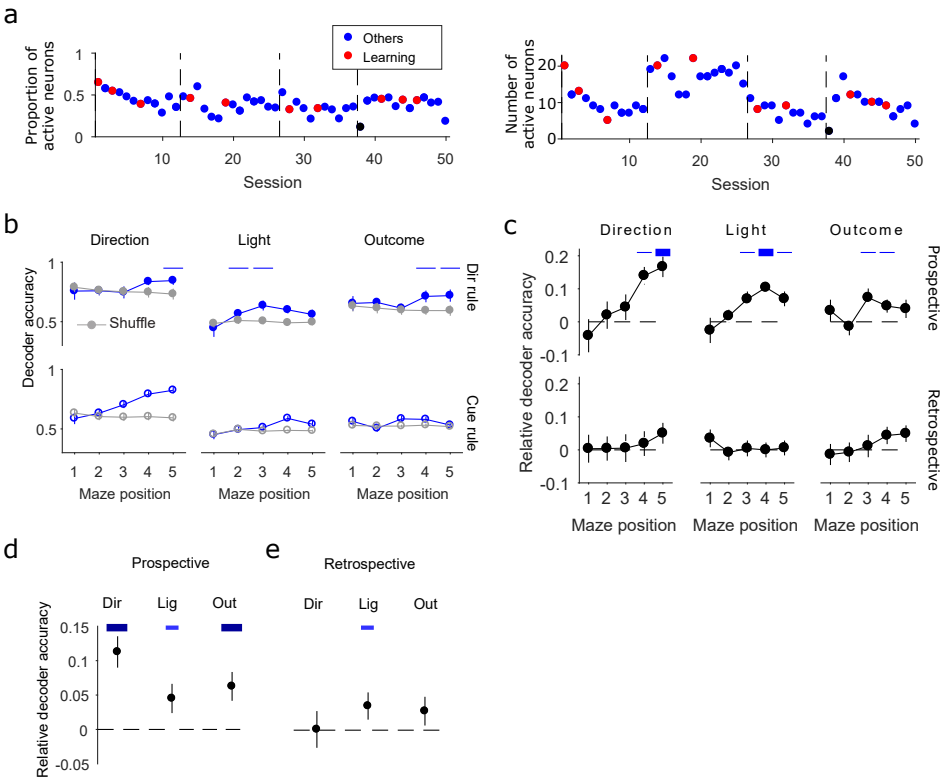
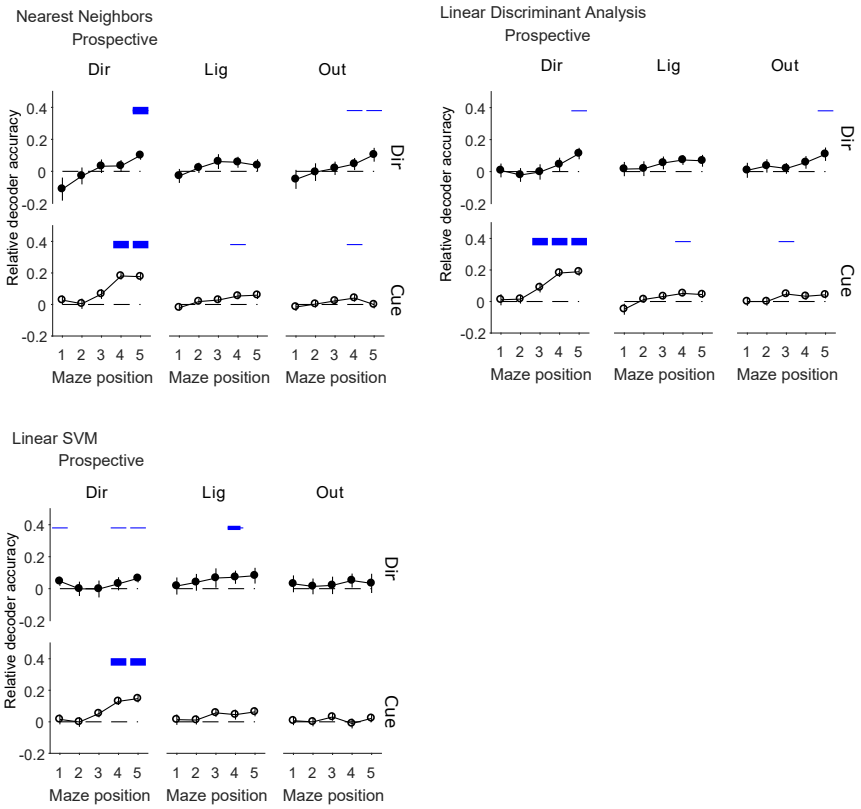
Independent population coding of the past and the present in prefrontal cortex during learning

