas were developed in the context of

the entorhinal cortex. Whilst we did not record from entorhinal cortex in the current study, recent fMRI evidence in humans in a conceptually similar experiment suggests entorhinal representations will also generalise the structure of reinforcement learning tasks[10]. It is also notable that abstract representations of trials are present in mPFC in purely spatial contexts[35,36]. It will therefore be intriguing to build an understanding of how these representations differ.

One possibility is that abstractions that affect behaviour over longer timescales will be preferentially represented in frontal regions[45], such as the policy representations described here. Indeed, the notion that policy representations may be abstracted aligns directly with recent ideas from computer science such as meta-reinforcement learning. When neural networks are trained with such algorithms, their internal representations resemble those seen in frontal cortex in a number of distinct tasks[6,46,47].

Along with extracting structure in regularities in the world the brain also processes ongoing experiences and creates memories of specific events. Memories are defined by vivid sensory representations, such as sounds, smells, tastes, and physical locations. The hippocampal formation play an important role in episodic memory, and contains neural representations relevant for memory encoding and recall[18,19,48]. In this broader context, it is perhaps unsurprising that hippocampus contain rich representations of the sensorimotor specifics of current experience. The fact that these coexisted with structural abstractions is consistent with the idea that hippocampus is modulated by the schema that underlie episodic experiences[19,49]. Notably, we also found that policy coding was not unique to prefrontal cortex, as hippocampus also contained policy representations, corroborating existing findings for the existence of signals relevant for decision-making in hippocampal formation[50,51]. We expand on these observations to provide further evidence that hippocampal activity might represent sensorimotor specifics of events in the context of broader memory schemas and task structures.

We do not perceive the world as it really is. Starting with the visual 2D inputs on the retina that we use along with prior experience to infer the 3D world around us[52], our brains likely develop structural placeholders for many of our experiences. In fact, we remember things more easily if we know the general schema or a script for a particular event[53], and often ignore information that does not align with our understanding of the world[54]. More broadly, here we demonstrate that mice also acquire sophisticated models of tasks they frequently experience in their environment and can apply this knowledge to solve new problems faster. We further show that prefrontal cortex contains representations of what can be thought of as a 'learning-set', or 'schema' of abstract relationships and variables needed to solve new related problems while hippocampus combines sensorimotor and abstract information to represent an interaction between the two, which might be crucial for both interpreting our ongoing experiences as well as encoding and recall of episodic memories.

Author Contributions

V.S., T.A., M.E.W. and T.E.J.B. designed the study; V.S., T.A. and J.L.B. acquired the data; V.S. and T.E.J.B analyzed the data with input from T.A. V.S., T.A. and T.E.J.B wrote and edited the manuscript with input from M.E.W.

Competing interests

Authors declare no competing interests.

Acknowledgments

We would like to thank Tom Jahans-Price for his help with setting up electrophysiology in our lab and training us to conduct our first recordings. We would also like to thank Tom Jahans-Price, Mohamady El-Gaby and Yves Weissenberger for providing helpful comments on the drafts of the manuscript. This work was funded by the following grants: Well-

come Principal Research Fellowship (219525/Z/19/Z), and JS McDonnell Foundation award (JSMF220020372) to T.E.J.B.; Wellcome Collaborator award (214314/Z/18/Z) to T.E.J.B., T. A., M.E.W. and Senior Research Fellowship (202831/Z/16/Z) to M.E.W. The Wellcome Centre for Integrative Neuroimaging and Wellcome Centre for Human Neuroimaging are each supported by core funding from the Wellcome Trust (203139/Z/16/Z, 203147/Z/16/Z).

Data and materials availability

All data, analysis and behavioural training code will be released on publication.

References

[1] Piaget, J. The theory of stages in cognitive development. In *Measurement and Piaget* (1971).

[2] Tse, D. *et al.* Schemas and memory consolidation. *Science* DOI: 10.1126/science.1135935 (2007).

[3] Tse, D. *et al.* Schema-dependent gene activation and memory encoding in neocortex. *Science* DOI: 10.1126/science.1205274 (2011).

[4] Harlow, H. F. The formation of learning sets. *Psychological Review* DOI: 10.1037/h0062474 (1949).

[5] Bozinovski, S. Reminder of the first paper on transfer learning in neural networks, 1976. *Informatica (Slovenia)* DOI: 10.31449/INF.V44I3.2828 (2020).

