

1 A Cellular Basis for Mapping Behavioural Structure

2 Mohamady El-Gaby^{1*}, Adam Loyd Harris^{1,2}, James C. R. Whittington³, William
3 Dorrell⁴, Arya Bhomick⁵, Mark E. Walton², Thomas Akam², Tim E. J. Behrens^{1,5*}

4

5 1 Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

6 2 Department of Experimental Psychology, University of Oxford, Oxford, UK

7 3 Department of Applied Physics, Stanford University, Palo Alto, USA

8 4 Gatsby Computational Neuroscience Unit, University College London, London, UK

9 5 Sainsbury Wellcome Centre for Neural Circuits and Behaviour, University College London, London,
10 UK

11 *corresponding author(s): mohamady.el-gaby@ndcn.ox.ac.uk and timothy.behrens@ndcn.ox.ac.uk

12

13 **ABSTRACT**

14

15 To flexibly adapt to new situations, our brains must understand the regularities in the world,
16 but also in our own patterns of behaviour. A wealth of findings is beginning to reveal the
17 algorithms we use to map the outside world^{1–6}. In contrast, the biological algorithms that
18 map the complex structured behaviours we compose to reach our goals remain enigmatic.
19 Here we reveal a neuronal implementation of an algorithm for mapping abstract behavioural
20 structure and transferring it to new scenarios. We trained mice on many tasks which shared a
21 common structure organising a sequence of goals, but differed in the specific goal
22 locations. Animals discovered the underlying task structure, enabling zero-shot inferences
23 on the first trial of new tasks. The activity of most neurons in the medial Frontal cortex tiled
24 progress-to-goal, akin to how place cells map physical space. These “goal-progress cells”
25 generalised, stretching and compressing their tiling to accommodate different goal
26 distances. In contrast, progress along the overall sequence of goals was not encoded
27 explicitly. Instead a subset of goal-progress cells was further tuned such that individual
28 neurons fired with a fixed task-lag from a particular behavioural step. Together these cells
29 implemented an algorithm that instantaneously encoded the entire sequence of future
30 behavioural steps, and whose dynamics automatically retrieved the appropriate action at
31 each step. These dynamics mirrored the abstract task structure both on-task and during
32 offline sleep. Our findings suggest that goal-progress cells in the medial frontal cortex may
33 be elemental building blocks of schemata that can be sculpted to represent complex
34 behavioural structures.

35

36 **Introduction**

37

38 Our behaviours are highly structured. From cooking a meal to solving a maths problem, we
39 compose elaborate sequences of actions to achieve our goals. When elements of this
40 structure are common across tasks, we can build a schema; a generalised representation of
41 task states that allows us to instantly compose new behavioural sequences, and infer steps
42 never taken before^{7,8}. These behavioural structures can be complex and hierarchical, where
43 sequences of actions leading to individual goals are nested within a higher order structure
44 relating the goals to each other^{9,10}. To successfully execute a hierarchical behavioural
45 sequence, one must simultaneously track their position at all levels of the task hierarchy.

47

48 Lesion, imaging and neurophysiological studies implicate the frontal areas of the neocortex
49 in mapping task structure. This involves roles in forming a schema of task structure^{11–17},
50 generating complex behavioural sequences^{18,19}, encoding goals^{20,21} and simultaneously
51 tracking a working memory of multiple task variables^{22–24}. These findings are consistent with
52 a role of the frontal cortex in generating a map of abstract task structure: a representation
53 that reflects the conserved relationships between task states and generalises across
54 similarly structured tasks with different sensorimotor specifics. A key remaining challenge is
55 to derive biological algorithms that mechanistically explain how frontal activity generates
56 such abstract maps.

57

58 What mechanisms mediate the mapping of abstract task structure? Such mapping should
59 comprise neuronal dynamics that evolve as a function of progress in a task, rather than
60 related variables such as elapsed time or the number of actions taken. This would naturally
61 enable representations to generalise across goal-directed behavioural sequences that differ
62 in length and duration. Frontal neurons have been found to track progress relative to
63 individual goals^{25–27}, regardless of the location of the goal or the distance covered to reach
64 it²⁷. However, behavioural tasks are complex and often composed of multiple, hierarchically
65 organised goals. It remains unclear how neurons track progress along such complex tasks.
66 One way this can be achieved is via neurons that invariantly track progress in a sequence of
67 goals regardless of their sensorimotor specifics, in direct analogy to the invariant tracking of
68 progress towards individual goals. Indeed, some findings suggest that frontal neurons are
69 tuned to general task states rather than specific stimuli or actions^{11,28–31}. This has led to the
70 view that, in each new task, neurons encoding abstract task states are flexibly bound via
71 synaptic plasticity to those representing the detailed behavioural sequences to be
72 executed^{31–34}. Alternatively, a separate line of work on recurrent neural networks suggests
73 that such binding is not necessary and schematic inferences can be made in new scenarios
74 with no new plasticity. Here, details of new task examples are stored as patterns of neural
75 activity using network dynamics sculpted, through the learning of previous examples, by the
76 abstract structure of the task³⁵. While the mechanistic details of how such dynamics support
77 task-schema formation remain unclear, this strategy relies on representations that track
78 memories of previous actions and rewards^{22–24,35}. Whether and how such representational
79 logic relates to generating a schema that tracks an animal's progress in task-space remains
80 an open question.

81

82 Here we sought to elucidate a neuronal algorithm for mapping abstract task structure. We
83 trained mice on a series of tasks, each of which required visiting 4 goal locations in a
84 repeating, loop-like sequence. The sequential loop structure relating the goals remained the
85 same across tasks, while the goal locations, and hence the behavioural sequences needed
86 to navigate between them, changed. Mice used this abstract structure to perform zero-shot
87 inferences on the first trial of new tasks. Using multi-unit silicon probe recordings, we found
88 that neurons in the medial Frontal Cortex (mFC) tracked progress to the next goal,
89 regardless of the behavioural sequences used to reach it. Crucially, these neurons were
90 further sculpted by the task structure. Neurons were arranged into modules, where activity
91 along each module tracked progress in the higher order sequence of goals. Individual
92 neurons on a given module had mnemonic fields at a fixed lag from a particular behavioural
93 step. Thus, a module behaved like a memory buffer, creating dynamics that track progress in
94 task-space from a particular behavioural step. Each of these memory buffers was shaped by

95 the abstract task structure, reflecting the 4-reward loop, and hence allowed predicting the
96 animals' future actions a long time before they were made. These findings point to an
97 algorithm that uses Structured Memory Buffers (SMBs) to encode new behavioural
98 sequences into the dynamics of neural activity without needing associative binding. More
99 broadly, we propose that a schema mapping any complex behavioural structure can be
100 generated by sculpting task-naive progress-to-goal tuning to represent task-structured
101 memories of individual behavioural steps.
102

103 **Results**

104

105 **The ABCD paradigm: an abstract task structure guides rapid sequence learning and** 106 **inference**

107

108 We developed a task wherein goal-directed behavioural sequences are hierarchically
109 organised by an abstract structure (the “ABCD” paradigm). Mice (N=11) learned to navigate
110 to identical water rewards arranged in a sequence of 4 locations (**a**,**b**,**c** and **d**) on a 3x3 grid
111 maze (Figure 1a,b). The reward at each location only became available after the previous
112 goal was visited, so the goals had to be visited in sequence to obtain rewards. Once reward
113 **d** was obtained, reward **a** became available again, allowing the animal to complete another
114 loop. Each rewarded location (**a**,**b**,**c** or **d**) defined the beginning of a task “state” (A,B C or D
115 respectively; Figure 1a) and a single ABCD loop defined a trial of a task. A brief tone was
116 played upon reward delivery in state A, marking the beginning of a loop on every trial.
117 Animals encountered multiple tasks where the reward locations changed but the general
118 ABCD loop structure remained the same (Figure 1a). Crucially, task structure was made
119 orthogonal to the structure of physical space: the physical distance between two rewards on
120 the maze was not correlated with their task-space distance (Figure 1a,b, Extended data
121 figure 1a). The task therefore encouraged animals to separately learn the spatial sequences
122 leading to individual goals, and the higher order task structure that organises these goal-
123 directed sequences (Figure 1b).
124

125

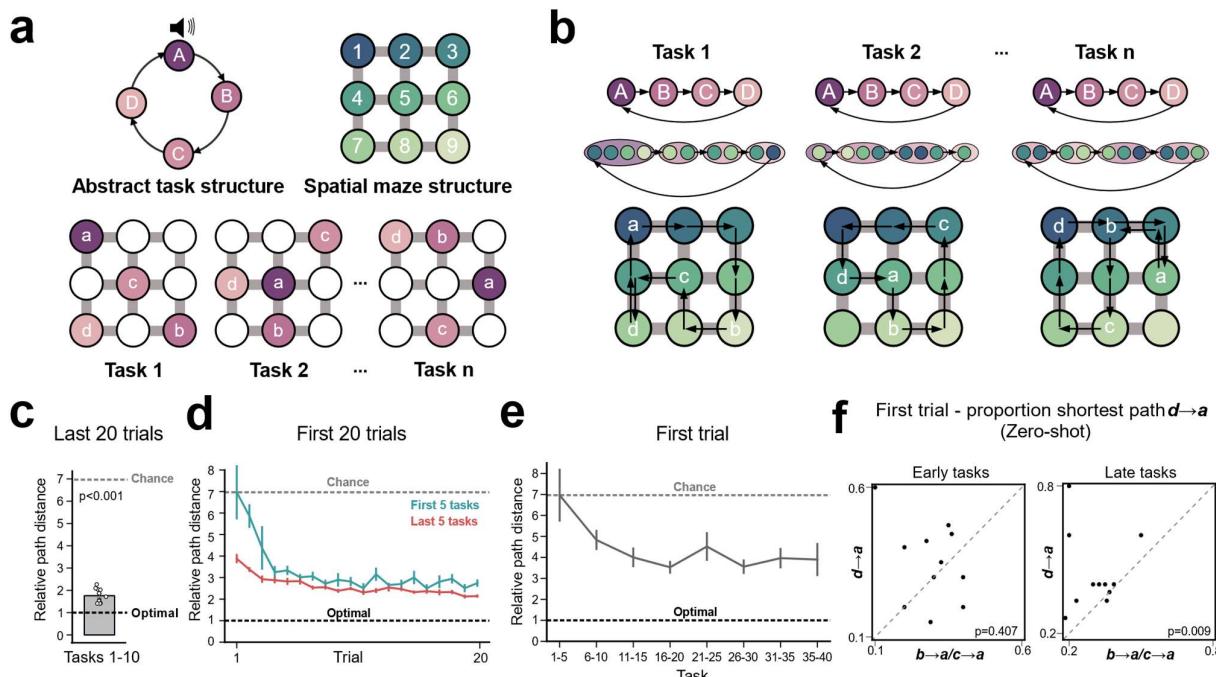
126 We first asked whether animals learned optimal sequences leading to individual goals. To
127 assess performance, we quantified either the ratio of the distance taken between two goal
128 locations compared to the shortest possible distance (“Relative path distance”), or the
129 proportion of trials where one of the shortest routes was taken (“Proportion shortest”). On
130 individual tasks, mice converged on a near-optimal policy that routinely took them between
131 goal locations via close-to-shortest routes (Figure 1c). This converged policy was highly
132 stereotyped, with animals taking only a subset of the available shortest routes (Extended
133 data fig 1b). In the first 10 tasks, we allowed animals to perform as many trials as needed to
134 converge on peak performance (Figure 1c; 70% shortest path transitions or 200 trial plateau;
135 mean number of trials per task: 325 ± 16). Subsequently, to encourage generalisation of task
136 structure, and to allow us to record from multiple tasks in a day, we moved animals to a high
137 task regime (tasks 11-40). Animals experienced 3 new tasks per day and hence could only
138 complete a fraction of the trials needed to reach peak performance (mean number of trials
139 per task: 38 ± 3). Despite this, animals still performed markedly above chance (Figure 1d).
140 Suboptimal performance was associated with a persisting preference for taking routes
around the maze for which animals had an *a priori* bias to take before exposure to the task

141 (Extended data figure 1c). Nevertheless, animals' performance improved across even the
 142 early trials of a given task (Figure 1d).

143

144 We next asked whether mice learn the higher order, abstract task structure. Initial
 145 performance on the earliest trials improved with the number of tasks completed (Figure 1d);
 146 Extended data figures 1d,e). This improvement was seen even on the very *first* trial of each
 147 new task (Figure 1e). This effect is consistent with animals transferring knowledge of task
 148 structure to rapidly learn new sequences. Notably, this task allows for a *direct* test of this
 149 structural knowledge transfer because of the ABCD loop. After discovering the 4 reward
 150 locations (**a**,**b**,**c** and **d**) on the first trial of a new task, animals that understand the task
 151 structure should then return directly to **a**. On the first trial, this transition (**d** → **a**) has never
 152 been experienced, so cannot be executed through repetitive learning or memory. Instead it
 153 must reflect a zero-shot inference based on abstract knowledge of the ABCD task structure.
 154 Remarkably, we found that experienced animals took the shortest path between **d** and **a** on
 155 the first trial more often than chance and more readily than premature returns to **a** from **b** or
 156 **c** (Figure 1f). This was not explained by any pre-existing biases in the animals' exploration of
 157 the maze (Extended data fig 1f), differences in analytical chance levels (Extended data fig
 158 1g) or differences in the distances of the **d**-to-**a** transition compared to those for **c**-to-**a** and
 159 **b**-to-**a** (Extended data fig 1h). Moreover, this zero-shot inference was associated with
 160 animals returning to the start of the loop after 4 rewards (**d**-to-**a**) rather than to later points of
 161 the loop (**d**-to-**b** or **d**-to-**c**; Extended data fig 1i). Thus animals not only waited until 4 rewards
 162 were obtained before returning to **a** (Figure 1f) but also more readily returned to **a** than to
 163 other reward locations after 4 rewards (Extended data fig 1i). Overall these findings suggest
 164 that mice learn an abstract, task-defined behavioural structure nesting multiple goals.

165



166

Figure 1 – Mice Learn an abstract task structure

167

a) Task design: animals learned to navigate between 4 sequential goals on a 3x3 spatial grid-maze. Reward locations changed across tasks but the abstract structure, 4 rewards arranged in an ABCD loop, remained the same.

171 b) Learning involved finding optimal sequential paths between each goal. Optimal paths differed
172 in length both within and across tasks but were always organised by an ABCD loop relating all
173 4 goals.

174 c) When allowed to learn across multiple sessions, animals readily reached near-optimal
175 performance in the last 20 trials, as demonstrated by comparing path length between goals to
176 the shortest possible path (i.e. computing a “relative path distance” measure). T-test against
177 chance (6.96): N=11 mice, statistic=-56.6 P=7.15x10-14 df=10. Chance level was calculated
178 empirically using the mean relative path distance across the first trial of the first 5 tasks.

179 d) Performance improved across the initial 20 trials of each new task. This improvement was
180 markedly more rapid for the last 5 tasks compared to the first 5 tasks. A two-way repeated-
181 measures ANOVA (N=11 mice) showed a main effect of Trial F=11.7 P=6.2x10⁻⁵, df1=19,
182 df2=190, Task F=26.8 P=4.2x10⁻⁵, df1=1, df2=10 and a Trial x Task interaction F=3.56
183 P=0.0201, df1=19, df2=190.

184 e) Performance on the very first trial improved markedly across tasks. A one-way repeated-
185 measures ANOVA (N=7 mice – N.B. only 7 of the 11 mice were presented with all 40 tasks)
186 showed a main effect of Task F=3.10 P=0.0100, df1=7, df2=42.

187 f) Animals readily performed zero-shot inference on the first trial of late tasks but not in early
188 tasks. The proportion of tasks in which animals took the most direct path from **d** to **a** on the
189 very first trial is compared to the same measure but for premature returns from **c** to **a** and **b** to
190 **a**. Wilcoxon test; Early tasks: N=11 mice, statistic=15.5, P=0.407; Late tasks: N=11 mice,
191 statistic=2.0, P=0.009

192 All error bars represent the standard error of the mean

193

194

195 **Progress to goal is a primary feature of Frontal task structure representations**

196

197 Animals learn the sequences leading to individual goals, and the abstract structure
198 organising these goals. What are the basic neural underpinnings of this hierarchical
199 learning? In order to perform such complex sequences, animals must track their position not
200 only in the physical space they are navigating, but crucially also their “progress” in task
201 space, that is the stage the animal has reached in a sequence of goal-directed behaviours.
202 Are mFC neurons tuned to task progress?

203

204 We used silicon probes to record the activity of mFC (prelimbic) neurons (Figure 2a;
205 Extended data figure 2a) from animals (N=5) performing late tasks (tasks 21-40), a stage
206 where we see robust evidence for task structure knowledge (Figure 1d-f). Each recording
207 day comprised 3 sessions with 3 new tasks (X, Y and Z) and a subsequent fourth session in
208 which the first task was repeated (X'). In addition, we recorded sleep sessions at the start
209 and end of the day and in between every session. We used a generalised linear model to
210 tease out the variables explaining mFC neuronal activity. The large majority (80%) of mFC
211 neurons were strongly and consistently tuned to the animal's progress towards a rewarded
212 goal, regardless of where the reward was spatially (Figure 2b-f). Neurons tiled goal-progress
213 space: some fired immediately after the animal reached its goal (early goal-progress cells),
214 others were most active between two goals (intermediate goal-progress cells) and others still
215 just before a goal was reached (late goal-progress cells; Figure 2c-e). This goal-progress
216 tuning was highly invariant across tasks with distinct reward locations (Figure 2e) and not
217 explained by simple monotonic tuning to speed nor acceleration (Figure 2f). Furthermore,
218 these neurons fired in relation to goal-progress relative to goals independently of elapsed
219 time or physical distance (Figure 2f). Thus, mFC neurons exhibit stimulus-invariant tracking
220 of progress towards individual goals.

221

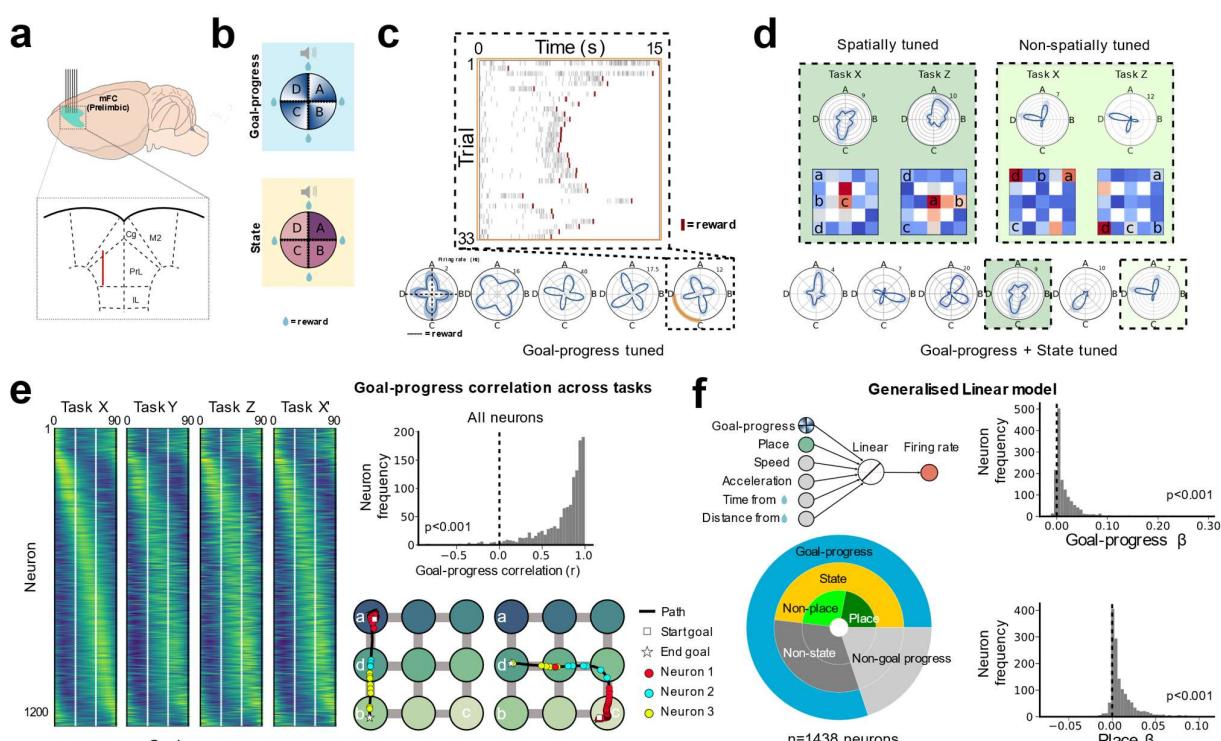
222 The invariant goal-progress tuning of mFC neurons is consistent with this region tracking
 223 progress in relation to *individual* behavioural goals, rather than physical distance or time.
 224 However, such goal-progress tuning alone is insufficient for tracking position in the higher
 225 order task structure organising multiple goals. We therefore leveraged the hierarchical
 226 structure of our ABCD task to ask whether mFC neurons are tuned to a given state in an
 227 individual ABCD task (e.g. state B). We found such “state” tuned neurons in abundance
 228 (56% of all neurons had state tuning in at least one task; Figure 2d,f; Extended data figure
 229 2b). Intriguingly, the large majority (87%) of these state tuned neurons were also goal-
 230 progress tuned (Figure 2f). Thus, task-state-tuned neurons are largely a subset of the more
 231 prevalent goal-progress-tuned neurons. A corollary of this is that a large proportion (60%) of
 232 goal-progress-tuned neurons also showed significant state preference in at least a single
 233 task (Figure 2d,f; Extended data figure 2b). Moreover, 46% of neurons tuned to both goal-
 234 progress and state were additionally tuned to place (Figure 2d,f; Extended data figure 2b).
 235 Thus, for 54% of cells with state and goal-progress tuning, state-preference was not
 236 explained by tuning to the animal’s current spatial location (Figure 2d,f; Extended data figure
 237 2b). These findings suggest that progress-to-goal is a key determinant of mFC neuronal
 238 firing and that a subset of these goal-progress cells are additionally tuned to a given state in
 239 a given task.

240

241

242

243



244

245 **Figure 2 – Progress-to-goal is a key feature of task-tuned neurons in the medial Frontal Cortex**

246 a) Multi-unit recording set-up: animals were implanted with silicon probes targeting 1mm of the
 247 anterior-posterior extent of the prelimbic region of the medial Frontal cortex. Inset: schematic
 248 of a coronal slice through the mFC showing electrode placement in the prelimbic region.

249 b) Schematics of polar plots showing a projection of neuronal activity onto the circular task
250 structure. “Goal-progress” refers to how much progress an animal has made towards a
251 rewarded goal location as a percentage of the time taken to reach this location, while “State”
252 refers to progress in the overall ABCD loop comprising all 4 goals. The radial axis represents
253 a neuron’s firing rate, while the angular axis represents progression along the task states.
254 Dashed lines along the cardinal directions represent the times reward was obtained (goal was
255 reached) in each state, with the vertical line at zero representing reward **a**, i.e. the start of
256 state A; going clockwise, the remaining dashed lines represent reward locations **b**, **c** and **d**,
257 and hence the starts of states B C and D respectively. To represent task position, bins along
258 the angular axis increment with relative progress between goals rather than raw elapsed time.
259 c) Neurons display consistent tuning to the progress of the animal’s relative to any goal (“goal-
260 progress tuned”). Inset: a raster plot of firing activity in one state “C” of a goal-progress cell,
261 showing firing consistently shortly before a goal is reached.
262 d) Some goal-progress-tuned cells are additionally modulated by the state in a given task (“goal-
263 progress + State tuned”). Inset: Polar plots and spatial maps for two goal-progress + State
264 tuned neurons across two distinct task configurations. The neuron on the left is spatially tuned
265 while the neuron on the right is non-spatially tuned.
266 e) Goal-progress tuning is consistent across tasks that differ in reward locations. Left: The
267 average firing rate vector of all neurons relative to an individual goal (from goal “n” to goal
268 “n+1”; averaged across all states). Animals experience 3 tasks per day (tasks X, Y and Z) and
269 then a further 4th session which task X is repeated (task X’). Each row represents a single
270 neuron and the neurons are arranged on the y axis by their peak firing goal-progress in task
271 X. This alignment is largely maintained in tasks Y and Z as well as a later session of the first
272 task (X’). White dashes indicate early intermediate and late goal-progress-cutoffs. Top right: A
273 histogram showing the mean goal-progress-vector correlation across tasks for each neuron.
274 One-sample T test against 0: N=1230 neurons; statistic=91.56; P=0.0, df=1229. Note that the
275 neurons used in this panel are those that were tracked for all 4 sessions on a given day.
276 Bottom right: two example paths, each between a pair of rewarded goals and overlaid spiking
277 of 3 goal-progress-tuned mFC neurons tuned to early goal-progress (neuron 1) intermediate
278 goal-progress goal-progress (neuron 2) and late goal-progress (neuron 3) regardless of the
279 goal locations or trajectory taken.
280 f) Top left: a schematic of the variables inputted into a generalised linear model that predicts
281 neuronal activity across tasks and states. Bottom left: Pie-chart showing the results of a
282 generalised linear model capturing variance as a function of goal-progress, place, speed,
283 acceleration, time from reward and distance from reward. Plot shows proportions of neurons
284 with significant regression coefficient values for goal-progress, place and their conjunctions. It
285 also shows proportions of state tuned neurons derived from a separate z-scoring analysis
286 (More details in Methods under “Tuning to basic task variables”). Proportion of all neurons
287 that are goal-progress cells: 80%; Two proportions test: N=1438 neurons, z=40.7, P<0.00.
288 Proportion of all neurons that had state tuning in at least one task: 56% Two proportions test:
289 N=1438 neurons, z=29.5, P<0.001. 87% of all state-tuned neurons were also goal-progress
290 tuned; Two proportions test: N=798 neurons, z=32.8, P<0.001. Proportion of goal-progress-
291 tuned neurons that also showed significant state preference in at least a single task: 60%;
292 Two proportions test: N=1152 neurons, z=28.2, P<0.001. Proportion of neurons tuned to goal-
293 progress and state that were also tuned to place: 46%, Two proportions test: N=692 neurons,
294 z=17.4, P<0.001.
295 Top right: A histogram showing the mean regression coefficient values for *goal-progress* as a
296 regressor across task/state combinations for each neuron. One-sample T test against 0:
297 N=1438 neurons; statistic=20.52; P=2.71x10⁻⁸², df=1437. Bottom right: A histogram showing
298 the mean regression coefficient values for *place* as a regressor across task/state
299 combinations for each neuron. One-sample T test against 0: N=1438 neurons; statistic=24.55;
300 P=1.96x10⁻¹¹¹, df=1437

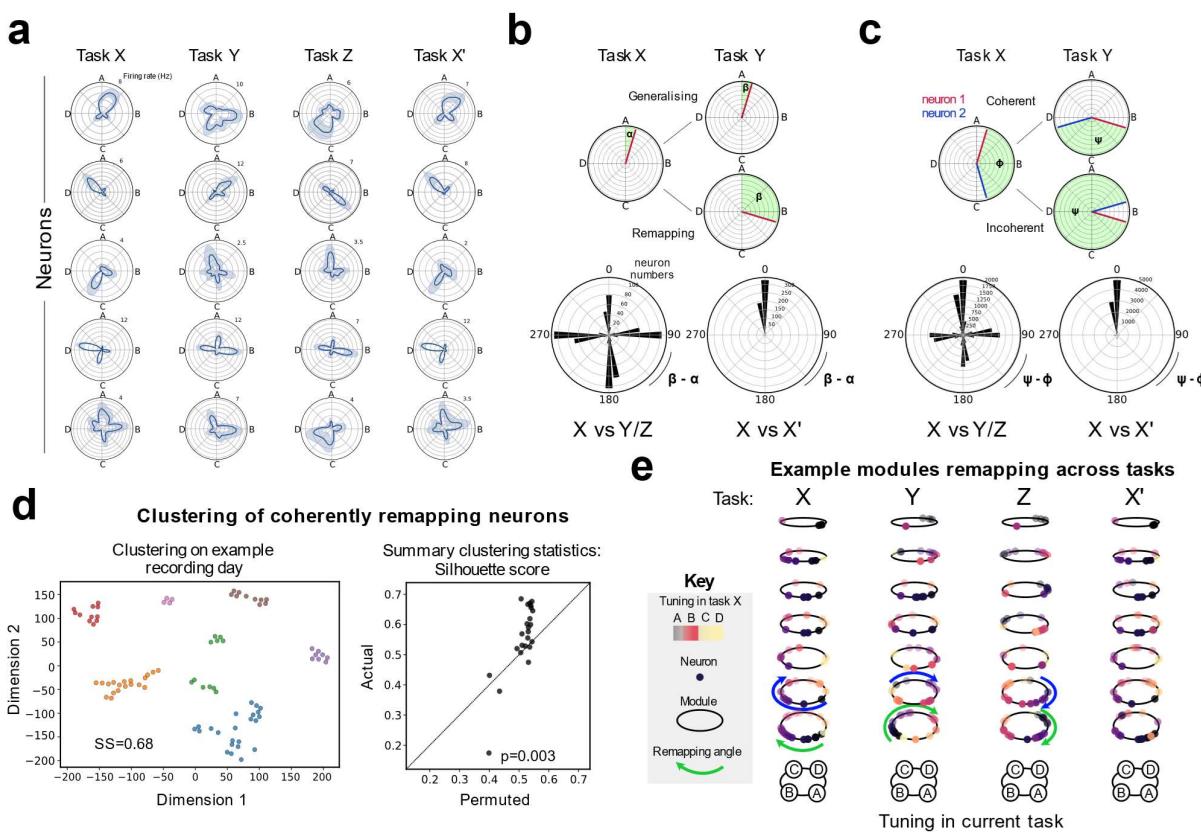
301

302 Modular organisation of mFC task structure mapping

303

304 Neurons in the mFC delineate task state on individual tasks. One view holds that, in order to
305 support transfer of task structure knowledge across distinct examples, such state tuning
306 should explicitly generalise across different tasks^{32,34}. This would then allow an associative
307 solution to the abstraction problem, whereby “abstract state” cells would (re)bind flexibly to
308 neurons representing specific spatial goals in order to encode new tasks^{32,34}. To investigate
309 this possibility, we asked whether the state tuning of mFC neurons, which reflects the
310 animal’s position in the ABCD loop of one task, is invariant across tasks. Intriguingly, rather
311 than invariant state tuning, we found that neurons consistently remapped in task space. In
312 contrast to their goal-progress tuning, which was invariant across tasks (Figure 2), mFC
313 neurons did not retain their state preference across tasks (Figure 3a,b). Notably, state-tuned
314 neurons with no discernable spatial tuning also remapped across tasks (Extended data
315 Figure 3b). Crucially, when we compared state tuning across different sessions of the same
316 task (X and X’), state tuning was highly conserved, despite two intervening sessions with
317 different tasks and different state tuning (Figure 3b). This was also seen when only
318 considering non-spatial neurons (Extended data Figure 3b). Overall these results indicate
319 that, while neurons in the mFC invariantly map progress to a given goal, they do not
320 invariantly map abstract task states in a higher order structure relating the sequential goals.
321

322 While remapping precludes a model in which neurons are invariantly tuned to the ordinal
323 position of each goal, it may nevertheless still reflect an invariant structure. Neurons could
324 maintain a consistent ring-like arrangement which maps the abstract task structure but
325 rotates coherently across different tasks (e.g. if all A neurons become C neurons then all D
326 neurons should become B neurons...etc). Such a representation could still allow tracking
327 abstract task position and the retrieval of concrete behavioural states through an associative
328 mechanism. We therefore asked whether the remapping of mFC neurons across tasks
329 reflects a coherent structure between neurons in the population. We measured the degree of
330 coherence between pairs of mFC state-tuned neurons: how likely is it that a pair of neurons
331 will maintain their relative state preference across tasks? Only a small proportion of pairs,
332 but significantly above chance level, showed coherent remapping across tasks (Figure 3c).
333 This pairwise coherence was true for non-spatial neurons (Extended data figure 3c) and
334 when neurons were tuned to distant points in task space (Extended data figure 3d).
335 Moreover, coherent pairs of neurons showed a trend towards being closer anatomically
336 (Extended data figure 3e). The partial pairwise coherence between mFC neurons suggests
337 that such neurons might be organised into modules; groups of neurons that maintain their
338 tuning relationships across tasks, akin to the modular arrangement of grid cells mapping
339 physical space³⁶. To investigate this, we used a clustering approach. We defined a distance
340 metric between pairs of cells which assigned low distances between cells that remapped
341 coherently between tasks, and high distance between cells that remapped incoherently. We
342 then applied a low dimensional embedding (tSNE) on the resulting distance matrix, followed
343 by hierarchical clustering. mFC neurons were significantly clustered (Figure 3d,e), indicating
344 that they were organised into modules which remap coherently across tasks. Overall, these
345 findings suggest that mFC neurons don’t generalise their state tuning relationships as a
346 coherent whole but are instead organised into modules that conserve their within-module,
347 but not between-module, tuning relationships across tasks.
348



349

350

351 **Figure 3 – Medial Frontal neurons are organised into task-space modules**

352 a) Polar plots showing state tuning of example neurons across multiple tasks. Each row is a
 353 neuron and each column is a session. Neurons readily remap their state tuning but maintain
 354 their goal-progress preference across tasks. State preference is maintained across different
 355 sessions of the same task (X vs X').

