Rat anterior cingulate neurons responsive to rule or strategy changes are modulated by the hippocampal theta rhythm and sharp-wave ripples

Running title: Hippocampus modulates rule- and strategy-shift responsive PFC neurons

Authors

M Khamassi[1,2], A Peyrache[1], K Benchenane[1], DA Hopkins[3], N Lebas[1], V Douchamps[1], J Droulez[1,2], FP Battaglia[1, 4], and SI Wiener[1,*]

[1]Center for Interdisciplinary Research in Biology (CIRB), College de France, CNRS, INSERM, PSL Research University, 75005 Paris, France.

[2]Sorbonne Université, CNRS, Institute of Intelligent Systems and Robotics, 75005 Paris, France

[3]Department of Medical Neuroscience, Dalhousie University, Halifax B3H 4R2, Canada

[4]Donders Institute for Brain, Cognition, and Behaviour, Radboud Universiteit Nijmegen, 6525 AJ, Nijmegen, The Netherlands

* Correspondence: Dr. Sidney Wiener, Center for Interdisciplinary Research in Biology (CIRB), College de France, 11 pl. Marcelin Berthelot, 75005 Paris, France

sidney.wiener@college-de-france.fr

64 pages

7 figures, no tables, 2 supplementary figures

Abstract

To better understand neural processing during adaptive learning of stimulus-response-reward contingencies, we recorded synchrony of neuronal activity in anterior cingulate cortex (ACC) with hippocampal rhythms in male rats acquiring and switching between spatial and visual discrimination tasks in a Y-maze. ACC population and single unit activity responded shortly after task rule changes, or just before the rats adopted different task strategies. Hippocampal theta oscillations (associated with memory encoding) modulated an elevated proportion of rule-change responsive neurons (70%), but other neurons that were correlated with strategy-change, strategy value, and reward-rate were not. However, hippocampal sharp wave-ripples modulated significantly higher proportions of rule-change, strategy-change and reward-rate responsive cells during post-session sleep but not pre-session sleep. This suggests an underestimated mechanism for hippocampal mismatch and contextual signals to facilitate ACC detection of contingency changes for cognitive flexibility, a function that is attenuated after it is damaged.

Introduction

Adopting an adaptive behavioral response policy, or strategy, in a novel situation requires observation, deduction and memorization of the rules that govern success as well as failure. Optimizing behavior in a changing world involves continuous selection of appropriate strategies, as well as inhibition of less successful ones. Furthermore, alternative, possibly more beneficial, behavioral strategies should be developed and tested. This all requires tracking outcomes over successive iterations, an application of working memory (Malenka, et al., 2009; Yu and Frank, 2015). This would benefit from neural coding for the presence or absence of rewards, which is found in anterior cingulate cortex (ACC) and upstream

structures like the hippocampus (HPC; e.g., Tabuchi et al., 2003; Hyman et al., 2011). Furthermore, the HPC is strongly implicated in spatial and contextual coding, but little is known about how hippocampal signal processing modulates neural representations of behavioral flexibility in ACC. Note that the nomenclature here follows van Heukelum, et al (2020): ACC for medial prefrontal cortex (mPFC), A32 for prelimbic area (PL), A25 for infralimbic area (IL), and A24 for anterior cingulate cortices Cg1 and Cg2, also referred to as dorsal anterior cingulate cortex (ACd). In addition to working memory, ACC and closely associated (Arikuni, et al., 1994) prefrontal cortex, (Goldman-Rakic, 1987, 1995; Lenartowicz and McIntosh, 2005; D'Esposito et al., 1995), mediate other cognitive processes underlying rule-based strategy selection: attention (Shallice, 1988), flexible adaptation of response-selection mechanisms in the presence of novel, fluctuating or ambiguous contexts (Dias et al., 1996; Robbins, 2007), and consolidation of memory (Frankland and Bontempi, 2005). Indeed, patients with prefrontal cortex damage are impaired in non-cued rule-shifting tasks, showing persistence errors (Milner, 1963; Drewe, 1974). Neurophysiological recordings, as well as neuropsychological studies, indicate a crucial role for the rodent ACC and its primate homolog in encoding task-relevant information such as reward (Pratt and Mizumori, 2001), reward rate (Genovesio, et al., 2005), and action-outcome contingencies (Killcross and Coutureau, 2003; Del Arco, et al., 2017). Further, ACC and closely associated areas can represent the current task rule (Mansouri et al., 2006; Rich and Shapiro, 2009; Bissonnette et al., 2013; Chiang et al., 2021, but see Malagon-Vina, et al., 2018) as well as the behavioral strategy adapted to the current task rule (Genovesio et al., 2005; Durstewitz et al., 2010). Moreover, in tasks involving purely spatial strategies, PFC activity has been reported to shift after instatement of a new rule, but before the animal' s volitional change in strategy (Karlsson, et al., 2012; Powell & Redish, 2016). However, it is unknown whether

this result generalizes to cue-guided strategies, whether distinct PFC populations encode distinct types of shifts, and how these behavioral flexibility-related prefrontal activities interact with the HPC during behavior and subsequent sleep.

It is not yet understood how executive and memory functions are linked and coordinated in the HPC-ACC axis. Specifically, how do populations of neurons within the ACC encode different types of task-relevant information in order to both detect sudden rule changes and participate in the learning a new behavioral strategy or switching to a previously acquired one following such changes? It is also unclear whether the same ACC neurons are coordinated with HPC during rule change detection and during initiation of new behavioral strategies. Furthermore, little is known about such cells' reactivation during sleep in synchrony with hippocampal ripple activity, when memory consolidation would occur (Girardeau, et al., 2009). Computational models of strategy selection in the ACC predict that distinct neuronal populations would subserve functions such as reward detection/processing, goal/strategy selection, and action planning (Hasselmo, 2005; Martinet et al., 2011; Dollé et al., 2018). Other models suggest that off-line replay may play a critical role in making context- or item-selective activity emerge during learning (e.g., Raudies and Hasselmo, 2014).

A possible substrate for memorizing the outcomes of previous actions in order to detect rule changes and implement rewarding strategies could involve synchronization of hippocampal and ACC activity by the theta rhythm, as has been observed during memory tasks (Hyman, et al., 2005; Jones and Wilson, 2005; Siapas, et al., 2005; Benchenane, et al. 2010; cf. Fries, 2015). A mechanism proposed to underlie consolidation of labile hippocampal memory traces into more stable cortical representations invokes the activation of ACC neurons in conjunction with hippocampal oscillations during training sessions. This coordinated activity would then be

reactivated in association with hippocampal memory replay during awake and asleep ripples (Jadhav, et al. 2016; Peyrache et al., 2009; Tang, et al., 2017; Yu, et al., 2018). SWR activity is coordinated among brain structures and this is associated with learning (Eschenko, et al., 2008; Ramadan, et al., 2009; Lansink et al., 2009; Girardeau et al., 2009; Roux, et al., 2017; Tatsuno, et al., 2020). Indeed, prefrontal cell assemblies form during periods of high HPC-ACC coherence upon initial acquisition of new rules (Benchenane, et al., 2010) and are reactivated during SWR in subsequent sleep (Peyrache, et al., 2009). Thus, the ACC can be viewed as a neural substrate for learning and evaluating the suitability of action-outcome contingencies (or rules), and orchestrating the execution of appropriate behavioral strategies (Bunge, 2004), benefitting from hippocampal contextual and memory signals.

In order to reveal neuronal activity profiles that could underlie key prefrontal functions, here we examined single unit activity in ACC subregions A32, A24 and A25 in relation to local field potential (LFP) oscillatory activity in the HPC of rats acquiring and shifting between spatial and cue-guided strategies in a Y-maze (original data set at https://crcns.org/data-sets/pfc/pfc-6/). Here we define strategies operationally as descriptions of the patterns of behavioral responses relevant to the structure of the task and the maze. We make no assumptions about the internal representations and neural mechanisms engaged when the animals perform the task. The task was designed to emulate the extradimensional set shifts of the Wisconsin Card Sorting Task employed to diagnose cingulate cortex deficits in human patients (Berg, 1948; Grant et al., 1949; Milner, 1963). Once the rats reached criterion performance in one task, the rule was changed ("RC sessions"). Then, the animal had to learn that the previous strategy was no longer optimal, and had to learn or shift to a new strategy. We examined sessions when the animal changed strategies ("SC sessions"), and these could include strategies not adherent to the current rule. Individual neurons with firing rate changes in these sessions were

identified as RC- or SC- responsive. We examined the modulation of these neurons by the hippocampal theta rhythm (8 Hz) during behavior, as well as by HPC sharp wave-ripple oscillations (SWRs) during sleep before and after the training sessions, since these would reflect preferential responsiveness to potentially informative hippocampal signals.

Results

Behavior

In the Y-maze, each trial began in the start arm. A central barrier was removed and a cue was lit in the left or right arm in pseudorandom sequence. To obtain a liquid reward, the rat had to go to the end of the appropriate reward arm according to the current rule. In the initial rule, the rewarded arm was on the right side (lit or unlit), and the next reward contingency required visits to the lit side (left or right), then the left side, and then the unlit ("dark") side (**Fig. 1A**). A different rule was applied (Rule Change: RC) after the rat reached a criterion level of 10 consecutive rewarded trials (or 11 rewards in the previous 12 trials). Learning was by trial and error. This stringent criterion was applied because some rats tended to explore other strategies after initial acquisition and this assured that the performance level had become rather stable.

The five rats performed a total of 3322 trials over 108 sessions. Two rats succeeded at learning only two of the rules (then recordings had to be stopped for technical reasons), and two others learned three, while one rat learned all four, and then continued on to make a total of 17 extradimensional (i.e., between spatial and visual discrimination) strategy switches between the four rewarded rules (e.g., **Fig. 1C**). Learning the first rule (Right) took about 10.2±3.7 (SEM) trials on average, while acquiring the Light rule took much longer, 104.2±50.1 trials. Switching

to the Left rule then required 34.3±7.8 trials, and then switching to the Dark rule (only by rat 20) took 319 trials. In most cases, performance did not appear to improve gradually from chance to criterion levels – rather, the animals' performance levels changed abruptly (**Fig. 1B**; cf., Gallistel et al., 2004). Sometimes animals also spontaneously performed alternation even though it was not rewarded, and this, as well as Right, Light, Left and Dark, are considered as "defined" strategies hereafter. Note, however, that when a Bayesian descriptive model of the behavioral data tested 256 rule combinations (see Supplementary Annex), the animals' strategies could be described in other frameworks, such as persistence to the same path (left or right) or cued arm (lit or unlit), as well as alternation. We refer to the defined strategies here for simplicity.

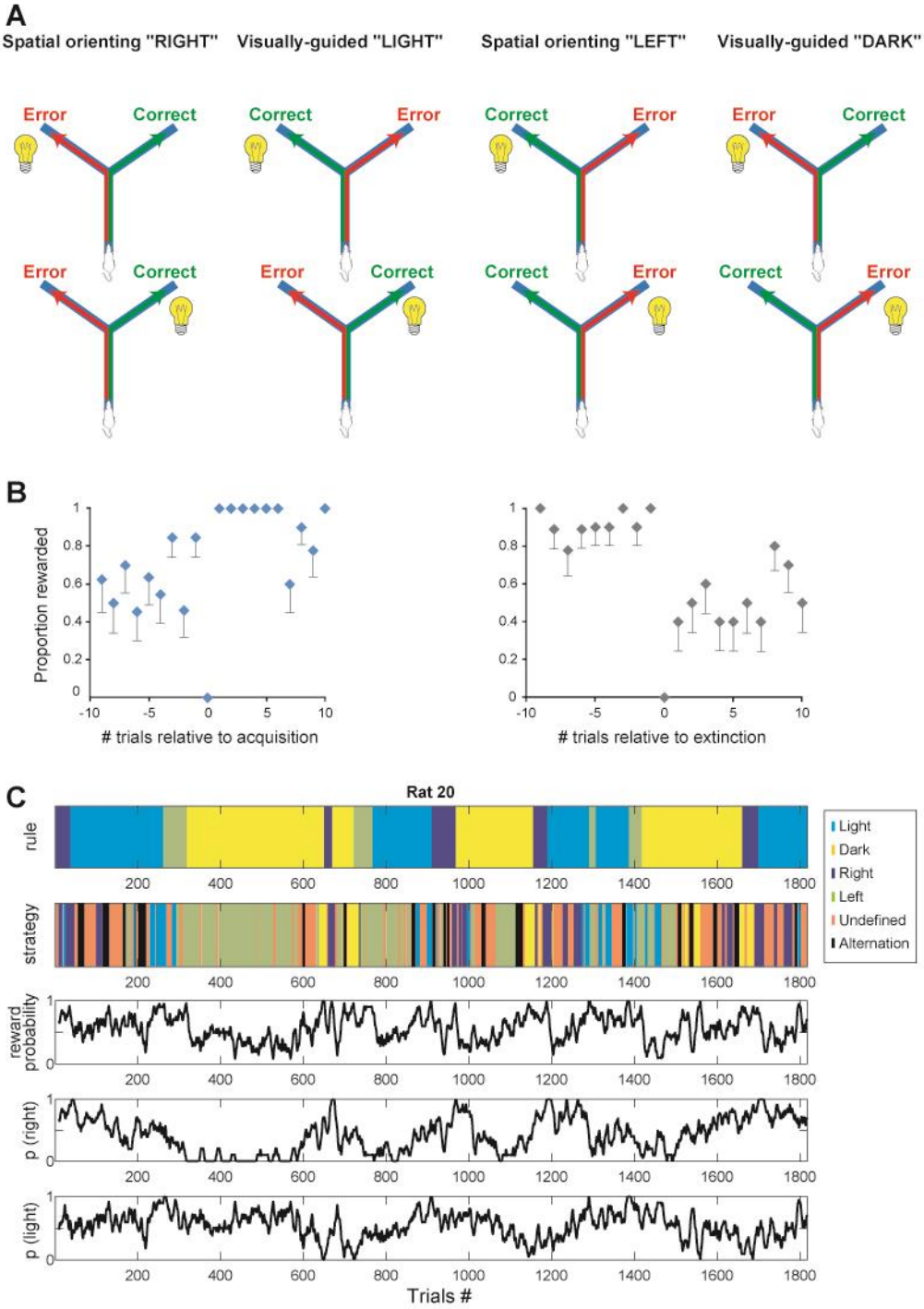


Figure 1. Task and behavior. A) Schema of the four Y-maze task rules. A visual cue is lit at the end of one of the two reward arms in a pseudo-random sequence. **B)** (Left) Reward rate changes upon first acquisition of each rule. The first correct trial is at trial zero. Blue diamond=mean; whiskers=SEM. If no whiskers, SEM=0. Data are from 13 sessions of initial

rule acquisition of various reward contingencies in the five rats. (Right) Reward rate changes relative to first unrewarded trial revealing change of the task rule (trial number zero). Chance levels of reward after rule change are consistent with initial persistence of the previously rewarded strategy (black diamonds=means; whiskers=SEM). (n=10 sessions). **C)** Behavior in rat 20 over all sessions. Top row: Task rule. Second row: Strategies used by the animal. Third row: Probability of reward. Fourth row: Probability of going to the right arm. Bottom row: Probability of going to the lit arm. Probabilities are calculated over sliding windows of 10 trials.