[6] Wang, J. X. *et al.* Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience* DOI: 10.1038/s41593-018-0147-8 (2018).

[7] Baraduc, P., Duhamel, J. R. & Wirth, S. Schema cells in the macaque hippocampus. *Science* DOI: 10.1126/science.aav5404 (2019).

[8] Zhou, J. *et al.* Evolving schema representations in orbitofrontal ensembles during learning. *Nature* DOI: 10.1038/s41586-020-03061-2 (2020).

[9] Baldassano, C., Hasson, U. & Norman, K. A. Representation of real-world event schemas during narrative perception. *Journal of Neuroscience* DOI: 10.1523/JNEUROSCI.0251-18.2018 (2018).

[10] Baram, A. B., Muller, T. H., Nili, H., Garvert, M. M. & Behrens, T. E. J. Entorhinal and ventromedial prefrontal cortices abstract and generalize the structure of reinforcement learning problems. *Neuron* DOI: 10.1016/j.neuron.2020.11.024 (2021).

[11] Wood, E. R., Dudchenko, P. A., Robitsek, R. J. & Eichenbaum, H. Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron* DOI: 10.1016/S0896-6273(00)00071-4 (2000).

[12] Guise, K. G. & Shapiro, M. L. Medial Prefrontal Cortex Reduces Memory Interference by Modifying Hippocampal Encoding. *Neuron* DOI: 10.1016/j.neuron.2017.03.011 (2017).

[13] Walton, M. E., Behrens, T. E., Buckley, M. J., Rudebeck, P. H. & Rushworth, M. F. Separable Learning Systems in the Macaque Brain and the Role of Orbitofrontal Cortex in Contingent Learning. *Neuron* DOI: 10.1016/j.neuron.2010.02.027 (2010).

[14] Takahashi, Y. K. *et al.* Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. *Nature Neuroscience* DOI: 10.1038/nn.2957 (2011).

[15] Wilson, R. C., Takahashi, Y. K., Schoenbaum, G. & Niv, Y. Orbitofrontal cortex as a cognitive map of task space. *Neuron* DOI: 10.1016/j.neuron.2013.11.005 (2014).

[16] Schuck, N. W., Cai, M. B., Wilson, R. C. & Niv, Y. Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron* DOI: 10.1016/j.neuron.2016.08.019 (2016).

[17] Dusek, J. A. & Eichenbaum, H. The hippocampus and memory for orderly stimulus relations. *Proceedings of the National Academy of Sciences of the United States of America* DOI: 10.1073/pnas.94.13.7109 (1997).

[18] Wood, E. R., Dudchenko, P. A. & Eichenbaum, H. The global record of memory in hippocampal neuronal activity. *Nature* DOI: 10.1038/17605 (1999).

[19] Manns, J. R. & Eichenbaum, H. Evolution of declarative memory, DOI: 10.1002/hipo.20205 (2006).

[20] Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B. & Botvinick, M. M. Neural representations of events arise from temporal community structure. *Nature Neuroscience* DOI: 10.1038/nn.3331 (2013).

[21] Constantinescu, A. O., O'Reilly, J. X. & Behrens, T. E. Organizing conceptual knowledge in humans with a gridlike code. *Science* DOI: 10.1126/science.aaf0941 (2016).

[22] Garvert, M. M., Dolan, R. J. & Behrens, T. E. A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *eLife* DOI: 10.7554/eLife.17086 (2017).

[23] Aronov, D., Nevers, R. & Tank, D. W. Mapping of a non-spatial dimension by the hippocampal-entorhinal circuit. *Nature* DOI: 10.1038/nature21692 (2017).

[24] Eichenbaum, H. Prefrontal-hippocampal interactions in episodic memory, DOI: 10.1038/nrn.2017.74 (2017).

[25] Knudsen, E. B. & Wallis, J. D. Closed-Loop Theta Stimulation in the Orbitofrontal Cortex Prevents Reward-Based Learning. *Neuron* DOI: 10.1016/j.neuron.2020.02.003 (2020).

[26] O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research* DOI: 10.1016/0006-8993(71)90358-1 (1971).

[27] Muller, R. U. & Kubie, J. L. The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *Journal of Neuroscience* DOI: 10.1523/jneurosci.07-07-01951.1987 (1987).