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

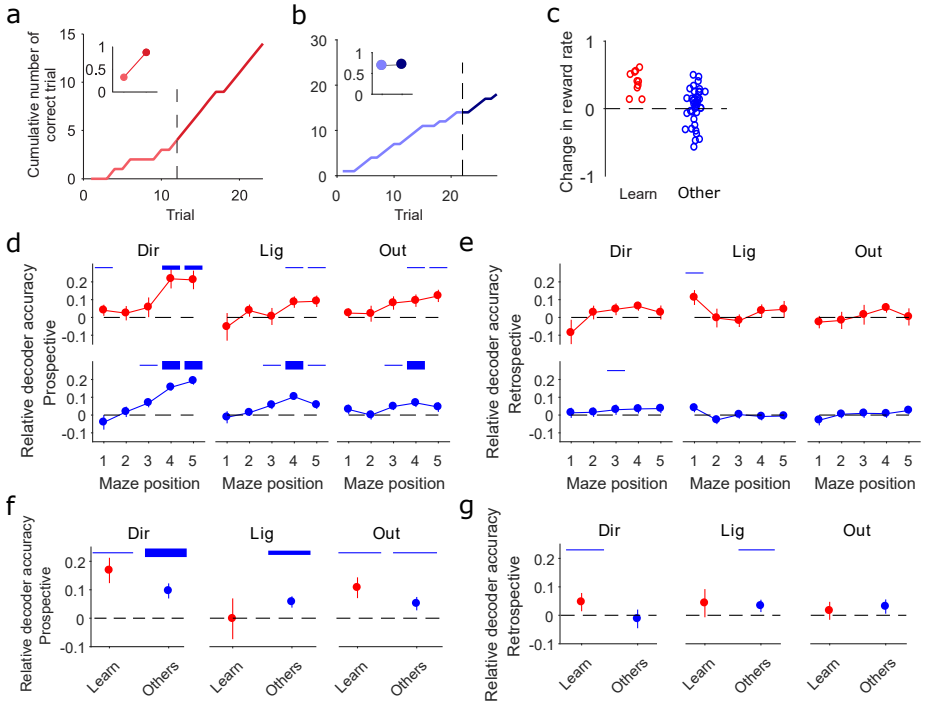
Supplementary Information



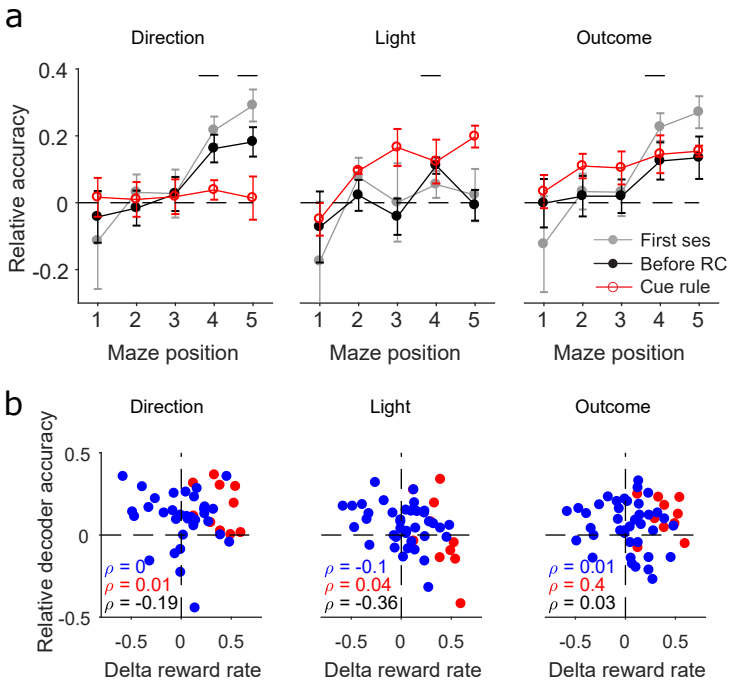
Supplementary Figure 1. Encoding of task features during trials.