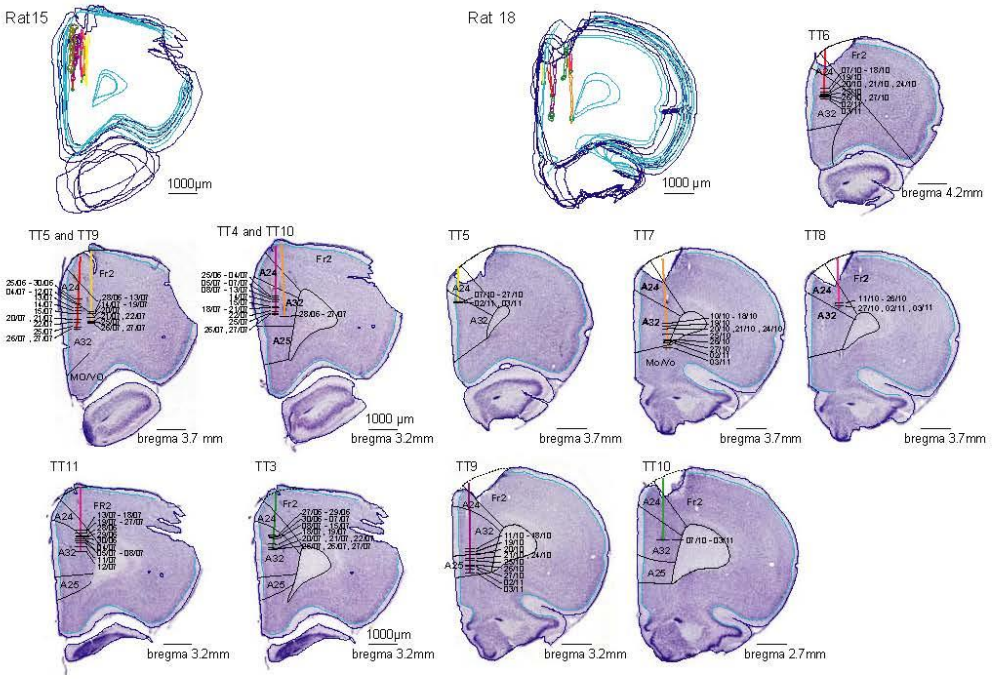
The task rule was changed in 24 sessions and, in 16 of these, the rats also made at least one strategy change, that is, extinction (cessation) of the previously rewarded strategy, or adoption of another defined strategy, rewarded or not. After rule changes, 11.4±8.4 trials were required for the animals to abandon the previous strategy, with the exception of one case where a rat persisted for 280 trials before acquiring the next rule. This suggests that, apart from this exception, the rats did not develop habitual behavior from overtraining. Rather this is consistent with them executing goal-oriented navigation strategies, which involve the hippocampus-ACC network (see e.g., Khamassi and Humphries, 2012). For analyses, the criterion of a minimum of six consecutive trials was selected to define a sequence (or “block") of trials with a given strategy. Simulations of random choices for this behavioral data set showed that this criterion is sufficient to exclude fortuitous inclusion of stochastic choices (see Methods). The animals followed a defined strategy in 170 blocks that included 2424 trials (for example, see **Fig. 1C**). In a total of 898 trials (27% of all trials) the strategy was classified as undefined. The rats consistently adhered to a defined strategy for the entire session in only five sessions. Prior to acquisition of a rewarded task rule, the rats did not tend to simply follow

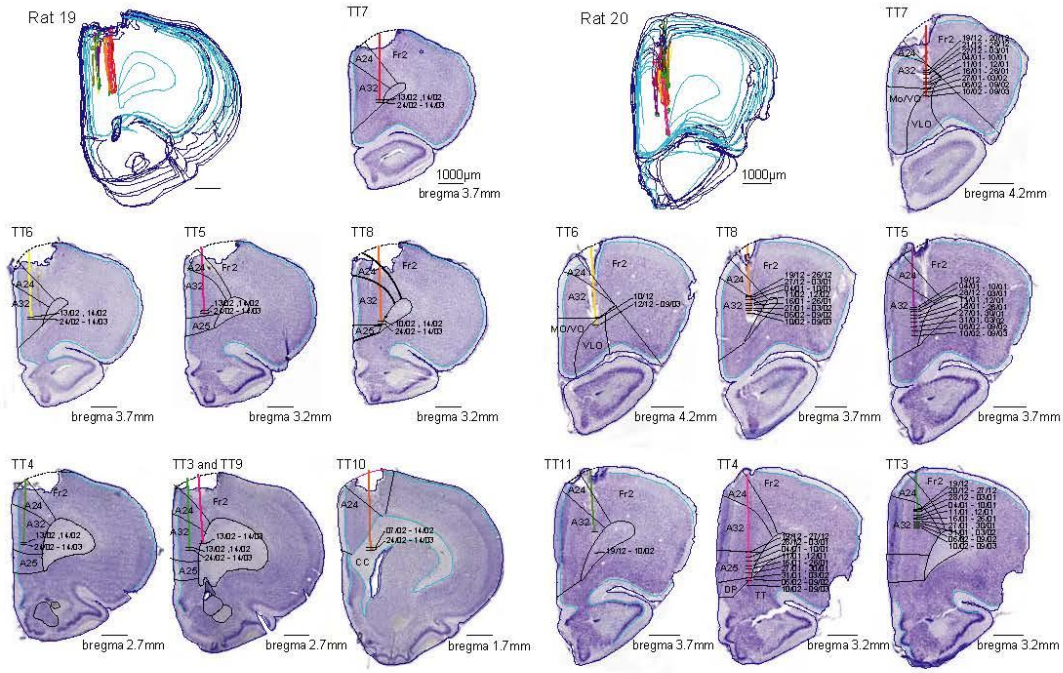
an undefined strategy. Rather, defined strategies other than the currently rewarded one were employed.

The rats changed strategies 118 times, sometimes more than once in a session. Of these 24 were shifts between the task-relevant spatial and visual strategies (Right, Left, Light or Dark). In 55 cases, rats shifted between a defined and an undefined strategy, and they shifted between a task-relevant strategy and the Alternation strategy 39 times. No spontaneous reversals were observed (that is, between Right and Left, or between Light and Dark strategies), nor were they ever rewarded.

In summary, among the 108 recording sessions, these analyses enable us to distinguish the following different types of sessions that are analyzed below:

- 6 *RC-only* sessions, when a rule change (RC) was imposed but the animal made no strategy change (SC) (rather it tended to persist in adhering to the previous rule);
- 65 *SC-only* sessions when a strategy change occurred (not necessarily to the currently rewarded strategy), but there was no RC;
- 18 *RC+SC* sessions, when at least one SC and one RC event occurred; and
- 19 *Training* sessions, when neither a SC nor a RC occurred.


Rat15
1000µm
Rat 18
1000 µm
TT6
bregma 4.2mm
TT5 and TT9
bregma 3.7 mm
TT4 and TT10
1000 µm
bregma 3.2mm
TT5
bregma 3.7mm
TT7
bregma 3.7mm
TT8
bregma 3.7mm
TT11
bregma 3.2mm
TT3
1000µm
bregma 3.2mm
TT9
bregma 3.2mm
TT10
bregma 2.7mm


Rat 19
TT7
1000µm
bregma 3.7mm
Rat 20
1000µm
TT7
bregma 4.2mm
TT6
bregma 3.7mm
TT5
bregma 3.2mm
TT8
bregma 3.2mm
TT6
bregma 4.2mm
TT8
bregma 3.7mm
TT5
bregma 3.7mm
TT4
bregma 2.7mm
TT3 and TT9
bregma 2.7mm
TT10
bregma 1.7mm
TT11
bregma 3.7mm
TT4
bregma 3.2mm
TT3
bregma 3.2mm

197

198

199

Figure 2. Anatomical localization of recording sites in the rat ACC. For each rat, the image to the upper left is a 3D reconstruction of color-coded tetrode (TT) tracks. The dates are indicated for each recording site. The histology for the fifth rat had technical problems and is not shown.

Electrophysiological recordings

The anatomical distribution of the recording sites of the ACC cells is shown in **Fig. 2**. The largest group was located in layer 5 of A32, although layers 2/3 and 6 of A32 and areas A24 and A25 were also represented.

Activity transitions at the population level in relation to RC and SC

We first tested if population activity changed relative to RC and SC, as shown previously for criterion performance in rewarded tasks (cited in the Introduction). Population vectors of overall activity of simultaneously recorded neurons were examined for each session to identify those trial(s) when changes in firing rate ("activity transitions") occurred. These transition points correspond to those trials where there is a major change in the population vector of activity of all recorded neurons. These activity transition trials were then compared to the trials when RC and SC events occurred and the delays were calculated. Specifically, a (z-scored) population vector correlation matrix was constructed with the element at row *i* and column *j* quantifying the correlation between the population vectors at trials *i* and *j* (cf. **Figure 3**). A matrix block decomposition algorithm detected the transition points by searching for the presence of distinct "sub-matrices" within the matrix. These correspond to sequences of trials with distinct patterns of population activity. This process identified and excluded inter-trial comparisons outside the blocks (e.g., in the top row of **Figure 3**, the blue submatrix blocks at the lower left and upper right, called the "off-diagonal" terms) while preserving the maximal

amount of information of the matrix (i.e., the highly correlated sub-matrix blocks at the upper left and lower right of Fig. 3). Block decomposition was performed when the cumulated correlation of off-diagonal terms represented less than 1/3 of the total correlation in the matrix. This analysis detected neural activity transitions in 100/108 (92.6%) sessions (see **Fig. 3A**). Of these, 81 sessions contained at least one behavioral event (RC or SC). Interestingly, a population activity transition occurred within four trials before or after a behavioral event (either a RC or a SC) in 68 of the 81 sessions (i.e., 84%; **Fig. 3B**). This is more than expected by chance since a Monte Carlo distribution of 1000 population transitions randomly generated in these sessions yielded only 28% of cases with only four or fewer trials between neural and behavioral events (binomial test, $z=11.0$, $p<.001$). Note, however, that there was also a population activity transition in all 19 Training sessions (where no RC or SC event occurred), consistent with previously reported lability of ACC single neuron activity (e.g., Rich and Shapiro, 2009, p. 7214; 12.6% of their neurons; Tanaka, 2007). This could reflect gradual neural processes underlying acquisition of new rules. (We speculate that this corresponds to our subsample of neurons transitioning between attractor states, but this alone may not be sufficient to alter behavior. Rather, this might need to be concerted with some critical number of other neurons, and perhaps in other brain structures as well. Thus, such a population ensemble could develop over the course of several training sessions as more cells acquire the appropriate behavioral correlate and become synchronous, leading to a behavioral change.) Conversely, in 8 (9%) of the 89 sessions where there was an RC or SC behavioral event, no neural activity transition was detected in the respective samples of neurons recorded then.

Since multiple behavioral events in the same session could lead to possible confounds, we first examined sessions with only a RC or a SC. In five of the six RC-only sessions a population vector transition occurred 2.3±1.6 trials (median 2.5 trials) *after* the first trial revealing that the RC

had occurred. Thus, this neuronal population activity responded rapidly to the RC, even though no explicit cue was presented to signal the RC's. Strikingly, in the 65 SC-only sessions, the neural activity transition occurred on average 1.0±0.86 trial (median = 1) *before* the first trial of the block of trials qualifying as an SC event. Thus, the transitions in ACC population activity tended to anticipate SCs. Analyses of RC+SC sessions were inconclusive with the population vector approach, and these will be examined further below at the single-cell activity level. In summary, the transitions in mPFC population activity tended to occur in conjunction with RC or SC events.

In order to examine the relation of hippocampal activity with RC and SC responses in ACC, further analyses focused on single neuron activity modulation by hippocampal theta and SWR. First, however, it was necessary to determine if the population responses also appeared in individual neurons. In contrast with the population activity, the individual neuron analyses distinguish firing rate changes at different points of the maze. This permits reward site activity to be set aside, since this could lead to potential confounds related to changes in reward consumption prior to versus after RCs and SCs.