[28] Leutgeb, J. K. *et al.* Progressive transformation of hippocampal neuronal representations in "morphed" environments. *Neuron* DOI: 10.1016/j.neuron.2005.09.007 (2005).

[29] Fyhn, M., Hafting, T., Treves, A., Moser, M. B. & Moser, E. I. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature* DOI: 10.1038/nature05601 (2007).

[30] Lever, C., Burton, S., Jeewajee, A., O'Keefe, J. & Burgess, N. Boundary vector cells in the subiculum of the hippocampal formation. *Journal of Neuroscience* DOI: 10.1523/JNEUROSCI.1319-09.2009 (2009).

[31] Barry, C., Ginzberg, L. L., O'Keefe, J. & Burgess, N. Grid cell firing patterns signal environmental novelty by expansion. *Proceedings of the National Academy of Sciences of the United States of America* DOI: 10.1073/pnas.1209918109 (2012).

[32] Yoon, K. *et al.* Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nature Neuroscience* DOI: 10.1038/nn.3450 (2013).

[33] Høydal, Ø. A., Skytøen, E. R., Andersson, S. O., Moser, M. B. & Moser, E. I. Object-vector coding in the medial entorhinal cortex. *Nature* DOI: 10.1038/s41586-019-1077-7 (2019).

[34] Morrissey, M. D., Insel, N. & Takehara-Nishiuchi, K. Generalizable knowledge outweighs incidental details in prefrontal ensemble code over time. *eLife* DOI: 10.7554/eLife.22177 (2017).

[35] Yu, J. Y., Liu, D. F., Loback, A., Grossrubatscher, I. & Frank, L. M. Specific hippocampal representations are linked to generalized cortical representations in memory. *Nature Communications* DOI: 10.1038/s41467-018-04498-w (2018).

[36] Kaefer, K., Nardin, M., Blahna, K. & Csicsvari, J. Replay of Behavioral Sequences in the Medial Prefrontal Cortex during Rule Switching. *Neuron* DOI: 10.1016/j.neuron.2020.01.015 (2020).

[37] Behrens, T. E. *et al.* What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior, DOI: 10.1016/j.neuron.2018.10.002 (2018).

[38] Whittington, J. C. *et al.* The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell* DOI: 10.1016/j.cell.2020.10.024 (2020).

[39] Bradfield, L. A., Leung, B. K., Boldt, S., Liang, S. & Balleine, B. W. Goal-directed actions transiently depend on dorsal hippocampus. *Nature Neuroscience* DOI: 10.1038/s41593-020-0693-8 (2020).

[40] Knudsen, E. & Wallis, J. Hippocampal neurons construct a map of an abstract value space. *bioRxiv* DOI: 10.1101/2020.12.17.423272 (2020).

[41] Park, A. J. *et al.* Reset of hippocampal–prefrontal circuitry facilitates learning. *Nature* DOI: 10.1038/s41586-021-03272-1 (2021).

[42] Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience* DOI: 10.3389/neuro.06.004.2008 (2008).

[43] Bernardi, S. *et al.* The Geometry of Abstraction in the Hippocampus and Prefrontal Cortex. *Cell* DOI: 10.1016/j.cell.2020.09.031 (2020).

[44] Mark, S., Moran, R., Parr, T., Kennerley, S. W. & Behrens, T. E. Transferring structural knowledge across cognitive maps in humans and models. *Nature Communications* DOI: 10.1038/s41467-020-18254-6 (2020).

[45] Murray, J. D. *et al.* A hierarchy of intrinsic timescales across primate cortex. *Nature Neuroscience* DOI: 10.1038/nn.3862 (2014).

[46] Frans, K. *et al.* Meta-learning shared strategies: OpenAI. *Iclr 2018* (2018).

[47] Dasgupta, I. *et al.* Causal reasoning from meta-reinforcement learning (2019). 1901.08162.

[48] Scoville, W. B. & Milner, B. Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry* DOI: 10.1136/jnnp.20.1.11 (1957).

[49] Sun, C., Yang, W., Martin, J. & Tonegawa, S. Hippocampal neurons represent events as transferable units of experience. *Nature Neuroscience* DOI: 10.1038/s41593-020-0614-x (2020).

[50] Masuda, A. *et al.* The hippocampus encodes delay and value information during delay-discounting decision making. *eLife* DOI: 10.7554/eLife.52466 (2020).