(**a**) Size of retained neural populations. Left: the proportion of neurons active in every trial of a session (dots). No differences between learning and other non-learning sessions was observed here. Right: the actual number of neurons contributing to the retained population varied between 2 and 22 (right panel; mean ± SEM, 11.3 ± 0.7). All sessions with 4 or more neurons were included in our analyses, so excluding just one session (black dot), giving 49 in total. Sessions are plotted in time order; the vertical dashed lines separate sessions by animal.
(**b**) Decoder accuracy for logistic regression classifiers tested on Other sessions (blue lines) and compared to the accuracy of a shuffled control model (grey). Filled symbols refer to direction rule sessions (upper panels), while empty symbols to cue rule sessions (bottom panels). Due to the unbalanced distribution of choices and cue stimuli, the accuracy of shuffled model was sometimes above 0.5. Here and in panels (c-e): symbols are mean ± SEM; blue bars indicate decoding performance significantly better than chance (Wilcoxson sign rank test, $p < 0.05$ thin light blue line; $p < 0.01$ medium thickness blue; line $p < 0.005$ thick dark blue line.).
(**c**) Populations encode information in the absence of neural tuning. Relative decoder accuracy of present task-features (upper panels) and previous trial task-features (bottom panels) along the maze, only including sessions that lack feature-selective neurons (e.g. neurons that significantly change firing rates between right and left choice direction).