Firing rate transitions in single ACC neurons in relation to RC or SC

Of the 2290 recorded ACC neurons, 1888 with average firing rates superior to 0.3 Hz in at least one of the task periods (pre-start, post-start, pre-arrival at reward site, post-arrival) were retained for these analyses. Putative interneurons comprised 19% of the cells recorded, and 79% were classified as putative pyramidal cells, while 2% could not be clearly identified. Firstly, 1337 (71%) of the 1888 cells showed behaviorally correlated activity, that is, the firing rate significantly varied across the four trial periods (ANOVA, $p<.05$). This proportion was similar considering only pyramidal cells as well, and also in the subgroup consisting of A32 layer 5

cells. These proportions are compatible with previous reports of data recorded in animals performing other tasks, employing analyses using different analytical criteria (50%: Pratt and Mizumori, 2001; 62%: Mulder et al., 2003; 46%: Rich and Shapiro, 2009).

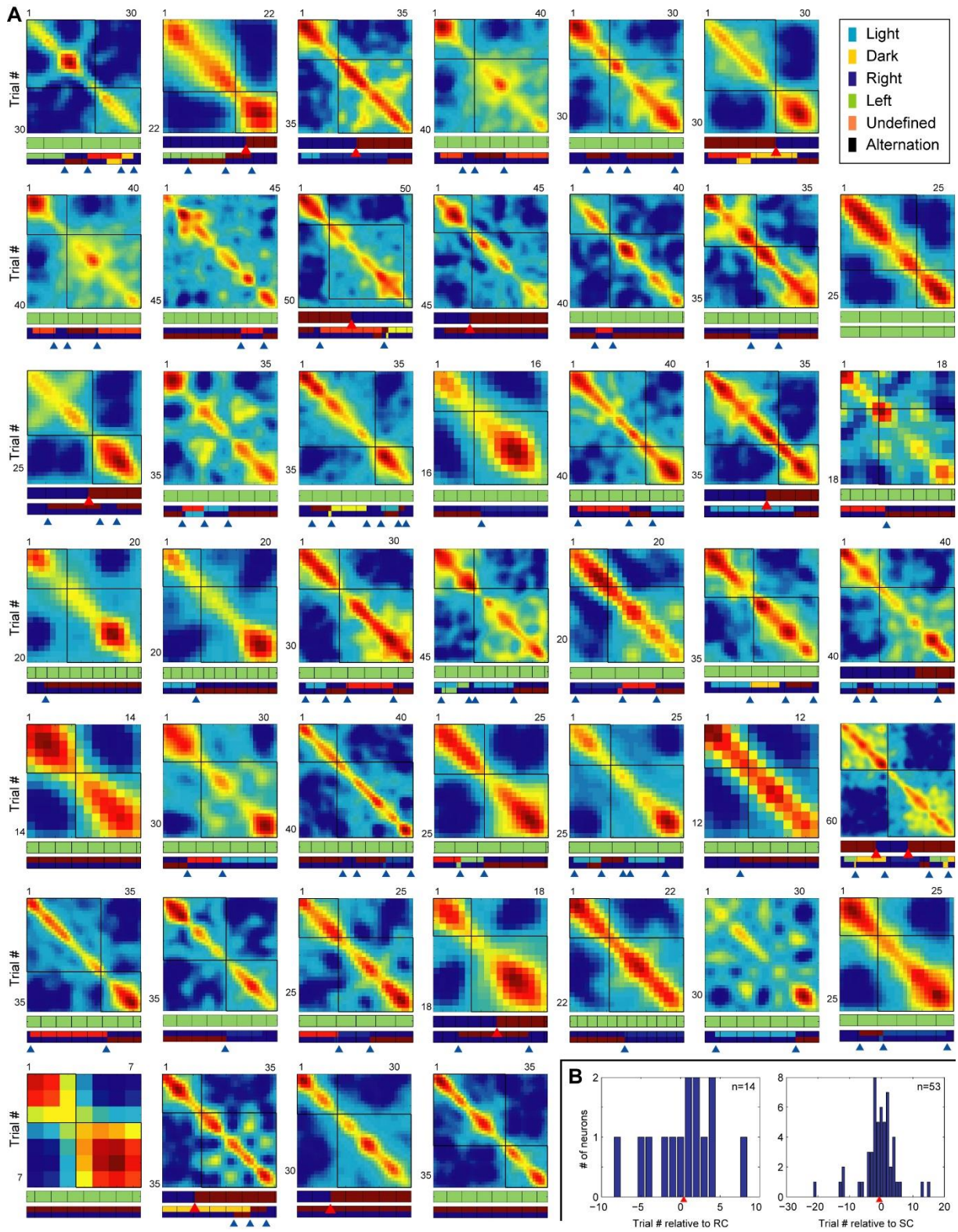


Figure 3. Transitions in population activity over the course of the sessions. A) Each box plot shows the correlation matrix of the population vectors between different trials (rows and columns) of an example session. (Hotter colors indicate higher correlations). Crossbars

represent the automatically determined decomposition of the matrix into sub-blocks along the diagonal. (Cases with no crossbars are presented to demonstrate how sub-blocks could be absent – in these cases, decomposition led to supra-threshold information loss since off-diagonal terms contained substantial correlations in population activity between trials). Below the matrices, top bars indicate when task rules changed (red arrowheads). Lower bars indicate when strategies changed (blue arrowheads). In the bars, dark blue represents undefined strategies while other colors are defined strategies with an arbitrary color code here. Within each plot, a given color indicates the same rule or strategy. **B)** Incidence of neurons with shifts in population activity at various delays relative to RC or SC behavioral events (which occur at trial zero). Red arrowheads point to means (SC: 0.36 trial; RC: -0.62 trial).

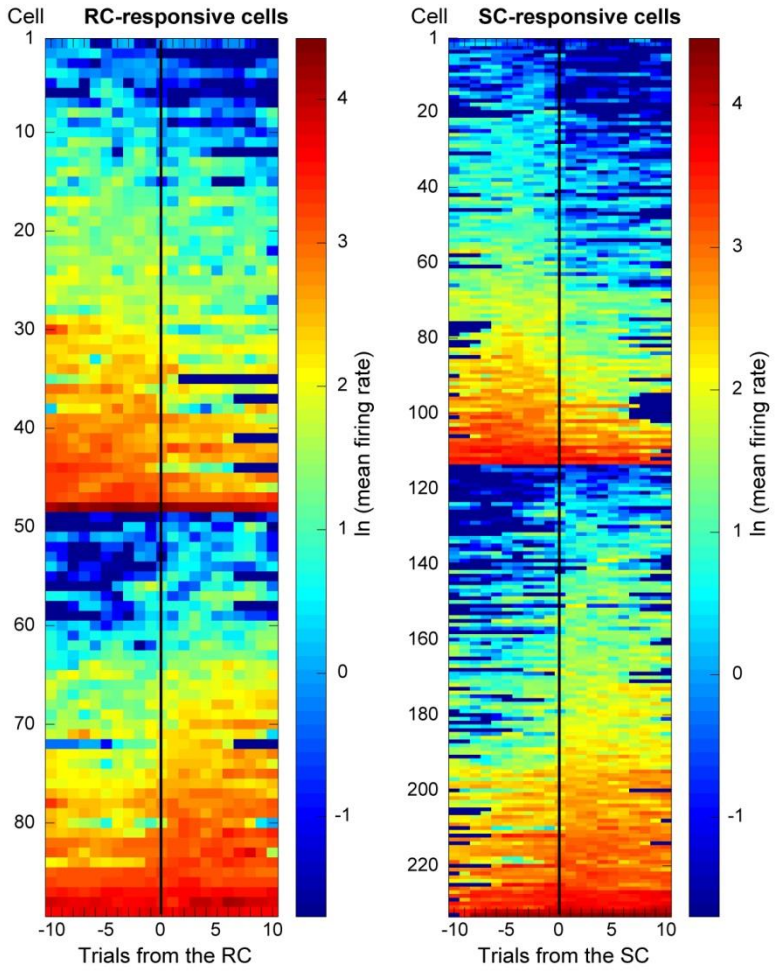


Figure 4. Overview of significant firing rate changes in representative individual neurons before and after RC and SC. Decreases in activity are at the top, and increases at the bottom. The color scale is the natural logarithm of the firing rate. Each row corresponds to one neuron. Plots are synchronized to the trial when the RC could be detected by reward absence, or the trial when the SC occurred.

To provide an overview, neuronal activity was plotted as a function of trials prior to or after RC and SC events. This revealed numerous neurons that appeared to change firing rate relative to these events (**Figure 4)**. These transitions could be either increases or decreases in firing rate after the RC or SC events, and occurred in cells with high or low firing rates. To more precisely compare the timing of transitions in firing rates of neurons relative to when RC's and SC's occurred, Monte Carlo analyses (Fujisawa et al., 2008) examined single unit activity along the maze. This only considered maze regions visited on all trials of the session. This analysis first searched for positions on the maze where actual spike activity exceeded shuffled data first for trials before vs after the 5th trial of the session, then before and after the 6th trial, etc., seeking the trial that had the greatest difference in average firing before vs. after it. These data are represented as the rows of color raster plots of 2nd column of **Figs. 5A**-**E**. For example, in a session with 40 trials, one of the random shuffles between trials 1 through 5 and trials 6 through 40 could exchange the sequential bins of trial 1 with that of trial 36, trial 2 with trial 13, trial 3 with trial 40, trial 4 with trial 6 and trial 5 with trial 27. The firing rates in the two surrogate groups were compared for each of the spatial bins, then accumulated with data from other such shuffles to create spike density functions for the shuffled data. Then, analogously to transition detection in the population analysis of **Fig. 3**, the greatest "peak transition trials" were determined, and then compared to the timing of the SC or RC trials. The

null hypothesis is that the timing of these peak transition trials is unrelated to these behavioral events.

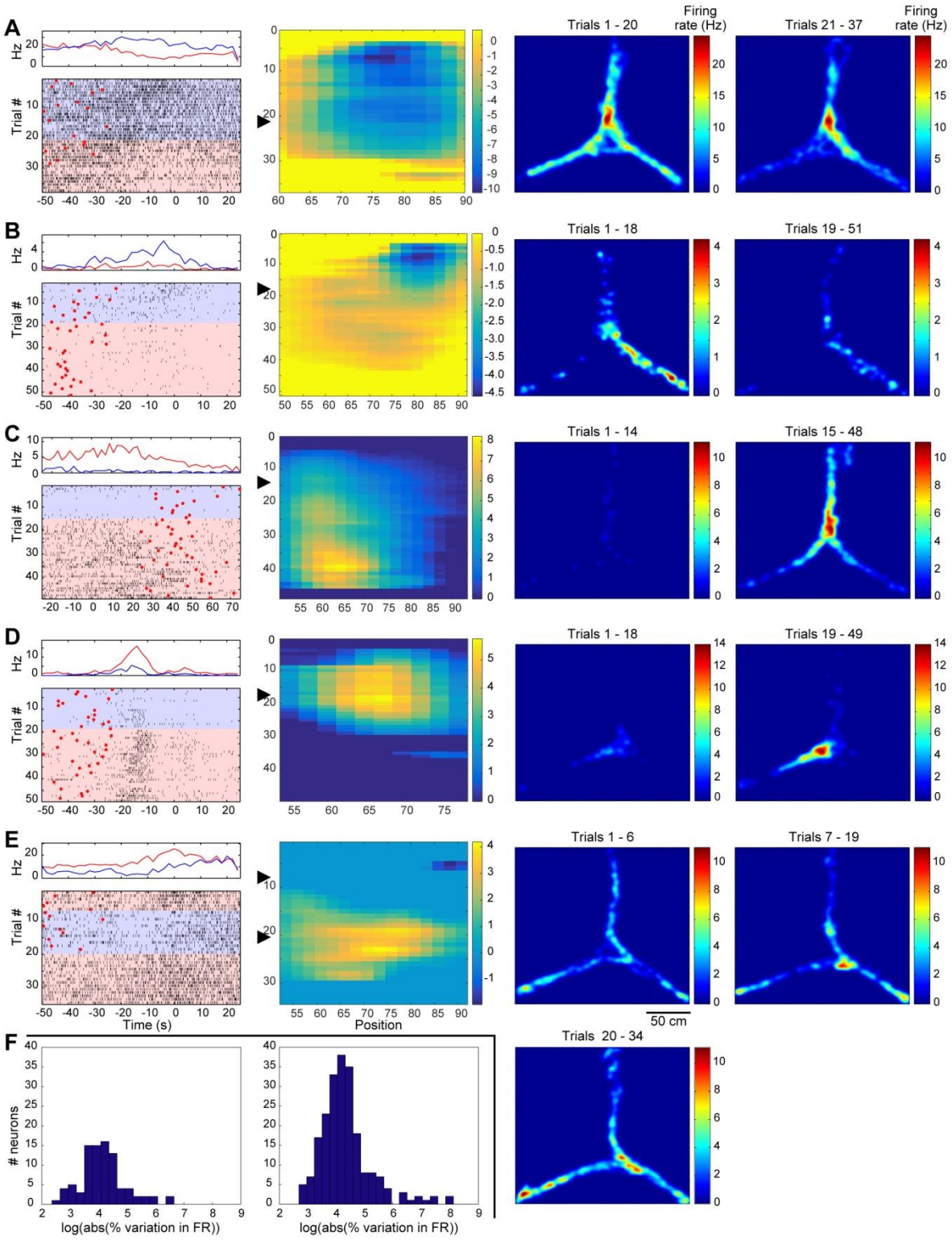


Figure 5. Examples of RC and SC associated changes in firing rate. A through E) Data from representative neurons. The left column contains raster displays and above them, the corresponding histograms of spiking activity, both as functions of time. Blue (or red) histogram traces represent the firing rate before (or after) the RC or SC, corresponding to the color-coded shading of the raster displays. Each row in the rasters is a single trial. In rasters, when the red circles are after zero they correspond to reward site arrival, and zero is the trial start event in these cases. When the red circles are before zero, they indicate the trial start events, and zero is the reward site arrival time. (In E, the histogram for trials 1-6 is not shown.) The plots in the second column show all significant Monte Carlo bootstrap SDF values (trial number on y-axis and linearized position on the maze on the x-axis). Only deep blue rectangles (zero) had non-significant SDFs. The color scale represents the magnitude (in units of spikes/s) of the difference between the SDF value and the upper or lower 2.5% confidence limits. Arrowheads indicate the RC or SC). The third and fourth columns show the spatial distribution of the neuron's activity before and after the RC or SC. The start arm is above. **A)** A rule-responsive neuron with stable activity at the decision point and a rule-modulated activity in the arrival arms. **B)** A rule-responsive neuron with rule-modulated activity in the left arrival arm. **C)** A strategy-responsive neuron with activity selective for the start arm and decision point. **D)** A strategy-responsive neuron with activity selective for the decision point and right arrival arm. **E)** A strategy-responsive neuron with transitions in activity at two strategy changes on trials #7 and #20 (hence the three spatial plots). The activity is reduced at the decision point while the animal follows the correct Light strategy between trials #7 and #19. The activity remains low during two trials after the task rule change (when the Light strategy is no longer rewarded), but resumes on trial #20 when the rat started a block of Alternation trials. **F)** Distribution of the incidence of neurons with respective variations in firing rate following

changes in rule (left) or strategy (right). FR= firing rate. X-axis ranges vary among Figures since they are limited to maze positions the rats occupied on every trial of the respective sessions.