[51] Wimmer, G. E., Daw, N. D. & Shohamy, D. Generalization of value in reinforcement learning by humans. *European Journal of Neuroscience* DOI: 10.1111/j.1460-9568.2012.08017.x (2012).

[52] von Helmholtz, H. *Treatise on Physiological Optics* (1866).

[53] Bartlett, F. F. Remembering: An experimental and social study (1932).

[54] Nickerson, R. S. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* DOI: 10.1037/1089-2680.2.2.175 (1998).

[55] Akam, T. *et al.* pyControl: Open source, Python based, hardware and software for controlling behavioural neuroscience experiments. *bioRxiv* DOI: 10.1101/2021.02.22.432227 (2021).

[56] Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M. & Kenneth D., H. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv* DOI: 10.1101/061481 (2016). Publisher: Cold Spring Harbor Laboratory _eprint: https://www.biorxiv.org/content/early/2016/06/30/061481.full.pdf.

Materials and Methods

Behavioural apparatus

Experiments were performed in custom made operant boxes (https://github.com/pyControl/hardware), controlled using pyControl[55]. The boxes used in the training phase of the experiment had six nose poke ports mounted on the back wall, each with infrared beam, stimulus LED and solenoid valve for dispensing liquid rewards, and a speaker for auditory stimuli. For recording experiments mice were transferred to operant boxes with nine nose poke ports located in electrically shielded sound attenuating chambers.

Subjects

Nine male C57BL/6J experimentally naïve mice bred in the Biomedical Sciences Facility at the University of Oxford were obtained for this experiment at six weeks of age. Animals were group housed prior to surgery, and individually housed post-surgery, in a humidity- and temperature-controlled vivarium, on a 12-hour light-dark cycle (7:00 to 19:00). All nine animals were implanted with silicon probes, but we only obtained data from seven animals, due to one probe being damaged during surgery and having to cull one animal prior to recordings. Experiments were carried out in accordance with the Oxford University animal use guidelines and performed under UK Home Office Project Licence P6F11BC25.

Behavioural Training

Mice were placed on water restriction 48 hours prior to starting behavioural training, with 1 hour water access provided 24 hours before the first session. Mice were trained six days per week, and on the day off they received 1 hour ad lib water access in their home cage. On training days, mice typically received all their water in the task, but were given additional water if required to maintain their body weight above 85 % of their pre-restriction baseline weight.

Mice were trained on a sequence of reversal learning tasks each with the same structure but a different physical port layout. Each reversal learning problem used three nose poke ports, out of the six or nine ports available in the operant box. One port was used for trial initiation, the other two were choice ports where reward could be obtained. During the initial training phase (Figure 1A) ports not used in the current task were covered. During recording sessions, ports used in all three tasks presented in the session were exposed throughout, and unused ports were covered.

Each trial started with the initiation port lighting up, until the subject poked it, after which two choice ports both lit up. Mice chose one of the choice ports which triggered a sound cue (250ms long) indicating the trial outcome, with a pure tone (5 kHz) indicating they will get a reward and white noise indicating reward omission. Reward was delivered at the termination of the auditory cue. A 2s inter-trial interval started once the animal left the port following reward consumption or a non-rewarded choice. One in four randomly selected trials was a forced choice trial, where a single randomly selected choice port lit up which the animals had to select. At any given point in time, one choice port had a high reward probability and the other one had low probability. Reward probability reversals were triggered 5-15 trials after the subject crossed a threshold of 75 % correct choices (exponential moving average, tau=8 trials).

In the initial training stage of experiment mice (Figure 1) encountered a single task (i.e., port layout) per session, and moved to the next task the session after they had completed 10 reversals on the current task. In each task, the first three reversals had reward probabilities of 0.9 and 0.1 at the good/bad choice ports. The fourth and fifth reversals had reward probabilities of 0.85 and 0.15, and the remaining reversals had reward probabilities of

0.8 and 0.2. In this phase each session was 30 minutes long and animals performed two sessions per day. The reward sizes during this stage were incrementally decreased from 15 ul in the beginning of the training to 4 ul, based on the current weight of the animal and its performance on the previous session. Each session started with a free reward given from each of the two choice ports. Mice were divided into three groups with each group starting on a different task layout. Sequentially presented layouts were chosen to be as different as possible, and the sequence of task layouts was counterbalanced across animals.