(**d**) Relative decoder accuracy of features of the current trial for population firing rate vectors constructed over the entire maze.
(**e**) Similar to panel (d), for decoding features of the previous trial.



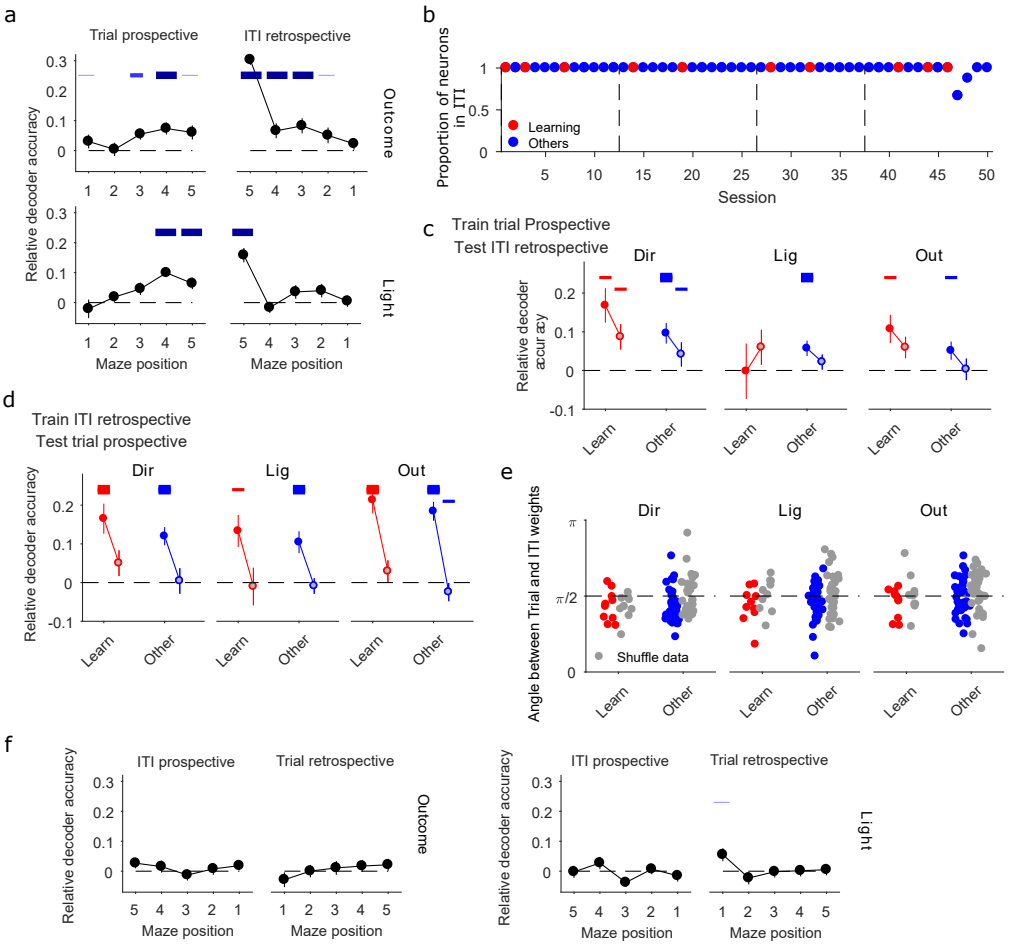
Supplementary Figure 2. Population vector encoding of relevant task features is robust across multiple methods. Three other linear classification methods have been used to test the prospective encoding of task features. Here respectively Nearest Neighbors (top left), Linear Discriminant Analysis (top right) and Linear SVM (bottom left). Blue bars on top of the panels decoding performance significantly better than chance (Wilcoxon sign rank test, $p < 0.05$ thin blue line; $p < 0.01$ medium thickness blue line; $p < 0.005$ thick blue line).



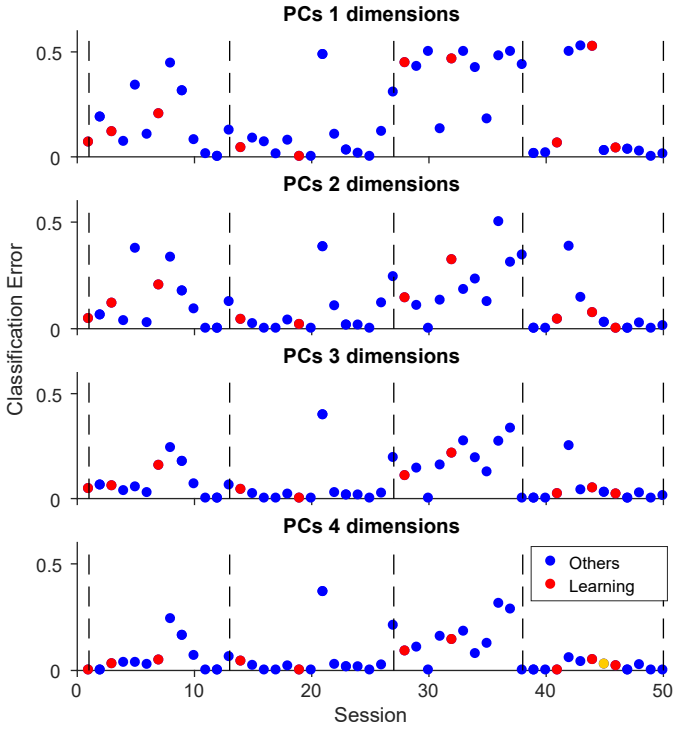
Supplementary Figure 3. Population encoding is consistent between learning and non-learning sessions
(**a**) Example learning curve from a learning session, plotting the cumulative number of correct trials. Black dashed line identifies the learning trial as the first of three consecutive correct trials followed by at least 80% correct trials. Inset: reward rates before (light red) and after (dark red) the learning trial. Reward rates were given by the slope of linear regressions fitted to the learning curve before and after the learning trial.
(**b**) Example learning curve from an "Other" session. The black dashed line identifies the trial of maximum change in reward accumulation between before and after. Inset: the reward rate before the identified trial (light blue) and after (dark blue).