Among the 114 neurons analyzed from the 6 RC-only sessions, 50 (43.9%) showed at least one spike density function (SDF) difference exceeding the shuffled distribution, indicating there was a significant difference in firing rate between at least one of the pairs of blocks of trials in at least one bin on the maze. Similarly, among the 846 neurons analyzed from the 65 SC-only sessions, 400 (47.3%) showed a significant SDF difference (Monte Carlo, $p<0.05$, two-tailed). **Figs. 5A** and **B** show data from two representative RC-responsive neurons and **Figs. 5C-E** show three representative SC-responsive neurons. **Fig. 5F** shows the wide range of firing rate increases or decreases in comparisons between blocks before vs. after the RC or SC.

To detect transitions in single neuron activity, the greatest values of these inter-block differences were determined. Then, for each neuron, the incidences of trials separating blocks with these peak SDF values were plotted as a function of the trial when that session's RC or SC occurred. Since these distributions did not pass the Lilliefors test for normality ($p>0.05$; **Fig. 6A**), they were fitted with beta distributions to determine normative values. Strikingly, in **Fig. 6B**, the mean delay was 3.9±2.5 trials after the trial when the RC became evident in RC-only sessions (median=3 trials after; significantly different from zero, t-test, $p<0.001$, n=8649). The mean was also 3.8±2.3 trials after the RC for RC+SC sessions (median=3 trials after, again significantly different from zero, t-test, $p<0.001$, n=9311).

In SC-only sessions, the mean of the beta distribution of the SDF values was 2.9±1.4 trials *before* the SC event; (**Fig. 6C**; median=1 trial before; significantly different from zero: t-test, $p<0.001$, N=95629). Similarly, in RC+SC sessions the mean was 2.5±1.3 trials earlier than the

SC (median=1 trial before; t-test, $p<0.001$). Thus, individual ACC neurons can respond rapidly to changes in action-outcome contingencies (i.e., RC's), or precede adaptive responses to these changes (i.e., SC's). These results are consistent with those of the population analyses (2.3±1.6 trials after RC and 1.0±0.86 trial before SC).

While it seems counter-intuitive that some transitions occur prior to RC (**Figs. 3B** and **6B**), this may be related to the lability of ACC neurons, or to anticipation of RC by the more experienced rats. The reduced incidence for separations being observed between blocks for early and for late trials could be related to fewer combinatorial samples there. However, the mean values following RC or preceding SC events cannot be due to these potential confounds.

One possible confound for interpreting RC responses in the SD tasks is that the neuronal activity could be correlated with action value, rather than the RC per se. For example, in the Right rule, each trial (rewarded or not) reinforces the value of the appropriate action. This value progresses over trials, but, after RC's, it decreases for the previous rule, and increments on each trial with the new rule. There is a similar issue with the visual discrimination rules, and thus we extended the concept of action value to "strategy value". This emulates, in a simple and parsimonious way, the strategy value learning process in relevant computational models (Dollé, et al., 2018; see Supplementary Figure 1). Strikingly, multiple regression analyses showed that the activity of only 8/50 (16%) of RC-responsive neurons was significantly correlated with one of the strategy values ($p<.05$, significance threshold set with bootstrapping with 10,000 shuffles). Thus, most RC responses cannot simply be a confound with strategy value. Interestingly, firing in 288 other neurons was exclusively correlated with strategy value, and their modulation by hippocampal rhythms is discussed below.

Another possible confound concerned reward rate (RR), which could toggle between 100% and 50% after RCs or SCs. But the firing rate of only 4/50 (8%) RC-responsive neurons was correlated with RR ($p<.05$, regression analyses with bootstrapping, with rate averaged over six trials). Similarly, only a minority of SC neurons was significantly correlated with RR (16/400, 4%), and thus this was not a major confound. However, 59 neurons were exclusively correlated with RR. RR correlates have been previously observed in ACC (e.g., Genovesio et al., 2005) and would be useful for continuous monitoring of task performance levels, complementing the RC-responses.

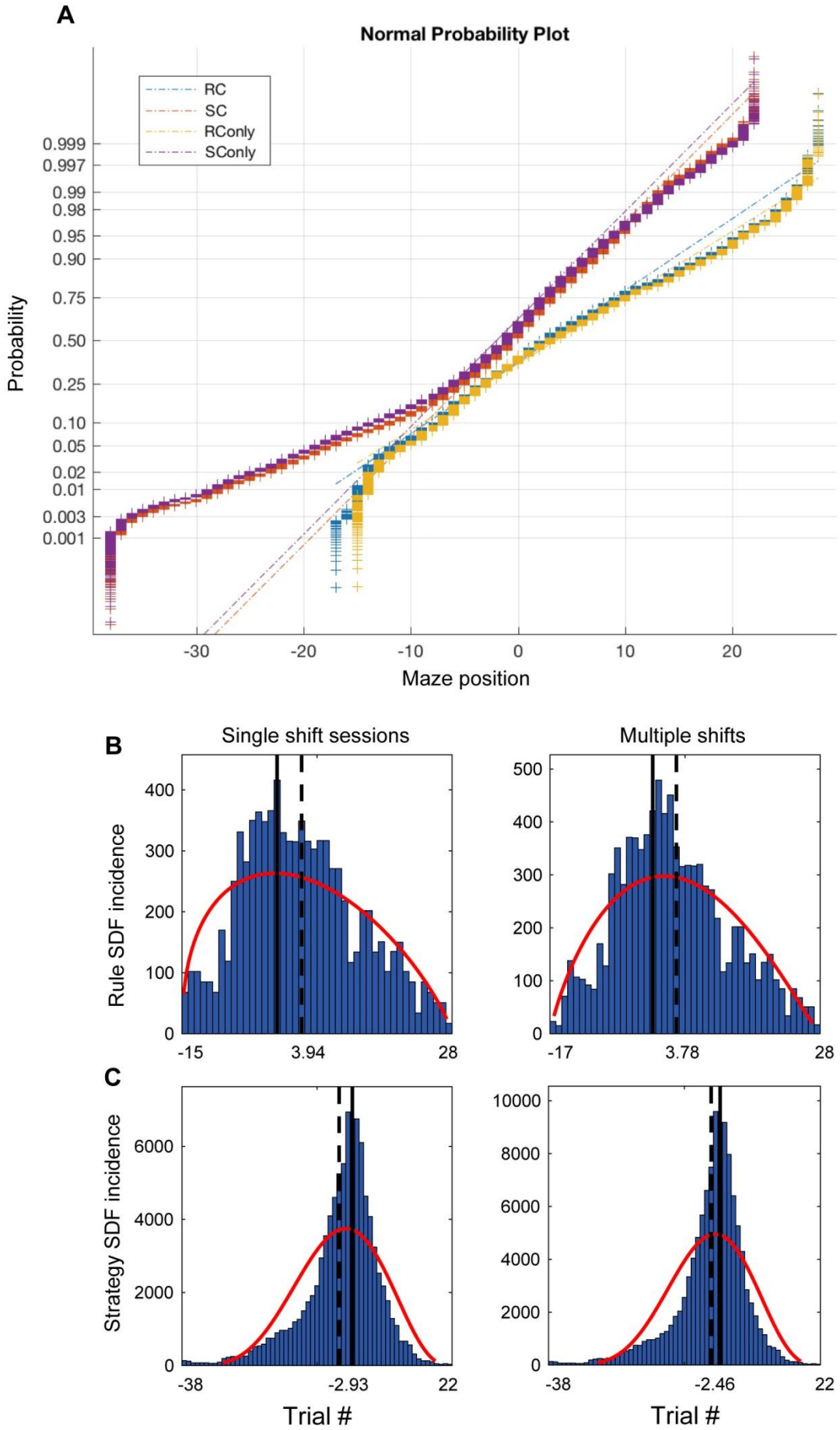

A
Normal Probability Plot
RC
SC
RConly
SConly
Probability
0.999
0.997
0.99
0.98
0.95
0.90
0.75
0.50
0.25
0.10
0.05
0.02
0.01
0.003
0.001
-30
-20
-10
0
10
20
Maze position
B
Single shift sessions
Multiple shifts
Rule SDF incidence
-15
3.94
28
-17
3.78
28
C
Strategy SDF incidence
-38
-2.93
22
-38
-2.46
22
Trial #
Trial #

Figure 6. Distributions of significant SDF values for RC-responsive and SC-responsive neurons. A) Spatial distributions of significant SDF values do not follow a normal distribution since they deviate from the best fit regression bars (dashed lines). **B and C)** Beta distribution fits (red curves) of incidences of significant SDF values for rule and strategy changes respectively in either Single shift sessions or Multiple shift sessions. Vertical lines represent the trials when the RC (top row) or SC (bottom row) occurred. Vertical dashed lines represent the means of the fitted beta distribution. The SDF distribution means occur after the RCs, but before the SCs.

Modulation of neurons by hippocampal SWRs in RC and SC sessions

Since coordinated HPC and PFC activity during sleep SWRs is implicated in the consolidation of memories (e.g., Peyrache, et al., 2009; Girardeau, et al., 2009; Maingret, et al., 2016; Rothschild, et al., 2017), we compared the incidence of neurons reactivated during post-session sleep (S2) after RC as well as SC sessions. The incidence of neurons modulated by SWR in S2 was significantly greater than S1 in RC sessions (18.4%) than in Training sessions (8.8%; Fig. 7A; χ^2=9.25, df=1, p=0.002) and this S2 value was significantly above baseline. The incidence of SWR modulation in S2 was also greater than S1 in sessions with an RC and an SC (Fig. 7A; χ^2=7.66, df=1, p=0.006). However, the S2 incidence was not significantly above baseline, while in S1 it was significantly lower. No significant increase in SWR modulation was observed during any sleep sessions prior to or after an SC session (Fig. 7A; χ^2 test, p>.05). A significantly higher proportion of all neurons were reactivated during SWRs during S2 than during S1, both in RC-only sessions (χ^2 = 5.16, df=1, p = 0.013; Fig. 7A) and in RC+SC sessions (χ^2 = 9.12, df=1, p=0.0014; Fig. 7A). However, this difference was not significant in SC-only

sessions (χ^2 = 1.41, df=1, p = 0.166) or in Training sessions (χ^2 = 2.22, df=1, p = 0.088; dashed lines in **Fig. 7A**).

The SWR modulation ratios of reactivated neurons during S1 vs. S2 were indistinguishable for RC sessions (1.41 and 1.39 respectively; rank sum test, p=0.85) and SC sessions (1.60 and 1.41; rank sum test, p=0.085) or RC+SC sessions (1.42 and 1.39; rank sum test, p=0.81). In summary, more ACC neurons were modulated by SWR during S2 than in S1 for RC and RC+SC sessions, but not SC sessions, and there was no significant change in the strength of the modulation.

Hippocampal SWR modulation of cells with identified responses

The next question concerned whether such activity modulations by hippocampal SWRs were specific to neurons with firing rate changes relative to RC or SC, or to those with other responses, and if, in these specific cases, this modulation was greater in S2 than in S1. Strikingly, a significantly higher proportion of RC-responsive neurons were modulated by ripples during S2 than during S1 ($\chi2$=4.23, df=1, p=0.040; Fig. 7B). This difference also held in RR responsive neurons ($\chi2$=3.85, df=1, p=0.05; Fig. 7B), and was close to significant in SC-responsive neurons ($\chi2$=3.53, df=1, p=0.06). In contrast, this was not significant in SV-responsive neurons ($\chi2$=0.634, df=1, p=0.43). Importantly, none of the groups had a significantly greater proportion of neurons modulated by SWRs during S1 than unresponsive neurons ($\chi2$ test, p>.05). In S2, the incidences of RC, SC and RR neurons were above baseline ($\chi2$=8.4, 7.1 and 4.4, df=1, p=0.004, 0.003 and 0.04). Thus, greater proportions of RC- and RR-responsive cells modulated during SWRs in S2 than S1 were not due to sub-baseline values in S1. Among neurons modulated by SWR during both S1 and S2, SC-and RR-responsive neuron proportions were significantly higher than non-responsive neurons ($\chi2$ = 8.3, df=1, p=0.004; Fig. 7B). This could contribute to maintaining task-relevant information between the two sleep periods.