Once mice had completed 10 tasks during this initial training phase, we started to present multiple tasks in each session, to prepare them for recording sessions where we sought to record neurons across multiple tasks. Initially, mice were trained on two tasks in a session, in the nine port operant boxes subsequently used for recordings. Mice completed 12 different tasks in this stage, with the port layout used in each chosen to be as different from the previous one as possible. The reward probabilities in this phase were always 0.8 and 0.2 and the reward size was 4 ul. After mice completed two reversal blocks on one layout, choice ports that were going to be a part of the new task layout both lit up. Mice received a free reward from each of the new choice ports. Next, the new initiation port lit up signalling mice where they could initiate a trial.

Behavioural Training during Recordings

During recordings, subjects completed four reversal blocks in each of three different task layouts in every session. All task parameters were kept the same as during the two-layout per session training stage, with the exception that now subjects needed to complete four blocks on each task before they were moved onto a new one. As before, the task change was signalled by the two new choice ports lighting up and staying lit up until the subject collected a reward from each port. This was followed by the new initiation port lighting up. Port layouts used for tasks during recording sessions were designed to allow us to ask specific questions of the neural activity. As described in the Results section, all layouts were reflections of three basic layout types, each of which was presented once each session, in a randomised order (Figure 2B).

Electrophysiological Recordings and Spike Sorting

The silicon probes used were Cambridge Neurotech 32 channel probes. F series probes were used for hippocampus, P series for mPFC. For hippocampal recordings we started the recordings only after we lowered the probe enough to detect characteristic of hippocampus sharp wave ripples in the local field potential while the animal was asleep in its home cage. For mPFC recordings we lowered the probe $\sim$ 100um on every recording day. Fore more details on recording sites see Supplementary Figure 1. Neural activity was acquired at 30kHz with a 32-channel Intan RHD 2132 amplifier board (hardware bandpass filtering between 1.1 and 7603.8 Hz; Intan Technologies, USA) connected to an OpenEphys acquisition board via a flexible serial peripheral interface cable ('Ultra Thin RHD2000 SPI cable', Intan Technologies). Behavioural and ephys data were synchronised by sending sync pulses from the pyControl system to the OpenEphys acquisition board. Electrophysiological recordings were then spike sorted offline using KiloSort[56] and manually curated using phy (https://github.com/kwikteam/phy). Clusters were classified as single units and retained for further analysis if they had a characteristic waveform shape, showed a clear refractory period in its autocorrelation, were stable over time and were present only on nearby channels. We merged clusters only if there was a high similarity in waveforms and channels they came from, had a refractory period in their cross-correlation histograms and occupied similar areas in feature space or appeared to drift into one another.

Surgery and Histology

Subjects were taken off water restriction 48 hours prior to surgery, then anaesthetised with isoflurane (3 % induction, 0.5–1 % maintenance), treated with buprenorphine (0.1 mg/kg) and meloxicam (5 mg/kg), and placed in a stereotactic frame. A silicon probe mounted on a

Microdrive (Ronal Tools) was implanted into either mPFC (AP:1.95, ML:0.4 , DV:-0.8), or dCA1 (AP:-2 , ML:1.7 , DV:-0.7), and a ground screw was implanted above the cerebellum. Both of the DV coordinates are relative to the brain surface. Mice were given additional doses of meloxicam each day for 3 days after surgery, and were monitored carefully for 7 days post-surgery, then placed back on water restriction 24 hours before restarting task behaviour. At the end of the experiment, electrolytic lesions were made under terminal pentobarbital anaesthesia to mark the probe location, animals were perfused, and the brains fixed in formal saline for subsequent histology to identify lesion locations.

Data analysis

All analyses were carried out using custom written code in Python.

Time in Trial Alignment

Activity was aligned across trials by warping the time interval between trial initiation and choice to match the median interval across all recorded trials. Activity prior to trial initiation or after choice was not warped. Spike times that occurred between initiation and choice were converted into the aligned reference frame by linear interpolation between initiation and choice time. The firing rate of each neuron was calculated in the aligned reference frame at time points evenly spaced every 40ms, from 1 second before trial initiation to 1 second after trial outcome, using a Gaussian kernel with 40ms standard deviation. To compensate for the change in spike density due to time warping, spikes in the warped interval between initiation and choice were weighted by the stretch factor applied, prior to evaluating the firing rate.