356 b) Top: A schematic showing how the difference in tuning angles for the same neuron across
 357 sessions is quantified. Bottom left: Polar histograms show that state-tuned neurons remap by
 358 angles close to multiples of 90 degrees, as a result of conserved goal-progress tuning and the
 359 4 reward structure of the task. No clear peak at zero is seen relative to the other cardinal
 360 directions when comparing sessions spanning separate tasks (Two proportions test against a
 361 chance level of 25% N=831 neurons; mean proportion of generalising neurons across one
 362 comparison (mean of X vs Y and X vs Z)=20%, $z=9.04$, $P<0.001$; i.e. significantly lower than
 363 chance). Bottom right: Neurons maintain their state preference across different sessions of
 364 the same task (X vs X' Two proportions test against a chance level of 25% N=624 neurons;
 365 proportion generalising=87%, $z=29.0$, $P<0.001$).

366 c) Top: A schematic showing how the difference in relative angles between pairs of neurons
 367 across sessions is quantified. Bottom left: Polar histograms show that the proportion of
 368 coherent pairs of state-tuned neurons (comprising the peak at zero) is higher than chance but
 369 less than 100%, indicating that the whole population does not rotate coherently. Two
 370 proportions test against a chance level of 25% N=14930 pairs; mean proportion of coherent
 371 neurons across one comparison (mean of X vs Y and X vs Z)=32%, $z=59.8$, $P<0.001$).
 372 Bottom right: As expected from panel b, the large majority of state-tuned neurons keep their
 373 relative angles across sessions of the same task (X vs X'; Two proportions test against a
 374 chance level of 25% N=10722 pairs; proportion coherent=79%, $z=109.2$, $P<0.001$).

375 d) Left: Example from a single recording day showing the result of tSNE embedding and
 376 hierarchical clustering derived from a distance matrix quantifying cross-task coherence

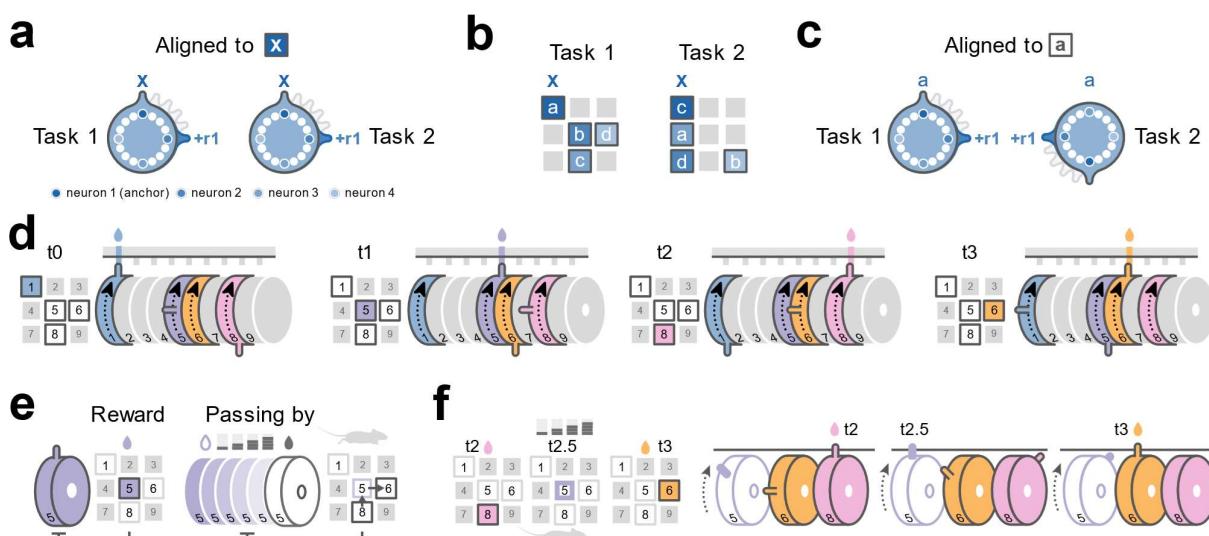
377 relationships between state-tuned neurons. Each dot represents a neuron. Right: Summary
378 silhouette scores for the clustering for real data compared to permuted data that maintains the
379 neuron's goal-progress preference and initial state distribution. Each dot is a recording day.
380 Wilcoxon test: N=24 recording days; statistic=50.0, P=0.003
381 e) Visualisation of tuning relationships between clusters computed in a single recording day.
382 Each dot is a neuron and each ring is a cluster derived from the analysis in panel d. The
383 colour code represents the tuning of the neurons in task X. The x,y position defines the tuning
384 in each task. The z position corresponds to cluster ID. Note that the ordering along the z axis
385 is arbitrary. Neurons rotate (remap) in task space while maintaining their within-cluster tuning
386 relationships but not cross-cluster relationships across tasks.
387
388

389 **Structured memory buffers: a unified model for behavioural schema and sequence
390 memory**

391
392 The modular arrangement of mFC neurons indicates that the entire population isn't anchored
393 to a single invariant reference point (e.g. state A). Instead, because they rotate
394 independently, each module is anchored to a distinct reference point. What could these
395 reference points be? We reasoned that the strong tuning of some frontal neurons to both
396 goal-progress and spatial location (Figure 2) could offer an answer. Each module of neurons
397 could be anchored to a particular conjunction of goal-progress and place (e.g. early goal-
398 progress at location 1; Figure 4a) through a subset of "anchor" neurons tuned to that specific
399 goal-progress/place combination. The other "non-anchor" neurons in the module would then
400 each fire at a specific lag in task-space from when the animal visits the goal-progress/place
401 anchor (Figure 4a). This amounts to a change in reference point for each module: from the
402 abstract states (ABCD) to a module-specific anchor. Under this scheme, the apparent
403 remapping seen when aligning activity to abstract states (ABCD; Figure 3) occurs because
404 animals visit the goal-progress/places in a different sequence in each task (Figure 4b,c). If in
405 one task location 1 is rewarded in state A and in another the same location is rewarded in
406 state C, a module anchored to early goal-progress in location 1 would appear to rotate by
407 180 degrees across tasks when aligned to the abstract states (Figure 4c). All neurons on this
408 module, not just the anchor neurons, would rotate by the same amount. The anchor
409 neurons, which are tuned to a particular goal-progress/place conjunction, would remap in a
410 way explained by their spatial tuning. However, neurons further from the anchor along a
411 given module would remap in one task in a way that is not explained by their spatial map in
412 another task (Figure 4 b,c) just as seen empirically (Extended data figure 3b,c). Each
413 module with a different anchor would rotate by a different amount depending on when the
414 goal-progress/place encoded by the anchors are encountered in the behavioural trajectory of
415 the current task. Thus, we posit that the true invariance of state neuron activity can only be
416 seen when realigning their activity to their putative goal-progress/place anchors (Figure 4a).
417

418 A key implication of this model is that each module is a *memory buffer* for visits to a
419 particular anchor. Because it is a combination of location and goal-progress, an anchor
420 represents a particular behavioural step. The flow of activity along a given module answers
421 the following question: how much of the overall *task* (i.e. the 4 reward sequence in our
422 ABCD task) has elapsed since visiting a given behavioural step? (for example since visiting
423 early goal-progress in location 1; Figure 4a,d) Because there are neurons for every task-
424 space lag from a given behavioural step within each module, and modules for every possible
425 behavioural step, there are always a subset of active neurons that are tied to each visited

426 behavioural step (Figure 4d). The instantaneous mFC activity therefore always encodes the
427 entire sequence of behavioural steps. Moreover, because there are anchors representing
428 behavioural steps at intermediate goal-progress, not only at rewards, this encoded sequence
429 is the entire behavioural sequence executed by the animal (i.e. the route taken through the
430 maze), not just the sequence of 4 reward locations (Figure 4e,f). These modules are
431 organised by task structure in two ways. Firstly, the strong goal-progress tuning of mFC
432 neurons means that activity on the module evolves as a function of the number of goals
433 obtained (i.e. the number of goal-progress cycles completed). The modules therefore track
434 true task-progress rather than other dimensions such as elapsed time or distance travelled.
435 Secondly, a module is shaped by the structure of the task, in our case a four reward ring.
436 Once an animal completes a full trial (i.e. traverses 4 goals in the ABCD task) since it last
437 visited a given behavioural step, activity along the module will circle back to the anchor point
438 (Figure 4d,f). This means that the memory buffers are “structured”: they are internally
439 organised to reflect the abstract structure of the task, a 4-reward loop in our case. We refer
440 to this over-arching mechanism as the Structured Memory Buffer (SMB) model.
441
442 The SMB model posits that activity along mFC modules could be used to guide the
443 execution of task-paced behavioural sequences. Once an animal completes a full trial since
444 it last visited a given behavioural step, either the anchor neurons themselves or neurons
445 close to them along a given module could be used as output neurons that bias the animal to
446 return to the behavioural step represented by the anchor. Collectively, activity dynamics
447 along *all* active modules would allow retrieval of a sequence of behavioural steps (policy) to
448 solve a given task. Consequently, any new sequence with the same structure can be
449 mapped in a programmable way, simply by reconfiguring the order in which modules are
450 activated. These network dynamics are sufficient to encode new tasks, without needing new
451 plasticity. For simplicity, we have assumed so far that each neuron has a single anchor with
452 a single lag. However, in principle the same computational logic can be used even when
453 individual neurons respond to a combination of different anchors and lags. The read-out in
454 such a scenario would involve combinatorial activity across anchor neurons across multiple
455 modules.
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470



471
472

473 **Figure 4 – The Structured Memory Buffers model**

474 a) A hypothetical ring-shaped module of neurons “anchored” to location 1 on the maze (marked
475 by an X) in early goal-progress (i.e. during reward consumption). The ring therefore
476 represents a memory buffer for this anchor. Aligning the ring of neurons by this anchor
477 reveals the invariant relationships between neurons across any two tasks where location 1 is
478 rewarded. 4 neurons are highlighted; the dark blue anchor neuron (neuron 1) and 3 other
479 neurons firing at lags of 90 degrees (i.e. one reward away (one state away); neuron 2), 180
480 degrees (i.e. two rewards away; neuron 3) and 270 degrees (i.e. three rewards away; neuron
481 4) from the anchor. Note that other neurons at all other task lags, not just multiples of 90
482 degrees, are also present on the ring (white circles). A bump of activity is initiated when the
483 goal-progress/place anchor is visited (top) and moves around the ring paced by the animal’s
484 progress in task space. In this task, because there are four rewards, the bump moves by a
485 total of 90 degrees around the ring after each reward (e.g. the bump reaches neuron 2 after
486 the animal has received 1 further reward since receiving reward in location 1).
487 b) Two example ABCD tasks that share one reward location (location 1, marked by an X). The
488 shaded regions correspond to the hypothetical spatial firing fields of each of the 4 neurons
489 shown in panel a - shading corresponds to the shading of the neurons on the ring in panel a.
490 While the anchor neuron (neuron1; dark blue) fires consistently at the same goal-
491 progress/place conjunction across tasks, other neurons (neurons 2-4; lighter shades of blue)
492 will fire in different locations in the two tasks. This is because they primarily encode elapsed
493 task progress from early reward in location X, rather than physical space. So for example,
494 neuron 4 (the lightest shade of blue) consistently fires 3 rewards from the time the animal got
495 reward in location 1: this means it fires in location 6 (reward **d**) in task 1 as it was 3 rewards
496 from reward in location 1 (reward **a**); in task 2 this neuron fires in location 9 (reward **b**) as this
497 is 3 rewards from reward in location 1 (reward **c**).
498 c) The same ring, when aligned by the abstract task states (e.g. aligned to start of state A; i.e.
499 reward in location **a**), appears to rotate by 180 degrees across tasks. This is a direct result of
500 reward in location X corresponding to the start of a different state in each task (state A in task
501 1 and state C in task 2).
502 d) A time series showing the flow of activity along 4 rings, each anchored to one of the 4
503 rewarded locations in task 1. A bump of activity is initiated when the goal-progress/place
504 anchor is visited (top) and moves around the ring paced by the animal’s progress in task
505 space. When it circles back close to the start, it biases the animal to return to this anchor.

506 Multiple rings have active bumps at any one time, thereby simultaneously tracking the history
507 of different goal-progress/place anchor visits (i.e. the history of previous behavioural steps).
508 e) As well as rings tracking task-progress from behavioural steps involving a rewarded place (a
509 conjunction of a place with early goal-progress), there are also rings tracking task-progress
510 from places conjoined with intermediate and late goal-progress. The anchors of these rings
511 are activated when the animal passes through a location, not when it is rewarded, but at a
512 defined, non-zero progress relative to the goal.
513 f) Non-zero goal-progress anchored rings (e.g. purple outline) allow tracking task-progress from
514 behavioural steps in between two goals. Hence, across all rings, a history of the entire
515 sequence of steps taken by the animal, not just the sequence of reward locations, is encoded
516 at any one point in time.
517

518 The SMB model is computationally attractive because it unifies frontal cortex functions in
519 behavioural schema formation and sequence memory, while offering a programmable way of
520 encoding and retrieving sequential policies. It is also empirically attractive because it
521 explains the spatial-tuning-independent state preference (Figure 2d,f), remapping (Figure
522 3a,b) and the modular arrangement (Figure 3c-e) of the neurons across tasks. Crucially, the
523 model makes a number of new empirical predictions about the tuning of single neurons, their
524 relationship to behavioural choices and their organisation at the population level. We explore
525 these predictions in turn below.

526

527 **1-Mnemonic fields in task space**

528 The SMB model proposes that, instead of neurons being anchored to the abstract task
529 states (A,B,C or D), they instead encode how much task-space has elapsed (i.e. how much
530 task progress has been made) since the animal visited a specific behavioural step. These
531 neurons should therefore maintain an invariant task-space lag from their module's anchor
532 (behavioural step) across tasks. For example, a neuron would always fire 2½ states from the
533 time the animal received reward (i.e. early goal-progress) at location 3, regardless of where
534 the animal is physically at this point. We tested this prediction using three complementary
535 approaches. In all of these analyses, we concatenated two recording days, giving a total of
536 up to 6 new tasks per neuron.

537

538 In the first approach, we implemented a linear regression model to predict the state tuning of
539 neurons across tasks. For each neuron, the model describes state-tuning activity as a
540 function of all possible behavioural steps (conjunctions of goal-progress and place) and task
541 lags from each possible behavioural step. Thus a neuron could fire at a particular
542 behavioural step but also at a non-zero lag in task space from this behavioural step. We
543 trained the model on all tasks but one (training tasks) and then used the resultant regression
544 coefficients to predict the activity of the neuron in a left out (test) task. To ensure our results
545 are due to task-lag preference and not the powerful effect of goal-progress tuning (which is
546 easy to predict across tasks due to its invariance; Figure 2 e,f), both training and cross-
547 validation were only done in the preferred goal-progress of each neuron. Using this
548 approach we were able to predict the state preference of most state-tuned neurons on
549 individual tasks as evidenced by a strong rightward (positive) shift in the correlation between
550 predicted and actual state tuning in the test task (Figure 5a,b; Extended data Figure 4a;
551 Extended data Figure 5a). Crucially, this rightward shift was also seen when only
552 considering activity at non-zero lags relative to all anchors (Figure 5b, Extended data Figure
553 5a,b). Neurons were distributed across all lags from the anchors, with an overrepresentation
554 of zero lag neurons, corresponding to the anchors (behavioural steps) themselves (Extended

555 data Figure 5c). Moreover, we found neurons anchored to all goal-progress/place
556 combinations (Extended data Figure 5d).

557

558 The second approach involves assessing the spatial tuning of mFC state-tuned neurons.
559 Classically, spatial maps relate the activity of neurons to where the animal is located in the
560 *present*, that is with zero lag in task space¹. Our model proposes the existence of neurons
561 tuned to where the animal was a set amount of task lag in the *past*. The logic of this is that,
562 while spatial neurons should consistently fire at the same locations(s) in the present (i.e. at
563 zero lag from the present), neurons that track a memory of the anchor will instead
564 consistently fire in relation to when the animal visited the anchor at a fixed lag in task space.
565 They will therefore have a peak in their cross-task spatial correlation at a fixed, non-zero
566 task lag in the past (Figure 5c; Extended data Figure 4b). Unlike the first analysis, this
567 approach assumes a single lag from an anchor per neuron, but still allows the anchor to be a
568 combination of locations. To quantify this effect, we again used a cross-validation approach,
569 this time using training tasks to calculate the lag at which cross-task spatial correlation was
570 maximal, and then measuring the correlation between the spatial maps in the left out (test)
571 task and the training tasks at this lag. This correlation was again strongly right-shifted
572 (Figure 5d), which was the case even when considering only neurons with non-zero lag
573 relative to the animal's current location (Figure 5d, Extended data Figure 5e,f).

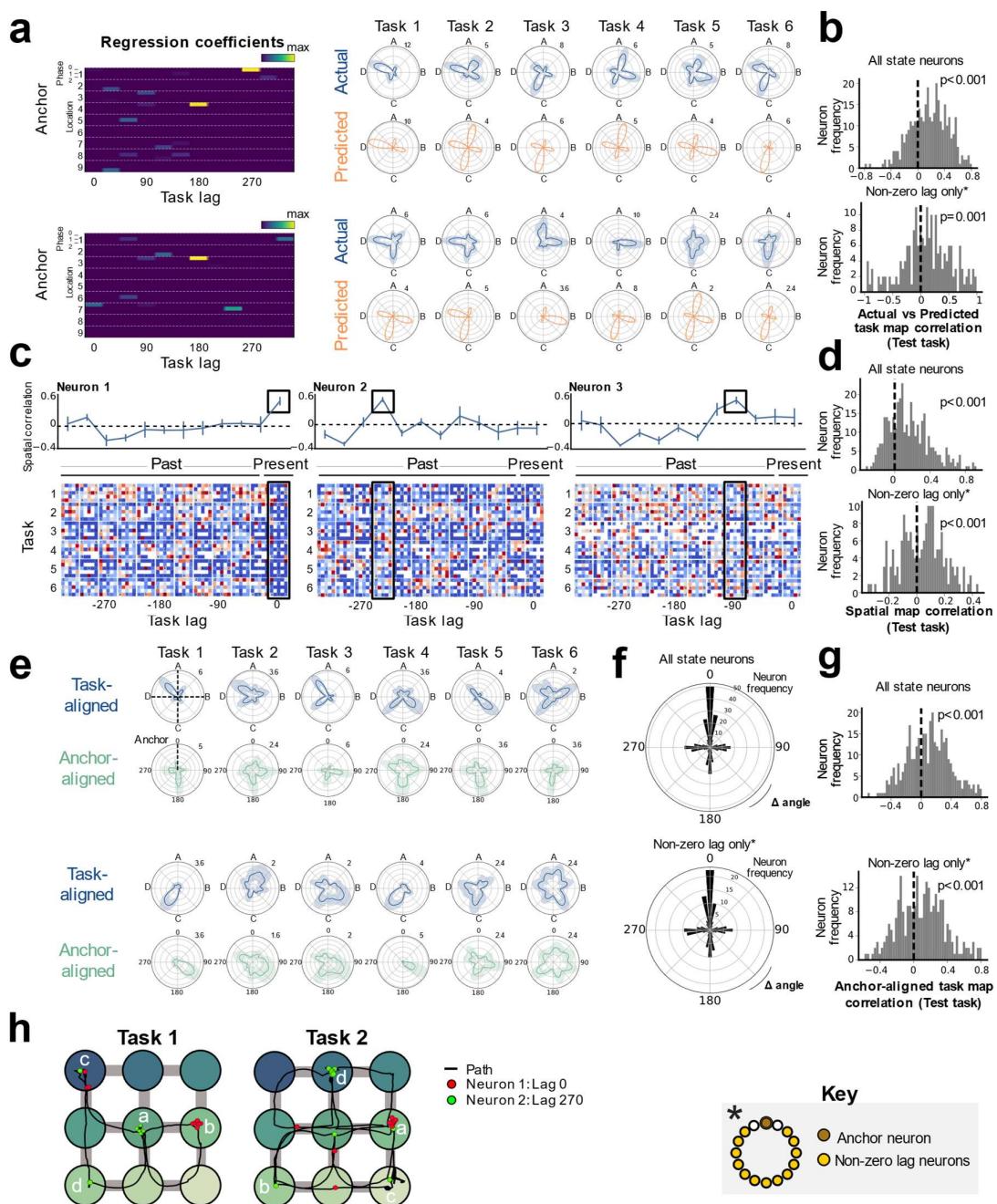
574

575 The third approach is effectively the reverse of the second. Instead of comparing spatial
576 tuning aligned to particular lags, it compares lag-tuning when aligned to particular anchor
577 visits (Figure 5e-h). Unlike the first two analyses, this approach assumes a single anchor per
578 neuron. We fitted the anchor by choosing for each neuron the goal-progress/place
579 conjunction which maximises the correlation between lag-tuning-curves in the training tasks,
580 and again used cross-validation by assessing whether this anchor leads to the same lag-
581 tuning in a left out task (Figure 5e; Extended data Figure 4c). Using this cross-validation
582 approach, we found significant alignment between firing distances from the best anchor in
583 the training and test tasks (Figure 5f), which was crucially seen even when only considering
584 non-zero lag neurons (Figure 5f). This was in stark contrast to aligning the activity by
585 abstract state (ABCD), which showed no peak at zero relative to other cardinal directions
586 (Figure 3b). To quantify the degree of alignment further, we measured the correlation
587 between the activity of neurons in their preferred goal-progress between the test and training
588 tasks. The resultant distribution was again right-shifted, even for non-zero lag neurons
589 (Figure 5g, Extended data Figure 5h,i). Thus, converging lines of evidence indicate that mFC
590 neurons form mnemonic fields that track *task-progress* from specific behavioural steps.

591

592

593



594

595

596 **Figure 5 – Medial Frontal neurons track task-progress from specific behavioural steps**

597 a) Regression analysis reveals neurons with mnemonic fields in task space from a given goal-
 598 progress/place conjunction, alongside neurons directly tuned to a goal-progress/place. This
 599 allows predicting state-tuning and its remapping across tasks. The regression coefficients are
 600 shown on the left with the actual and predicted activity of the neurons shown on the right.
 601 Left: The top neuron has its highest coefficients at 180 degree task lag from early goal-
 602 progress in location 4 and 270 degrees task lag from early goal-progress in location 1. The
 603 bottom neuron has its highest coefficients at 180 degree task lag from early goal-progress in
 604 location 3 and 240 degrees task lag from intermediate goal-progress in location 7.
 605 Right: Actual (blue) and model-predicted (orange) activity across 6 tasks

606 b) Histograms showing the right shifted distribution of mean cross-validated correlation values
607 between model-predicted (from training tasks) and actual (from a left out test task) activity.
608 Top: this correlation is shown for all state-tuned neurons and Bottom: only state-tuned
609 neurons with non-zero-lag firing from their anchors (i.e. state-tuned neurons with all of the
610 three highest regression coefficient values at non-zero lag (30 degrees either side of the
611 anchor). To avoid contamination due to potential residual spatial-tuning, for non-zero lag
612 neurons, we only use regression coefficient values more than 30 degrees in task space either
613 side of the anchor point to predict the state tuning of the cells. T test against 0: All state-tuned
614 neurons N=340 neurons, statistic=11.1, P=1.36x10⁻²⁴, df=339; Non-zero lag state-tuned
615 neurons N=194 neurons, statistic=3.26, P=0.001, df=193.

616 c) Lagged spatial field analysis. Example plots showing spatial maps for 3 neurons. Bottom:
617 Each row represents a different task and each column a different lag in task space. The
618 activity of each neuron is plotted as a function of the animal's current location (far right
619 column for each cell) and at successive task space lags in the past for the remaining
620 columns. For example, the second to last column is the firing of the cell in relation to where
621 the animal was 30 degrees (1/3rd of a state) back from the current point in task space, 7th to
622 last column is the firing of the cell in relation to where the animal was 180 degrees (2 states)
623 back from the current point in task space...etc. Top: the correlation of spatial maps across
624 tasks at each lag. The neuron on the left is an anchor cell (goal-progress/place cell), as seen
625 by the correlation peak at zero lag, while the middle and right-most neurons are neurons
626 lagged by 240 and 90 degrees from their anchors respectively.

627 d) Histograms showing the right shifted distribution of the mean cross-validated spatial
628 correlations between maps at the preferred lag (from training tasks) and the spatial map at
629 this lag from a left out test task for all state-tuned neurons (top) and only non-zero lag
630 neurons (bottom). Non-zero lag neurons are those with a peak spatial map correlation more
631 than 30 degrees in task space either side of zero lag in the training tasks. T test against 0: All
632 state-tuned neurons N=350 neurons, statistic=13.0, P=1.24x10⁻³¹, df=349; Non-zero lag state-
633 tuned neurons N=179 neurons, statistic=3.48, P=6.39x10⁻⁴, df=178.

634 e) Single anchor alignment analysis. Top (blue) plots for each neuron shows activity aligned by
635 the abstract states (with the dashed vertical line at zero representing reward **a**, i.e. the start of
636 state A; going clockwise, the remaining dashed lines represent reward locations **b**, **c** and **d**,
637 and hence the starts of states B, C and D respectively). Neurons appear to remap in task
638 space across tasks. Bottom (green) plots for each cell show that it is possible to find a goal-
639 progress/place anchor that consistently aligns neurons across tasks (the zero dashed vertical
640 line corresponds to visits to the goal-progress/place anchor). The top Neuron aligns best to its
641 anchor close at 180 degrees while bottom neuron aligns best to its anchor at 120 degrees.

642 f) Polar histograms showing the cross-validated alignment of neurons by their preferred goal-
643 progress/place anchor (calculated from training tasks) in a left-out test task. The top plot is for
644 all state-tuned neurons while the bottom plot only includes non-zero lag neurons. Non-zero
645 lag neurons are those with a peak firing rate more than 30 degrees in task space either side
646 of the anchor. Two proportions test against chance (25%): All state-tuned neurons: Proportion
647 generalising: 38.1% N=350 neurons, z=10.66 P<0.001; Non-zero lag state-tuned neurons:
648 Proportion generalising: 38.5% N=269 neurons, z=9.43, P<0.001.

649 g) Histograms showing the right shifted distribution of the mean cross-validated task map
650 correlations between state-tuned neurons aligned to their preferred goal-progress/place
651 anchor (from training tasks) and the task map aligned to this same goal-progress/place
652 anchor from a left out test task. This is shown for all state-tuned neurons (top) and only non-
653 zero lag state-tuned neurons (bottom). Non-zero lag neurons are those with a peak firing rate
654 more than 30 degrees in task space either side of the anchor. P values are from T tests
655 relative to 0. T test against 0: All state-tuned neurons N=350 neurons, statistic=6.12,
656 P=2.53x10⁻⁹, df=349; Non-zero lag state-tuned neurons N=269 neurons, statistic=5.44,
657 P=1.89x10⁻⁷, df=268.

658 h) Two example paths each spanning an entire trial across two distinct tasks and overlaid
659 spiking of 2 mFC state-tuned neurons anchored to early goal-progress in location 6 (middle
660 right location on the maze). Neuron 1 is an anchor neuron and is hence tuned to this goal-
661 progress/place. Neuron 2 fires with a lag of roughly 270 degrees in task space from the
662 anchor and so fires 3 states after the animal visits early goal-progress in location 6. Thus
663 neuron 2 fires when animals get reward in location 5 in task 1 (reward **a**) as this is 270
664 degrees (3 states) in task space from reward in location 6 (reward **b**). In task 2, this neuron
665 fires most in location 2 (reward **d**) which is also 270 degrees (3 states) from reward in location
666 6 (reward **a**). For visualisation purposes, spikes were jittered to ensure directly overlapping
667 spikes are distinguishable.