There were two types of SCs: to rewarded or to unrewarded strategies. The proportions of SWR-modulated neurons in these two groups were not significantly different (S1: 6.7% vs 3.9%; S2: 4.5% vs 8.6%, S1 & S2: 5.6% vs 6.6%, $\chi2$ = 2.99, df=3, p=0.394). Similarly, the incidence of SWR modulation in SV-responsive neurons was not significantly different for rewarded vs non-rewarded strategies (S1: 6.0% vs. 2.6%, S2: 8.0% vs. 2.6%, S1 & S2: 1.2% vs 5.3%; $\chi2$ = 5.20, df=3, p=0.158). As observed in computational models of strategy learning (e.g., Dollé et al., 2018), this is consistent with reinforcing values of correct strategies or weakening values of unrewarded strategies each playing symmetrical roles in ACC-dependent learning, and would thus be equally reactivated during sleep.

A possible confound is that cells with higher firing rates have a greater chance to be reactivated and modulated by SWRs during sleep (Fernando-Ruiz, et al., 2019). Indeed, overall, ripple-modulated neurons here had higher average firing rates (11.9 Hz) during behavior than non-ripple-modulated neurons (5.6 Hz; Kruskal-Wallis test, $\chi2$=155.5, df=1, p=0). However, a two-way ANOVA, with neuron category (SC, RC, SV, RR, non-responsive cells in RC-SC sessions, and non-responsive cells in training sessions) and sleep period as factors, revealed no firing rate difference between neurons modulated by ripples in S1 and in S2 (F=0.13, df=1, p=0.72), and no significant interaction (df=5, p=0.53), despite the differences between S1 and S2 reported above. The neuron category factor was significant (F=3.3, df=5, p=0.006). But the only group with a significantly lower firing rate (RR selective neurons) was the one with a significantly greater incidence of ripple modulation (Tukey-

Kramer post-hoc test, p<.05; Supp. Figure 2).

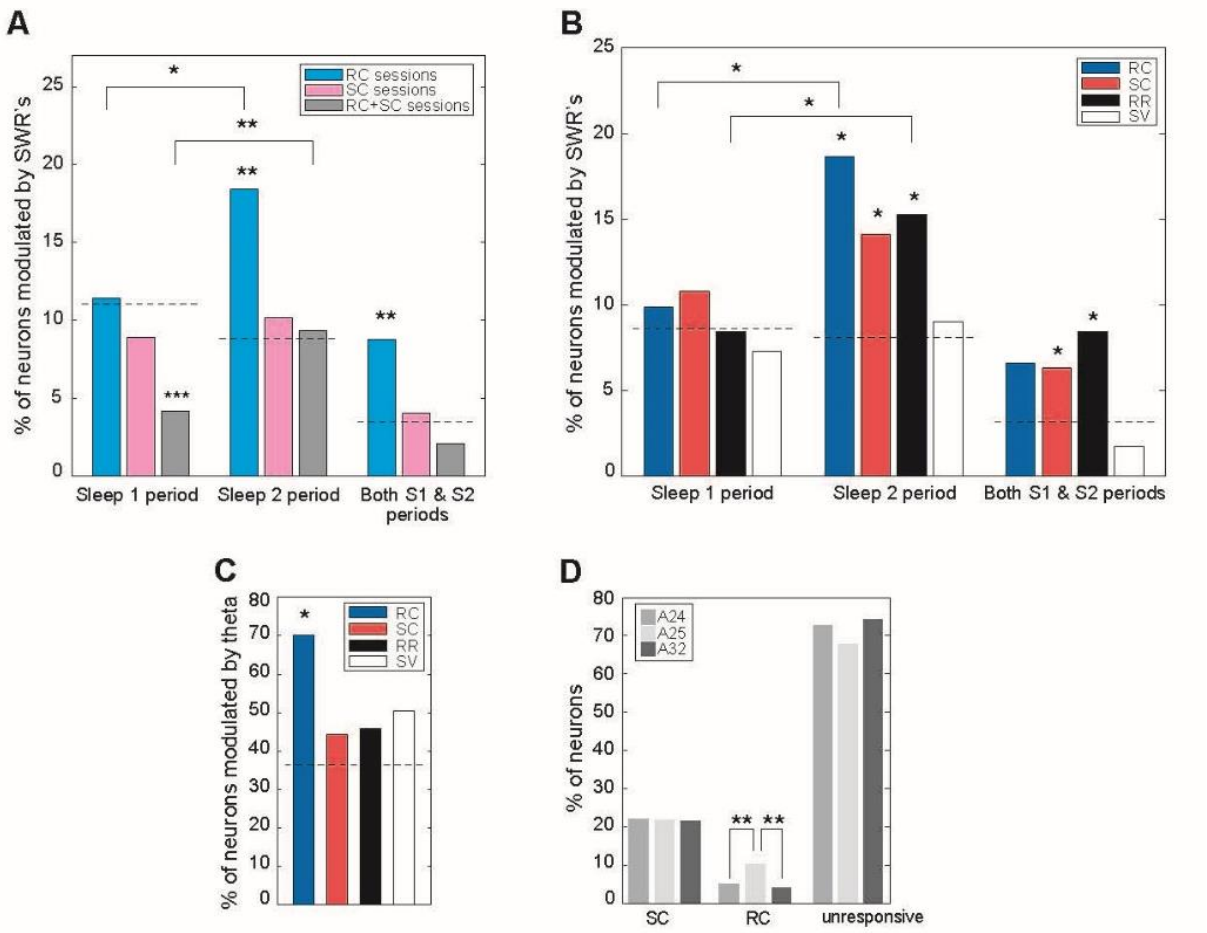


Figure 7. Modulation of ACC activity by hippocampal sharp-wave ripples (SWR) and theta oscillations. A) Incidence of neurons with SWR modulation in sleep before and after sessions with a RC, a SC, and both. Dashed lines indicate reference baseline (here, neurons in Training sessions). "Sleep 1 period" includes neurons modulated by SWR in S1 only, as well as in both S1 & S2. Same for "Sleep 2 period". Stars above single bars indicate a significant difference from relative to baseline (dashed line; unresponsive cells recorded in all sessions). **B)** Incidence of SWR modulation for SC-, RC-, strategy value, and RR responsive cells. Dashed line indicates reference baseline (here, from unresponsive neurons in training sessions). **C)** Incidence of theta modulation according to cell response category. Baseline is proportion of unresponsive

neurons in all sessions. **D)** Anatomical distribution of responses. * - p<0.05; ** - p<0.01; *** - p<0.001.

Modulation of RC and SC responsive neurons by the hippocampal theta rhythm

During all sessions with a RC and/or a SC, about 39% (838/2127) of the ACC cells analyzed were modulated by the HPC theta rhythm (Rayleigh Z tests: p<.05). Importantly, HPC theta modulated a significantly greater proportion of RC-selective cells (70%) than unresponsive cells from all sessions, while SC-, RR-, and strategy value neurons did not (**Fig. 7C**). However there was no significant difference in the strength of modulation between RC- and SC-responsive cells (RC: 6.0%, SC: 6.5%; Mann-Whitney test, zval 0.096, rank sum=3329, p=0.92). Differences in incidence of theta modulation could not be explained by differences in average firing rates in these groups (Supp. Figure 2; Mann-Whitney test, zval=1.234, rank sum=3692, p=0.22). (Differences in running speed were not related to differences in incidence of theta modulation; see Supplemental Data).

Anatomical distribution of neurons with activity transitions in relation to RC and SC

The proportion of RC-responsive neurons was higher in A25 than in A24 or A32 (**Figure 7D**; χ^2=10.7, df=2, p=0.0024). Furthermore, a significantly lower proportion of RC-responsive neurons (7.7%) was located in the dorsal third of A24 (near A32), compared to SC-responsive neurons (21.3%) and unresponsive neurons (19.6%; χ^2=7.73, df=2, p=0.0105). The ratio of pyramidal cells to interneurons did not vary significantly among neurons with the respective response types (74% pyramidal cells for RC-responsive, 80% for SC-responsive, 79% for unresponsive; χ^2=2.47, df=2, p=0.15). Nor did the concentration in superficial vs deep ACC layers vary for RC or SC (38% were in deep layers).

In summary, more ACC neurons were modulated by HPC SWRs during S2 in RC sessions than baseline, and the incidence in S2 was greater than S1 in these sessions too. Furthermore, modulation by HPC SWRs preferentially occurred in neurons selective for RC, SC and RR, with greater incidence in S2 than S1 for RC and RR. RC-responsive neurons also had a greater incidence of theta modulation than other response types, although all groups had high values. Overall, these results show that modulation by hippocampal theta and SWR was preferentially associated with neurons responsive to changes in task contingencies.

Discussion

Here, populations and substantial proportions of individual ACC neurons responded rapidly after changes in reward contingencies (RC), or in a fewer trials prior to a SC, and were modulated by SWRs or hippocampal theta rhythms. The experimental design permitted characterization and comparison of the neurons with these two response types in the same data set. In effect, the difficulty of the task permitted comparisons with activity in sessions with no RCs, when the animals spontaneously shifted between different strategies in search for the correct one. Thus, delays between RC and SC were often sufficiently long to permit analyses to distinguish activity changes relative to these respective events, reducing possible confounds. In comparable studies that addressed other questions, RC and SC responses could not always be clearly distinguished (Rich and Shapiro, 2009; Durstewitz, et al., 2010; Karlsson, et al., 2012; Powell and Redish, 2016; Malagon-Vina, et al. 2018; Trouche et al, 2018). Here, we extend upon these and other results by showing that ACC activity also makes abrupt transitions for extradimensional shifts, here between spatial vs. sensory discrimination rules and strategies. Trial-by-trial analyses permitted greater precision than trial block analyses.

Thus, the population analyses typical of many studies were complemented here by analyses of individual neurons permitting detection of hippocampal theta and SWR modulation.

Theta and SWR modulation

Strikingly, a dramatically higher proportion of RC-responsive cells (than SC-, RR-, strategy value-responsive, or unresponsive cells) were modulated by hippocampal theta. This is consistent with a theoretical framework wherein the hippocampal theta rhythm coordinates coherent activity in HPC and ACC for registering experiences. This would lead to synaptic modifications underlying labile memory traces through hippocampal neurons firing sequentially in successive phases of theta cycles (Skaggs et al., 1996; Wagatsuma and Yamaguchi, 2004), and associated PFC activity. SWR events correspond to the replay of sequences of hippocampal activity from the preceding behavioral session, and this modulation could correspond to consolidation of memory traces via transmission to the cerebral cortex, and ACC in particular (Buzsáki, 1989; review: Girardeau and Zugaro, 2011).

Experimental evidence for this is provided by increased reactivation of neurons during SWR's in S2 compared to S1. During S1, the incidence of hippocampal SWR modulation was indistinguishable among cell groups as well as from baseline values. Furthermore, a significantly higher proportion of in RC- and RR- responsive cells were modulated by SWRs in S2 than S1. This is consistent with a hippocampal contribution to ACC responses to changes in task contingencies, since the hippocampus responds to mismatches between expectations and observations, and novelty (Gray, 1982; O'Keefe, 1976; Vinogradova, 2001; Kumaran and Maguire, 2006; Kafkas and Montaldi, 2018). The proportion of SWR-modulated SC cells was greater than baseline in S2, but was not significantly greater than S1. The greater incidence of hippocampal SWR modulation of RR-responsive neurons is consistent with the enhanced

replay of rewarded trajectories in post-learning sleep (Michon, et al., 2019). Interestingly, in an elevated plus maze anxiety paradigm, Adikhari et al (2011) also found increased ventral hippocampal theta modulation of ACC neurons with task-related responses.

In summary, both theta and SWR modulation could reflect selective transmission of hippocampal system signals that would contribute to these ACC neuronal responses. The high incidence of RC- and RR-responsive neurons suggests a role for HPC-ACC coordination in signaling new environmental contingencies. In SC-responsive neurons, hippocampal signals could contribute to elaborating appropriately adaptive behavioral responses.

Those responsive cells not modulated by sleep SWRs might be expected to be modulated by awake SWRs (Tang, et al., 2017). Unfortunately, we did not make reliable recordings of awake SWRs, and thus we could not test this. Furthermore, since modulation was only investigated in neurons with behavioral correlates during mobility, the present work would not have detected cells active during awake immobility when they would be modulated by SWRs (Yu, et al., 2017).

ACC neuronal response correlates

The RC- and SC-responsive neurons, as well as the RR and strategy value responsive neurons, could participate in ACC's role in promoting flexible behavior (Granon and Poucet, 2000; Cardinal et al., 2002; Gisquet-Verrier and Delatour, 2006). ACC lesions impair behavioral flexibility in response to a change in the task rule (de Bruin et al., 1994; Birrell and Brown, 2000; Colacicco et al., 2002; Salazar et al., 2004; Lapiz and Morilak, 2006; review: Hamilton and Brigman, 2015). Moreover, ACC damage particularly impairs performance of tasks requiring shifting from one strategy to another, whether the initial strategy has been learned (Granon and Poucet, 1995; Ragozzino et al., 1999a, b) or is spontaneously used by the animal

(Granon et al., 1994). Interestingly, here RC-responsive cells were significantly more concentrated in A25 (formerly, IL) than SC-responsive cells and other unresponsive cells. This is consistent with lesion studies showing that A25 lesions impair visual cue reversal learning, while sparing initial acquisition of visual cue discriminations (Chudasama and Robbins, 2003; Li and Shao, 1998). For goal-directed behavior (as well as conditioned fear extinction), A25 has been considered to exert a "stop signal", and also to suppress representations of action-outcome contingencies from influencing behavior (Barker, et al., 2014). Indeed, an optimal response to a rule change would require cessation of adherence to the previous rule, consistent with the increased incidence of A25 RC responses here.