Statistical Significance

The significance of the differences between brain areas in analyses reported throughout the paper was computed by shuffling the sessions of CA1 and PFC animals to obtain null distributions. Real differences in the data were compared against the 95th and 99th percentiles of such null distributions. To correct for multiple comparisons, the maximum differences between CA1 and PFC across time points was taken as a threshold for multiple comparison correction, such that value at each time step was compared not to its respective shuffled value at the same time step, but the biggest value at any time step. All comparisons also survived a group test obtained by shuffling animal identities between regions.

Representational Similarity Regression Analysis

We created representational similarity matrices which consisted of the Pearson correlation coefficients of neurons in 15 different task condition, defined by the trial stage, choice, outcome and task number (see Results section and Figure 3). Because neurons were not simultaneously recorded, we collapsed data across recording sessions for each brain region into a single matrix (cells x task events) and then calculated the correlation matrix across cells between different task events (i.e., representational similarity). We used a linear regression to model the patterns of representation similarity in the data as a linear combination of representation similarity design matrices (RDMs):

$$r_{i,j} = \beta_0 + \sum_{n=1}^{9} \beta_n RDM_{n(i,j)} + \epsilon_{i,j}$$

Where $r_{(i,j)}$ are elements of the RSA matrix and $RDM_{n(i,j)}$ are elements of the nth RDM. The set of RDMs used is shown in Figure 3D. Before regressing the correlation matrices onto the RDMs the diagonal elements from both were deleted and a constant matrix of ones was added to the design matrix to account for any condition independent correlation between neurons. We plotted the coefficients of partial determination (CPDs) from the regression model described above. The CPD was defined as:

$$CPD\left(RDM_i\right) = \left(SSE_{\sim i} - SSE_{\text{full model}}\right) / SSE_{\sim i}$$

Where $SSE_{\sim i}$ refers to sum of squares from a regression model excluding the RDM_i of interest and $SSE_{\text{full model}}$ is the sum of squares from a regression model including all the RDMs. CPDs describe how much unique variance does each RDM account for in the RSA matrix calculated from firing rates.

Surprise Measure

To investigate the time course of how quickly the firing rates of neurons change in response to layout changes (Supplementary Figure 4), we used the 'surprise' measure from the information theory:

$$s\left(x_{ij}\right) = \left(x_{ij} - \mu_k\right)^2 / \sigma_k^2$$

where x_{ij} is the firing rate of one neuron on a given trial i and task layout j, μ_k and σ_k are the baseline mean and the standard deviation of the firing rate of that neuron on a particular task layout. If $j = k$, then the $s(x_{ij})$ on each trial i is calculated based on the mean firing rate μ and standard deviation σ of the withheld trials from the same task. More precisely, to calculate how much the firings rates change during the same task layout $s(x_{ij})$ was calculated on the 10 trials before the task layout switch ('test' within task), where μ_k and σ_k were calculated on the 10 trials before those 'test' trials ('train' within task). If $j \neq k$, then the $s(x_{ij})$ on each trial i was calculated based on the mean firing rate μ and standard deviation σ of the withheld trials from a different task. So, to estimate how much the firings rates change after the task layout switch $s(x_{ij})$ was calculated on the 20 trials after the task layout switch ('test' between tasks), where μ_k and σ_k were calculated from the 'train' trials from a different task layout. This measure was calculated for each neuron separately and then averaged across all neurons for each brain region.

Singular Value Decomposition

Singular value decomposition (SVD) was performed using the numpy linalg.svd function in Python. SVD is a principal component analysis technique that decomposes any n x m matrix into a product of three matrices:

$$D = U\Sigma V^T$$

where D comprises the data matrix to be decomposed and the U Σ and V^T matrices have specific interpretations depending on the type and organisation of data in matrix D. The U Σ and V^T are computed based on the non-normalised covariances in the column space:

$$\begin{gathered} DD^T = \left(U\Sigma V^T\right)\left(U\Sigma V^T\right)^T \\ DD^T = \left(U\Sigma V^T\right)\left(V\Sigma U^T\right) \\ DD^T = U\Sigma^2 U^T \end{gathered}$$

and row space:

$$\begin{gathered} D^T D = \left(U\Sigma V^T\right)^T\left(U\Sigma V^T\right) \\ D^T D = \left(V\sum U^T\right)\left(U\Sigma V^T\right) \\ D^T D = V\Sigma^2 V^T \end{gathered}$$