668 All error bars represent the standard error of the mean

669

670

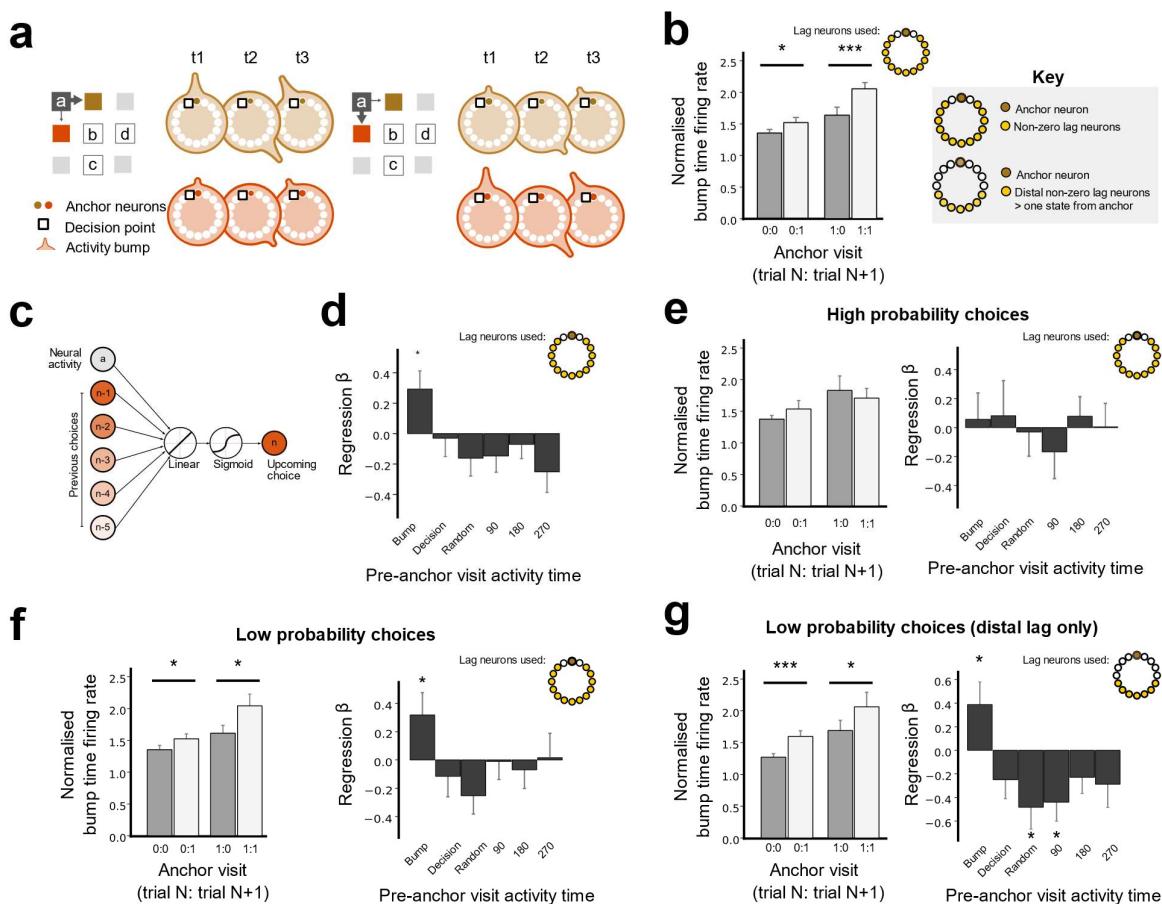
671 **2-Distal prediction of behavioural choices**

672 The SMB model posits that mFC neurons represent a memory of the animal's past policy
673 and provides a mechanistic explanation for this mnemonic function that we empirically
674 validate above. The model also proposes a mechanism for how such neurons can be used
675 to guide future behaviour. This is because the modules are proposed to be shaped by the
676 structure of the task, in our case a loop of four rewards. This ring structure means modules
677 can be used to repeat an effective policy once it is found (Figure 6a). If in trial N an animal
678 makes a choice to visit a given behavioural step (e.g. early goal-progress location 2), then a
679 bump of activity is initiated on the module that is anchored to this behavioural step at this
680 point in the task (e.g. start of state B). This activity bump moves around the ring paced by
681 progress in the task until it circles back close to the anchor point (e.g. towards the end of
682 state A). This close-to-anchor activity biases the animal to return back to this same
683 behavioural step in trial N+1 that it visited at the same point in the task in trial N (e.g. start of
684 state B), thereby repeating the previous trial's policy (Figure 6a). Thus, at any point in the
685 sequence, the animal can choose among options for the next step in the maze by "listening"
686 to the ring with the largest bump near its anchor point. The result is a sequential policy
687 paced by the task periodicity. Crucially, the same memory buffers can be activated in a
688 different sequence to encode a different task, without needing to build or bind new
689 representations: thereby providing a programmable encoding of new task sequences.

690 This model makes a unique prediction: the choice an animal will make to visit a particular
691 maze location should be predictable from the activity of neurons anchored to the possible
692 choices several steps and 10s of seconds before the animal actually makes a decision. This
693 is because the bump of activity is constrained to move around a given ring. Hence, bump
694 size at an early point in the trial should strongly correlate with its size just before the anchor
695 point (Figure 6a). The time point just before the anchor point is the "decision time" for a given
696 ring where the output neurons of the ring can be read out to bias the animal's choice towards
697 the behavioural step represented by the anchor (Figure 6a). The time at which the putative
698 bump of activity passes through a given neuron (the neuron's "bump-time") reflects the
699 neuron's task-space lag from the behavioural step to which its anchored: i.e. its position on
700 the ring relative to the anchor (Figure 4a, Figure 6a). The SMB model proposes that activity
701 restricted to this "bump-time" can be used to predict whether an animal will return back to
702 this same behavioural step at the same point in the next trial. Because a trial is on average
703 ~30 seconds in length, this means an animal's choices can be predicted 10s of seconds
704 before they are made. To test this prediction, we related the trial-by-trial "bump-time" activity
705 of neurons at non-zero lag to their anchor with subsequent visits to this same anchor.

706 Intriguingly, while controlling for previous choices, we found that the “bump-time” activity of
707 neurons was higher before animals visited the neurons’ anchor (Figure 6b). To investigate
708 this further, we ran a logistic regression on the trial by trial activity of neurons at non-zero lag
709 from their anchor to predict upcoming behavioural choices by the animal. To control for the
710 autocorrelation in animals’ behaviour, we regressed out previous choices going all the way
711 back to 5 trials in the past (Figure 6c). We found that mFC neurons significantly predicted
712 future choices only when taking their activity at the correct “bump-time”. Crucially, activity of
713 a given neuron at the “wrong” times, whether at random timepoints, times shifted by 90
714 degree intervals relative to the “bump-time” to preserve goal-progress tuning, or even at the
715 “decision time” itself, did not predict subsequent choices (Figure 6d). This prediction held
716 even when only considering choices to intermediate, non-rewarded locations (Extended data
717 figure 6a).

718 Notably, animals are not performing optimally (Figure 1d). Whilst performance is well above
719 chance, they continue to be biased by pre-existing policies that existed before they learned
720 any ABCD task (Extended Data Figure 1c). We reasoned that if the frontal cortex exerts goal
721 directed control²¹, then we might only be able to predict choices where animals take actions
722 that go against their pre-existing biases. We therefore divided choices into high and low
723 probability ones, based on the probability animals made a given choice in an exploration
724 session on the same maze but before the animals had been exposed to any ABCD task.
725 Intriguingly, we found that frontal activity was not predictive of high probability choices
726 (Figure 6e). In contrast, mFC activity predicted low probability choices (Figure 6f). This is in
727 line with the view that frontal neurons are preferentially implicated in altering behaviour
728 against stereotyped policies²¹. Moreover, “bump-time” activity of neurons with mnemonic
729 fields that are lagged by more than one state away from the anchor also predicted choices
730 (Extended data figure 6b). This again was seen for low (Figure 6g) but not high probability
731 choices (Extended data figure 6c). Overall these results, showing a distal prediction of
732 behavioural choices, empirically validate the SMB model as a mechanism for the retrieval of
733 task-paced policies by the mFC.



734
735

736 **Figure 6 – Medial Frontal activity predicts distal behavioural choices**

737 a) Schematic showing distal prediction of animals choices from memory buffers. When the
738 animal visits a goal-progress/place ($t=1$) in trial N, a bump of activity is initiated in the memory
739 buffer that is anchored to this goal-progress/place. This bump travels around the buffer (e.g.
740 $t=2$), paced by progress in the task. When the activity bump circles back to a point close to
741 the anchor ($t=3$), it can be read out to bias the animal to return back to the same goal-
742 progress/place in trial N+1 that was visited in the same task state in trial N. This read-out time
743 defines a “decision point” that is specific for each memory buffer. Left: If, at $t=3$ in the example
744 given, the bump on the buffer anchored to intermediate goal-progress in location 1 (brown
745 square) is larger than that for the other option (intermediate goal-progress in location 4; red
746 square) the animal will choose location 1. Right: Location 4 (red square) is chosen if the bump
747 anchored to intermediate goal-progress in location 4 is larger at $t=3$. This choice could have
748 been predicted from the bump sizes at an earlier time point (e.g. $t=2$) as the bump size will
749 remain highly stable for the duration of a single trial, hence allowing distal prediction of
750 choices from the memory buffers.

751 b) Prediction of behaviour. Normalised firing rates of neurons during their “bump time”: i.e. the
752 lag at which they are active relative to the anchor. X-axis labels denote visits to a goal-
753 progress/place anchor in the current (N) and upcoming trial (N+1). For example a value of 0:1
754 means the anchor was not visited in trial N but visited in trial N+1. Bump time activity is higher
755 before visits to the neuron’s anchor in trial N+1 whether the anchor was not visited in trial N
756 (left) or when it was visited in trial N (right). Wilcoxon tests: Anchor not visited in trial N: $n=78$
757 sessions, statistic=1090, $P=0.025$. Anchor visited in trial N: $n=72$ sessions, statistic=666,
758 $P=2.76 \times 10^{-4}$. In addition, an ANOVA on all data ($N=72$ sessions) showed a main effect of

759 Past F=20.6 P=2.2x10-5, df1=1, df2=71, main effect of Future F=9.5 P=0.003, df1=1, df2=71
760 and no Past x Future interaction F=0.41 P=0.525, df1=1, df2=71.

761 c) Design of logistic regression to assess effect of activity of the neuron on future visits to the
762 anchor. To control for any autocorrelation in the choices of the a mouse, previous choices
763 (binary values: 1s and 0s indicating visits) as far back as 5 trials in the past are added as
764 regressors. Separate regressions are done for activity at different times: "bump time" (as in b);
765 random times, decision time (the last 30 degrees of task space before the time where the
766 animal could have visited the anchor of a given neuron, i.e. was one goal-progress-place
767 away from it), and times shifted by 90 degree intervals relative to the neuron's "bump time".
768 d) Regression coefficients are significantly positive for the bump time but not any of the other
769 control times. T tests against 0: "bump time": N=78 sessions, statistic=2.43, P=0.017, df=77;
770 "decision time": N=78, statistic=-0.26, P=0.792, df=77; "random time" N=78, statistic=-0.127,
771 P=0.899, df=77; "90 degree shifted time" N=78, statistic=-1.34, P=0.183, df=77; "180 degree
772 shifted time" N=78, statistic=-0.700, P=0.486, df=77; "270 degree shifted time" N=78,
773 statistic=-1.86, P=0.066, df=77.
774 e) Left: Distal prediction of behaviour to intermediate (non-rewarded) locations including choices
775 to reward locations for only the high probability choices (choices that animals made with a
776 probability in top half of maze transition probabilities during pre-task exploration). Bump time
777 activity is indistinguishable before visits to the neuron's anchor compared to non visits in trial
778 N+1 whether the anchor was not visited in trial N (left) or when it was visited in trial N (right).
779 Wilcoxon tests: Anchor not visited in trial N: n=59 sessions, statistic=728, P=0.236. Anchor
780 visited in trial N: n=53, statistic=637, P=0.487. In addition, an ANOVA on all data (N=53
781 sessions) showed a trend towards a main effect of Past F=3.26 P=0.077, df1=1, df2=52, no
782 main effect of Future F=0.026 P=0.871, df1=1, df2=52 and no Past x Future interaction
783 F=1.60 P=0.211, df1=1, df2=52. Right: Regression coefficient is not significant for the bump
784 time. T tests against 0: "bump time": N=59 sessions, statistic=0.324, P=0.747, df=58;
785 "decision time": N=59, statistic=0.329, P=0.743, df=58; "random time" N=59, statistic=0.106,
786 P=0.916, df=58; "90 degree shifted time" N=59, statistic=-0.918, P=0.362, df=58; "180 degree
787 shifted time" N=59, statistic=0.611, P=0.544, df=58; "270 degree shifted time" N=59,
788 statistic=0.052, P=0.959, df=58.
789 f) Left: Distal prediction of behaviour to intermediate (non-rewarded) locations including choices
790 to reward locations for only the low probability choices ((choices that animals made with a
791 probability in bottom half of maze transition probabilities during pre-task exploration).
792 Wilcoxon tests: Anchor not visited in trial N: n=64 sessions, statistic=744, P=0.048. Anchor
793 visited in trial N: n=54, statistic=486, P=0.027. In addition, an ANOVA on all data (N=54
794 sessions) showed a main effect of Past F=5.66 P=0.021, df1=1, df2=52, a main effect of
795 Future F=6.40 P=0.014, df1=1, df2=52 and no Past x Future interaction F=1.67 P=0.201,
796 df1=1, df2=52. Right: Regression coefficients are significantly positive for the bump time but
797 not any of the other control times. T tests against 0: "bump time": N=64 sessions,
798 statistic=2.03, P=0.047, df=63; "decision time": N=64, statistic=-0.806, P=0.423, df=63;
799 "random time" N=64, statistic=0.883, P=0.380, df=63; "90 degree shifted time" N=64,
800 statistic=-0.079, P=0.937, df=63; "180 degree shifted time" N=64, statistic=-0.537, P=0.593,
801 df=63; "270 degree shifted time" N=64, statistic=0.100, P=0.920, df=63.
802 g) Prediction of behaviour using neurons anchored at distal points (one state away or more)
803 from the anchor for only the low probability choices (choices that animals made with a
804 probability in bottom half of maze transition probabilities during pre-task exploration).
805 Normalised firing rates of neurons during their "bump time": i.e. the lag at which they are
806 active relative to the anchor. Bump time activity is higher before visits to the neuron's anchor
807 in trial N+1 whether the anchor was not visited in trial N (left) or when it was visited in trial N
808 (right). Wilcoxon tests: Anchor not visited in trial N: n=67, statistic=463, P=2.41x10-5. Anchor
809 visited in trial N: n=51, statistic=430, P=0.029. In addition, an ANOVA on all data (N=51
810 sessions) showed a main effect of Past F=5.29 P=0.026, df1=1, df2=50, no main effect of

811 Future $F=10.19$ $P=0.002$, $df1=1$, $df2=50$ and no Past x Future interaction $F=7.73 \times 10^{-4}$
812 $P=0.978$, $df1=1$, $df2=50$. Regression coefficients are significantly positive for the bump time
813 but not any of the other control times. T tests against 0: "bump time": $N=67$ sessions,
814 $statistic=2.02$, $P=0.048$, $df=66$; "decision time": $N=67$, $statistic=-1.54$, $P=0.128$, $df=66$;
815 "random time" $N=67$, $statistic=-2.60$, $P=0.011$, $df=66$; "90 degree shifted time" $N=67$,
816 $statistic=-2.73$, $P=0.008$, $df=66$; "180 degree shifted time" $N=67$, $statistic=-1.65$, $P=0.103$,
817 $df=66$; "270 degree shifted time" $N=67$, $statistic=-1.46$, $P=0.148$, $df=66$.
818 All error bars represent the standard error of the mean
819

820 **3-Internally organised memory buffers**

821 Our findings so far support the view that mFC modules invariantly track task-progress from
822 different behavioural steps, and use this memory to retrieve a sequential policy that is paced
823 by the task's periodicity. To retrieve such task-paced sequences, the SMB model proposes
824 that memory buffers are shaped by the structure of the task. In the ABCD task, the structure
825 is a loop consisting of 4 rewarded goals (4 goal-progress cycles). Thus the neuronal state
826 space should curve back onto itself after 4 goals, creating an internally organised ring
827 structure. While the policy retrieval results (Figure 6; Extended Data figure 6) are consistent
828 with this view, we sought a more direct test of this task-structuring hypothesis. To this end,
829 we investigated whether pairwise coactivity during pre-task sleep reflects a ring-like neuronal
830 state space. We regressed the circular distance between pairs of neurons that share the
831 same anchor, and hence belong to the same module, against their cross-correlation during
832 sleep (Figure 7a). This was done while co-regressing the forward distance from the anchor:
833 the distance between two neurons in a state space shaped into a line that begins with the
834 anchor (Figure 7a). We found that regression coefficient values were significantly negative
835 for circular distance, while controlling for forward distance, indicating that neurons closer on
836 a circular state space (smaller neuron-neuron distances on a ring) are more coactive (Figure
837 7b). These results held while controlling for pairwise spatial map similarity and goal-
838 progress-tuning proximity between neurons, which were added as co-regressors. Forward
839 distance did not predict co-activity when controlling for circular distance (Figure 7b).
840 Moreover, we isolated pairs of neurons that had systematically opposed circular and forward
841 distances (i.e. pairs of neurons separated by a forward distance of 180 degrees or more).
842 Regressing circular distance against sleep coactivity for only those neuron pairs also
843 showed a significantly negative regression coefficient, further supporting a ring-shaped state
844 space (Figure 7c). In addition, pairs of consistently anchored neurons that share the same
845 anchor showed significantly more negative regression coefficient values than neuron pairs
846 across anchors (Figure 7d). This was true both for pre-task and post-task sleep (Figure 7d).
847 Taken together, these findings provide offline evidence that mFC modules are internally
848 organised by the structure of the task.
849

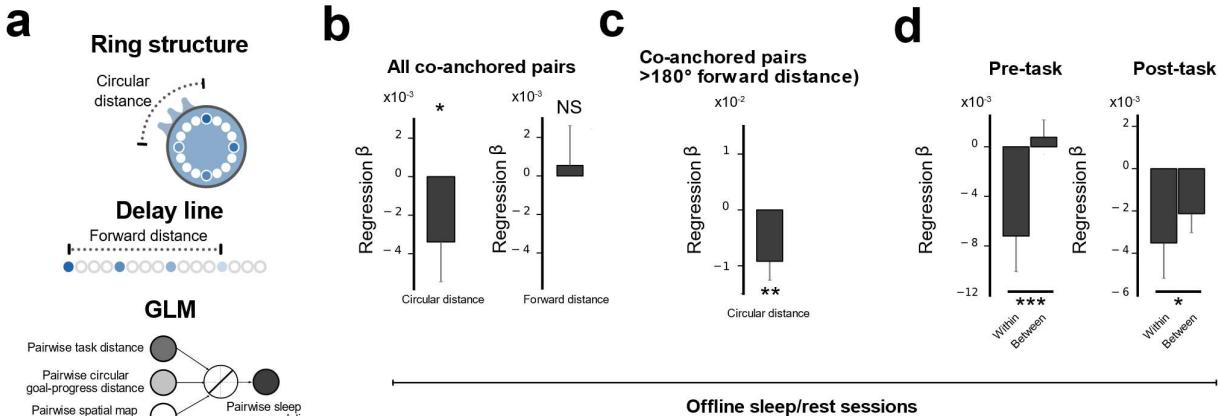


Figure 7 – Offline activity of mFC neurons is internally organised by the Task structure

- a) Top: Schematic showing potential neuronal state spaces - if neurons are arranged on a ring, then circular distance is a better description of how close two neurons are in state space than forward distance relative to the anchor. Conversely, if neurons lie on a delay line, forward distance is a better description of neuron-neuron co-firing relationships. Bottom: Schematic showing the inputs and outputs of linear regression model relating pairwise task-distance (either circular or forward distance) with coactivity during sleep while regressing out pairwise goal-progress tuning distance and spatial map similarity.
- b) Regression coefficient values for circular (left) or forward (right) distance regressed against sleep cross-correlation for co-anchored neurons - T test relative to 0 (One-tailed): circular distance: N=612 pairs t=-1.656 P=0.049; forward distance: t=0.266 P=0.395, df=612
- c) Regression coefficient values for circular distance for only co-anchored pairs of neurons separated by a forward distance of 180 degrees or more in task space: T test relative to 0 (One-tailed): N=149 pairs t=-2.770 P=0.003, df=148 Note that forward distance is the exact inverse of circular distance for forward distances of 180 degrees or higher, and hence forward distance regression coefficients will be exactly the same magnitude, but opposite sign, to circular distance regression coefficients.
- d) Regression coefficient values for circular distance against sleep cross-correlation using pairs of neurons consistently anchored to the same anchor (within) vs pairs with different anchors (between). Regression coefficient values were significantly more negative for within compared to between comparison for pre-task (left) and post-task (right) sleep. Two-tailed unpaired T test: Pre-task: t = 7.7216, N=268 pairs (within), 2722 pairs (between), df = 2988, P<0.0001; Post-task: N=256 pairs (within), 2774 pairs (between), df = 3028, t = 2.3145, P=0.0207

All error bars represent the standard error of the mean

Discussion

Our findings identify a cellular algorithm for mapping abstract behavioural structure. We found that mice can learn an abstract structure organising multiple goals and use it to rapidly learn complex behavioural sequences. The large majority of mFC neurons tiled progress-to-goal, generalising across behavioural sequences of different distances, times and locations. This goal-progress tuning was further elaborated to form representations shaped by the overall structure of the task, a four-goal loop in our case. The use of goal-progress tuning as a scaffold for building task-structure representations ensured that mFC neuronal dynamics evolved as a function of true progress in the overall task as defined by the goals, rather than other physical dimensions. Using goal-progress-tuned neurons also meant that frontal

888 representations tracked task progress not only from each goal location but also from
889 intermediate behavioural steps *en route* between goals. This allowed encoding long
890 sequences of behaviours that were parsed by the task structure, guided primarily by the
891 network dynamics of pre-learned memory buffers. The resulting algorithm provides a unified
892 account of mFC roles in schema formation and sequence memory, by internally organising
893 mnemonic activity according to an abstract structure shared by many tasks. It further
894 suggests a general mechanism for building behavioural schema by using “goal-progress”
895 neurons as a building block that can be sculpted to represent the structure of any complex
896 task.
897

898 The SMB algorithm we discover here can potentially reconcile a number of findings
899 concerning the mFC. Lesions to the Frontal cortex across species affect the sequencing of
900 complex, goal-directed behaviours³⁷. Here we show that if mnemonic representations
901 resembling those implicated in sequential working memory²² are shaped by an abstract task
902 structure, they could explain another key function of the mFC: the formation of generalised
903 task schema¹². Moreover, the differences in the representational logic employed at each
904 level of the task hierarchy seen in our study could potentially reconcile seemingly disparate
905 views on task structure generalisation. At the level of tracking progress towards a single
906 goal, we find that neurons tile this space using fixed sequences that generalise across
907 different behavioural trajectories. These goal-progress cells are similar in spirit to neurons in
908 the mFC that encode general task states regardless of sensorimotor specifics^{11,28–31}.
909 However, at the level of tracking task-progress in a sequence comprising multiple goals,
910 network dynamics are not universally coherent across all neurons, but are instead
911 apportioned into modules, each tracking a memory of a particular behavioural step. Such
912 representational logic is consistent with recent empirical findings in sequence working
913 memory tasks^{22,38} and tasks where tracking previous choices and/or reward history is
914 necessary for optimal behaviour^{23,24} as well as computational work on recurrent neural
915 networks trained to generalise across tasks³⁵. Moreover, while such SWM tasks are not
916 explicitly circular, maintenance of the experienced sequence in the delay period might be
917 implemented via similar ring attractor dynamics. These dynamics would later guide retrieval
918 of the sequence in the same way we describe here, through activating anchor neurons
919 representing the next step in the sequence via simulation rather than direct experience.
920 More generally, we propose that invariant representations may track progress towards a
921 single goal while stimulus-specific working-memory-like representations are used for tracking
922 progress in a complex task comprising multiple goals of equal valence.
923

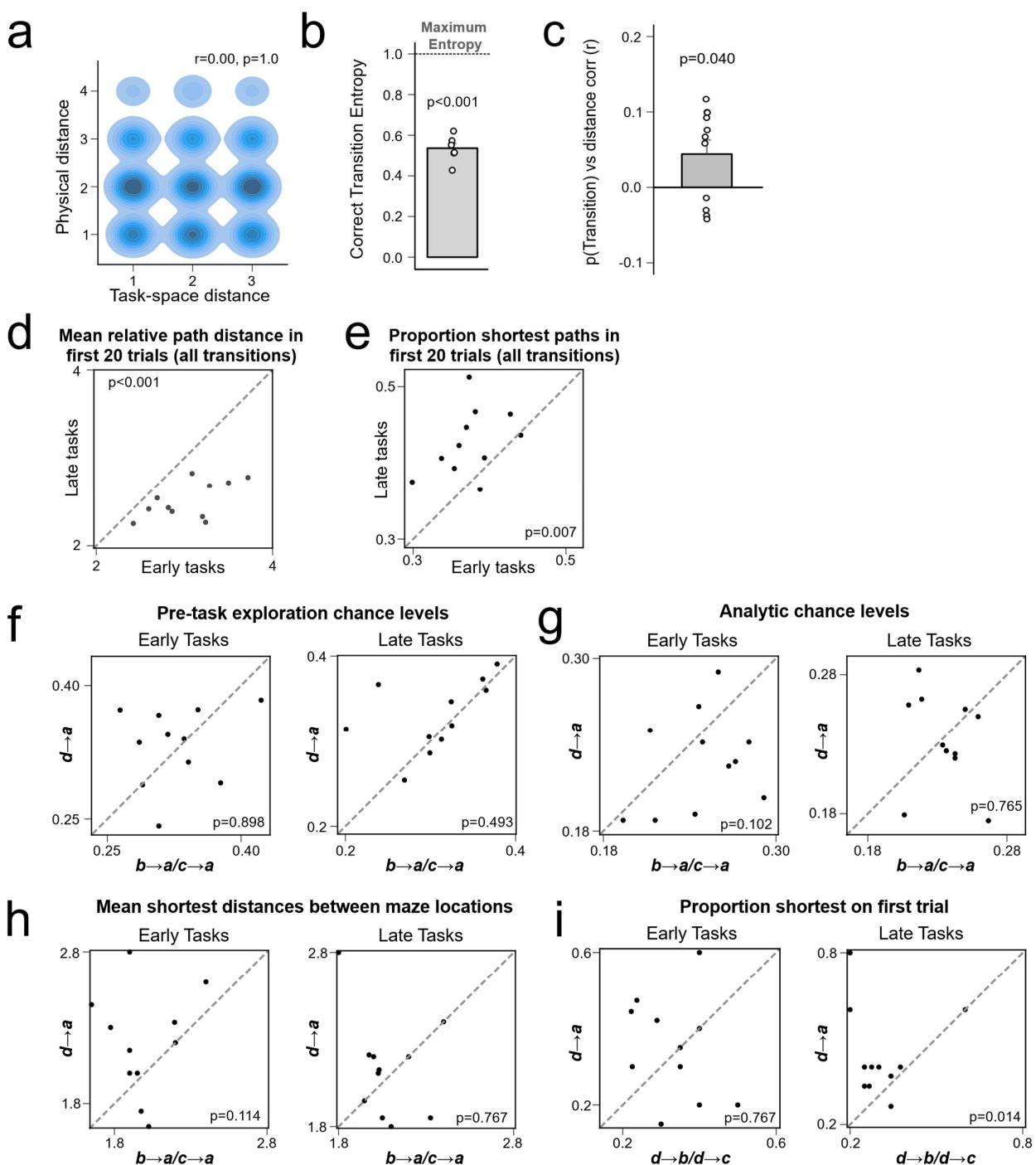
924 How does the organisation of neurons in the mFC mapping “behavioural structure” compare
925 with the cellular bases of cognitive maps of “world structure”? Like the grid-cells that underlie
926 world-structure mapping, frontal neurons are internally organised according to the structure
927 they map. Neurons maintain this structure when the contents of the sequence they are
928 mapping change in new tasks in mFC (Figure 3) and in new spatial contexts in mEC³⁶, and
929 even in the absence of any structured sensory input during sleep (Figure 7 for mFC; ^{39,40} for
930 mEC). We also note the modularity that underlies both types of maps (Figure 3 for mFC; ³⁶
931 for mEC), which may be a general feature of cortical representations that allows mapping
932 spaces from different reference points⁴¹. There are also key differences. One difference
933 relates to the velocity signal driving dynamics in both maps. By virtue of needing to map
934 behavioural space rather than physical space, activity bumps along attractors representing
935 behavioural structure evolve as a function of progress towards a goal (Figure 2), while grid

936 cells use the velocity of the animal in physical space to track its spatial position⁴². This basic
937 goal-progress scaffold in the Frontal cortex enables building a representation that evolves as
938 a function of task stages, allowing animals to reliably map true task-progress independently
939 of other variables such as elapsed time or distance. Another apparent difference relates to
940 how maps are used to guide flexible behaviour in new contexts. For maps of world structure,
941 convergent input from medial Entorhinal grid cells and sensory input from the lateral
942 Entorhinal cortex into the Hippocampus is thought to give rise to the canonical “place cell” in
943 a new spatial context⁴³. This grounding of abstract, relational “world-structure” knowledge to
944 concrete experiences allows us to navigate, path integrate and infer new routes to
945 goals^{3,4,44,45}. Conversely, by virtue of having fixed anchoring to concrete behavioural steps,
946 frontal neurons store new behavioural sequences in the dynamics of the network, without
947 needing access to an external memory and new plasticity. This offers a programmable
948 solution to mapping new sequences, which avoids the need for associative binding on new
949 tasks. Intriguingly, a recent study suggests that similar fixed anchoring of grid-cells to
950 landmarks may allow rapid, plasticity-free mapping of new spatial environments⁴⁶, akin to
951 what we report here for frontal maps of behavioural structure. How associative and network
952 dynamics-based solutions compare computationally, how they coexist in the same and
953 distinct circuits and when they are used remain open questions. We address some of these
954 computational questions in a recent theoretical study⁴⁷.
955
956 The programmable solution employed to map behavioural structure in expert animals places
957 the burden of plasticity on sculpting attractor dynamics when learning early tasks. How is
958 such learning achieved? A potential answer comes from considering that memory buffers
959 representing the abstract task structure are expressed on top of lower level goal-progress
960 tuning. Goal-progress-tuned neurons may have been initially part of task-naive goal-
961 progress sequences that then became sculpted, through learning of many examples, to form
962 structured memory buffers. These naive goal-progress sequences may develop first to map
963 sequences to any behavioural goal independently of any higher order structure organising
964 the goals. Input from the Hippocampus and/or medial Entorhinal cortex could provide spatial
965 information to the mFC⁴⁸ to generate the conjunctive goal-progress/place anchors. Plasticity
966 at recurrent mFC-thalamus-Basal ganglia loops could then allow learning task-structured
967 memory buffers that track task-progress from these anchors^{49,50}. Goal-progress sequences
968 could therefore provide an inductive bias which facilitates the formation of a schema
969 encoding any task-structure in the mFC. Monitoring and manipulating mFC neurons and
970 their inputs during learning of the ABCD task and equivalent hierarchical learning paradigms
971 will allow understanding the inductive biases, learning rules and circuit mechanisms that
972 generate maps of behavioural structure.