Powell & Redish (2016) showed that ACC activity transitions were more related to detecting a task rule change, and initiating new behavioral strategies than to initiation of the animal's behavioral response per se. Here, we extend these results by showing that the communication between HPC and ACC via theta and SWRs is even greater for the detection of task rule changes than for the initiation of new behavioral strategies. Hippocampal contextual information sent to the ACC (and their mutual output structure, the ventral striatum) would be integrated with reward history information to detect task rule changes, perhaps in cooperation with the RR-responsive neurons, as well as A24 (Hyman et al., 2017). The rule shifts would be experienced as the failure to continue to consistently receive reward for a previously successful strategy. This is consistent with the concept of the HPC as a detector of mismatch and novelty, as evoked above. Note that the RC-responsive activity occurred on parts of the maze visited prior to the reward site, thus corresponding to contextual rather than simple post-reward responses. Furthermore, RC responses occurred in a population of neurons with no significant activity correlates with reward rate.

Here, RC and SC associated activity changes in individual neurons were restricted to limited portions of the maze. It is possible that such effects could be masked in population analyses, which average firing over all cells recorded over entire trials. Overall, the population analyses of all neurons showed shifts occurring about 2.5 trials after RC's and about 1 trial before SC's. In contrast, the analyses of individual neurons with behavioral correlates modulated by RC or SC showed averaged responses at about 4 trials after the RC and about 3 trials before the SC's. Since the sampled data sets, calculations and averaging methods are quite different, it is difficult to directly compare the results of the two approaches which suggest that, overall, the population activity responds more rapidly to RC, but that individual neuronal activity "predicts" SCs earlier. Nonetheless, they agree that PFC neurons do anticipate strategy shifts and respond rapidly to rule changes.

In order to learn and perform goal directed choice tasks, one must learn from recent experience concerning the presence or absence of reward and the association of cues and actions. The durations of the behaviorally correlated ACC activity observed here are suited for this since they could be rather brief, or could extend over the order of seconds (cf., Fig. 5). Such activity could serve as buffers for working memory and data processing – not only for cues but also for movements, behavioral sequencing and cue-choice contingencies (Procyk & Goldman-Rakic, 2006; Wang & Hayden, 2017). The ACC could be involved in learning associations of concurrent and sequential events including actions (and not just simple cue-action associations) (Del Arco et al., 2017). These processes could lead to development of reward-based strategy learning. The signals recorded here could hence be considered as instrumental for cerebral processing of the stimulus-action-outcome relations that rules represent.

The RC and SC responses could be informed by modality-specific neocortical inputs relevant to the respective strategies (e.g., visual vs spatial cues for the light-dark vs left-right tasks respectively), as well as by thalamic inputs through cortico-striatal loops (Alexander and Crutcher, 1990; Haber, 2003). The changes in activity levels and formation of assemblies of co-activated ACC neurons can be considered to correspond to functionally active modules that would synchronize with, and activate subsets of downstream basal ganglia neurons to participate in strategy and behavioral choices. Such selection among different rules could also be viewed in terms of neural, attentional and behavioral rechanneling. Indeed, other electrophysiological studies have also reported rodent ACC activity to reflect crucial parameters underlying flexible goal-directed behaviors, such as task-related locomotion (Poucet, 1997; Jung et al., 1998; Fujisawa et al., 2008), reward (Pratt and Mizumori, 2001; Miyazaki et al., 2004), working memory (Baeg et al., 2003), and action-outcome contingencies (Mulder et al., 2003; Kargo et al., 2007). Importantly, the RC-responsive and SC-responsive subpopulations in the ACC are relevant to computational studies implicating rat ACC in action selection (e.g., Hasselmo, 2005; Martinet et al., 2011), and in detecting environmental changes (Caluwaerts et al., 2012). Indeed, our results suggest that the rat ACC is well-positioned to combine contingency change detection and strategy selection functions (Holroyd and McClure, 2015), integrating them during behavior to help build lasting memory traces in conjunction with hippocampal SWRs during sleep.

Materials and Methods

Rats

Five Long-Evans male adult rats (225 to 275 g; from the Centre d'Elevage René Janvier, Le Genest-St-Isle, France; RRID: RGD 18337282) were maintained in clear plastic cages bedded with wood shavings. A 12 h/12 h light/dark cycle was applied and all manipulations took place during the light part of the cycle. The rats were housed in pairs while habituating to the animal facility. They were weighed and handled each workday. Prior to pre-training they were placed in separate cages and 14 g of rat chow were provided daily after the pre-training or recording session in order to maintain body weight at not less than 85% of normal values (as calculated for animals of the same age provided *ad libitum* food and water). The rats were examined daily for their state of health and were fed to satiation at the end of each work week. This level of food deprivation was necessary to motivate performance in the behavioral tasks, and the rats showed neither obvious signs of distress (e.g., excessive or insufficient grooming, hyper- or hypo-activity) nor health problems. The rats were kept in an approved (City of Paris Veterinary Services) animal care facility in accordance with institutional (CNRS Operational Committee for Ethics in the life sciences), national (French Ministère de l'Agriculture, de la Pêche et de l'Alimentation No. 7186) and international (US National Institutes of Health, Helsinki Declaration) guidelines.

Surgery

From the start of the pre-training, rats were placed on the mild food deprivation regime. For at least a week before surgery, rats were habituated to running on the Y-maze by allowing them to forage there for 5-25 minutes daily, with reward available at the end of each arm. Once habituated to the experimental environment (after 7 to 10 days), rats were anesthetized

with intramuscular xylazine (Rompun, 0.1 ml) and intraperitoneal pentobarbital (35 mg per kg of body weight). Xylocaine solution was injected under the scalp. A drive containing seven tetrodes (six for recording, plus one as reference) was implanted on the skull above the right medial PFC (anterior-posterior, 3.5–5 mm; medial-lateral, 0.5–1.5 mm). Each tetrode was placed in a 30-gauge stainless steel tube, and the tubes were assembled together in two adjacent rows. Tetrodes were twisted bundles of polyimide coated nichrome wire (13 µm in diameter, Kanthal, Palm Coast, FL). Microdrives allowed independent adjustment of tetrode depth. After retraction of the dura, the rows of cannulae were implanted parallel to the sagittal sinus so that they targeted the superficial and deep layers of the medial bank of the cortex. A separate microdrive containing three tetrodes targeted the ventral HPC (anterior-posterior, -5.0 mm; medial-lateral, 5.0 mm) since the principal projections to ACC originate there (Thierry, et al., 2000). A screw implanted on the occipital bone above the cerebellum served as the reference electrode. The hippocampal tetrodes were lowered to the CA1 pyramidal layer; the depth was adjusted with the help of LFP signs (flat sharp waves, strong ripple oscillations). After surgery, rats recovered for at least 2 weeks while the tetrodes were lowered to reach A24, A32, and A25, as well as the ventral hippocampal CA1 pyramidal layer. Between sessions, tetrodes were gradually lowered to probe different dorso-ventral levels in the ACC.

Behavioral task

The Y-maze (**Fig. 1A**) was formed by three arms (85 cm long, 8 cm wide with 2 cm high borders) separated by 120 degrees, with a cylindrical barrier at the center. The maze area was surrounded by a cylindrical curtain with no explicit cues on it. The rats were habituated to the maze. The rats had to learn by trial-and-error to adapt their behavior to the current rule in

order to maximize the amount of reward received for a given amount of effort. Rats started all trials from the *departure* arm, and after the central barrier was lowered, a light went on at the end of one of the two goal arms selected in a pseudo-random sequence. The rats then had to select one of the two choice arms and then go to the end for reward. Chocolate milk rewards (30 µl) were delivered for choices adhering to the current rule. A photodetector signaled arrivals at the reward trough and triggered a solenoid to release reward when appropriate. On each trial, the reward was available on only one arm. Initially, the animal had to learn that the reward was always located on the right arm, no matter which arm was lit (*Right rule*). Once the animal entered a reward arm the entrance was blocked and it was not unblocked until the rat had visited the reward trough (rewarded or not). The rats had been trained to return to the departure arm after the outcome of the trial (rewarded or not), and the barrier was raised until the next trial. To avoid extensive overtraining, a different rule was applied (Rule Change: RC) after the rat had clearly acquired the current rule (i.e., performance reached a criterion level of 10 consecutive rewarded trials), or 11 rewards in the previous 12 trials. This stringent criterion was applied because some rats tended to engage other strategies after initial acquisition and this assured that the performance had become stable. The rule change was not explicitly signaled to the rat in any way, and could only be detected by the pattern of unrewarded trials. After acquiring the Right rule, the rat then had to learn to go to the lit arm, regardless of whether it was on the left or on the right (*Light rule).* After that was learned, the rats were successively challenged with a *Left rule* and a *Dark rule* (where the unlit arm was rewarded), and then, in the only rat that succeeded in reaching this level (rat 20), the sequence of task rules was started again from the beginning (*Right*, *Light*, etc.). All rule changes thus qualified as extra-dimensional, engaging spatial or visual cues respectively. No intra-dimensional shifts (reversals) were imposed such as *Right rule* -> *Left*

rule, or *Dark rule -> Light rule*, or their inverses, since these have been reported to not require medial PFC function (Birrell and Brown, 2000).

Rats were trained in daily sessions. Each session consisted of 10 to 60 consecutive trials, stopping when the rat ceased performing. Each session began with the same rule as the final trials of the previous session. Since several sessions were sometimes required to learn certain task rules, there were sessions where no shift in the task rule was imposed. Thus, we will distinguish *shift sessions* (i.e., sessions where a shift in the task rule occurred) from *non-shift sessions*.

Data acquisition

Spike waveforms were filtered between 600 and 6000 Hz, digitized with a Power1401 interface (Cambridge Electronic Design, UK) and time-stamped. For this, 32 samples at 32 kHz (1 ms total) were recorded whenever the signal exceeded a manually set threshold (8 pre-trigger and 24 post-trigger samples). The signal recorded from the same tetrodes was also passed into a low-band filter (between 0.1 and 475 Hz) in order to extract LFPs. The timestamps of behavioral events (e.g., photodetector crossings) were integrated with the spike data online. A video camera was synchronized with the data acquisition software and monitored the consecutive positions of the animal during the experiment. The video tracking data of the instantaneous position of the animal were acquired using MaxTRAQ® (Innovision Systems, Columbiaville, MI) software.

For action potential spike discrimination, the data were processed with a custom Python script for Principal Component Analysis. Then, the Klusters and Klustakwik software (Harris et al., 2000; Hazan et al, 2006; Buzsáki Lab, RRID:SCR_008020) carried out the spike-sorting using the expectation-maximization (EM) algorithm (Celeux and Govaert, 1992). The classifications

were then revised manually. Parameters of the EM algorithm were intentionally chosen to extract a high number of clusters (typically from 15 to 60 from the multiple tetrode array). Then those clusters likely to belong to the same unit were merged in manual online analyses.

At the end of experiments, a small electrolytic lesion was made with cathodal current through one wire of each tetrode (25 µA for 10 s). Rats were euthanized with sodium pentobarbital and perfused with saline then formaldehyde solution (4% v/v). Lesion sites were detected in Nissl stained 60 µm sections. Positions of recorded neurons were determined with respect to the lesion site by interpolating the measured descent of the electrodes. Electrode sites were reconstructed in three dimensions with the Neurolucida® system.

Identification of putative pyramidal cells and interneurons

Putative interneurons and pyramidal cells in ACC were distinguished according to spike width with the method from Barthó et al. (2004). The distribution of spike widths for all ACC neurons was recorded and was bimodal (see Supplementary Figure 6 in Benchenane et al, 2010). Cells with spike widths inferior to 0.3 ms were classified as putative interneurons, whereas those superior to 0.35 ms were considered as putative pyramidal cells.

Behavioral analyses

Each trial was characterized by parameters including the current task rule (right, light, left or dark), the position of the light (right or left), the arm chosen by the animal (right or left), and the outcome of the trial (rewarded or not).

Two principal events were the bases for the analyses of behavioral and neurophysiological data from each trial: *Start* and *Arrival*. The trial Start time was extracted from the video tracked position data and was defined as the onset of initial forward acceleration in the

departure arm. The Arrival event was defined as the instant when the animal first blocked the photodetector at the reward reservoir of the selected arm.

Next, to systematically categorize the animal's behavior in terms of the "strategy" it followed, the behavioral data from all of the sessions were concatenated for each animal. A custom script scanned these files and extracted series of trials where the animal's behavioral strategy was consistently stable ("blocks of trials"). The strategies considered in the behavioral analyses included: 1) Going to the right arm regardless of the visual cue ("Right" strategy); 2) "Left"; 3) Going to the lit arm, regardless of whether it was to the right or left ("Light" strategy); 4) "Dark"; and 5) Spatial "Alternation" (not rewarded) (see Figure 1A). Trials not included in such blocks are considered as a sixth "Undefined" strategy, while the latter five are termed "Defined". We systematically searched for higher order strategies (e.g., left-left-right-right-left-left, etc.) but found insufficient evidence for them. Note also that a single trial could be compliant with two strategies (for instance Left and Dark) or even three strategies (for instance Left, Dark and Alternation), but the randomization of the light positions assured that blocks of 4 or more trials could not comply with more than one strategy.