where $DD^T U = U\Sigma^2$ and $D^T DV = V\Sigma^2$ are analogous to eigenvalue decomposition $AQ = Q$. These equations provide an intuition for what the U, Σ and V^T matrices mean. In the analyses of our data, matrices D were of neuron x timepoints* trial type dimensions. As DD^T is a non-normalised covariance in the column space, this means that the U singular vectors come from the eigendecomposition of the covariances between neurons (as column space in D is neuron number) and thus describe the neural patterns in the data (i.e., neurons that are active/silent together). $D^T D$ is non-normalised covariance in the row space, meaning the V singular vectors come from the eigen decomposition

of the covariances between time and trial type (as D row space is time and trial type) and thus describe the trial and time modes in the data (i.e., trial times/types that are represented similarly). The Σ is diagonal matrix and captures the overall strength of association between each U and V^T vectors in the data matrix D, hence how much loading there is of a particular neural mode together with its respective trial x time mode in the data.

Our goal was to use the SVD to test how well cellular and temporal patterns generalise across different tasks. To make the D matrix we averaged time warped trial firing rates for each neuron in A choice rewarded, A choice non-rewarded, B choice rewarded and B choice non-rewarded conditions and concatenated the data from all sessions for each region separately such that each matrix was had the neurons x time point in trial and condition dimensions. We performed the SVD on demeaned firing rates separately for each task and for cross-validation purposes performed the decomposition separately on the first half and second half of the task:

$$D_{ij} = U_{ij}\Sigma_{ij}V_{ij}^T$$

where i is the task number $i = 1, 2, 3$ and j is the half of the task the data is taken from $j = 1, 2$.

To test how well the neural and temporal patterns generalised between pairs of tasks we used the U_{i2}, V_{i2}^T from the second half of the first task but the activity matrix from the first half of the next task D_{i+11} to compute the $\Sigma_{\text{pred } i+1}$:

$$\Sigma_{\text{pred } i+1} = U_{i2}^T D_{i+11} V_{i2}$$

Cross-validation was computed in an analogous manner but based on the data from the same task. Selecting the second versus first half of the task data ensured there was no time confound in cross-validated results, as the between task analysis would have analogous time effects.

Since we had different number of neurons in each brain region, each Σ was normalised by the number of neurons recorded from the respective brain region. Computing the $\Sigma_{\text{pred } i}$ for the new D using U^T and V from a decomposition of a different D matrix results in a Σ matrix that is no longer diagonal. However, by looking at the diagonal elements we can estimate how much the U^T and V from one task explain the activity of neurons from a different layout or in the cross-validated version – same layout but second half of the task. More specifically, the diagonal elements tell us how strong the association between each U^T and V vectors computed on one of the D matrix is in a different data matrix D.

Hence, when we looked at how much variance the combination of neural and temporal components from one task explain in a different task, we looked at the cumulative diagonal elements in Σ_{pred}. Selecting only the diagonal elements from the Σ_{pred} also means that the meaningful comparison is between the cross-validated within task $\Sigma_{\text{pred } i}$ and between task $\Sigma_{\text{pred } i+1}$ as the cumulative sum of the singular values in either $\Sigma_{\text{pred } i+1}$ or $\Sigma_{\text{pred } i}$ will not add up to a 100 % because the matrix is no longer diagonal in either cross-validated or cross-layout conditions because the singular vectors U and V were computed on a different data matrix data matrix D. Thus, we normalised the test weights by the peak of the cross-validated cumulative weights.

To investigate how much variance either U or V singular vectors independently explain in the data matrix from a different task we removed the constraint for any Us or Vs to be linked to each other. We estimated how much variance the temporal components V on their own explain in the new task:

$$M_{\text{pred } i+1=}D_{i+11}V_{i2}$$

and cross-validated analogously:

$$M_{\text{pred } i=}D_{i2}V_{i1}$$

Similarly, to estimate how much variance the neural components U explained in a different task we computed:

$$M_{\text{pred } i+1} = U_{i2}^T D_{i+11}$$

And cross-validated analogously:

$$M_{\text{pred } i=} U_{i2}^T D_{i1}$$

To determine the significance of the differences between two regions we compared differences in the data between PFC and CA1 against a null distribution of differences between areas under the curve by shuffling the sessions between CA1 and PFC animals.