973 **Extended Data Figures**

974

975 **Extended Data Figure 1**

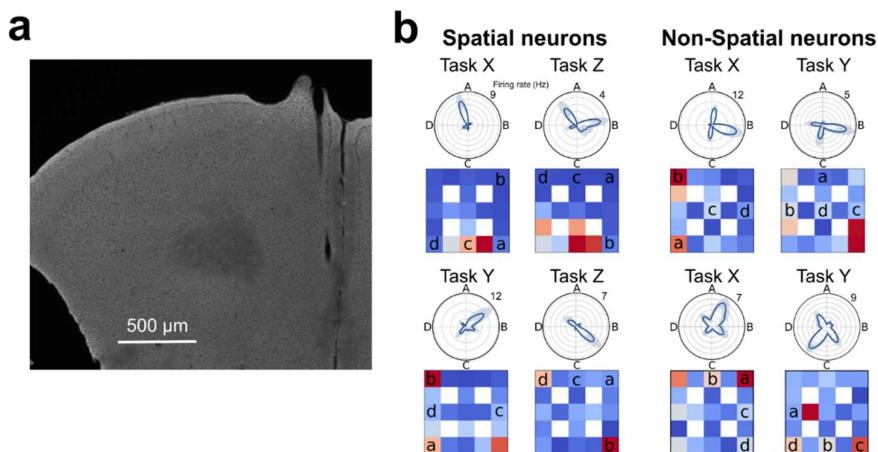


976

977 a) Tasks were designed such that task space and physical space are orthogonal to
978 each other. A kernel density estimate plot showing that distances between reward
979 locations in task space (how many states are between the rewards) are not

980 correlated with the physical distances in the maze. Data points are individual tasks.
981 Pearson correlation: $r=0.00$ $P=1.0$
982 b) Animals used stereotyped routes when taking the shortest route to a goal. The
983 entropy of correct transitions taken is lower than expected if animals took all shortest
984 routes equally. T-test against 1 – N=11 animals, statistic=-20.38, $P=9.07\times10^{-7}$, df=10
985 c) Suboptimal performance was associated with persisting behavioural biases from
986 before exposure to the task. Y-axis shows the r value calculated from a correlation
987 between the mean relative path distance taken between goals and the probability the
988 transitions within this trajectory would have been taken when the animal was naive to
989 any ABCD task (when the animal explored the arena before any rewards or tasks
990 were presented). A net positive correlation indicates that when animals take longer
991 routes (i.e. perform less optimally) they take these routes through transitions that
992 they were more likely to take before exposure to any ABCD task. T-test against 0 –
993 N=11 animals, statistic=2.36, $P=0.040$, df=10
994 d) Mean relative path distance travelled by the mice between goals in the first 20 trials
995 of early vs late tasks. Wilcoxon test N=11 animals, Statistic=0.0, $P=9.77\times10^{-4}$
996 e) Mean proportion of transitions where one of the shortest routes was taken in the first
997 20 trials of early vs late tasks. Wilcoxon test N=11 animals, Statistic=4.0 $P=0.007$
998 f) No difference in the empirical chance levels (baseline transition probabilities
999 calculated when animals explored the maze before experiencing any ABCD tasks:
1000 see Methods under “Behavioural Scoring”) between **d** to **a** and **c-to-a/b-to-a**
1001 transitions on the *first* trial in early (left) and late (right) tasks. Wilcoxon test; Early
1002 tasks: N=11 animals, statistic=31.0, $P=0.898$; Late tasks: N=11 animals,
1003 statistic=23.0, $P=0.413$
1004 g) No difference in the analytical chance levels (see Methods under “Behavioural
1005 Scoring”) between **d** to **a** and **c-to-a/b-to-a** transitions on the *first* trial in early (left)
1006 and late (right) tasks. Wilcoxon test; Early tasks: N=11 animals, statistic=14.0,
1007 $P=0.102$; Late tasks: N=11 animals, statistic=29.0, $P=0.765$
1008 h) No difference in the shortest physical maze distances between **d** to **a** and **c-to-a/b-to-**
1009 **a** transitions on the *first* trial in early (left) and late (right) tasks. Wilcoxon test; Early
1010 tasks: N=11 animals, statistic=12.0, $P=0.114$; Late Tasks: N=11 animals,
1011 statistic=20.0, $P=0.767$
1012 i) Zero-shot inference on the first trial of late tasks is associated with animals returning
1013 from **d** to **a** more often than **d-to-b** or **d-to-c**. The proportion of tasks in which animals
1014 took the most direct path from **d-to-a** on the *first* trial is compared to the same
1015 measure but for premature returns from **d-to-b** and **d-to-c**. Early tasks are shown on
1016 the left and late tasks on the right. Wilcoxon test; Early tasks: N=11 animals,
1017 statistic=20.0, $P=0.767$; N=11 animals, Late tasks: statistic=6.0, $P=0.014$
1018 All error bars represent the standard error of the mean

1019 **Extended Data Figure 2**



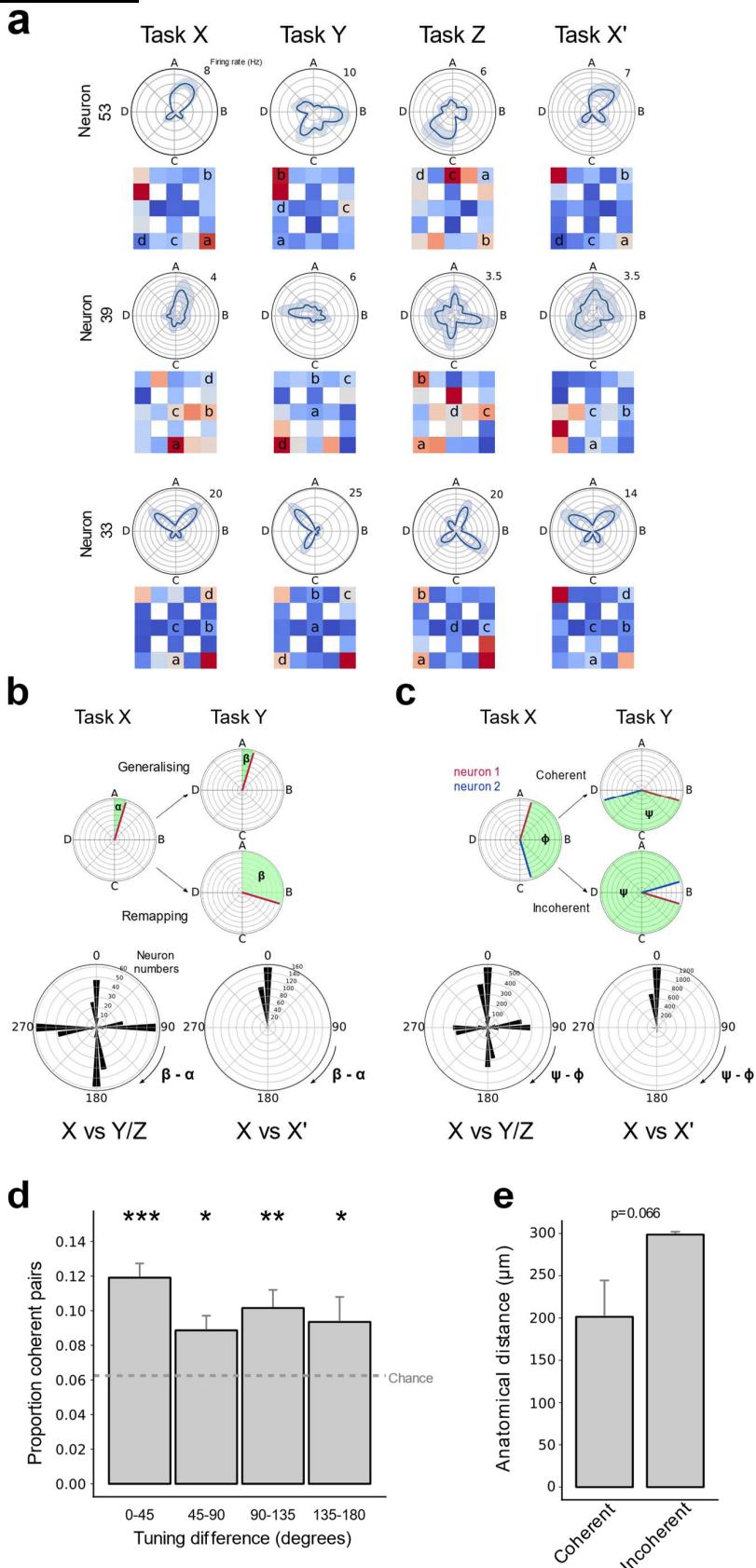
1020

1021

1022 a) Coronal slice from an implanted mouse showing silicon probe track terminating in the
1023 prelimbic region of mFC.

1024 b) Polar plots of task tuning and spatial maps for four example neurons that are tuned to
1025 both goal-progress and state. Each neuron is plotted across two tasks to illustrate
1026 spatial tuning (left two neurons) and lack thereof (right two neurons).

1027 **Extended Data Figure 3**



1029 a) Three example state-tuned neurons remapping across tasks. The top two neurons
1030 remap in a way that is not related to their spatial maps in any given task. The bottom
1031 neuron remaps in accordance to its spatial map.

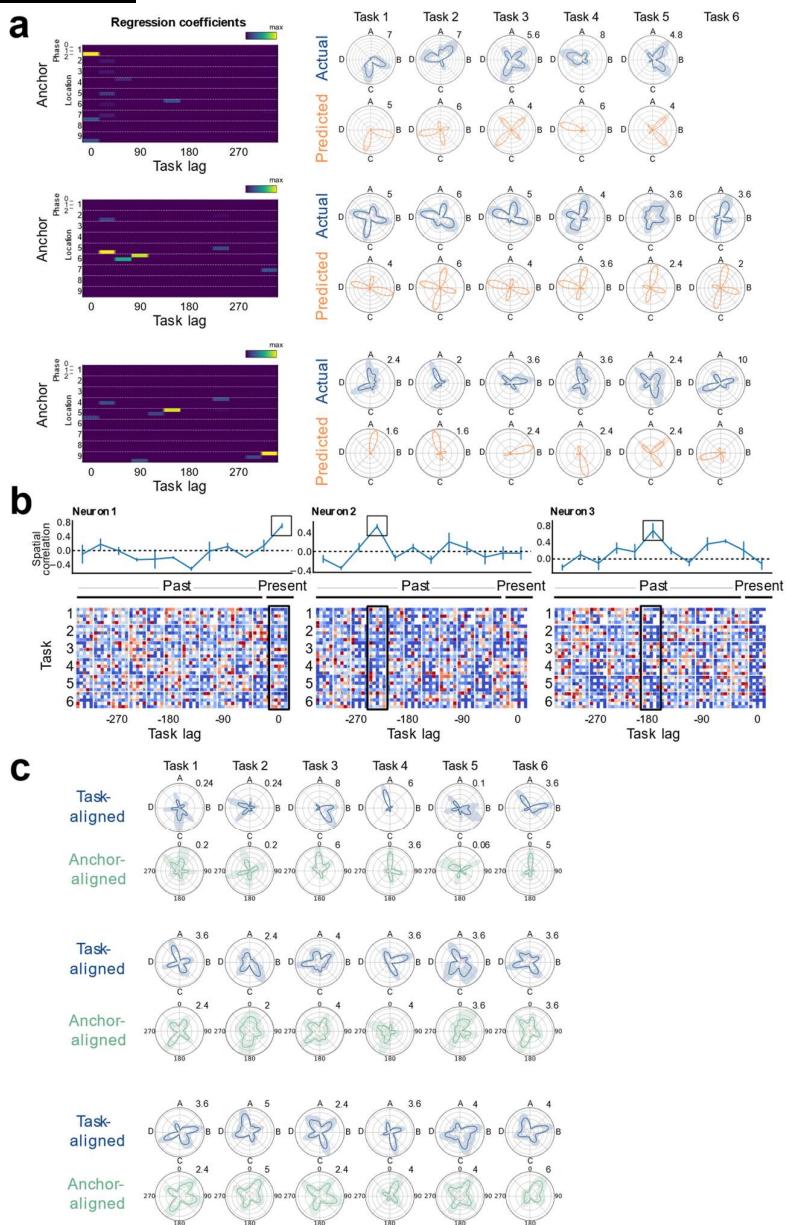
1032 b) Remapping of non-spatial neurons. Top: Quantifying the difference in tuning angles
1033 for the same neuron across tasks (top). Bottom left: Polar histograms show that non-
1034 spatial state-tuned neurons remap by angles close to multiples 90 degrees, as a
1035 result of conserved goal-progress tuning and the 4 reward structure of the task. No
1036 clear peak at zero is seen relative to the other cardinal directions when comparing
1037 sessions spanning separate tasks (Two proportions test against a chance level of
1038 25% N=432 neurons; mean proportion of generalising neurons across one
1039 comparison (mean of X vs Y and X vs Z)=21%, z=7.01, P=2.33x10-12: i.e.
1040 significantly lower than chance). Bottom right: Non-spatial state-tuned neurons
1041 maintain their state preference across different sessions of the same task (bottom
1042 right). Two proportions test against a chance level of 25% N=324 neurons; proportion
1043 generalising=84%, z=20.3, P<0.001).

1044 c) Coherence of non-spatial neuron pairs. Top: Quantifying difference in relative angles
1045 between neurons across tasks. Bottom: Only non-spatial state-tuned neurons are
1046 shown. Bottom left: Polar histograms show that the proportion of coherent pairs of
1047 non-spatial state-tuned neurons (comprising the peak at zero) is higher than chance
1048 but far from 1, indicating that the whole population does not rotate coherently (Two
1049 proportions test against a chance level of 25% N=3946 pairs; mean proportion of
1050 coherent neurons across one comparison (mean of X vs Y and X vs Z)=31%, z=29.8,
1051 P<0.001). Bottom right: As expected from panel b, the large majority of non-spatial
1052 state-tuned neurons keep their relative angles across sessions of the same task (X
1053 vs X'; Two proportions test against a chance level of 25% N=2805 pairs; proportion
1054 coherent=75%, z=53.6, P<0.001).

1055 d) Proportion of coherent pairs per recording day (pairs of state-tuned neurons where
1056 the relative angle doesn't change by more than 45 degrees across *both* X to Y and X
1057 to Z comparisons) relative to all pairs across different pairwise task space angles. T
1058 tests with Holm-Sidak correction against chance level of 1/16 (probability of neuron
1059 pair rotating coherently across two comparisons (i.e. 0.25²)): N=21 recording days,
1060 pairwise circular distance difference: 0-45 degrees statistic=6.89, P=4.33x10⁻⁶, df=20;
1061 45-90 degrees statistic=3.10, P=0.011, df=20; 90-135 degrees statistic=3.74,
1062 P=0.004, df=20; 135-180 degrees statistic=2.14, P=0.044, df=20.

1063 e) Coherent pairs show a trend towards being closer anatomically than incoherent pairs.
1064 Independent T test: N=4138 pairs, statistic=-1.84, P=0.065, df=4136
1065 All error bars represent the standard error of the mean

1066 **Extended Data Figure 4**



1067

1068

1069

1039

1070

1071

1072

1073

1074

1075

1076

1078

1077

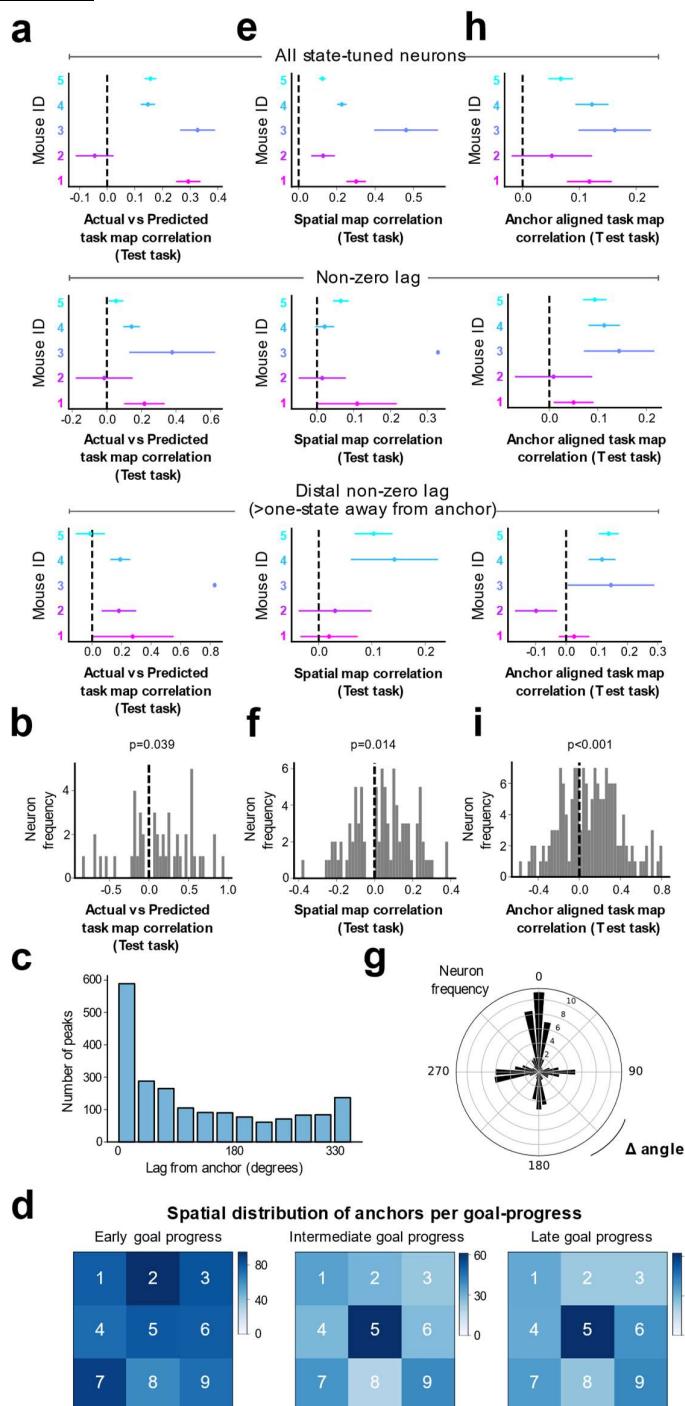
1078

1079

- a) Regression analysis reveals neurons with mnemonic fields lagged in task space from a given goal-progress/place anchor (bottom two neurons), alongside neurons directly tuned to a goal-progress/place (top neuron). The regression coefficients are shown on the left with the actual (blue) and predicted (orange) activity of the neurons shown on the right.
- b) Lagged spatial field analysis. Example plots showing spatial maps for 3 neurons. Each row represents a different task and each column a different lag in task space. Bottom: Activity of each neuron is plotted as a function of the animal's current location (far right column for each cell) and at successive task space lags in the past for the remaining columns. Top: the correlation of spatial maps across tasks at each lag. The neuron on the left is an anchor cell (goal-progress/place cell), as seen by the correlation peak at zero lag, while the middle and right-most neurons are neurons lagged by 240 and 180 degrees from their anchors respectively.

1081 c) Single anchor alignment analysis. Top (blue) plots for each neuron shows activity
1082 aligned by the abstract states (with the dashed vertical line at zero representing
1083 reward **a**, i.e. the start of state A; going clockwise, the remaining dashed lines
1084 represent reward locations **b**, **c** and **d**, and hence the starts of states B C and D
1085 respectively). Neurons appear to remap in task space across tasks. Bottom (green)
1086 plots for each cell show that it is possible to find a goal-progress/place anchor that
1087 consistently aligns neurons across tasks (the zero line corresponds to visits to the
1088 goal-progress/place anchor). Top neuron has a peak at zero relative to the anchor,
1089 while the middle and the bottom neurons peak at 200 degrees and 45 degrees from
1090 their anchors respectively.

1091 **Extended Data Figure 5**



1092

1093 a) Mean, per mouse distribution of cross-validated correlation values between model-
1094 predicted (from training tasks) and actual activity (from a left out test task) for:
1095 Top: All state-tuned neurons,
1096 Middle: non-zero lag state-tuned neurons (30 degrees or more away from anchor),
1097 Bottom: distal non-zero lag state-tuned neurons (90 degrees or more away from
1098 anchor).

1099 One-tailed binomial test against chance (chance being mean values equally likely to
1100 be above or below 0): All state-tuned neurons: 4/5 mice with mean positive
1101 correlation P=0.1875; Non-zero lag state-tuned neurons: 4/5 mice with mean positive
1102 correlation P=0.1875; Distal Non-zero-lag state-tuned neurons: 4/5 mice with mean
1103 positive correlation P=0.1875).

1104 b) Histogram showing the right shifted distribution of mean cross-validated correlation
1105 values between model-predicted (from training tasks) and actual activity (from a left
1106 out test task) for only non-zero lag state-tuned neurons with one of the top 3
1107 maximum regression coefficient values a whole state (90 degrees) or more either
1108 side of the anchor. To avoid contamination due to potential residual spatial-tuning,
1109 only regression coefficient values more than 90 degrees in task space either side of
1110 the anchor point are used for the prediction. T test against 0: N=46 neurons,
1111 statistic=2.13, P=0.039, df=45.

1112 c) Distribution of task space lags from anchor for all state-tuned neurons.

1113 d) 2D histograms showing spatial and goal-progress distributions of anchors. The colour
1114 bar represents the number of neurons anchored to each maze location at each goal-
1115 progress (maze repeated 3 times to display results for early, intermediate and late
1116 goal-progress anchors).

1117 e) Mean, per mouse distribution of cross-validated spatial correlations between spatial
1118 maps at the preferred lag (from training tasks) and the spatial map at this lag from a
1119 left out test task for:
1120 Top: All state-tuned neurons
1121 Middle: non-zero-lag state-tuned neurons (30 degrees or more away from anchor)
1122 Bottom distal non-zero-lag state-tuned neurons (90 degrees or more away from
1123 anchor).
1124 One-tailed binomial test against chance (chance being mean values equally likely to
1125 be above or below 0): All state-tuned neurons: 5/5 mice with mean positive
1126 correlation P=0.03125; Non-zero lag state-tuned neurons: 5/5 mice with mean
1127 positive correlation P=0.03125; Distal Non-zero lag state-tuned neurons: 4/4 mice
1128 with mean positive correlation P=0.0625).

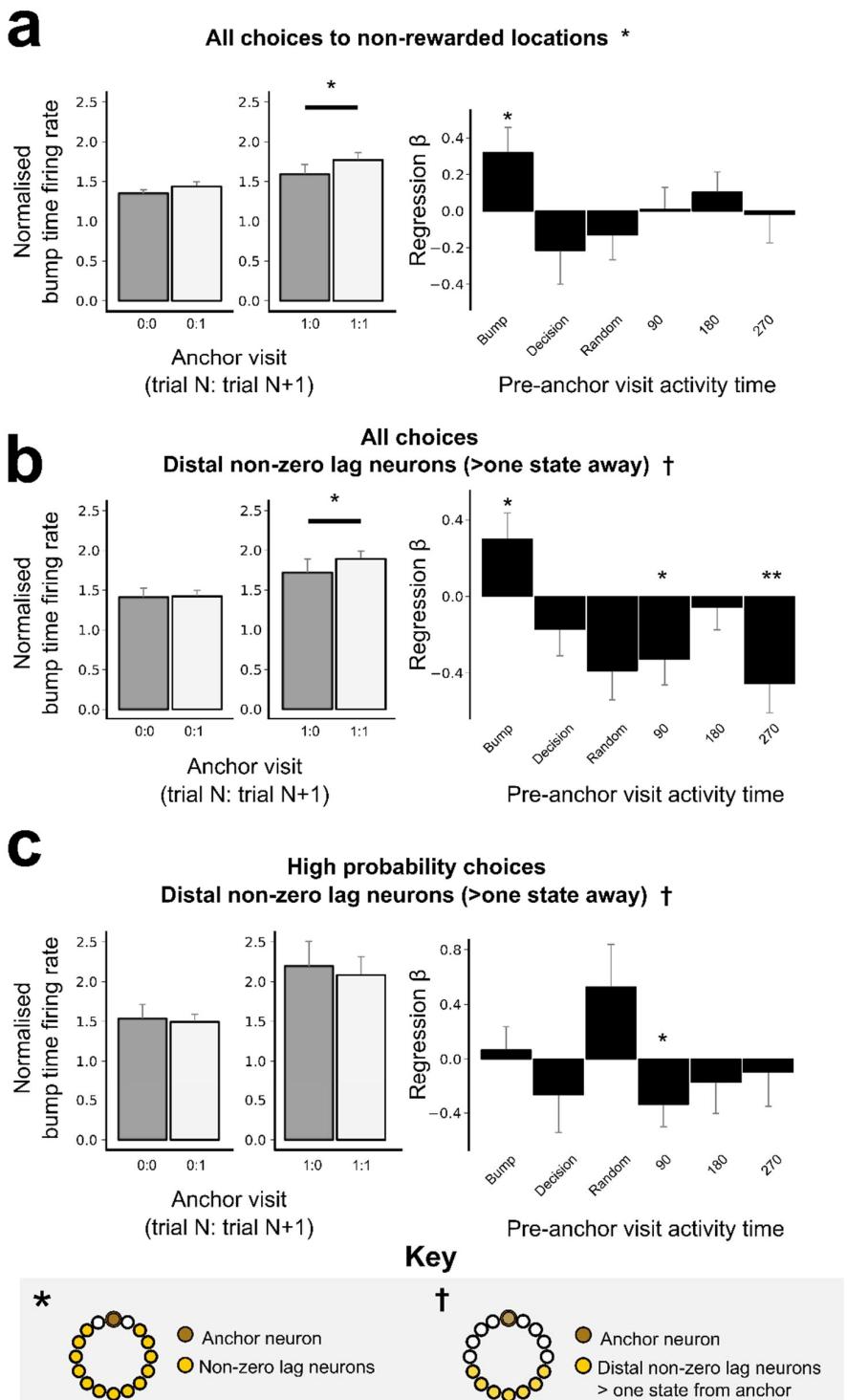
1129 f) Histogram showing the right shifted distribution of the mean cross-validated spatial
1130 correlations between spatial maps at the preferred lag in training tasks and the
1131 spatial map at this lag from a left out test task for only non-zero-lag state-tuned
1132 neurons with spatial correlation peaks a whole state (90 degrees) or further either
1133 side of zero-lag. T test against 0: N=90 neurons, statistic=2.49, P=0.015, df=89.

1134 g) Polar histogram showing the cross-validated alignment of non-zero-lag neurons by
1135 their preferred goal-progress/place (calculated from training tasks) in a left-out test
1136 task. Only neurons with a lag of 90 degrees (one state) or more either side of their
1137 anchor are shown. Two proportions test against chance (25%): Proportion
1138 generalising=39.2%; N=154 neurons, z=7.23, P=4.98x10⁻¹³

1139 h) Mean, per mouse distribution of cross-validated task map correlations between
1140 neurons aligned to their preferred goal-progress/place anchor (from training tasks)
1141 and the task map aligned to this goal-progress/place from a left out test task for:
1142 Top: all state-tuned neurons
1143 Middle: non-zero-lag state-tuned neurons (30 degrees or more away from anchor)
1144 Bottom distal non-zero-lag state-tuned neurons (90 degrees or more away from
1145 anchor). One-tailed binomial test against chance (chance being mean values equally
1146 likely to be above or below 0): All neurons: 5/5 mice with mean positive correlation

1147 P=0.03125; Non-zero-lag neurons: 5/5 mice with mean positive correlation
1148 P=0.03125; Distal Non-zero-lag neurons: 4/5 mice with mean positive correlation
1149 P=0.188).
1150 i) Histogram showing the right shifted distribution of the mean cross-validated task map
1151 correlations between neurons aligned to their preferred goal-progress/place anchor
1152 (from training tasks) and the task map aligned to this goal-progress/place from a left
1153 out test task for only non-zero-lag state-tuned neurons with a lag of 90 degrees or
1154 more either side of their anchor. T test against 0: N=154 neurons, statistic=4.59,
1155 P=9.38x10⁻⁶, df=153.
1156
1157 All error bars represent the standard error of the mean

1158 **Extended Data figure 6**



1159
1160

1161 a) Left: Prediction of behaviour to intermediate (non-rewarded) locations i.e. excluding
1162 choices to reward locations. Normalised firing rates of neurons during their “bump
1163 time”: i.e. the lag at which they are active relative to the anchor. Bump time activity is
1164 higher before visits to the neuron’s anchor in trial N+1 whether the anchor was not

1165 visited in trial N (left) or when it was visited in trial N (right). Wilcoxon tests: Anchor
1166 not visited in trial N: n=76 sessions, statistic=1093, P=0.055. Anchor visited in trial N:
1167 n=71 sessions, statistic=880, P=0.023. In addition, an ANOVA on all data (N=71
1168 sessions) showed a main effect of Past: F=11.0 P=0.001, df1=1, df2=70, a trend
1169 towards an effect of Future: F=2.74 P=0.103, df1=1, df2=70 and no Past x Future
1170 interaction F=0.001 P=0.973, df1=1, df2=70. Right: Regression coefficients were
1171 significantly positive for the bump time but not all other control times. T tests against
1172 0: "bump time": N=76 sessions, statistic=2.32, P=0.023, df=75; "decision time": N=76,
1173 statistic=-1.17, P=0.246, df=75; "random time" N=76, statistic=-0.702, P=0.485,
1174 df=75; "90 degree shifted time" N=76, statistic=0.082, P=0.935, df=75; "180 degree
1175 shifted time" N=76, statistic=0.956, P=0.342, df=75; "270 degree shifted time" N=76,
1176 statistic=-0.112, P=0.911, df=75.