To determine the minimum number of strictly compliant trials to be considered a strategy block, we performed simulations consisting of 1.6×10^7 random runs. Each of these runs included the 3322 trials of the data set, with the same sequence of rules and lit arm positions, but the arm choices were randomly generated. These simulations showed that for blocks of length 6 or greater the only 5.4% of random arm choices would belong to such a block for each single strategy. Thus, 95% of random moves would not belong to such a block. Hence, we considered a block to be strictly compliant to one of the five Defined strategies if it contained

a continuous sequence of at least 6 trials such that all the rat's choices were compliant with that strategy. This is the same, or even more stringent criterion than used in this type of study (e.g., Kaefer, et al., 2020; Lapiz-Bluhm, et al., 2009, Tait, et al., 2017).

Since rats occasionally made non-compliant choices within long runs of compliant trials, we also considered as strategy blocks those sequences containing at least 9 trials with only one non-compliant (NC) trial. For example, if the animal's choices were compliant (C) with a strategy for 6 trials, NC for 1 trial, and then C again for 2 trials, the whole block of 9 trials is considered as a "leniently compliant" strategy block. Following the same principle, we included blocks of CCCCC-NC-CCC (and its inverse), or CCCC-NC-CCCC. Longer sequences might include several NC trials providing that these rules were respected. In simulations as described above, the probability that a random choice would belong to a leniently compliant block for each of the strategies was only 8.6%; 91.4% of random moves would not belong to such a block.

	Simulation results (% trials in blocks)				Actual values	
Strategy	Mean	Median	Lower bound (2.5%)	Upper bound (97.5%)	%	n
Left	8.6	8.6	5.8	11.6	27.0	896
Right	8.6	8.6	5.8	11.6	22.0	730
Light	8.6	8.6	5.8	11.6	23.4	776
Dark	8.6	8.6	5.8	11.6	6.9	229
Alternation	8.6	8.6	5.8	11.6	13.8	459
Defined	38.9	38.9	34.2	43.7	78.2	2599
Undefined	61.1	61.1	56.3	65.8	21.8	723

Table 1. Incidences of trials in leniently compliant blocks of length >5. At left are results for the simulations described in the text. Actual values are from animal data. Note that blocks

compliant with different specific strategies could partially overlap, so that the sum of trials for the five strategies do not equal that of Defined trials.

To gauge the risk that sequences detected with these criteria could have occurred stochastically, the incidence of leniently compliant sequences was computed in simulations of 1.6x107 random runs. Each random run included the 3322 trials of the data set, with the same sequence of rules and lit arm positions, but the arm choices were randomly generated. These simulations yielded values of the mean, median and 95% confidence intervals for the percentages of trials that would have stochastically appeared in leniently compliant blocks of length greater than or equal to 6 trials (Table 1). For all of the strategies, the incidence of experimental trials in blocks compliant to the respective strategies are outside of the confidence interval limits, inconsistent with fortuitous compliance by means of stochastic choices. Similarly, the incidence of trials following any defined strategy (78.2%) greatly exceeded the confidence interval upper bound (43.7%).

To verify that the results of these analyses were not dominated by very long sequences that were clearly not stochastic, we further examined smaller blocks of 11 trials or less (which composed 99.1% of the trials in randomly generated choices). Over these respective block lengths, 40% of the actual values fell outside the $p<.05$ confidence limits, again providing evidence against stochastic choices.

The performance of the animals prior to initial rule acquisition was also inconsistent with stochastic behavior. Four of the five animals reached criterion for the R rule in the very first training session (cf., values in Results “Behavior”), corresponding to fewer trials than expected by chance. During the initial acquisitions, the incidence of occurrence of sequences compliant

with a rule currently not being rewarded also extended beyond the lower or upper bounds of the confidence intervals of the randomized data simulation.

Strategy changes (SC) are defined as two consecutive blocks where the rat followed different strategies (e.g., the animal performs a series of 10 trials following the *Light* strategy, always going to the lit arm, and then performs a series of 6 trials following the *Alternation* strategy). We define *extradimensional shifts* as SC between spatial and visual strategies (or vice versa), such as Left then Dark strategy.

Experimental design and statistical analyses

Neuronal activity was recorded as the rats performed in a completely automated maze, or during sleep immediately preceding and following this. Each neuron served as its own control and activity levels were compared before vs. after events with analyses and statistical tests as described in the following sections. The tests include bootstrap Monte Carlo analyses, ANOVA, t-tests, Kruskal-Wallis, Mann-Whitney, $\chi 2$, and Rayleigh Z tests. The specific applications of these tests are explained below and in the Results section. (Insufficient sampling precluded analyses of REM sleep).

Activity transition analyses at the population level

An unbiased automatic decomposition of the trials in each session into blocks determined whether the onsets of transitions in neural activity are most closely temporally associated to behavioral (SC) or task (RC) changes. Similar to Powell & Redish (2016), for each trial in a given session, we computed a population vector from the firing rate of each given neuron averaged over the entire trial (from *Start* until *Arrival*) from all PFC cells recorded simultaneously. Each population vector was z-scored in order to give the same weight to cells with low and high

firing rates. We then computed the correlation matrix where each element E_{ij} is the correlation between population vectors of trials *i* and *j* for all trials of the session.

When displayed graphically, blocks of high values in the correlation matrix indicate series of trials with correlated population activity. We performed an automated search for such blocks by detecting the best decomposition of the matrix into blocks, i.e., finding the optimal partition of blocks of trials along the diagonal of the matrix, so that eliminating elements outside the blocks minimizes the amount of information lost. Information loss was measured as the percentage by which the sum of absolute values of the elements in the matrix is reduced after exclusion of the elements outside the blocks. To prevent the trivial decomposition (i.e., a single block encompassing everything) from being systematically considered as the one that minimizes the loss (since it has no loss of information), we rejected it if at least one decomposition into 2 or more blocks led to a loss of less than 1/3 of the information (i.e., if the information present within the blocks was superior or equal to 2/3 of total information within the matrix). Otherwise, no decomposition was retained, which was interpreted as no substantial change in population activity having occurred during that session.

In order to find the best decomposition into two or more blocks, we tested all possible decompositions of each matrix into blocks (with each block including at least 3 trials and at most 2/3 of the trials of the session). Then for each decomposition, we projected the original matrix onto the considered blocks (which are concentrated around the diagonal of the matrix), hence setting to zero the elements outside the blocks. This tests the hypothesis that the correlations of activity between blocks are negligible, as if the population activity has sharp state changes between blocks of trials. Then, for each decomposition, we measured how much information was lost, and retained the one which minimized this amount.

We developed this approach to decompose the population cross-correlation matrix rather than use k-means clustering employed by Powell and Redish (2016) because it was not clear that the data satisfied the requirements of k-means. Firstly, k-means is based on the principle of spherical clusters that can be separated in such a manner that the means approach the centers of the respective clusters. Furthermore, the clusters should have comparable sizes. But, here, we expected sharp transitions between series of trials with respectively homogeneous firing rates, and the transitions could occur early or late in the session, creating groups of unequal size. Furthermore, it was not clear how many clusters to preset in the k-means algorithm since RC and SC could each occur one or more times in the same session, and the recorded populations only had limited numbers of neurons, leading some of these to not respond to some RCs or SCs. There was also concern that the algorithm would stop at local minima, rather than at more optimal solutions.

Estimating timing of population activity transitions relative to RC or SC events. In order to quantify the possible relationship between an activity transition detected on a particular trial *i* with a RC or SC event occurring on trial *j*, the difference between *i* and *j* was taken if there was only one such event. If there were two or more events of the same or different types (i.e., RC or SC), the one occurring the fewer number of trials from the activity transition was selected. Possible ambiguities could occur when a RC was rapidly followed by a SC. As a control, analyses were first performed only on sessions where only one such event occurred. Activity changes associated with RC were only considered for the first trial after a conflict between the reward contingency and the previous rule was detectable (since individual behavioral choices can be consistent with more than one rule).

Event-locked analyses for activity transitions. Activity in time windows bracketing salient trial events were compared in trials prior to and after RC and SC. The four periods included: *pre-start*, in the time window [*Start* – 2.5 sec; *Start*]; *post-start* [*Start*; *Start* + 2.5 sec]; *pre-arrival* [*Arrival* – 2.5 sec; *Arrival*]; and *post-arrival* [*Arrival*; *Arrival* + 1.25 sec]. The post-arrival period was restricted to 1.25 s in order to restrict to the period when the rats were immobile irrespective of whether the trial was rewarded or not. This avoided possible confounds with different motor behaviors in rewarded and non-rewarded trials. In 96.4% of all trials recorded in the five animals (including rewarded and unrewarded trials), the rats remained at the reward site for at least 1.25 s.

The Wilcoxon-Mann-Whitney test examined differences in neuronal firing rate around RC and SC events. This was tested for each of the four trial event windows (with Bonferroni corrections for multiple comparisons).

To confirm and expand upon these analyses, a bootstrap Monte Carlo analysis was also performed for each cell, again with separate tests for each of the four periods. In all cases, the firing rates were compared between trial blocks before and after the RC or SC. Then the activity of these trials was randomly re-attributed to two surrogate sets with the same numbers of trials, and the difference was taken. This shuffling procedure was repeated 5000 times and yielded a distribution from which only those differences in firing rate with p-values falling within the upper or lower 2.5% of the shuffled distributions were considered as significant transitions in neural activity. To quantify the incidence of spontaneous changes of firing rate of PFC neurons, in sessions where the rat consistently performed a single strategy, the first half of the session was compared to the second half with the same bootstrap Monte Carlo analysis. Since the latter tests (reported in Results) showed that a non-negligible number

of significant transitions in firing rate occurred in the sessions with no strategy or rule changes (or other apparent behavioral differences), further complementary analyses were performed.

Estimating timing of individual neuron activity transitions relative to RC or SC events.

The preceding population analyses group data from different maze positions. It presumes that activity shifts will only take place at RCs or SCs. Another analysis without this *a priori* assumption examined each neuron's activity to determine on which trial (if any) there was the maximal significant difference between firing rates in the block of trials before and after that trial. The nonparametric analysis developed by Fujisawa et al. (2008) constructs bootstrap confidence intervals of positional spike density functions. The null hypothesis posits no difference in firing between trials before and after the event of interest (here, RC and SC). Sessions were divided into all possible pairs of sets of contiguous trials, with a minimum of four trials in each set. Thus, if a session had 50 trials, trials 1 to 4 were compared with 5 to 50, 1 to 5 vs 6 to 50, 1 to 6 vs 7 to 50, [...], through trials 1 to 46 vs 47 to 50. Spike density functions were determined for each of the pairs of sets of trials.

Here data could only be analyzed for positions that the rat occupied on every trial of the session, and other data (particularly on the start arm, where the rats did not always go to the end) were excluded. Data were divided into spatial bins of length 1.75 cm along the linearized trajectory from starting point to reward site (each corresponding to 2.5 pixels in the video image; x-axis of 1st and 2nd columns of **Fig. 5**).

For each pair of trial blocks, a bootstrap Monte Carlo analysis randomly redistributed all of the trials in the session to surrogate pairs of sets with the same respective numbers of trials as the original sets. For each surrogate set a spike rate function across the spatial bins was taken as a set of points on the maze where the spikes occurred and this was transformed to a spike

count function with a Gaussian kernel whose bandwidth is 2 bins (= 3.5 cm). Dividing the spike count functions by the time spent in each position over all of the trials for the respective parts of the session yielded the spike rate functions. The pre- versus post-transition difference was then computed for each block of trials.

To determine the statistical significance of these rate differences, the distribution of the rate difference statistics was estimated by randomly reassigning the data from each trial into two surrogate groups 5000 times. For each of the spatial bins, a pointwise confidence band with a p-value of 0.05 was assigned to the 125 greatest positive and negative differences respectively (a two-way test). Since the pointwise confidence band is computed at multiple points, the confidence level must be corrected for these multiple comparisons. This is achieved by computing the global confidence band: first, we calculate the “global confidence level”, taken as the percentage of resampled rate differences whose values are not all inside the area limited by the pointwise band. The procedure is repeated to build the pointwise confidence band by gradually decreasing the value of the pointwise confidence limit until the global limit is equal to 0.05. Then, a maze zone is considered to have significant difference in pre- versus post-transition firing only if this difference crosses both the global and pointwise bands, but the extent of this zone is determined only by the points where it lies beyond the pointwise band (Fujisawa et al, 2008; also see Catanese et al., 2012).

To control for possible spatial selectivity in the ACC responses (Euston et al., 2007), analyses were repeated separately using only data from either left or right choice trials in sessions where sample sizes were large enough (see also Lindsay, et al., 2018). This showed that 43 (73%; or 34% of the total of 127 neurons) retained significant SDF differences when either left

or right choice trials were analyzed alone, and thus their transition related activity changes are not confounded with location-selective firing.

Multiple regression analysis of single-unit activity with strategy value and reward rate. In order to control for potential confounds in putative RC- and SC-responses, we constructed several regressors, and then tested for correlation with single cell activity with a multiple regression analysis.

We first built a simple and parsimonious (without any free parameters) regressor mimicking progressively learned strategy values in computational models (Dollé et al., 2018). This consisted of the estimated value of the three observed strategies: V_{Right}, V_{Light} and $V_{Alternation}$ (V_{Left} being equal to $-V_{Right}$, and $V_{Dark} = -V_{Light}$). These values were initialized at 0 at the beginning of each new session. After each trial, the value of a strategy X was increased by 1 if the animal's behavior was consistent with it and got rewarded, decreased by 1 if it was inconsistent and got rewarded. Furthermore, the value was decreased by 1 if it was consistent but unrewarded, and increased by 1 if it was both inconsistent and unrewarded. For example, if at trial N the animal went to the left arm which was lit, and received a reward, then V_{Light} was increased by 1 while V_{Right} was decreased by 1.