(**c**) Change in reward rate during learning sessions (red) or "Other" sessions (blue). Each symbol is a session.
(**d**) Decoding accuracy for features of the present trial in learning and Other sessions, as a function of maze position. Here and in panels (e-g): symbols give means ± SEM; blue bars indicate decoding performance significantly better than chance (Wilcoxon sign rank test, $p < 0.05$ thin blue line; $p < 0.01$ medium thickness blue; line $p < 0.005$ thick blue line).
(**e**) As for (d), for decoding features of the preceding trial.
(**f**) Decoding accuracy for features of the present trial in learning and Other sessions, using firing vectors from the whole maze.
(**g**) As for (f), for decoding features of the preceding trial.



Supplementary Figure 4. Interaction between learning and encoding
(**a**) Changes in decoding accuracy over sessions. We plot the relative decoder accuracy for the first session of each animal (grey, $N = 4$), for all the sessions before the first rule change (black, $N = 9$ direction rule sessions), and for the sessions in which the cue rule was learned (red, $N = 3$). All values are mean ± SEM. Black bars indicate significant departure from shuffled control for the all sessions before the rule change (Wilcoxon sign rank test, $p < 0.05$). Note that there are too few first sessions and cue-learning sessions to meaningfully interpret an hypothesis test like the Wilcoxon; we note though that the error bars giving an approximate 68% confidence interval are narrow and far from zero at key maze positions.
(**b**) Relative decoder accuracy (on the whole maze) as a function of the change in reward rate during a session. Red: learning sessions; blue: Other sessions. Spearman's correlation coefficient ρ is given for: black, all sessions ($N = 49$); red, learning ($N = 10$); blue: Other ($N = 39$).



Supplementary Figure 5. Independent encoding of the past and present.
(**a**) Similar to Figure 2a, relative decoder accuracy of present outcome and light position during trials followed by the decoding of the same task features in the past during inter-trial interval. Here and in panels (c-d) and (g): symbols give means ± SEM; bars on top of the panels indicate decoding performance significantly better than chance (Wilcoxon sign rank test, $p < 0.05$ thin light blue line; $p < 0.01$ medium thickness blue line; $p < 0.005$ thick dark blue line.)
(**b**) The proportion of neurons active in all trials that are also active in all inter-trial intervals, and so define a common population. Vertical dashed lines separate sessions across the four rats.
(**c**) Decoder accuracy for leave-one-out cross-decoding between trials and inter-trial intervals (ITIs). The first symbol of each pair shows the decoder accuracy of training and testing on trials, using leave-one-out cross-validation. The second of each pair (grey filled circles) shows the accuracy of the classifier trained on $n-1$ trials and predicting the inter-trial interval corresponding to the excluded trial.
(**d**) As for panel (c), for leave-one-out cross-decoding between inter-trial intervals and trials. First symbol of each pair gives within-inter-trial interval decoding accuracy; the second symbol (grey filled circle) gives the accuracy of a classifier trained on $n-1$ inter-trial intervals and tested on the trial corresponding to the excluded inter-trial interval.
(**e**) Breakdown of angles between decoding vector weights by type of session (results in Figure 2d). For each session we plot the angle between its trial and inter-trial interval decoding weight vectors. Red, learning sessions; blue, Other sessions; grey, shuffled label data.
(**f**) Similar to Figure 2e, the relative accuracy of decoding outcome and light position in the next trial during inter-trial interval (left), followed by the decoding of the preceding trial's outcome and light position during the trial after the inter-trial interval (right).



Supplementary Figure 6. Trial and inter-trial interval population activity is easily separable.
Error in separating each session's trial and inter-trial interval population vectors, when projected onto increasing numbers of dimensions. Each symbol is the classification error for a session, divided into the four rats by the vertical dashed lines. Chance is 0.5. Almost all sessions have near-perfect separation of trial and inter-trial interval population activity even when projected into just three dimensions. Rat 3's population vectors are consistently less separable.