1177 b) Prediction of behaviour using neurons anchored at distal points (one state away or
1178 more) from the anchor. Normalised firing rates of neurons during their "bump time":
1179 i.e. the lag at which they are active relative to the anchor. Bump time activity is not
1180 higher before visits to the neuron's anchor in trial N+1 when the anchor was not
1181 visited in trial N (left) but was higher before visits to anchor in trial N+1 when the
1182 anchor was visited in trial N (right). Wilcoxon tests: Anchor not visited in trial N: n=74
1183 sessions, statistic=1130, P=0.225. Anchor visited in trial N: n=66 sessions,
1184 statistic=709, P=0.011. In addition, an ANOVA on all data (N=66 sessions) showed a
1185 main effect of Past F=16.9 P=1.1x10⁻⁴, df1=1, df2=65, no main effect of Future
1186 F=1.49 P=0.226, df1=1, df2=65 and no Past x Future interaction F=0.006 P=0.939,
1187 df1=1, df2=65. Regression coefficients were significantly positive for the bump time
1188 but not all other control times. T tests against 0: "bump time": N=74 sessions,
1189 statistic=2.20, P=0.031, df=73; "decision time": statistic=-1.24, P=0.216, df=73;
1190 "random time" N=74, statistic=-1.86, P=0.067, df=73; "90 degree shifted time" N=74,
1191 statistic=-2.43, P=0.017, df=73; "180 degree shifted time" N=74, statistic=-0.479,
1192 P=0.633, df=73; "270 degree shifted time" N=74, statistic=-2.98, P=0.004, df=73.

1193 c) Prediction of behaviour using neurons anchored at distal points (one state away or
1194 more) from the anchor for only the high probability choices (choices that animals
1195 made with a probability in top half of maze transition probabilities during pre-task
1196 exploration). Normalised firing rates of neurons during their "bump time": i.e. the lag
1197 at which they are active relative to the anchor. Bump time activity was
1198 indistinguishable before visits to the neuron's anchor compared to non visits in trial
1199 N+1 whether the anchor was not visited in trial N (left) or when it was visited in trial N
1200 (right). Wilcoxon tests: Anchor not visited in trial N: n=62 sessions, statistic=808,
1201 P=0.323. Anchor visited in trial N: n=50 sessions, statistic=622, P=0.886. In addition,
1202 an ANOVA on all data (N=50 sessions) showed a main effect of Past F=10.3
1203 P=0.002, df1=1, df2=49, no main effect of Future F=0.249 P=0.620, df1=1, df2=49
1204 and no Past x Future interaction F=2.09 P=0.155, df1=1, df2=49. Regression
1205 coefficients were not significant for the bump time. T tests against 0: "bump time":
1206 N=62 sessions, statistic=0.40, P=0.694, df=61; "decision time": N=62, statistic=-0.98,
1207 P=0.331, df=61; "random time" N=62, statistic=1.70, P=0.095, df=61; "90 degree
1208 shifted time" statistic=-2.11, P=0.039, df=61; "180 degree shifted time" N=62,
1209 statistic=-0.785, P=0.435, df=61; "270 degree shifted time" statistic=-0.379, P=0.706,
1210 df=61.
1211 All error bars represent the standard error of the mean

1212 **Methods**

1213

1214 **Animals**

1215 Experiments used adult male C57BL/6J mice (n = 11; Charles River Laboratories). Animals
1216 were run in 2 cohorts of 4 and 1 cohort of 3, and preselected based on criteria outlined in the
1217 Behaviour section below. Animals were housed with their littermates up until the start of the
1218 experiment with free access to water in a dedicated housing facility with a 12-h light/12-h
1219 dark cycle (lights on at 7:00). Food was available *ad libitum* throughout the experiments, and
1220 water was available *ad libitum* before the experiments (see below). Mice were 2–5 months
1221 old at the time of testing. Experiments were carried out in accordance with Oxford University
1222 animal use guidelines and performed under UK Home Office Project Licence P6F11BC25.

1223

1224 **Behavioural training**

1225

1226 **Definitions**

1227 **The ABCD paradigm**

1228 A set of tasks where subjects must find a sequence of four **reward locations** (termed **a, b, c**
1229 and **d**) in a 3x3 grid maze that repeat in a loop. Once the animal receives reward **a** the next
1230 available reward is in location **b** ...etc once the animal gets reward in location **d** then reward
1231 **a** becomes available again, hence creating a repeating loop.

1232

1233 **Location**

1234 Where the animal is in the physical maze

1235 The animal could be at a **node**: i.e. one of the circular platforms where reward could be
1236 delivered: coded 1-9 as shown below - they could also be at an **edge** which is a bridge
1237 between nodes. (Figure 1a)

1238

1239 **Task**

1240 An example of the ABCD loop with a particular arrangement of reward locations (e.g. 1-9-5-
1241 7; reward **a** is in location 1; reward **b** in location 9 ...etc; Figure 1a)

1242

1243 **Session**

1244 An uninterrupted block of trials of the same task. We used 20 minute sessions. Note that
1245 subjects could be exposed to 2 or more sessions of the same task on a given day. Animals
1246 were allowed to complete as many trials as they could in those 20 minutes. Animals were
1247 removed from the maze at the end of a session and either placed back in their home-cage or
1248 into a separate enclosed rest/sleep box.

1249

1250 **Trial**

1251 A single run through an entire ABCD loop for a particular task, starting with reward in
1252 location **a** and ending in the next time the animal gets reward in location **a** again (e.g. trial 12
1253 of a task with the following configuration: 1-9-5-7 starts with the 12th time the animal gets
1254 reward in location 1 and ends with the 13th time animal gets reward in location 1)

1255

1256 **State**

1257 The time period between an animal receiving reward in a particular location and receiving
1258 reward in the next rewarded location. State **A** (upper case) starts when animal receives

1259 reward **a** (lower case italicised) and ends when animal receives reward **b**...state **D** starts
1260 when animal gets reward **d** and ends when animal gets reward **a** ...etc

1261

1262 **Transition**

1263 A generalised definition of state. e.g. progressing from **a** to **b** is a transition, and from **c** to **d**
1264 is also a transition.

1265

1266 **Goal-progress**

1267 How much progress an animal has made between rewarded locations as a percentage of
1268 the time taken between them. Unless otherwise stated, we operationally divide this into
1269 early, intermediate and late goal-progress - which correspond to $\frac{1}{3}$ increments of the time
1270 taken to get from one reward location to the next: e.g. if the animal takes 9 seconds between
1271 one reward and another, then early goal-progress spans the first 3 seconds, intermediate
1272 goal-progress the next 3 seconds and late goal-progress the last 3 seconds. This would
1273 scale with the length of time it takes for the animal to complete a single state: e.g. if it takes
1274 15 seconds between two rewards, each of the goal-progress bins would be 5 seconds long.
1275 In the ABCD paradigm goal-progress repeats 4 times because there are 4 rewards (so there
1276 will be an early goal-progress for reward **a**, and early goal-progress for reward **b** ...etc).

1277

1278 **Choice**

1279 We use this to refer to one-step choices in the maze. At every node in the maze the animal
1280 has a choice of 2 or more immediately adjacent nodes to visit next. For example, when in
1281 node 1 the animal could choose to move to node 2 or node 4 (Figure 1a).

1282

1283 **Pre-selection**

1284 A total of 11 mice, across 3 cohorts, were used for experiments. For each cohort, 3-4 Mice
1285 were pre-selected for the experiment from 10-20 mice based on the following criteria:

1286 1-Weight above 22g

1287 2-No visible signs of stress upon first exposure to the maze in the absence of rewards.

1288 Stress was evidenced by thigmotaxis or defecation in a 20 minute exploration session with
1289 no rewards delivered

1290 3-On a partially connected version of the maze with only 6 accessible ports out of 9, Animals
1291 learned that poking in wells delivered water reward and that after gaining reward they must
1292 go to another port. Animals that obtained 40 or more rewards per 20 minute session were
1293 taken forward to pretraining.

1294 4-Final selection

1295 For a given cohort, if more than 4 animals satisfied these criteria, animals that explored the
1296 maze with the highest entropy were selected (see “Behavioural Scoring” below for entropy
1297 calculation).

1298

1299 **Habituation/Pretraining**

1300 After at least one week of post-surgery recovery (see “Surgeries” section below), animals
1301 were placed on water restriction. Animals were maintained at a weight of 88-92% of their
1302 baseline weight, which was calculated before water restriction but after implantation and
1303 recovery from the surgery (see below under “Surgeries”). This is to ensure that they
1304 remained motivated to collect water rewards during the task but not overly so, as excessive
1305 motivation is known to negatively affect model-based performance⁴⁴. Animals were
1306 habituated to being tethered to the electrophysiological recording wire while moving on the

1307 maze, as well as in sleep boxes for at least 3 days prior to the start of the experiment. During
1308 this period, animals were reintroduced to a partially connected maze (only 6 out of the 9
1309 ports available, and not all connected) while tethered to the electrophysiology wire, where
1310 water reward was delivered if the animal poked its nose at any port. Reward drops were only
1311 available once the animal poked its nose at the port. At this stage, there was no explicit task
1312 structure, except that once reward was obtained at one port, animals had to visit a different
1313 port to gain further reward (i.e. exactly as in step 3 of pre-selection criteria above but while
1314 implanted and tethered). Thus, animals (re)learned that poking in wells delivered reward and
1315 that after gaining reward they must go to another port. Animals were transitioned to the task
1316 when they obtained >40 rewards in a 20 minute session. Note that 3 animals (from cohort 1)
1317 were not implanted and so for these the rewarded sessions for pre-selection and pre-training
1318 were one and the same (see section below under “Numbers” for more details on mouse
1319 cohorts). The volume of water delivered during pretraining/preselection was higher than
1320 training, typically 10-15 μ L.

1321

1322 **Training**

1323 Animals navigated an automated 3x3 grid maze in search of rewards (Figure 1a), controlled
1324 using pyControl⁵¹. Water rewards (3-4 μ L) were presented sequentially at 4 different
1325 locations. Animals had to poke in a given reward port, breaking an infrared beam that
1326 triggered the release of the reward drop in this well. After reward **a** was delivered, reward
1327 was obtainable from location **b**, but only after the animal poked in the new location. Once the
1328 animal received reward in locations **a**, then **b**, then **c** and then **d**, reward **a** becomes
1329 available again, thus creating a repeating loop. Animals have 20 minutes to collect as many
1330 rewards as possible and no time outs are given if they make any mistakes. They are then
1331 allowed at least a 20 minute break away from the maze (in the absence of any water) before
1332 starting a new session. For each session, animals were randomly entered from a different
1333 side of the square maze, using custom made electromagnetic field shielding curtains
1334 (https://www.electrosmogshielding.co.uk/product.asp?P_ID=650&CAT_ID=104). This was to
1335 ensure that all sides of the maze are equivalent in terms of being entry and exit points from
1336 the maze, thereby minimising any place preference/aversion and minimising the use of
1337 different sides as orienting cues. One cue card was placed high up (at least 50 cm vertically
1338 from the maze) on one corner of the maze to serve as an orienting cue. No cues were visible
1339 at head level.

1340

1341 While all locations were rewarded identically, a brief pure tone (2 seconds at 5 kHz) was
1342 delivered when the animal consumed reward **a**. This ensured that task states were
1343 comparable across different task sequences. White noise was present throughout the
1344 session to avoid distraction from outside noises.

1345

1346 Task configurations (i.e. the sequence of reward locations) were selected pseudo-randomly
1347 for each mouse, while satisfying the following criteria:

1348 1-The distance between rewarded locations in physical space (number of steps between
1349 rewarded locations) and task space (number of task states between rewarded locations)
1350 were orthogonal for each mouse (Extended Data Figure 1a)
1351 2-The task can't be solved (75% performance or more) by moving in a clockwise or anti-
1352 clockwise circle around the maze.

1353 3-The first two tasks have location 5 (the middle location) rewarded - this is to ensure the
1354 first tasks the animals are exposed to can't be completed by circling around the outside N

1355 times to collect all rewards. (Note that all “late” tasks and those used in electrophysiological
1356 recordings are not affected by this criterion).
1357 4-Consecutive tasks don't share a transition (i.e. two or more consecutively rewarded
1358 locations)
1359 5-Chance levels are the same for all task transitions (***ab***, ***bc***, ***cd***, ***da***), and control transitions
1360 ***ca*** and ***ba*** transitions - whether determined analytically by assuming animals diffuse around
1361 the maze or empirically by using animal-specific maze-transition statistics from an
1362 exploration session before any rewards were delivered on the maze (see “Behavioural
1363 Scoring” below for chance-level calculations).
1364
1365 For the first 10 tasks, animals were moved to a new task when their performance reached
1366 70% (i.e. took one of the shortest spatial paths between rewards for at least 70% of all
1367 transitions) on 10 or more consecutive trials or if they plateaued in performance for 200 or
1368 more trials. For these first 10 tasks, animals were given at most 4 sessions per day, either all
1369 of the same task or, when animals reached criteria, two sessions of the old task and two
1370 sessions of the new task. From task 11 onwards, animals learned 3 new tasks a day with the
1371 first task being repeated again at the end of each day giving a total of 4 sessions with the
1372 pattern X-Y-Z-X'.
1373

1374 **Surgeries**

1375 Subjects were taken off water restriction 48 hours before surgery and then anaesthetised
1376 with isoflurane (3% induction, 0.5–1% maintenance), treated with buprenorphine
1377 (0.1 mg kg⁻¹) and meloxicam (5 mg kg⁻¹) and placed in a stereotactic frame. A silicon probe
1378 mounted on a microdrive (Ronal Tool) and encased in a custom made recoverable
1379 enclosure (ProtoLabs) was implanted into mFC (AP: 2.00, ML: -0.4, DV: -1.0), and a ground
1380 screw was implanted above the cerebellum. AP and ML coordinates are relative to bregma
1381 while DV coordinates are relative to the brain surface. Mice were given additional doses of
1382 meloxicam each day for 3 days after surgery and were monitored carefully for 7 days after
1383 surgery and then placed back on water restriction 24 hours before pretraining. At the end of
1384 the experiment, animals were perfused; and the brains were fixed-sliced and imaged to
1385 identify probe locations (Extended data figure 2a).
1386

1387 **Electrophysiology, spike sorting and behavioural tracking**

1388 Cambridge NeuroTech F-series 64 silicon channel probes (6 shanks spanning 1 mm
1389 arranged front-to-back along the anterior posterior axis) were used for all recordings. To
1390 record from the mFC, we lowered the probe ~100 µm during the pre-habituation period to
1391 reach a final DV position of between -1.3 and -1.5 mm below the brain surface (i.e. between
1392 -2.05 and -2.25 mm from bregma). This places most channels in the prelimbic cortex
1393 (<http://labs.gaidi.ca/mouse-brain-atlas/>). Neural activity was acquired at 30 kHz with a 32-
1394 channel Intan RHD 2132 amplifier board (Intan Technologies) connected to an OpenEphys
1395 acquisition board. Behavioural, video and electrophysiological data were synchronised using
1396 sync pulses output from the pyControl system. Recordings were spike sorted using
1397 Kilosort⁵², versions 2.5 and 3, and manually curated using phy
1398 (<https://github.com/kwikteam/phy>). Clusters were classified as single units and retained for
1399 further analysis if they had a characteristic waveform shape, showed a clear refractory
1400 period in their autocorrelation and were stable over time. We performed tracking of the mice
1401 in the video data using DeepLabCut⁵³, a Python package for marker-less pose estimation

1402 based in the TensorFlow machine learning library. Positions of the back of a mouse's head
1403 in x, y pixel coordinates were converted to region of interest information (which maze node
1404 or edge the animal is in for each frame) using a set of binary masks defined in ImageJ that
1405 partition the frame into its sub components.

1406

1407 **Behavioural Scoring**

1408 Performance was assessed by quantifying the percentage of transitions where animals took
1409 one of the shortest available routes between two rewards. We also quantified the path length
1410 taken between rewards and divided this by the shortest length to give the "relative path
1411 distance" covered per transition.

1412

1413 When using "percentage of shortest path transitions" as a criteria, chance levels were
1414 calculated either analytically or empirically. Analytical chance levels were calculated by
1415 assuming a randomly diffusing animal and calculating the chances the animal will move from
1416 node X to node Y in N steps. Empirical chance levels were calculated by using the location-
1417 to-location transition matrix recorded for each animal in the Exploration session before any
1418 exposure to reward on the maze. When using the "relative path distance" measure, chance
1419 was calculated empirically using the mean relative path distance for transitions in the first
1420 trial averaged across the first 5 tasks.

1421

1422 Correct transition entropy (The animal's entropy when taking the shortest route between
1423 rewards) was calculated for transitions where there was more than one shortest route
1424 between rewards. We calculated the probability distribution across all possible shortest
1425 paths for a given transition and calculated entropy as follows:

1426

1427 **Entropy = $\sum pk \log_x(pk)$**

1428

1429 Where **x** is the logarithmic base which is set to the number of shortest routes and **pk** is the
1430 probability of each transition. Thus an entropy of 1 signifies complete absence of a bias for
1431 taking any one path and an entropy of 0 means only one of the paths is taken (i.e. maximum
1432 stereotypy).

1433

1434 **Activity Normalisation**

1435 We aimed to define a task space upon which to project the activity of the neurons. To
1436 achieve this, we aligned and normalised vectors representing neuronal activity and maze
1437 location to the task states. Activity was aligned such that the consumption of reward **a**
1438 formed the beginning of each row (trial) and consumption of the next reward **a** started a new
1439 row. Normalisation was achieved such that all states were represented by the same number
1440 of bins (90) regardless of the time taken in each state. Thus, the first 90 bins in each row
1441 represented the time between rewards **a** and **b**, the second between **b** and **c**, the third
1442 between **c** and **d** and the last between **d** and **a**. We then computed the averaged neuronal
1443 activity for each bin. Thus the activity of each neuron was represented by an Nx360 matrix,
1444 where N is the number of trials and 360 bins represent task space for each trial. This activity
1445 was then averaged by taking the mean across trials, and smoothed by fitting a Gaussian
1446 kernel (sigma=10 degrees). To avoid edge effects when smoothing, the mean array was
1447 concatenated to itself three times, then smoothed, then the middle third extracted to
1448 represent the smoothed array. To reflect the circular structure of the task, the mean and

1449 standard error of the mean of this normalised and smoothed activity were projected on polar
1450 plots (e.g. Figure 2b-d).

1451

1452 **Tuning to basic task variables**

1453

1454 Generalised linear model

1455 To assess the degree to which mFC neurons are tuned to task space, we used a linear
1456 regression to model each neuron's activity. Specifically, we aimed to quantify the degree to
1457 which goal-progress and location tuning of the neurons is consistent across tasks and
1458 states. For this we used a leave-one-out cross-validation design: we divided all tasks into the
1459 time periods spanned by each of the 4 states and used all data except one Task/State
1460 combination to train the model. The remaining task/state (e.g. task 3, state B) was used to
1461 test the model. This was repeated so that each task/state combination had been left out as a
1462 test period once. The training periods were used to calculate mean firing rates for 5 levels of
1463 goal-progress relative to reward (5 goal-progress bins) and each maze location (9 possible
1464 node locations). Edges were excluded from analyses since they are systematically not
1465 visited at the earliest goal-progress bin. The mean firing rates for goal-progress and place
1466 from the training sessions were used as (separate) regressors to test against the binned
1467 firing rate of the cell in the test data (held out task-state combination). To assess the validity
1468 of any putative task tuning, a number of potentially confounding variables were added to the
1469 model. These were: acceleration, speed, time from reward, and distance from reward. This
1470 procedure was repeated for all Task/State combinations and a separate regression
1471 coefficient value was calculated for each. To assess significance per neuron, we repeated
1472 the regression but with random circular shifts of each neuron's activity array and computed
1473 regression coefficient values for each iteration (100 iterations) and then used the 95th
1474 percentile of this distribution as the regression coefficient threshold. Neurons with an actual
1475 regression coefficient value higher than this threshold were considered to be tuned for this
1476 variable. Two proportions Z-tests were performed to assess whether the proportion of
1477 neurons with significant regression coefficient values for a given variable were statistically
1478 higher than a chance level of 5%.

1479

1480 State tuning

1481 For state tuning, we first wanted to test whether neurons were tuned to a given state in a
1482 given task. We therefore analysed state tuning separately from the GLM above, which
1483 explicitly tests for the consistency of tuning across tasks. Instead, we used a z-scoring
1484 approach. First we took the peak firing rate in each state and trial, giving 4 values per trial:
1485 i.e. a *maximum activity* matrix with dimensions Nx4 where N is the number of trials. Then we
1486 z-scored each row of this maximum activity matrix (i.e. giving a mean of 0 and standard
1487 deviation of 1 for each trial). We then extracted the z-scores for the preferred state across all
1488 N trials and subsequently conducted a T test of this array against 0. Neurons with a P value
1489 of < 0.05 for a given task were taken to be state tuned in that task.

1490

1491 **Neuronal generalisation**

1492 To assess whether individual neurons maintained their state preference across tasks we
1493 quantified the angle made between a neuron in one task and the same neuron in another
1494 task. Only state-tuned neurons were used in these analyses. To ensure we captured
1495 robustly state tuned neurons, we restricted analyses to neurons state tuned in more than

1496 1/3rd of the recorded tasks. This subsetting is used throughout the manuscript where state
1497 tuned cells are investigated. Quantifying the angle between neurons was achieved by
1498 rotating the neuron in task Y by 10 degree intervals and then computing the Pearson
1499 correlation between this rotated firing rate vector and the mean firing rate vector in task X.
1500 Using this approach, we found for each neuron the rotation that gave the highest correlation.
1501 For cross-task comparisons we calculated a histogram of the angles across the entire
1502 population and averaged this across both comparisons (X vs Y and X vs Z). Within task
1503 histograms were computed by comparing task X to task X' (Figure 3b). To compute the
1504 proportion of neurons that generalised their state tuning, we found the maximum rotation
1505 across both comparisons (X vs Y and X vs Z). We then set 45 degrees either side of 0
1506 rotation across all tasks as the generalisation threshold. Because this represents $\frac{1}{4}$ of the
1507 possible rotation angles, chance level is equal to $\frac{1}{4}^m$ where m is the number of comparisons.
1508 When calculating generalisation across one comparison, chance level is therefore 25%,
1509 whereas when two comparisons are taken, the chance level is 1/16 (6.25%).
1510
1511 Generalisation could also be expressed at the level of tuning relationships between neurons.
1512 For example two neurons that are tuned to A and C in one task could then be tuned to B and
1513 D in another, thereby maintaining their task-space angle (180 degrees) but remapping in
1514 task space across tasks. To test for this, we computed the tuning angle *between* pairs of
1515 neurons and assessed how consistent this was across tasks. This angle was computed by
1516 rotating one neuron by 10 degree intervals and calculating the Pearson correlation between
1517 the mean firing vector of neuron k and the rotated firing vector for neuron j. The rotation with
1518 the highest Pearson correlation gave the between-neuron angle (Figure 3c). Thus, we
1519 compared the angle between a pair of neurons in task X to the same between-neuron angle
1520 in tasks Y and Z. Again histograms were averaged across both comparisons (X vs Y and X
1521 vs Z) for cross-task histograms while within-task histograms were computed by comparing
1522 task X to task X' (Figure 3c). To compute the proportion of neuron pairs that were coherent
1523 across tasks, we found the maximum rotation of the angle between each pair across both
1524 comparisons (X vs Y and X vs Z). We then set 45 degrees either side of 0 rotation across all
1525 tasks as the coherence threshold. Because this represents $\frac{1}{4}$ of the possible rotation angles,
1526 chance level is equal to $\frac{1}{4}^m$ where m is the number of comparisons and therefore=25% for
1527 one comparison and 1/16 (6.25%) for two comparisons.
1528
1529 To assess whether the mFC population was organised into modules of coherently rotating
1530 neurons, we used a clustering approach. The first step was to take the *maximum* difference
1531 in pairwise, between-neuron angles (across all comparisons) and convert this into a
1532 maximum circular distance (1- cosine(angle)) thereby generating a distance matrix reflecting
1533 coherence relationships between neurons (incoherence matrix). The second step is to
1534 compute a low dimensional embedding of this incoherence matrix, using t-distributed
1535 stochastic neighbour embedding (t-SNE; using the TSNE function of scikit learn manifold
1536 library, with perplexity=5). Finally, the third step is to use hierarchical clustering on this
1537 embedded data (using the AgglomerativeClustering function of scikit learn cluster library,
1538 with distance threshold=300). This procedure sorts neurons into clusters reflecting
1539 coherence relationships between neurons. We quantified the degree of clustering by
1540 computing the Silhouette Score for the clusters computed in each recording day:
1541
1542 **Silhouette Score=(b-a)/max(a,b)**
1543

1544 Where a =mean intracluster distance, b =mean intercluster distance. We repeated the same
1545 procedure but for permuted data, where state tuning in task X and goal-progress tuning in all
1546 tasks was identical to the real data but the state preference of each neuron remapped
1547 randomly across tasks. This allowed us to compare the Silhouette Scores for the real and
1548 permuted data (Figure 3d). To visualise clusters and the tuning of neurons within them in the
1549 same plot, we plotted some example neurons from a single recording day where the x and
1550 the y axes represented state tuning and the y axis arranged neurons based on their cluster
1551 id (the ordering along the z axis is arbitrary; Figure 3e).

1552

1553 **Mnemonic task space tuning**

1554 The Task structured memory buffers (SMBs) model predicts the existence of neurons that
1555 maintain an invariant task space lag from a particular anchor representing a behavioural
1556 step, regardless of the task sequence. Concretely, behavioural steps are conjunctions of
1557 goal-progress (operationally divided into early, intermediate or late) and place (nodes 1-9).
1558 To test this prediction, we used three complementary analysis methods. All of these
1559 analyses were conducted on data where two recording days were combined and spike-
1560 sorted concomitantly, giving a total of 6 unique tasks per animal (with two exceptions that
1561 had 4 and 5 tasks each; see exclusions under the “Numbers” section below). For all of these
1562 analyses, only state tuned neurons were used (see “Tuning to basic task variables”).

1563

1564 Method 1: Model fitting

1565 For each neuron we computed a regression model that described state-tuning activity as a
1566 function of all possible combinations of goal-progress/place and all task lags from each
1567 possible goal-progress/place. Thus a neuron could fire at a particular goal-progress/place
1568 but also at a particular lag in task space from this goal-progress/place. We used an Elastic
1569 Net (using the ElasticNet function from the scikit learn linear_model package) that included a
1570 regularisation term which was a 1:1 combination of L1 and L2 norms. The alpha for
1571 regularisation was set to 0.01. A total of 9x3x12 (312) regressors were used for each
1572 neuron, corresponding to 9 locations, 3 goal-progress bins (so 27 possible goal-
1573 progress/place anchor points) and 12 lags in task space from the anchor (4 states x 3 goal-
1574 progress bins). We trained the model on 5 (training) tasks and then used the resultant
1575 regression coefficients to predict the activity of the neuron in a left out (test) task. Both
1576 training and cross-validation were only done in the preferred goal-progress of each neuron to
1577 ensure our prediction results are due to state preference and not the strong effect of goal-
1578 progress preference (Figure 2 e,f). For non-zero-lag neurons, we only used state-tuned
1579 neurons with all of the three highest regression coefficient values at non-zero lag from an
1580 anchor (lag from anchor of 30 degrees or more for Figure 5b bottom; 90 degrees (one state)
1581 or more for Extended data figure 5b) in the training tasks. Also, for non-zero lag neurons, we
1582 only use regression coefficient values either 30 degrees (Figure 5b) or 90 degrees
1583 (Extended data figure 5b) either side of the anchor point to predict the state tuning of the
1584 cells. This ensured that the prediction was only due to lagged activity and not direct tuning of
1585 the neurons to the goal-progress/place conjunction.

1586

1587 Method 2: Lagged spatial similarity

1588 To detect putative mnemonic task space neurons, we calculated spatial tuning to where the
1589 animal was at different task lags in the past (Figure 5c). While spatial neurons should
1590 consistently fire at the same locations(s) at zero lag, neurons that track a memory of the

1591 goal-progress/place anchor will instead show a peak in their cross-task spatial correlation at
1592 a non-zero task lag in the past (Figure 5c). To quantify this effect, we used a cross-validation
1593 approach, using all tasks but one to calculate the lag at which cross-task spatial correlation
1594 was maximal, and then measuring the Pearson correlation between the spatial maps in the
1595 left out task and the training tasks at this lag (Figure 5d). To account for the strong goal-
1596 progress tuning of the neurons, all maps were computed in each neuron's preferred goal-
1597 progress.
1598

1599 **Method 3: Single anchor alignment**

1600 This approach assumes each neuron can only have a single goal-progress/place anchor and
1601 quantifies the degree to which task-space lag for this neuron is conserved across tasks. We
1602 fitted the anchor by choosing the goal-progress/place conjunction which maximises the
1603 correlation between lag-tuning-curves in all but one (training) tasks, and again used cross-
1604 validation by assessing whether this anchor leads to the same lag tuning in the left out (test)
1605 task. The fitting was conducted by first identifying the times an animal visited a given goal-
1606 progress/place and sampling 360 bins (1 trial) of data starting at this visit, then averaging
1607 activity aligned to all visits in a given task, and smoothing activity as described above (Under
1608 "Activity Normalisation"). This realigned activity is then compared across tasks to compute
1609 the angle (θ) between the neuron's mean aligned/normalised firing rate vector across tasks.
1610 This involves essentially doing all the steps for the "Neuronal generalisation analysis" but for
1611 the anchor-aligned activity instead of state-A -aligned activity. This was done for all possible
1612 task combinations and then a distance matrix (M) was computed by taking distance=1-
1613 cosine(θ). This distance matrix M has dimensions Ntraining_tasks x Ntraining_tasks x
1614 Nanchors (typically 5x5x27 as there are usually 6 tasks, meaning 5 training tasks are used,
1615 and 3x9 possible anchors corresponding to 3 possible phases (early, intermediate and late)
1616 and 9 possible maze locations). The distance can then be averaged across all comparisons
1617 to find the mean distance between all comparisons for a given anchor for a given training
1618 task - generating a mean-distance matrix (M_{mean}). This M_{mean} matrix has the
1619 dimensions Ncomparisons x Nanchors; where Ncomparisons= Ntraining_tasks – 1; typically
1620 this will be 4 x 27). The entry with the minimum value in this M_{mean} matrix gives the
1621 combination of training task and goal-progress/place anchor that best aligns the neuron –
1622 the training task selected is referred to as the reference task. Next, the neuron's mean
1623 activity in the test task is aligned to visits to the best anchor calculated from the training
1624 tasks. This allows calculating how much this aligned activity array has remapped relative to
1625 the aligned activity in the reference training task - if it has remapped by 0 degrees or close to
1626 0 degrees (within a 45 degree span either side of zero) then the neuron is spatially anchored
1627 (i.e. maintains a consistent angle with its anchor across tasks). For a given test-train split, we
1628 computed a histogram of the angles across all the neurons and then we averaged the
1629 histograms across all test-train splits to visualise the overall distribution of angles between
1630 training and test tasks (e.g. Figure 5f). To quantify the degree of alignment further, we
1631 measured the correlation between the anchor-aligned activity of neurons in the test task
1632 versus reference training task; only considering activity of neurons in their preferred goal-
1633 progress bin (Figure 5g). The neuron's lag from its anchor was identified by finding the lag at
1634 which anchor-aligned activity was maximal. This lag is used below for "Predicting
1635 Behavioural Choices".
1636

1637 For per mouse effects reported in Extended data figure 5, we tested whether the number of
1638 mice with a mean cross-validated correlation above 0 is higher than chance, chance level

1639 being a uniform distribution (50:50 distribution of per mouse correlation means above and
1640 below zero). We used a one-tailed binomial test against this chance level. All 5 mice need to
1641 have mean positive values for this test to yield significance.

1642

1643 **Predicting Behavioural Choices**

1644

1645 The SMB model proposes that behavioural choices should be predictable from bumps of
1646 activity along specific memory buffers long before an animal makes a particular choice. By
1647 "choice" here we mean a decision to move from one node to one of the immediately adjacent
1648 nodes on the maze (e.g. from location 1 should I go to location 2 or 4?; Figure 6a). To test
1649 whether these choices are predictable from distal neuronal activity we used a Logistic
1650 regression model. For this analysis we used only consistently anchored neurons, that is
1651 neurons that had the same anchor and same lag to anchor in at least half of the tasks. This
1652 relied on the single-anchor analysis (Figure 5e-g). Furthermore, to avoid contamination of
1653 our results due to simple spatial tuning, we only used neurons with activity lagged far from
1654 their anchor (at least 30 degrees in task space either side of the anchor e.g. Figure 6b,d,e,f;
1655 and this was also repeated for lags of at least 90 degrees (one whole state) either side of the
1656 anchor: e.g. Figure 6g and Extended data 6b,c). We measured the activity of a given neuron
1657 during its "bump" time, that is the time at which a neuron is lagged relative to its anchor.
1658 Precisely, this is the mean firing rate from a period starting with the lag time from the anchor
1659 and ending 30 degrees forward in task space from that point (i.e. 1/3rd of a state). This
1660 mean activity was inputted on a trial by trial basis every time the animal was at a goal-
1661 progress/place conjunction that was one step before the goal-progress/place anchor in
1662 question (e.g. if the anchor is at early goal-progress in place 2, the possible goal-
1663 progress/places before this are: late goal progress in place 1, late goal-progress in place 3
1664 and late goal-progress in place 5; see maze structure in Figure 1a). We used this activity to
1665 predict a binary vector that has 1s when the animal visits the anchor and 0 when the animal
1666 could have visited the anchor (i.e. was one step away from it) but did not chose to visit the
1667 anchor. To remove confounds due to the autocorrelated previous behavioural choices, we
1668 added previous choices up to 5 trials in the past into the regression model. Furthermore, to
1669 assess whether any observed prediction was specific to the "bump" time as predicted by the
1670 SMB model, we repeated the logistic regression for other times (random times, decision time
1671 (30 degrees from the potential anchor visit) and times shifted by 90, 180 and 270 degrees
1672 from the bump time). We further repeated this regression only for visits to non-zero goal-
1673 progress anchors (i.e. non-rewarded locations) and also only taking neurons that are more
1674 distal from their anchor (i.e. at least 90 degree separation either side of the anchor). We also
1675 conducted this analysis separately for high probability choices (choices that were in the top
1676 50% of the maze transition probability distribution calculated in a baseline exploration
1677 session before any reward or task was ever experienced on the maze; e.g. Figure 6e) and
1678 low probability choices (bottom 50% of the same transition probability distribution; e.g.
1679 Figure 6f). This allowed us to test whether frontal neurons generally predict all choices or
1680 more specifically predict choices which animals were not already predisposed to make.

1681

1682 **Sleep/Rest analysis**

1683 To investigate the internal organisation of task-related mFC activity we recorded neuronal
1684 activity in a separate enclosure containing bedding from the animal's home cage but no
1685 reward or task-relevant cues. Animals were pre-habituated to sleep/rest in these "sleep

1686 boxes" before the first task began. We measured neuronal activity across sleep/rest
1687 sessions both before any tasks on a given day and after each session. The first (pre-task)
1688 sleep/rest session was 1 hour long, inter-session sleep/rest sessions were 20 minutes long
1689 and the sleep/rest session after the last task was 45 minutes long. All sessions except the
1690 first sleep session were designated as "post-task" sleep sessions.
1691
1692 We wanted to assess whether the memory buffers were organised by the task structure,
1693 which in our case is a ring. In other words, do neurons internally maintain this ring-like
1694 organisation even in the absence of externally structured sensory/behavioural input? Activity
1695 was binned in 250 ms bins and cross-correlations between each pair of neurons were
1696 calculated using this binned activity. We then regressed the awake angle difference between
1697 pairs of neurons sharing the same anchor against this sleep cross-correlation. This angle
1698 was taken from the first task on a given day for pre-task sleep, and from the task
1699 immediately before the sleep session for all post-task sleep sessions. The idea is that
1700 neurons closer to each other in a given neuronal state-space should be more likely to be
1701 coactive within a small time window compared to neurons farther apart. To control for place
1702 and goal-progress tuning, we added the spatial map correlation and circular goal-progress
1703 distance as co-regressors. Thus we assessed the degree to which the regression
1704 coefficients were negative (i.e. smaller distances correlate with higher coactivity).
1705
1706 We first sought to identify whether the neurons were functionally organised into rings or
1707 delay lines. For this we measured the forward distance between pairs of co-anchored
1708 neurons in reference to their anchor. If neurons are internally organised on a line, then the
1709 larger the forward distance between a pair of neurons the further away two neurons are from
1710 each other in neuronal state space, and hence the less coactive they will be (Figure 7a). If
1711 neurons are arranged on a ring, forward distance should also positively correlate with
1712 neuronal state space distance for pairs of neurons separated by less than 180 degrees in
1713 task space. Beyond 180 degrees, forward distance should instead negatively correlate with
1714 neuronal state space distance, as neurons circle back towards each other (Figure 7a). In this
1715 situation, circular distance, not forward distance, would be the best description of the
1716 neuronal state space. Thus, for a delay line, the higher the forward distance beyond 180 the
1717 further the neurons should be from one another, and hence the coefficient for the regression
1718 between forward distance and coactivity should be negative (Figure 7a). For a ring, the
1719 coefficient value for forward distance against coactivity should conversely be positive. This
1720 pattern inverts when considering the circular distance: positive coefficient for delay line and
1721 negative coefficient for rings. We used a linear regression to compute the regression
1722 coefficients for circular and forward distances in the same regression for *all* neuron pairs
1723 (Figure 7b). To control for place and goal-progress tuning, we again added the spatial map
1724 correlation and circular goal-progress distance as co-regressors. We also used a linear
1725 regression to compute this coefficient for only the circular distance for pairs of neurons
1726 separated by 180 degrees or more, while again co-regressing spatial correlations and goal-
1727 progress distances (Figure 7c). We further analysed whether neurons that shared the same
1728 anchor showed stronger state-space relationships than those that have different anchors.
1729 For this we restricted our analysis to consistently anchored neurons, that is neurons that had
1730 the same anchor and same lag to anchor in at least half of the tasks. We conducted the
1731 regression of circular distance against sleep cross-correlation either for pairs of neurons that
1732 share the same anchor, or those that have different anchors (Figure 7d). As before, we co-
1733 regressed spatial correlations and goal-progress distances.

1734
1735 Note that, arguably, the last analysis (Figure 7d) could be seen as sufficient to indicate a
1736 ring-like structure. However, this is only true when assuming a uniform distribution of
1737 pairwise angles, which we know not to be the case given the overrepresentation of neurons
1738 with a small lag from the anchor (Extended data Figure 5c), thus necessitating the first two
1739 analyses (Figure 7b,c).

1740

1741 **Numbers**

1742 Animals

1743 11 animals in total were used for behavioural recordings across 3 separate cohorts
1744 conducted by 2 different experimenters (AH and ME) - 4 of these animals only completed 10
1745 tasks as part of the first cohort and the remaining 7 completed 40 tasks
1746 Of the 11 animals, 5 animals in total were used for electrophysiological recordings, the
1747 remaining 6 animals are accounted for below:

1748 • 3 animals (in cohort 1) were not implanted at any point
1749 • 1 animal was implanted with silicon probes but was part of the first cohort so did not
1750 get to the 3 task days (i.e. only completed the first 10 tasks)
1751 • 2 animals were implanted but their signal was lost before the 3 task days

1752

1753 Exclusions: No animals were excluded from analyses: All animals (11) were included in the
1754 behavioural analyses, and all animals for which there was an electrophysiological signal by
1755 the 3 task days (5) were included in the electrophysiological analyses.

1756

1757 Electrophysiological recording days

1758 All electrophysiological data was obtained from days where animals completed 3 tasks per
1759 day (cohorts 2 and 3) and in late tasks (tasks 21-40) where animals showed robust evidence
1760 for having learned the abstract task structure (Figure 1). 33 recording days were spike sorted
1761 as individual days and were used in Figure 2 and Figure 3. In 23 of those 32 days, animals
1762 completed more than 2 trials in the last session (the repeat of task X), hence fewer neurons
1763 appear in analyses comparing session X with session X' (Figure 3b,c and Extended data
1764 figure 3b,c). 14 recording days were merged to create double-days, allowing tracking of
1765 neuronal activity across up to 6 tasks. Although 6 tasks were always attempted across each
1766 double day, of these 14 double days, 12 spanned 6 tasks, 1 spanned 5 tasks and 1 spanned
1767 4 tasks. This was due to file corruption (1 task) or animals lacking motivation (less than 2
1768 trials completed per task; 2 tasks). Double days were used in Figures 3,5,6 and 7. Only
1769 single days were used in Figure 2.

1770

1771 Data was obtained from all possible double-days from late tasks (tasks 21-40). This means 3
1772 double days per mouse (spanning the last 18 tasks) - except for two mice where each had
1773 one double-day where the signal had dropped, and one mouse where 4 double days were
1774 used (including a day spanning tasks 20-22) - thus a total of 14 double-days were used.

1775

1776 Total number of neurons

1777 We report “neuron-days” - i.e. by summing up each day’s neuron yield below:

1778

1779 1807 total “neuron-days” (i.e. while splitting each double-day into two and summing across
1780 days)

1781 1438 separately sorted neuron days (i.e. using a single yield for each double day), of which
1782 there were:

- 1783 • 747 neuron days on single days
- 1784 • 691 neuron days on double days

1785

1786 Exclusions:

- 1787 • Recording days with 10 or fewer state-tuned neurons were excluded from analyses
1788 of pairwise coherence between neurons.
- 1789 • Neurons whose mean firing rate during sleep dropped to less than 20% of their
1790 awake firing rate were excluded from preplay/replay analysis - this was assumed to
1791 be largely a spike sorting artefact.
- 1792 • Only neurons that were significantly state tuned (neurons with significant state tuning
1793 in more than 1/3rd of all tasks) and were sorted across double days were used in
1794 Figure 5 - this amounted to 350 neurons - although for the regression analysis, 10
1795 additional neurons were excluded due to having all zero regression coefficient values
1796 (caused by the regularisation), giving a total of 340 neurons for this analysis (Figure
1797 5a,b).

1798

1799 **Acknowledgements**

1800 We would like to thank Gil Costa for designing the model schematics; Diksha Gupta, Michael
1801 Bukwich, Thomas Mrsic-Flogel, Ray Dolan and Masahiro Nakano for valuable input on the
1802 manuscript; Raquel Pinacho, Lauren Burgeno, Beatriz Godinho, Marta Blanco Pozo,
1803 Veronika Samborska, for guidance on data collection and data preprocessing; All members
1804 of the Behrens and Walton labs for their input and support during the project.

1805

1806 ALH is supported by a Wellcome Trust PhD studentship (220047/Z/19/Z), TEJB is
1807 supported by a Wellcome Principal Research Fellowship (219525/Z/19/Z) and a Wellcome
1808 Collaborator award (214314/Z/18/Z). The Wellcome Centre for Integrative Neuroimaging and
1809 Wellcome Centre for Human Neuroimaging are each supported by core funding from the
1810 Wellcome Trust (203139/Z/16/Z, 203147/Z/16/Z). JCRW.; is supported by the Sir Henry
1811 Wellcome Post-doctoral Fellowship (222817/Z/21/Z). WD is supported by the Gatsby
1812 Charitable Foundation. TA is supported by the Wellcome Trust career development award:
1813 225926/Z/22/Z. MEW is supported by a Wellcome Collaborator award (214314/Z/18/Z) and a
1814 Wellcome trust SRF (202831/Z/16/Z).

1815

1816 **Author Contributions**

1817 ME, JCRW, MEW, TA and TEJB conceptualised the study and designed the ABCD task;
1818 ME, ALH and AB performed the surgeries and collected the behavioural and
1819 electrophysiological data with input from TA and MEW; ME and ALH analysed and
1820 interpreted the data with input from TEJB and TA; ME conceptualised the model with input
1821 from TEJB; ME, ALH, JCRW, WD, MW, TA and TEJB elaborated on the core model; ME,
1822 ALH, TA and TEJB wrote the manuscript and made the figures with input from all other
1823 authors.

1824

1825 **Competing interests**

1826 The authors declare no competing interests.

1827

1828 **Materials & Correspondence**

1829 Correspondence to Mohamady El-Gaby (mohamady.el-gaby@ndcn.ox.ac.uk) and Tim EJ
1830 Behrens (timothy.behrens@ndcn.ox.ac.uk).

1831

1832 **Data Availability**

1833 All data will be made available upon publication.

1834

1835 **Code Availability**

1836 All code will be made available upon publication.

1837

1838 **References**

1. O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* **34**, 171–175 (1971).
2. Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005).
3. Whittington, J. C. R. *et al.* The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell* **183**, 1249–1263.e23 (2020).
4. Behrens, T. E. J. *et al.* What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron* **100**, 490–509 (2018).
5. George, D. *et al.* Clone-structured graph representations enable flexible learning and vicarious evaluation of cognitive maps. *Nat. Commun.* **12**, 2392 (2021).
6. Banino, A. *et al.* Vector-based navigation using grid-like representations in artificial agents. *Nature* **557**, 429–433 (2018).
7. Bartlett, F. C. *Remembering: A study in experimental and social psychology*. xix, 317 (Cambridge University Press, 1932).
8. Piaget, J. *The origins of intelligence in children*. 419 (W W Norton & Co, 1952).
doi:10.1037/11494-000.
9. Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. How to Grow a Mind: Statistics, Structure, and Abstraction. *Science* **331**, 1279–1285 (2011).
10. Tomov, M. S., Yagati, S., Kumar, A., Yang, W. & Gershman, S. J. Discovery of hierarchical representations for efficient planning. *PLOS Comput. Biol.* **16**, e1007594 (2020).
11. Wallis, J. D., Anderson, K. C. & Miller, E. K. Single neurons in prefrontal cortex encode abstract rules. *Nature* **411**, 953–956 (2001).
12. Tse, D. *et al.* Schemas and memory consolidation. *Science* **316**, 76–82 (2007).
13. Gershman, S. J. & Niv, Y. Learning latent structure: carving nature at its joints. *Curr. Opin. Neurobiol.* **20**, 251–256 (2010).
14. Baldassano, C., Hasson, U. & Norman, K. A. Representation of Real-World Event Schemas during Narrative Perception. *J. Neurosci.* **38**, 9689–9699 (2018).
15. 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* **109**, 713–723.e7 (2021).
16. Zhou, J. *et al.* Evolving schema representations in orbitofrontal ensembles during learning. *Nature* **590**, 606–611 (2021).
17. Bernardi, S. *et al.* The Geometry of Abstraction in the Hippocampus and Prefrontal Cortex. *Cell* **183**, 954–967.e21 (2020).
18. Shallice, T. Specific impairments of planning. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **298**, 199–209 (1982).
19. Averbeck, B. B., Sohn, J.-W. & Lee, D. Activity in prefrontal cortex during dynamic selection of action sequences. *Nat. Neurosci.* **9**, 276–282 (2006).

1877 20. Balleine, B. W. & Dickinson, A. Goal-directed instrumental action: contingency and incentive
1878 learning and their cortical substrates. *Neuropharmacology* **37**, 407–419 (1998).

1879 21. Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annu. Rev.*
1880 *Neurosci.* **24**, 167–202 (2001).

1881 22. Chiang, F.-K. & Wallis, J. D. Spatiotemporal encoding of search strategies by prefrontal
1882 neurons. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 5010–5015 (2018).

1883 23. Barraclough, D. J., Conroy, M. L. & Lee, D. Prefrontal cortex and decision making in a mixed-
1884 strategy game. *Nat. Neurosci.* **7**, 404–410 (2004).

1885 24. Seo, H. & Lee, D. Cortical mechanisms for reinforcement learning in competitive games.
1886 *Philos. Trans. R. Soc. B Biol. Sci.* **363**, 3845–3857 (2008).

1887 25. Rubin, A. *et al.* Revealing neural correlates of behavior without behavioral measurements.
1888 *Nat. Commun.* **10**, 4745 (2019).

1889 26. Kaefer, K., Nardin, M., Blahna, K. & Csicsvari, J. Replay of Behavioral Sequences in the
1890 Medial Prefrontal Cortex during Rule Switching. *Neuron* **106**, 154–165.e6 (2020).

1891 27. Basu, R. *et al.* The orbitofrontal cortex maps future navigational goals. *Nature* 1–4 (2021)
1892 doi:10.1038/s41586-021-04042-9.

1893 28. Shima, K. & Tanji, J. Neuronal activity in the supplementary and presupplementary motor
1894 areas for temporal organization of multiple movements. *J. Neurophysiol.* **84**, 2148–2160 (2000).

1895 29. Xie, J. & Padoa-Schioppa, C. Neuronal remapping and circuit persistence in economic
1896 decisions. *Nat. Neurosci.* **19**, 855–861 (2016).

1897 30. Reinert, S., Hübener, M., Bonhoeffer, T. & Goltstein, P. M. Mouse prefrontal cortex represents
1898 learned rules for categorization. *Nature* **593**, 411–417 (2021).

1899 31. Samborska, V., Butler, J. L., Walton, M. E., Behrens, T. E. & Akam, T. Complementary Task
1900 Representations in Hippocampus and Prefrontal Cortex for Generalising the Structure of Problems.
1901 *bioRxiv* 2021.03.05.433967 (2021) doi:10.1101/2021.03.05.433967.

1902 32. Whittington, J. C. R., McCaffary, D., Bakermans, J. J. W. & Behrens, T. E. J. How to build a
1903 cognitive map. *Nat. Neurosci.* **25**, 1257–1272 (2022).

1904 33. Barone, P. & Joseph, J.-P. Prefrontal cortex and spatial sequencing in macaque monkey.
1905 *Exp. Brain Res.* **78**, 447–464 (1989).

1906 34. Swaminathan, S. *et al.* Schema-learning and rebinding as mechanisms of in-context learning
1907 and emergence. at <http://arxiv.org/abs/2307.01201> (2023).

1908 35. Wang, J. X. *et al.* Prefrontal cortex as a meta-reinforcement learning system. *Nat. Neurosci.*
1909 **21**, 860–868 (2018).

1910 36. Stensola, H. *et al.* The entorhinal grid map is discretized. *Nature* **492**, 72–78 (2012).

1911 37. Petrides, M. & Milner, B. Deficits on subject-ordered tasks after frontal- and temporal-lobe
1912 lesions in man. *Neuropsychologia* **20**, 249–262 (1982).

1913 38. Xie, Y. *et al.* Geometry of sequence working memory in macaque prefrontal cortex. *Science*
1914 **375**, 632–639 (2022).

1915 39. Gardner, R. J. *et al.* Toroidal topology of population activity in grid cells. *Nature* **602**, 123–128
1916 (2022).

1917 40. Trettel, S. G., Trumper, J. B., Hwaun, E., Fiete, I. R. & Colgin, L. L. Grid cell co-activity
1918 patterns during sleep reflect spatial overlap of grid fields during active behaviors. *Nat. Neurosci.* **22**,
1919 609–617 (2019).

1920 41. Hawkins, J., Lewis, M., Klukas, M., Purdy, S. & Ahmad, S. A Framework for Intelligence and
1921 Cortical Function Based on Grid Cells in the Neocortex. *Front. Neural Circuits* **12**, (2019).

1922 42. McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. & Moser, M.-B. Path integration
1923 and the neural basis of the ‘cognitive map’. *Nat. Rev. Neurosci.* **7**, 663–678 (2006).

1924 43. Manns, J. R. & Eichenbaum, H. Evolution of declarative memory. *Hippocampus* **16**, 795–808
1925 (2006).

1926 44. Tolman 1948 Cognitive Maps in Rats and Men. *Psychol. Rev.*, 55 (1948), pp. 189–208

1927 45. Eichenbaum, H. Hippocampus: Cognitive Processes and Neural Representations that
1928 Underlie Declarative Memory. *Neuron* **44**, 109–120 (2004).

1929 46. John H. Wen, Ben Sorscher, Surya Ganguli, & Lisa M Giocomo. One-shot entorhinal maps
1930 enable flexible navigation in novel environments. *bioRxiv* 2023.09.07.556744 (2023)
1931 doi:10.1101/2023.09.07.556744.

1932 47. Whittington JCR, Dorrell W, Behrens TEJ, Ganguli S, El-Gaby M On prefrontal working
1933 memory and hippocampal episodic memory: Unifying memories stored in weights and activation slots.
1934 BioRxiv Preprint (2023)

1935 48. Ayaka Bota *et al.* Shared and unique properties of place cells in anterior cingulate cortex and
1936 hippocampus. *bioRxiv* 2021.03.29.437441 (2021) doi:10.1101/2021.03.29.437441.

1937 49. Dominey, P., Arbib, M. & Joseph, J. P. A model of corticostriatal plasticity for learning
1938 oculomotor associations and sequences. *J. Cogn. Neurosci.* **7**, 311–336 (1995).

1939 50. Beiser, D. G. & Houk, J. C. Model of cortical-basal ganglionic processing: encoding the serial
1940 order of sensory events. *J. Neurophysiol.* **79**, 3168–3188 (1998).

1941 51. Akam, T. *et al.* Open-source, Python-based, hardware and software for controlling
1942 behavioural neuroscience experiments. *eLife* **11**, e67846 (2022).

1943 52. Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M. & D, H. K. Kilosort: realtime spike-
1944 sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv* 061481 (2016)
1945 doi:10.1101/061481.

1946 53. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with
1947 deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).

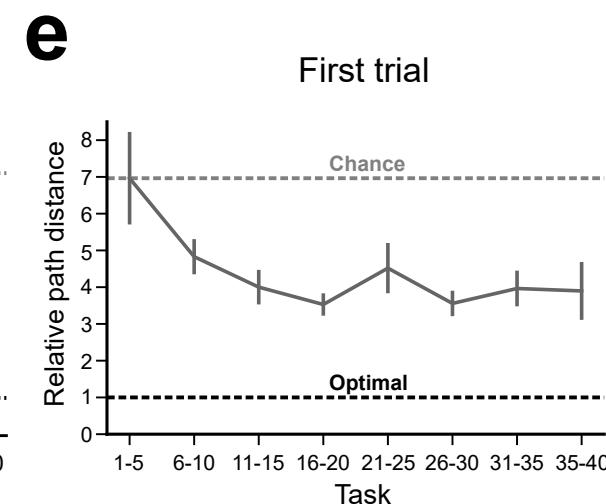
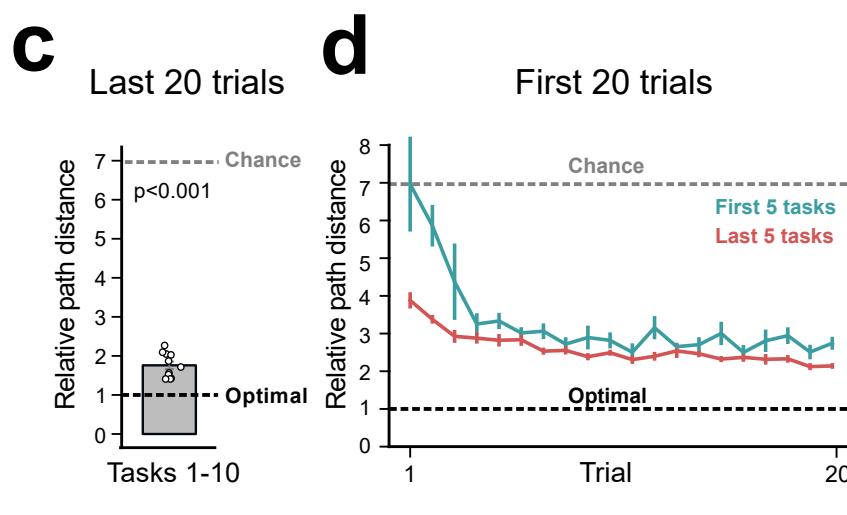
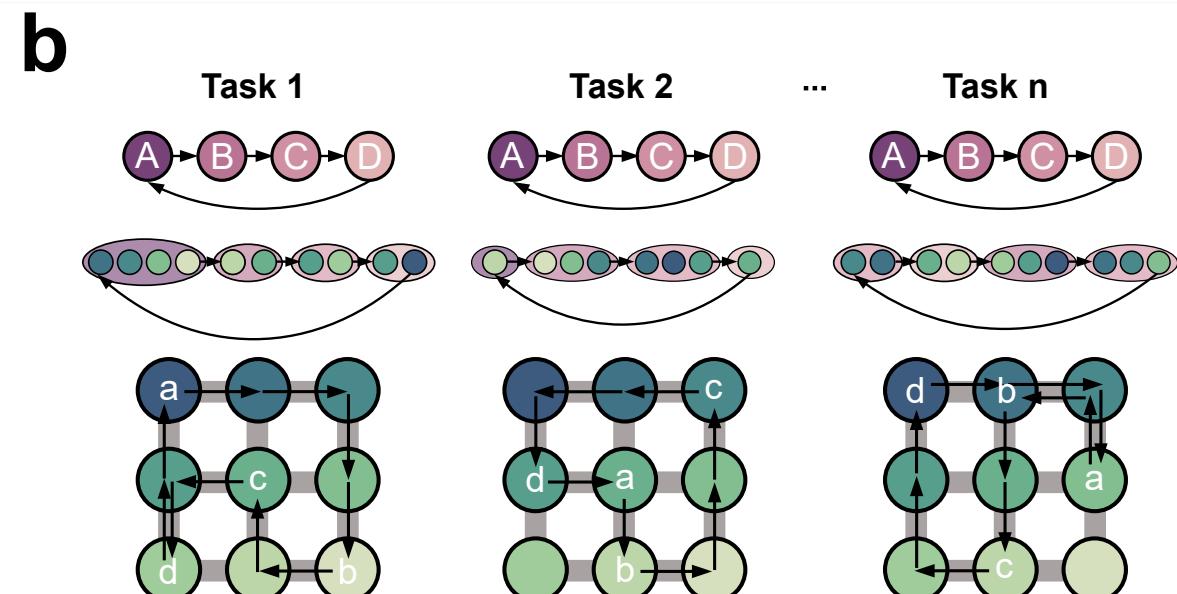
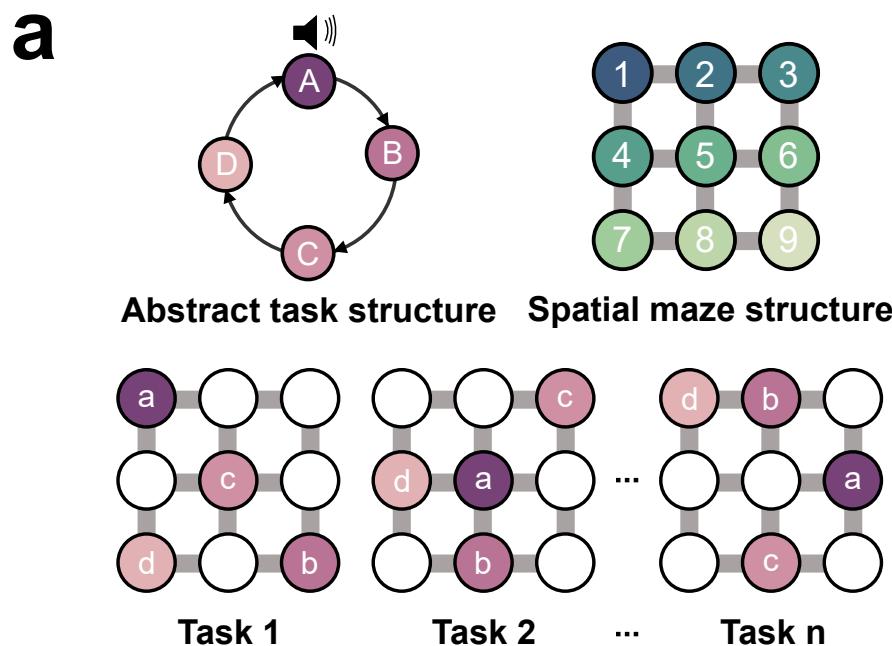
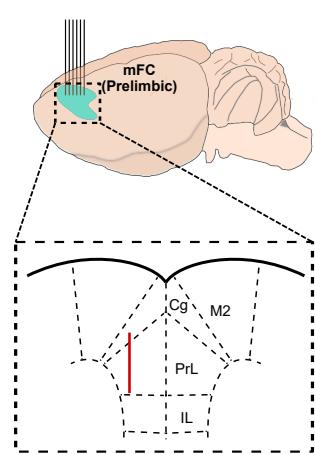
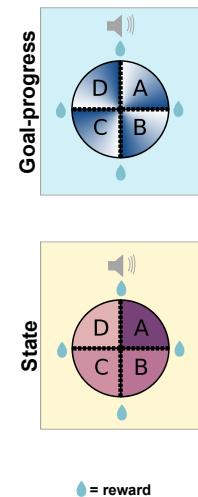
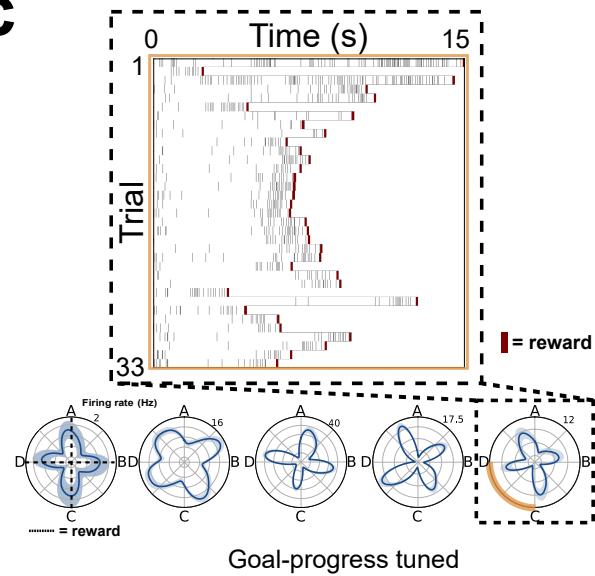
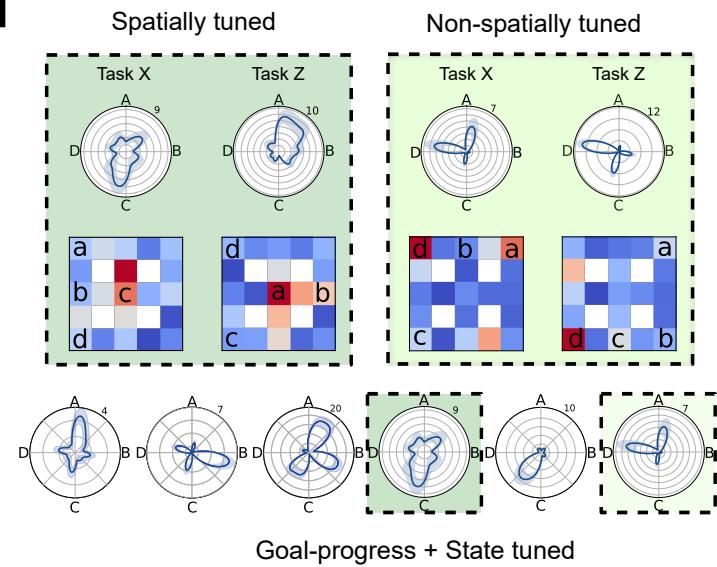
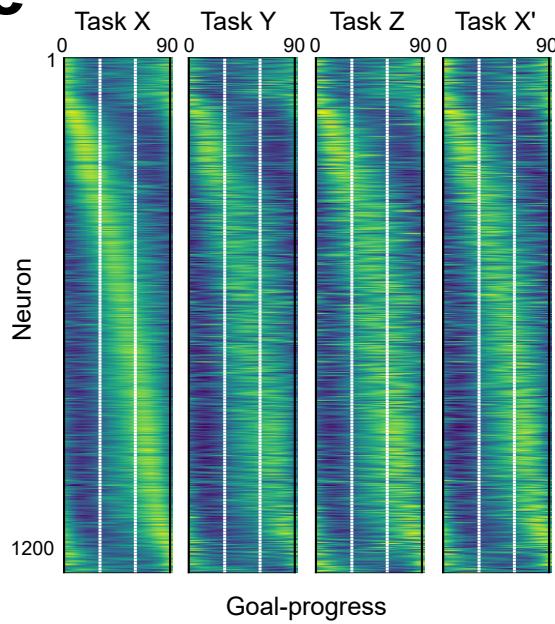
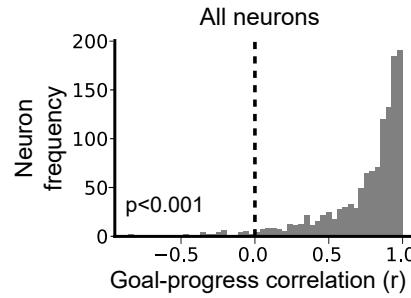
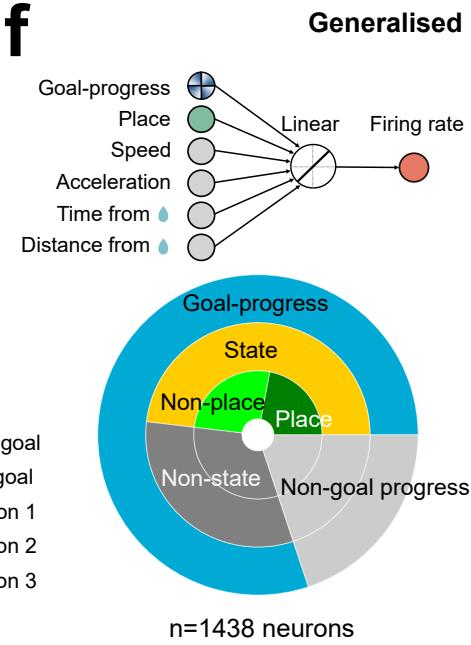
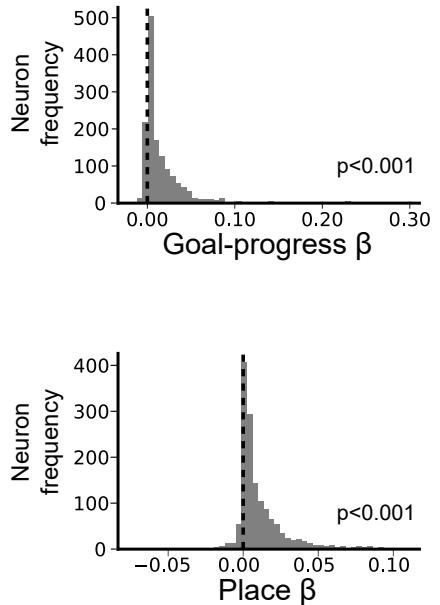
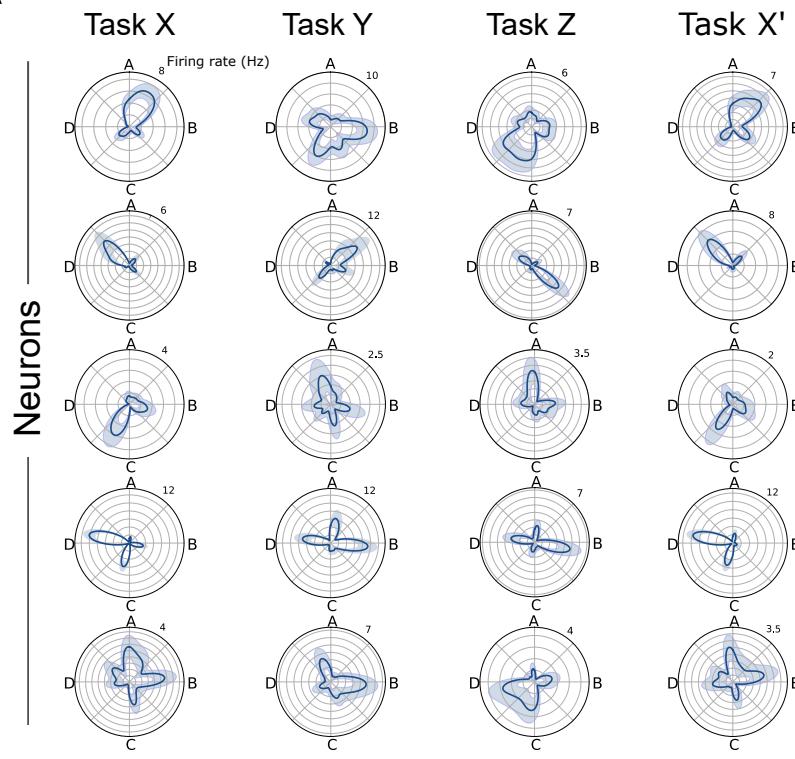
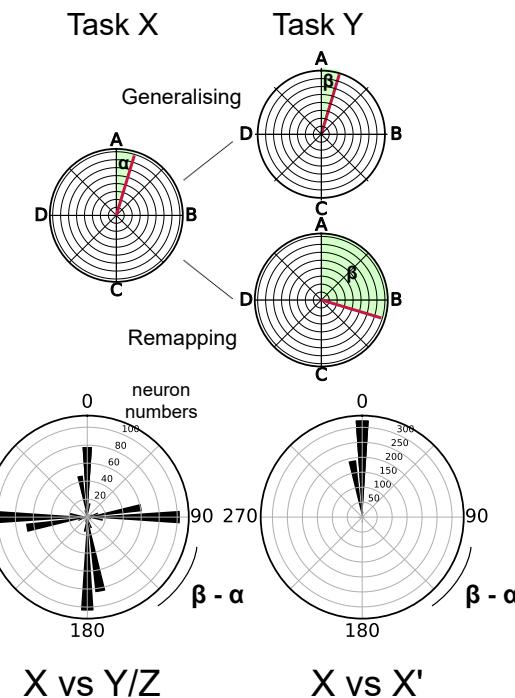
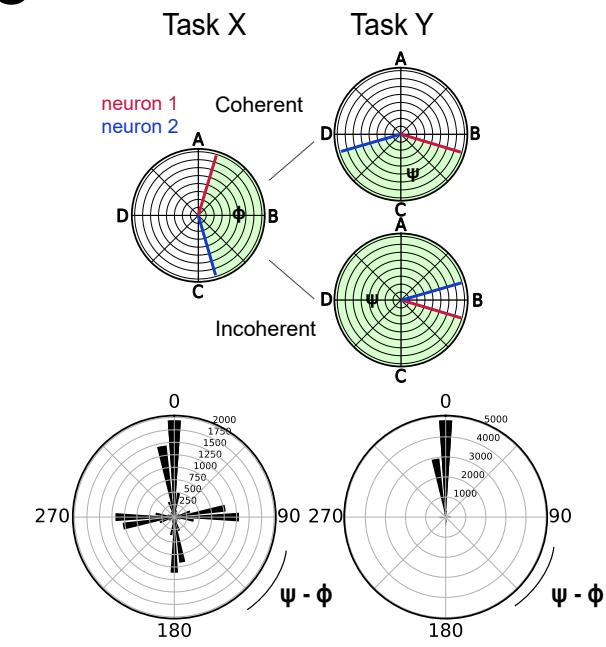


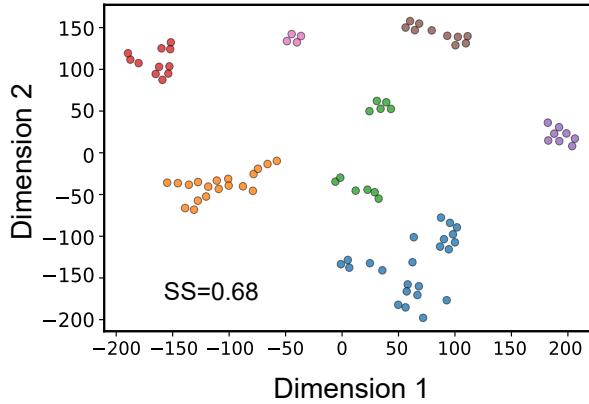
Figure 1

a**b****c****d****e****Goal-progress correlation across tasks****f****Generalised Linear model****Figure 2**

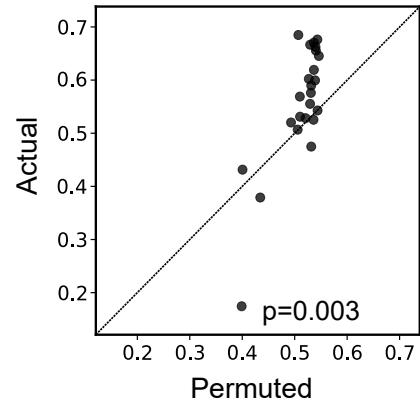
a**b****c****d**

Clustering of coherently remapping neurons

Clustering on example recording day



Summary clustering statistics:
Silhouette score

**e**

Example modules remapping across tasks

Task:

Key

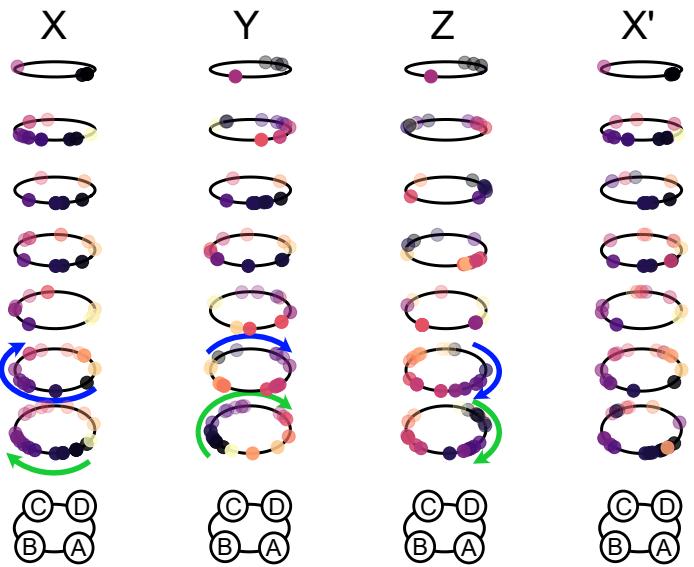
Tuning in task X

A B C D

Neuron

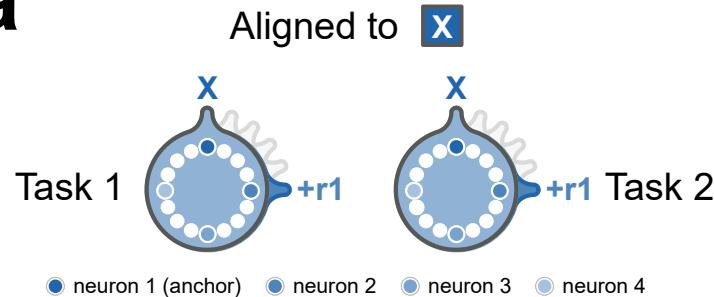
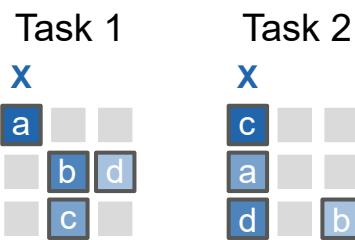
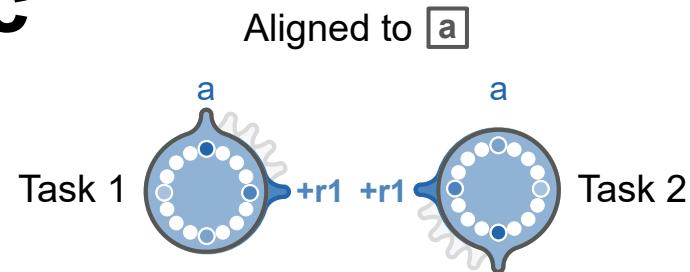
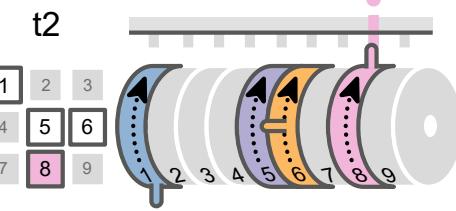
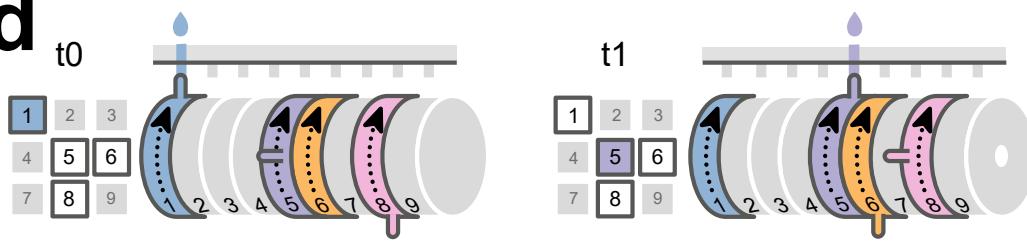
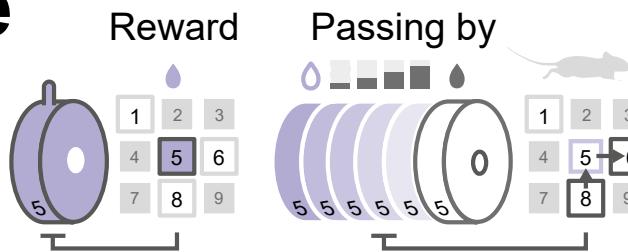
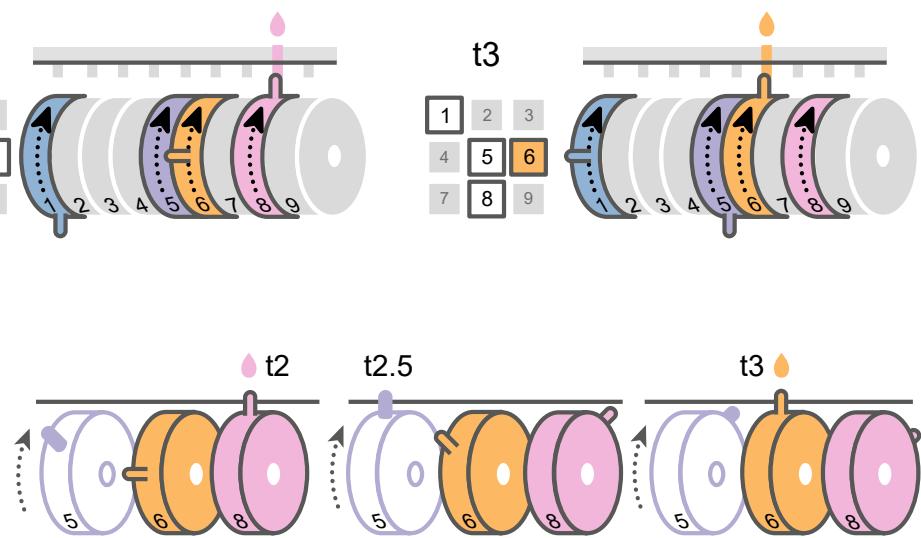
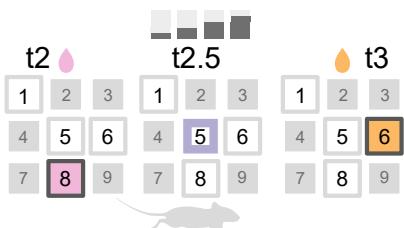
Module

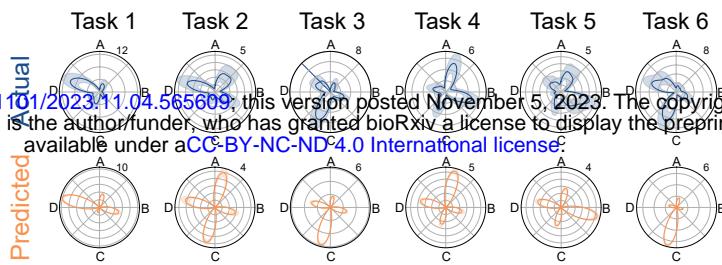
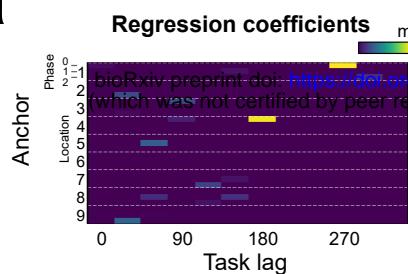
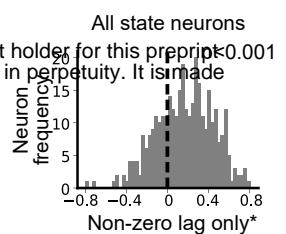
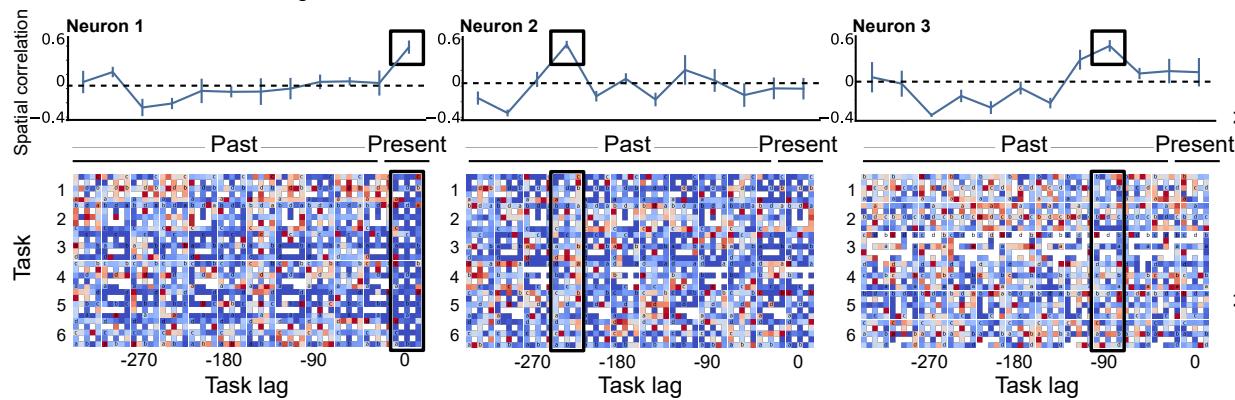
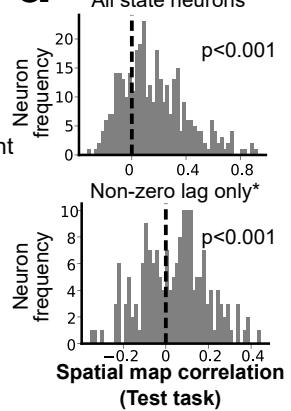
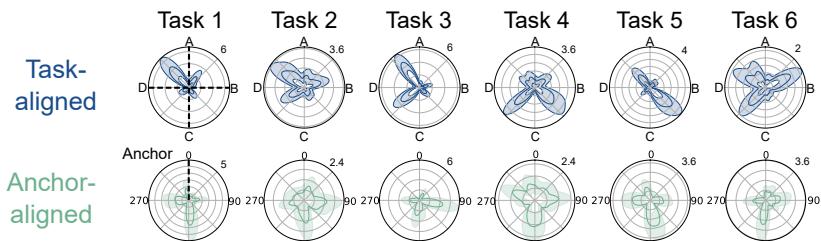
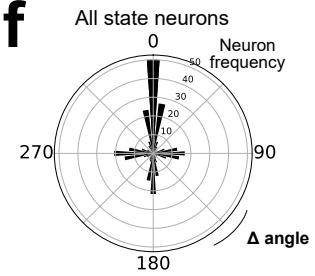
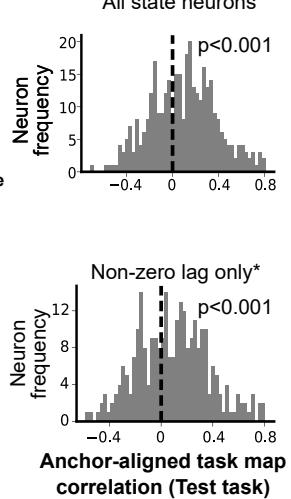
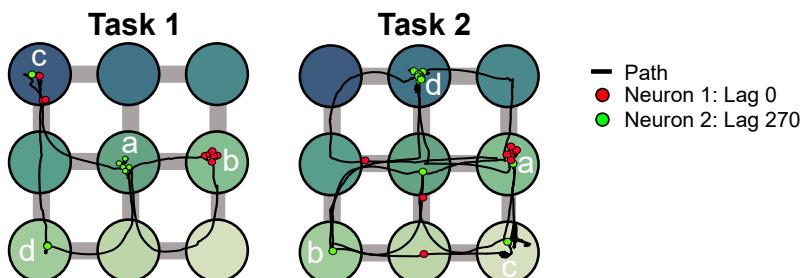
Remapping angle



Tuning in current task

Figure 3

a**b****c****d****e****f****Figure 4**

a**b****c****d****e****f****g****h****Key****Figure 5**

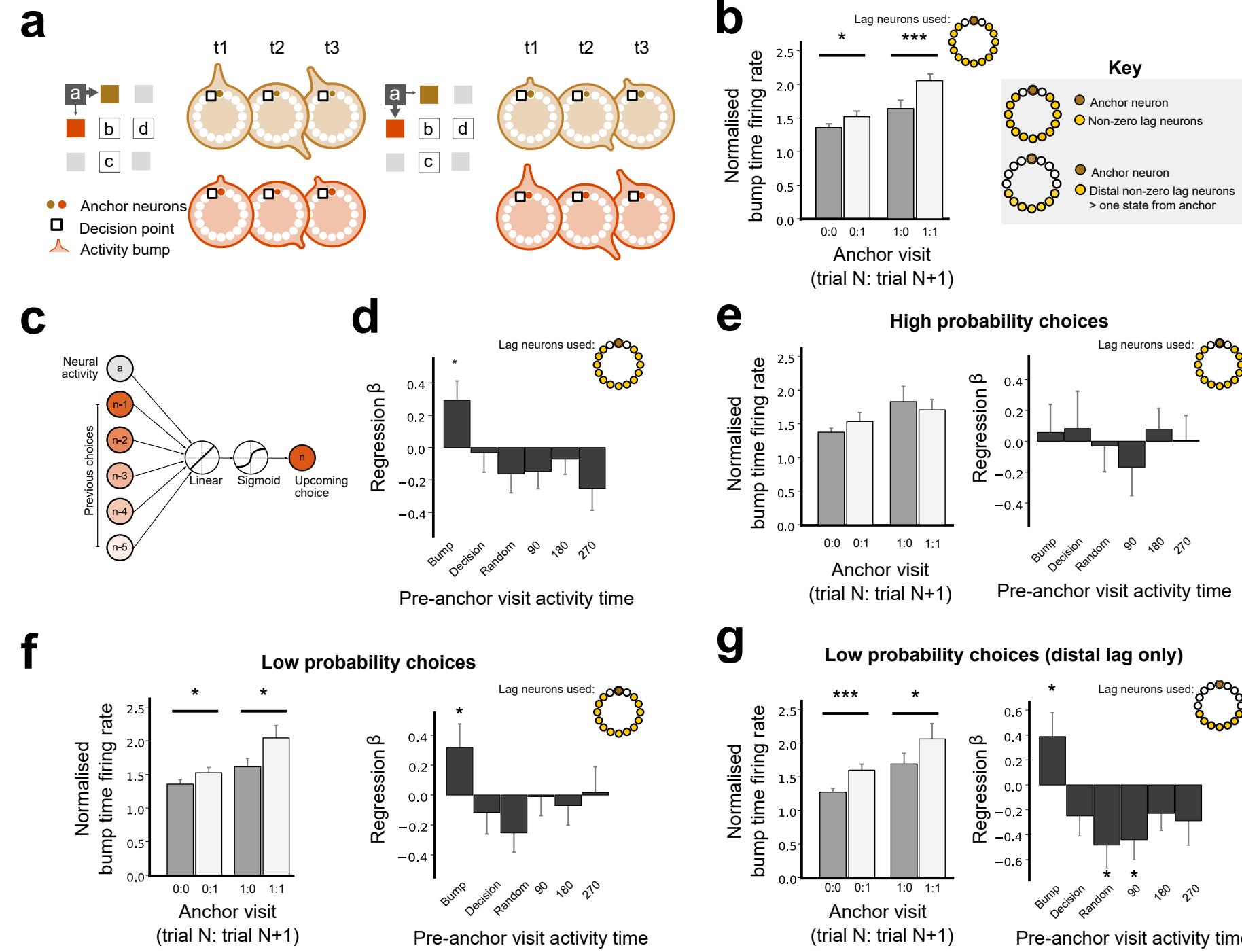
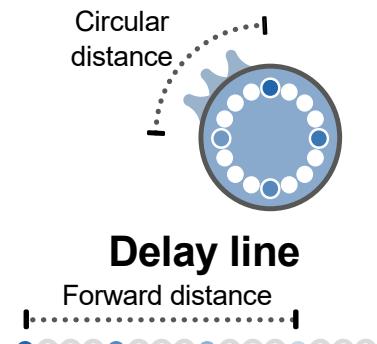
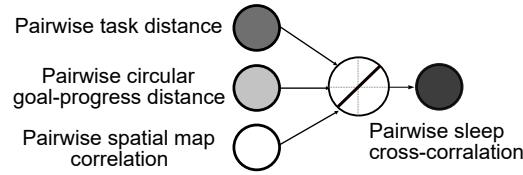
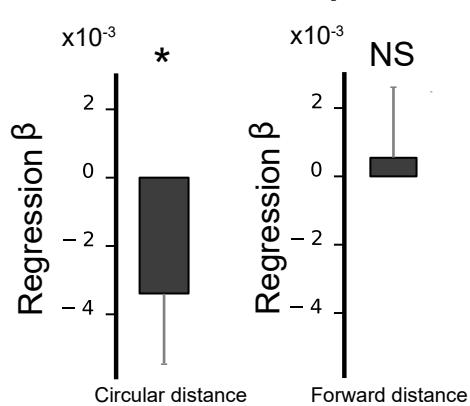
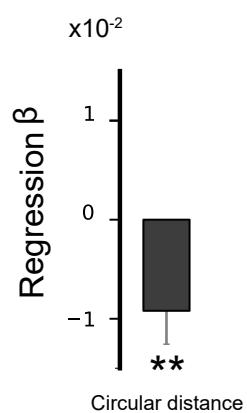
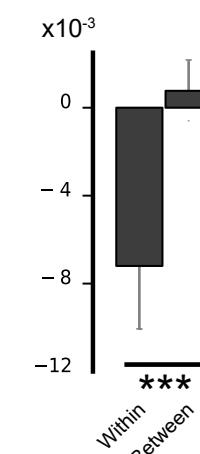
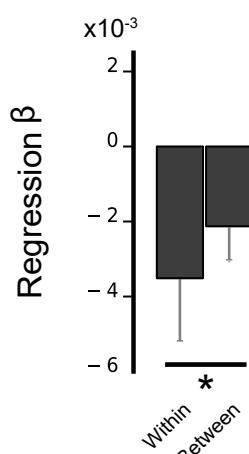
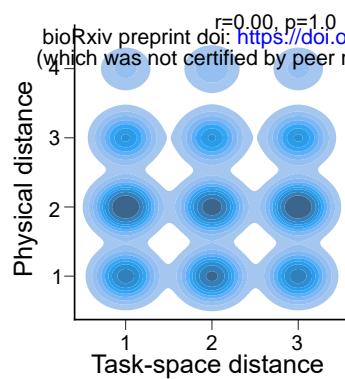
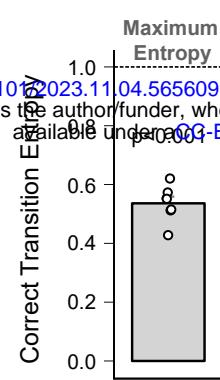
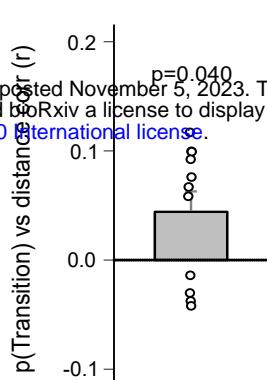
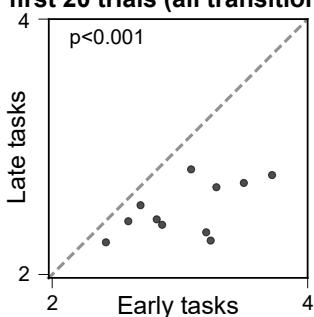


Figure 6

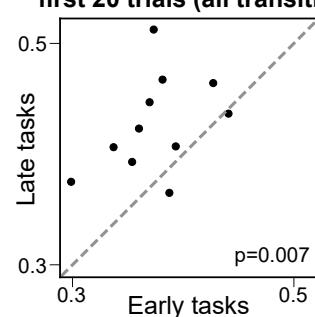
a**Ring structure****GLM****b****All co-anchored pairs****c****Co-anchored pairs
>180° forward distance)****d****Pre-task****Post-task****Offline sleep/rest sessions****Figure 7**

a**b****c****d**

Mean relative path distance in first 20 trials (all transitions)

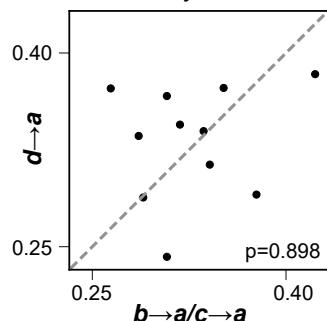
**e**

Proportion shortest paths in first 20 trials (all transitions)

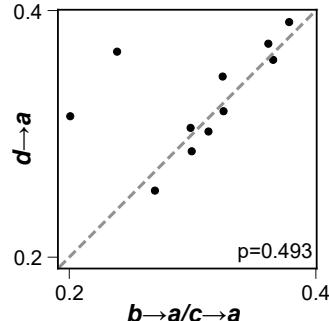
**f**

Pre-task exploration chance levels

Early Tasks

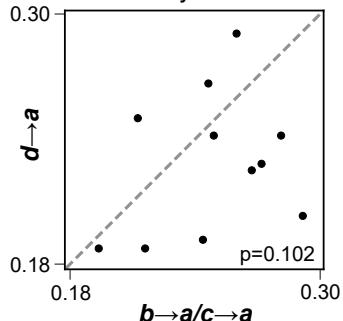


Late Tasks

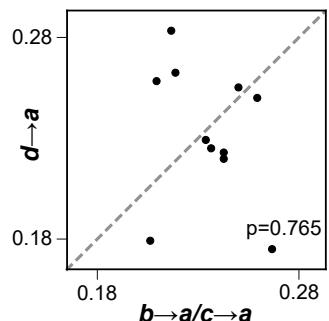
**g**

Analytic chance levels

Early Tasks

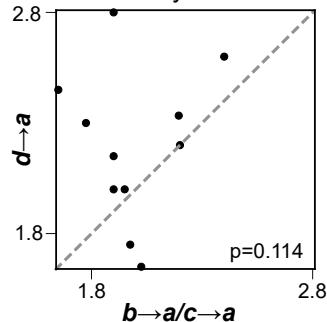


Late Tasks

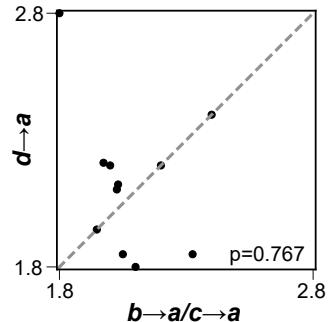
**h**

Mean shortest distances between maze locations

Early Tasks

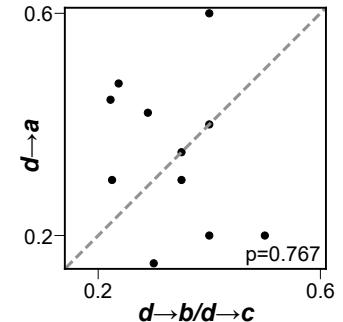


Late Tasks

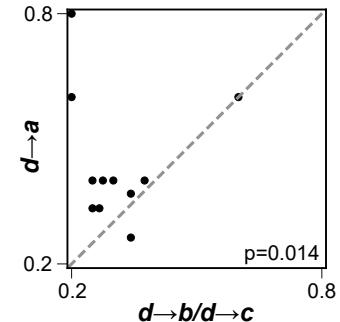
**i**

Proportion shortest on first trial

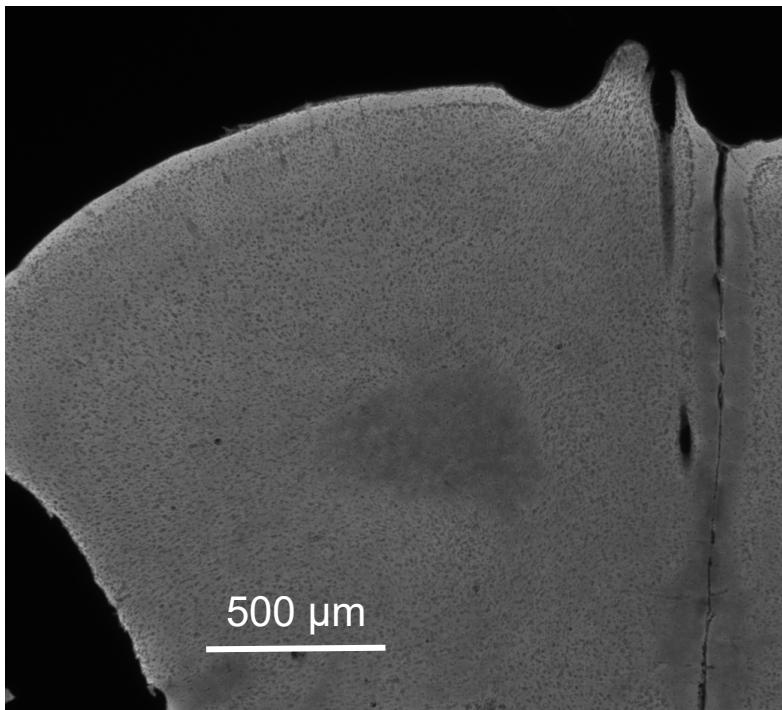
Early Tasks



Late Tasks



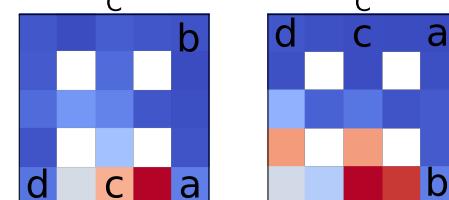
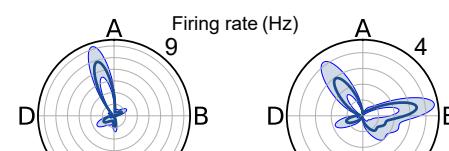
a



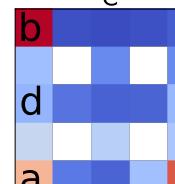
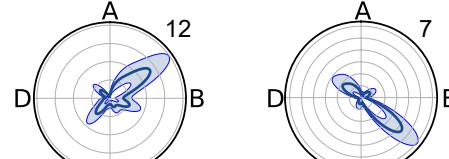
b

Spatial neurons

Task X

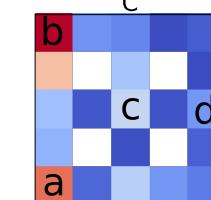
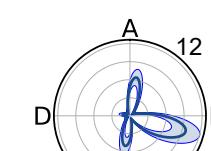


Task Y

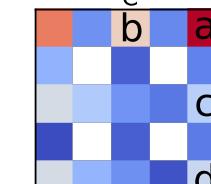
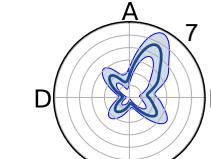


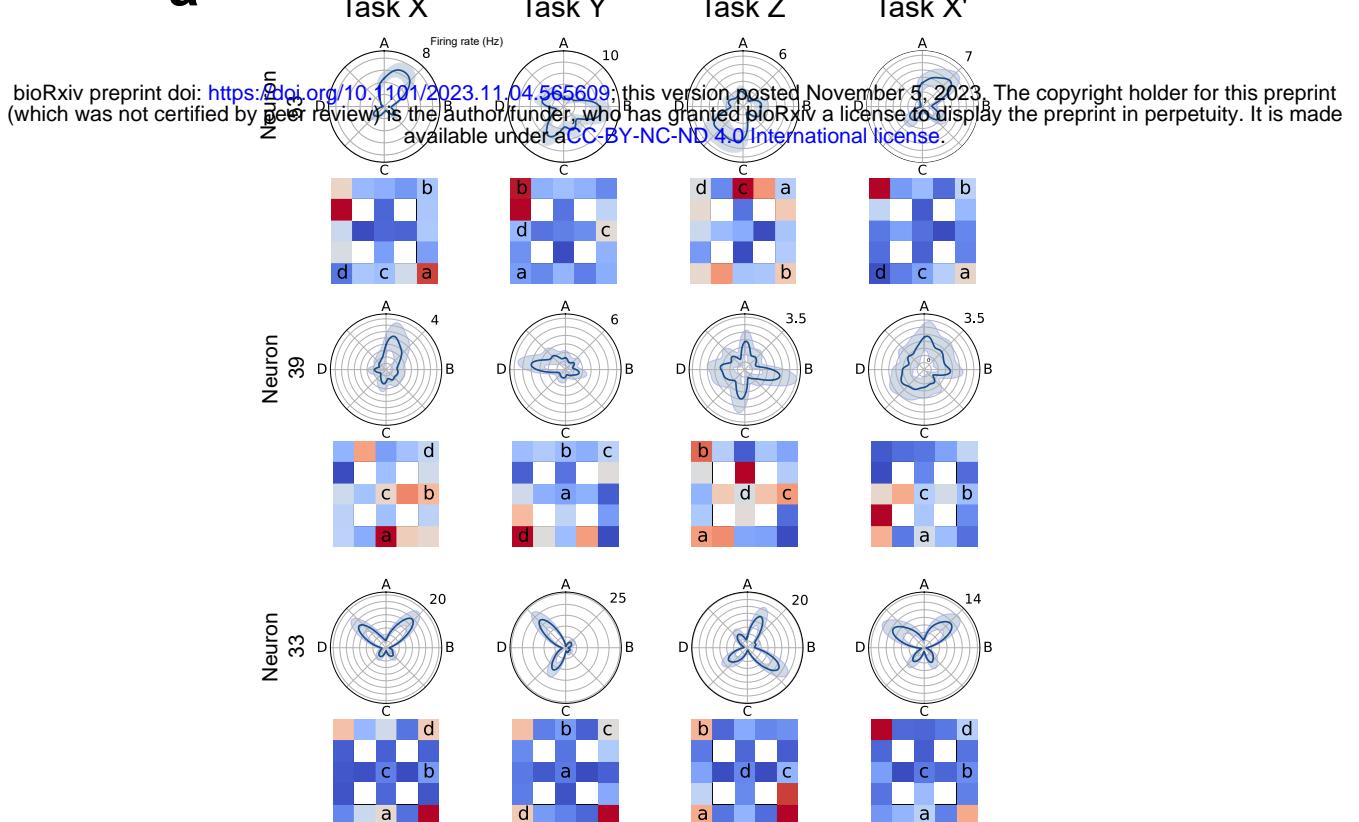
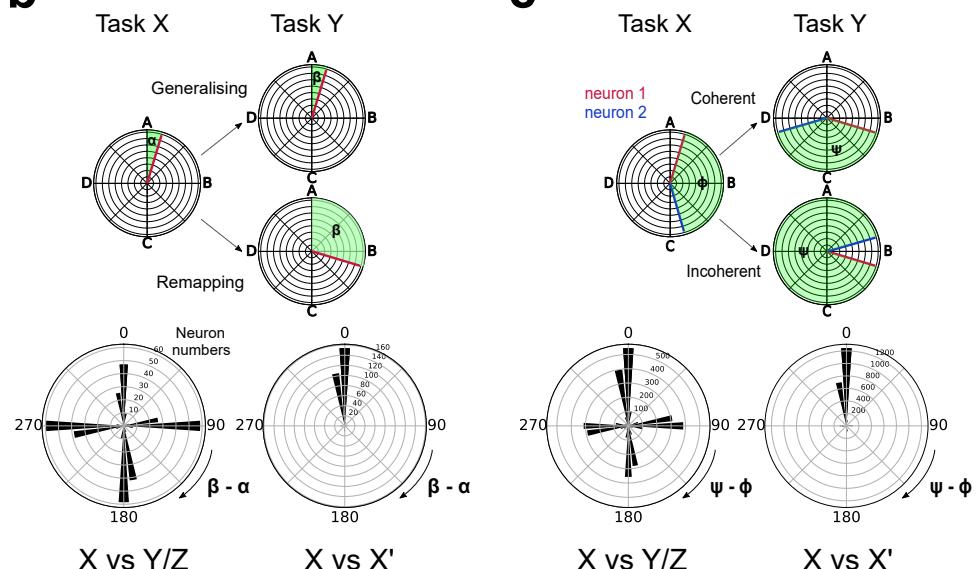
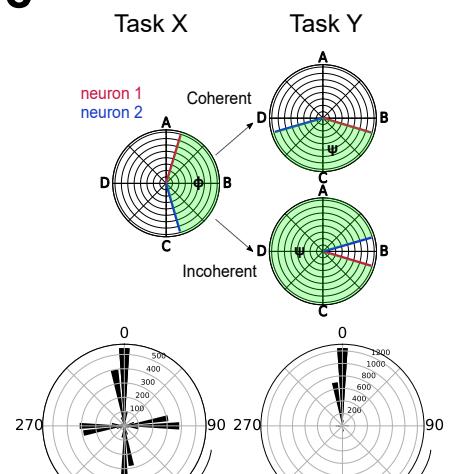
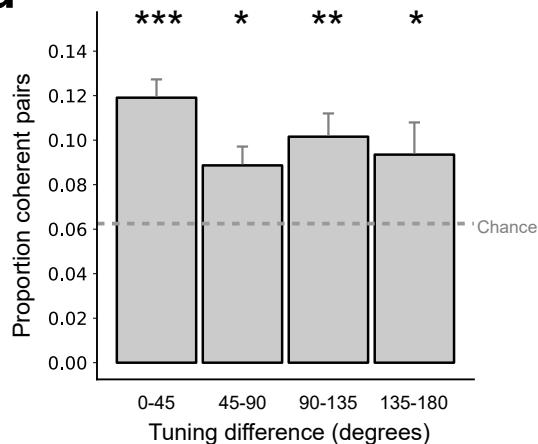
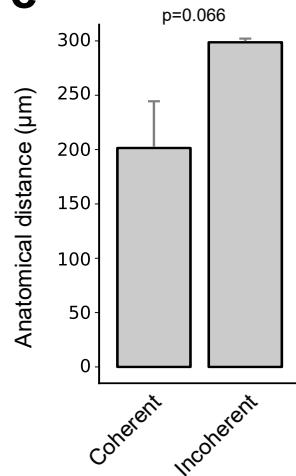
Non-Spatial neurons

Task X

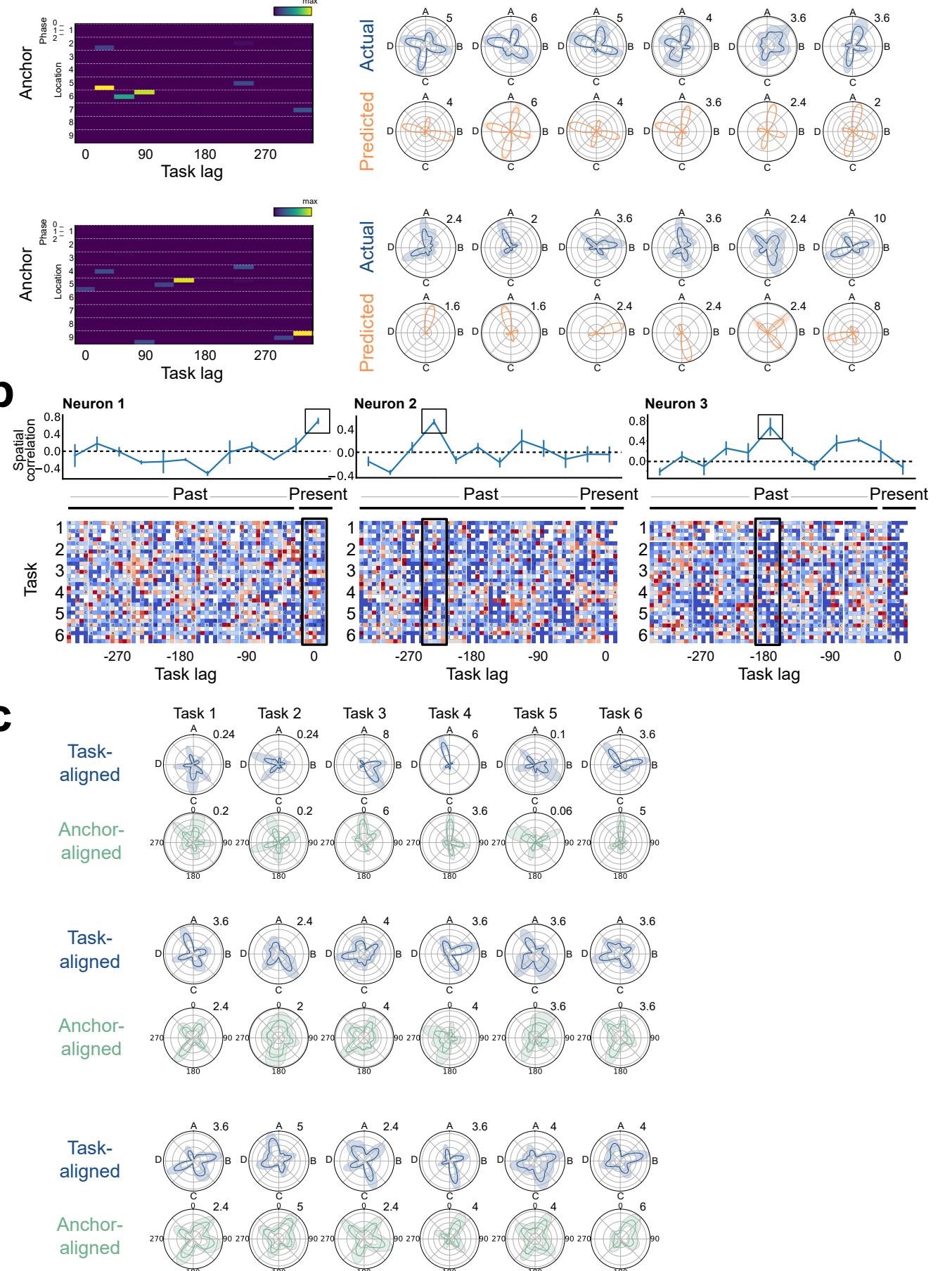
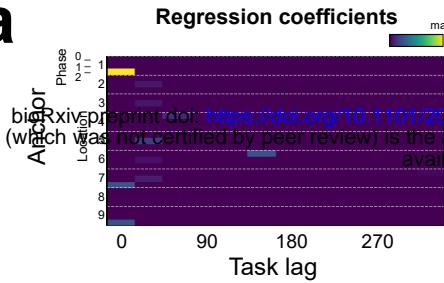


Task X



a**b****c****d****e**

Extended data figure 3

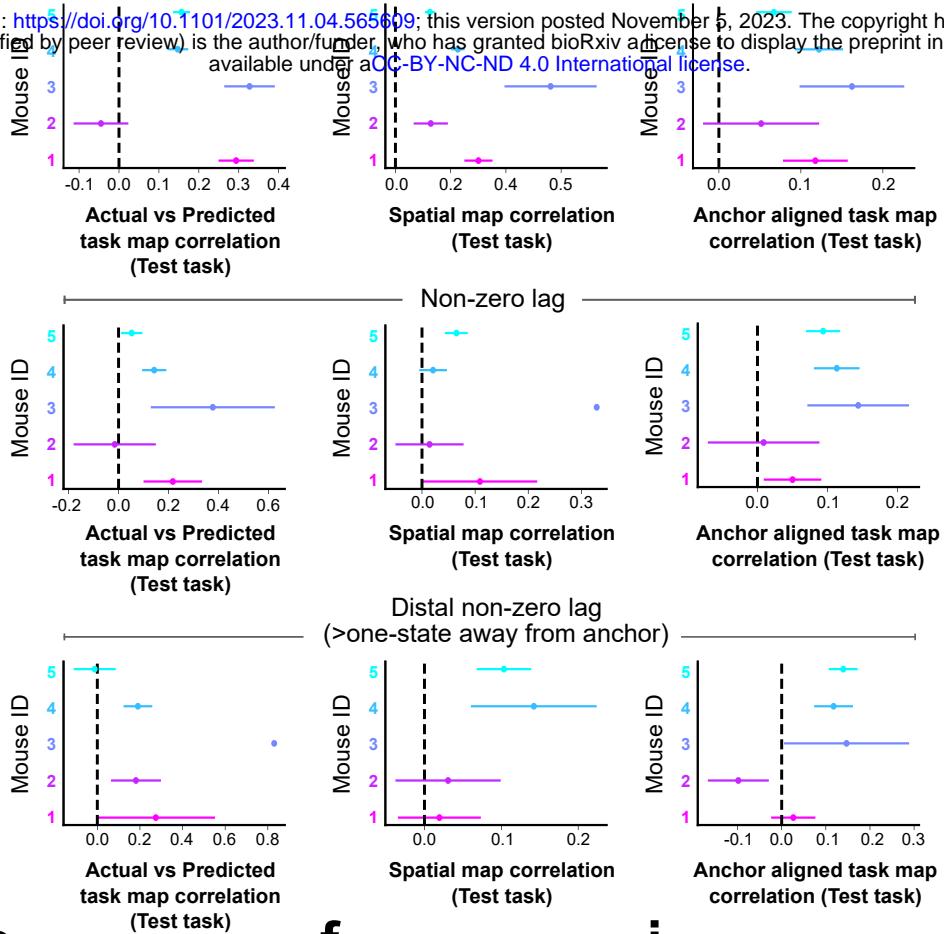
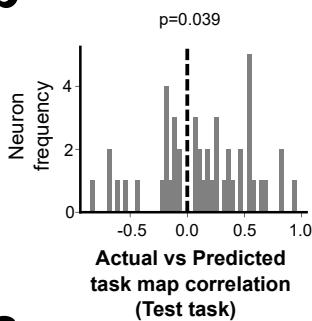
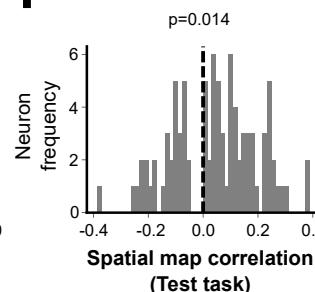
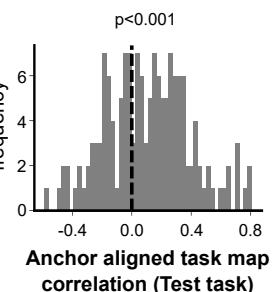
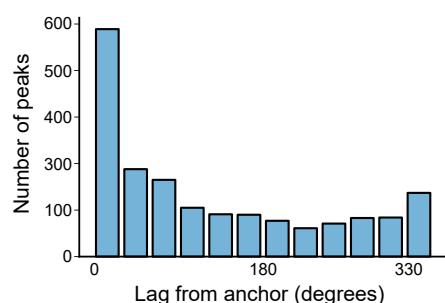
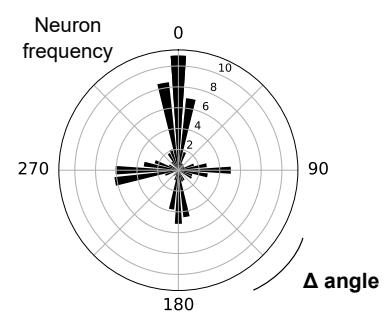
a

a

bioRxiv preprint doi: <https://doi.org/10.1101/2023.11.04.565609>; this version posted November 5, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

e

All state-tuned neurons

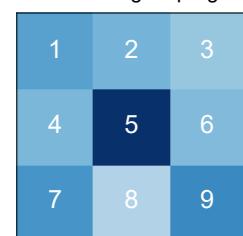
h**b****f****i****c****g****d**

Spatial distribution of anchors per goal-progress

Early goal progress

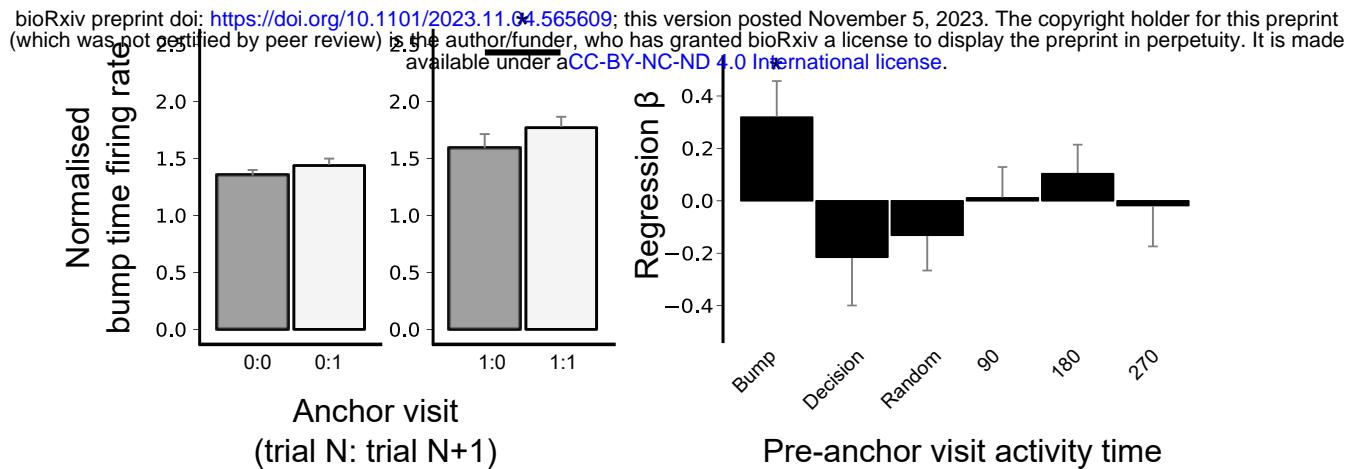
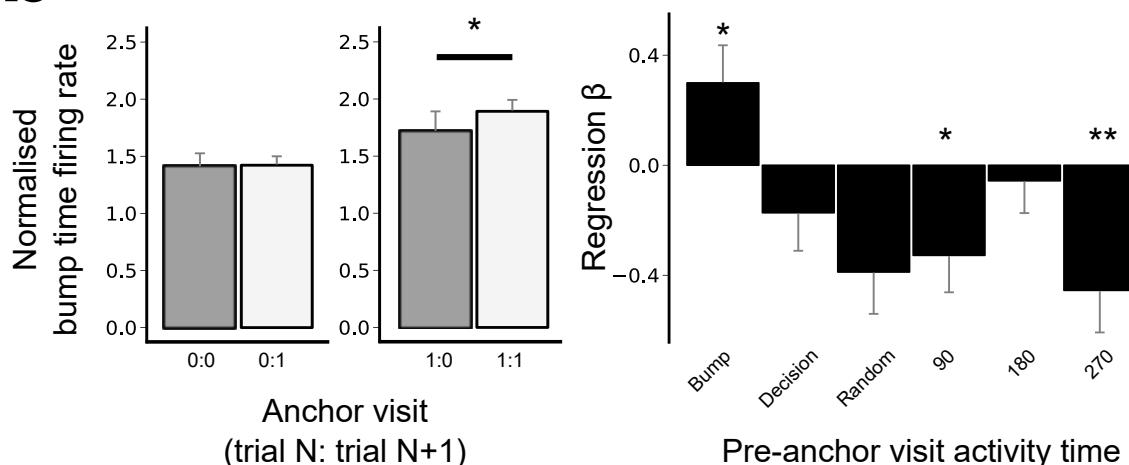
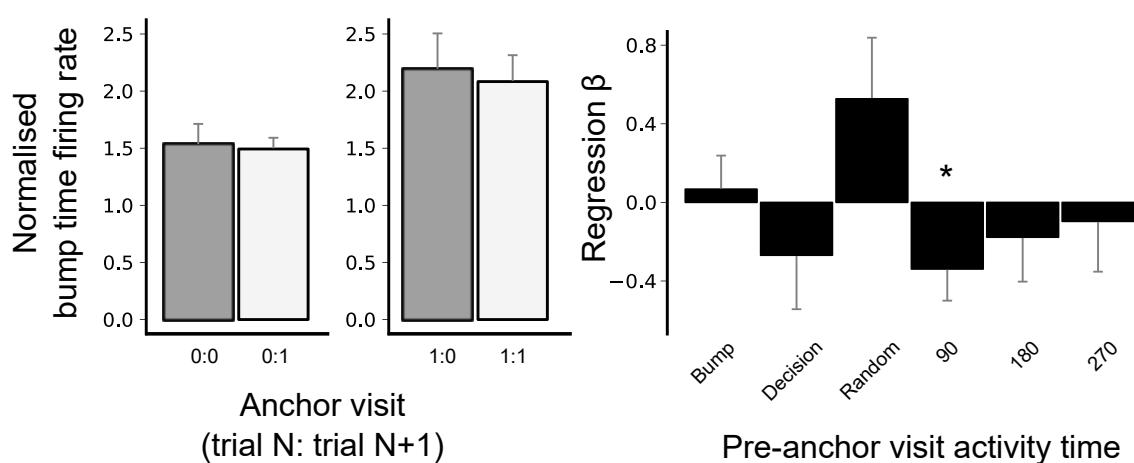


Intermediate goal progress

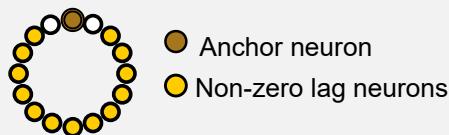


Late goal progress

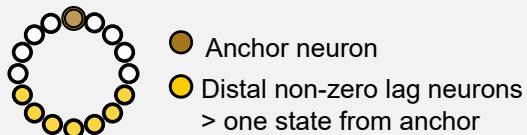


a**All choices to non-rewarded locations *****b****All choices****Distal non-zero lag neurons (>one state away) †****c****High probability choices****Distal non-zero lag neurons (>one state away) †****Key**

*



†

**Extended data figure 6**