We then built a second regressor representing the reward rate (RR), computed over a 6-trial sliding window. We finally also included regressors representing the lit arm on the current trial (L; 0 for left or for 1 right), arm choice (C; 0 for left or 1 for right), and reward (R; 0 or 1).

The spike rate y(t) in trial t was then analyzed using the following multiple linear regression model:

$$y(t) = \rho_0 + \rho_1 V_{Right}(t) + \rho_2 V_{Light}(t) + \rho_3 V_{Alternation}(t) + \rho_4 RR(t) + \rho_5 L(t) + \rho_6 C(t) + \rho_7 R(t)$$

where ρ_i ($i \in \{1..7\}$) are the regression coefficients.

Following Seo and Lee (2009) and Khamassi et al. (2015), we applied a restrictive significance threshold permitting only 5% significance after permuting the trial order 10,000 times (bootstrap method) so that this regression model would not suffer from violation of independence between regressors.

LFP analyses

Each of the mid-ventral hippocampal tetrodes was electrically connected in a single-electrode configuration (all channels shorted together) and used for LFP recordings. The screw implanted in the occipital bone above the cerebellum was the reference. Local field potentials were sampled and stored at 2 kHz. Theta was filtered at 5-10 Hz. Theta modulation of cell activity was detected with the Rayleigh test with a criterion of $p<.05$. Detecting SWRs and measuring SWR modulation of cell activity employed the methods of Peyrache, et al. (2009). For the latter, PETH's plotted cell activity relative to ripples with 100 ms bins. A t-test compared the central bin to the baseline value.

For information about and access to the recording data base, see https://crcns.org/data-sets/pfc/pfc-6/about-pfc-6

Acknowledgments

We thank Prof. J.-M. Deniau, Drs. A.-M. Thierry, Y. Gioanni, M.B. Zugaro, R. Todorova, and E. Cerasti for valuable discussions, S. Doutremer for histology, F. Maloumian for help with figures, N. Quenech'du for help with the 3-D anatomical reconstructions, Y. Dupraz for mechanical engineering and Prof. A. Berthoz for support throughout the project.

Competing interests: None of the authors have any financial or non-financial competing interests.

Data availability: The electrophysiological data are available at Adrien Peyrache, Mehdi Khamassi, Karim Benchenane, Sidney I Wiener, Francesco Battaglia (2018); Activity of neurons in rat medial prefrontal cortex during learning and sleep. CRCNS.org http://dx.doi.org/10.6080/K0KH0KH5

Reference list

Adhikari A, Topiwala MA, Gordon JA. (2011) Single units in the medial prefrontal cortex with anxiety-related firing patterns are preferentially influenced by ventral hippocampal activity. Neuron 71(5):898-910. doi: 10.1016/j.neuron.2011.07.027.

Alexander GE, Crutcher MD. (1990) Functional architecture of basal ganglia circuits: neural substrates of parallel processing. TINS 13(7):266-271.

Arikuni T, Sako H, Murata A (1994) Ipsilateral connections of the anterior cingulate cortex with the frontal and medial temporal cortices in the macaque monkey. Neurosci Res 21(1):19-39. https://doi.org/10.1016/0168-0102(94)90065-5.

Baeg EH, Kim YB, Huh K, Mook-Jung I, Kim HT, Jung MW. (2003) Dynamics of population code for working memory in the prefrontal cortex. Neuron 40(1):177-188.

Barker JM, Taylor JR, Chandler LJ. (2014) A unifying model of the role of the infralimbic cortex in extinction and habits. Learn Mem 21(9): 441–448. doi: 10.1101/lm.035501.114.

Barthó P, Hirase H, Monconduit L, Zugaro MB, Harris KD and Buzsáki G. (2004) Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. J Neurophysiol 92:600–608.

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

Berg EA. (1948) A simple objective test for measuring flexibility in thinking. J Gen Psychol 39:15-22.

Birrell JM, Brown VJ. (2000) Medial frontal cortex mediates perceptual attentional set shifting in the rat. J Neurosci 20(11):4320–4324.

de Bruin, JPC, Sanchez-Santed F, Heinsbroek RPW, Donker A, Postmes P. (1994) A behavioural analysis of rats with damage to the medial prefrontal cortex using Morris water-maze: Evidence for behavioural flexibility, but not for impaired spatial navigation. Brain Res 652:323-333.

Bunge SA. (2004) How we use rules to select actions: a review of evidence from cognitive neuroscience. Cogn Aff Behav Neurosci 4(4):564-579.

Caluwaerts K, Staffa M, N'Guyen S, Grand C, Dollé L, Favre-Félix A, Girard B, Khamassi M. (2012) A biologically inspired meta-control navigation system for the Psikharpax rat robot. Bioinsp Biomim *7*(2):025009.

Cardinal RN, Parkinson JA, Hall J, Everitt BJ. (2002) Emotion and motivation: The role of the amygdala, ventral striatum and prefrontal cortex. Neurosci Biobehav Rev 26(3):321-352.

Catanese J, Cerasti E, Zugaro MB, Viggiano A, Wiener SI. (2012) Dynamics of decision-related activity in hippocampus. Hippocampus 22(9) 1901-1911.

Celeux G, Govaert G. (1992) A classification EM algorithm for clustering and two stochastic versions. Comput Stat Data Anal 14(3):315-332.

Chudasama Y, Robbins TW. (2003) Dissociable contributions of the orbitofrontal and infralimbic cortex to Pavlovian autoshaping and discrimination reversal learning: further evidence for the functional heterogeneity of the rodent frontal cortex. J Neurosci 23:8771–8780.

Chiang F-K, Wallis JD, Rich EL. (2022) Cognitive strategies shift information from single neurons to populations in prefrontal cortex. Neuron 110(4):709-721. doi:10.1016/j.neuron.2021.11.021.

Colacicco G, Welzl H, Lipp H, Würbel H. (2002) Attentional set-shifting in mice: modification of a rat paradigm, and evidence for strain-dependent variation. Behav Brain Res 132:95-102.

Del Arco A, Park J, Wood J, Kim Y, Moghaddam B. (2017) Adaptive encoding of outcome prediction by prefrontal cortex ensembles supports behavioral flexibility. J Neurosci 37(35):8363-8373. doi: 10.1523/JNEUROSCI.0450-17.2017.

D'Esposito M, Detre JA, Alsop DC, Shin RK, Atlas S, Grossman M. (1995) The neural basis of the central executive system of working memory. Nature 378(6554):279-281.

Dias R, Robbins TW, Roberts AC. (1996) Dissociation in prefrontal cortex of affective and attentional shifts. Nature 380(6569):69-72.

Dollé L, Chavarriaga R, Guillot A, Khamassi M. (2018) Interactions between spatial strategies producing generalization gradient and blocking: a computational approach. PLoS Comput Biol 14(4):e1006092.

Drewe EA. (1974) The effect of type and area of brain lesion on Wisconsin card sorting test performance. Cortex 10:159-170.

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

Eschenko O, Ramadan W, Mölle M, Born J, Sara SJ. (2008) Sustained increase in hippocampal sharp-wave ripple activity during slow-wave sleep after learning. Learn Mem. 15(4):222-8. doi: 10.1101/lm.726008.

Frankland PW, Bontempi B. (2005) The organization of recent and remote memories. Nature Rev Neurosci 6(2):119-130.

Fujisawa S, Amarasingham A, Harrison MT, Buzsáki G. (2008) Behavior-dependent

short-term assembly dynamics in the medial prefrontal cortex. Nat Neurosci 11:823-833.

Gallistel CR, Fairhurst S, Balsam P. (2004) The learning curve: Implications of a quantitative analysis. PNAS 101:13124-13131.

Genovesio A, Brasted PJ, Mitz AR, Wise SP. (2005) Prefrontal cortex activity related to abstract response strategies. Neuron 47:307-320.

Girardeau G, Benchenane K, Wiener SI, Buzsáki, Zugaro MB. (2009) Selective suppression of hippocampal ripples impairs spatial memory. Nature Neurosci 12(10):1222-1223.

Girardeau G, Zugaro M. (2011) Hippocampal ripples and memory consolidation. Curr Opin Neurobiol 21(3):452-459.

Gisquet-Verrier P, Delatour B. (2006) The role of the rat prelimbic/infralimbic cortex in working memory: not involved in the short-term maintenance but in monitoring and processing functions. Neurosci 141(2):585-596.

Goldman-Rakic PS. (1987) Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In: Higher functions of the brain: The nervous system; Handbook of Physiology. Volume V (Plum F, ed), pp 373–417. Bethesda, MD: American Physiological Society.

Goldman-Rakic PS. (1995) Cellular basis of working memory. Neuron 14(3):477-485.

Granon S, Vidal C, Thinus-Blanc C, Changeux JP, Poucet B. (1994) Working memory, response selection, and effortful processing in rats with medial prefrontal lesions. Behav Neurosci 108(5):883-891.

Granon S, Poucet B. (1995) Medial prefrontal lesions in the rat and spatial navigation: evidence for impaired planning. Behav Neurosci 109(3):474-484.

Granon S, Poucet B. (2000) Involvement of the rat prefrontal cortex in cognitive functions: A central role for the prelimbic area. Psychobiol 28(2):229-237.

Grant DA, Jones OR, Tallantis B. (1949) The relative difficulty of the number, and color concepts of a Weigle-type problem. J Exp Psychol 39:552-557.

Gray JA. (1982) The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system. New York, NY, US: Clarendon Press/Oxford University Press.

Haber SN. (2003) The primate basal ganglia: parallel and integrative networks. J Chem Neuroanat 26(4):317-330.

Hamilton DA, Brigman JL. (2015) Behavioral flexibility in rats and mice: contributions of distinct frontocortical regions. Genes, Brain Behav 14:4–21

Harris KD, Henze DA, Csicsvari J, Hirase H, Buzsáki G. (2000) Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. J Neurophysiol 84(1):401-414.

Hasselmo ME. (2005) A model of prefrontal cortical mechanisms for goal-directed behavior. J Cogn Neurosci 17(7):1115-1129.

Hazan L, Zugaro MB, Buzsáki G. (2006) Klusters, NeuroScope, NDManager: a free software suite for neurophysiological data processing and visualization. J Neurosci Meth 155:207-216.

Holroyd CB, McClure SM. (2015) Hierarchical control over effortful behavior by rodent medial frontal cortex: A computational model. Psychol Rev 122(1):54-83.

Hyman JM, Hasselmo ME, Seamans JK. (2011) What is the functional relevance of prefrontal cortex entrainment to hippocampal theta rhythms? Front Neurosci 5:24.

Hyman JM, Holroyd CB, Seamans JK. (2017) A novel neural prediction error found in anterior cingulate cortex ensembles. Neuron 95(2):447-456.e3. doi: 10.1016/j.neuron.2017.06.021.

Hyman JM, Zilli EA, Paley, AM, Hasselmo ME. (2005). Medial prefrontal cortex cells show dynamic modulation with the hippocampal theta rhythm dependent on behavior. Hippocampus 15, 739–749.

Jadhav SP, Rothschild G, Roumis DK, Frank LM. (2016) Coordinated excitation and inhibition of pefrontal ensembles during awake hippocampal sharp-wave ripple events. Neuron 90(1):113-27. doi: 10.1016/j.neuron.2016.02.010.

Jay TM, Witter MP. (1991) Distribution of hippocampal CA1 and subicular efferents in the prefrontal cortex of the rat studied by means of anterograde transport of Phaseolus vulgaris-leucoagglutinin. J Comp Neurol 313(4):574-86. doi: 10.1002/cne.903130404.

Jones MW, Wilson MA. (2005) Theta rhythms coordinate hippocampal–prefrontal interactions in a spatial memory task. PLoS Biology 3(12):e402.

Jung MW, Qin Y, Lee D, McNaughton BL. (1998) Firing characteristics of deep layer neurons in prefrontal cortex in rats performing spatial working memory tasks. Cereb Cortex 8:437-450.

Kaefer K, Nardin M, Blahna K, Csicsvari J. (2020) Replay of behavioral sequences in the medial prefrontal cortex during rule switching. Neuron 106(1):154-165.e6. doi: 10.1016/j.neuron.2020.01.015

Kafkas A, Montaldi D. (2018) How do memory systems detect and respond to novelty? Neurosci Lett 680:60-68. doi: 10.1016/j.neulet.2018.01.053.

Kargo WJ, Szatmary B, Nitz DA. (2007) Adaptation of prefrontal cortical firing patterns and their fidelity to changes in action-reward contingencies. J Neurosci 27(13):3548-3459.

Karlsson MP, Tervo DG, Karpova AY. (2012) Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. Science 338(6103):135-9. doi: 10.1126/science.1226518.

Khamassi M, Humphries MD. (2012) Integrating cortico-limbic-basal ganglia architectures for learning model-based and model-free navigation strategies. Front Behav Neurosci 6:79. doi:10.3389/fnbeh.2012.00079.

Khamassi M, Quilodran R, Enel P, Dominey PF, Procyk E (2015). Behavioral regulation and the modulation of information coding in the lateral prefrontal and cingulate cortex. Cereb Cortex 25(9):3197-3218.

Killcross AS, Coutureau E. (2003) Coordination of actions and habits in the medial prefrontal cortex of rats. Cereb Cortex 13(4):400-408.

Kumaran D, Maguire EA. (2006) An unexpected sequence of events: mismatch detection in the human hippocampus. PLoS Biol. 4(12):e424. doi: 10.1371/journal.pbio.0040424.

Lansink CS, Goltstein PM, Lankelma JV, McNaughton BL, Pennartz CM. (2009) Hippocampus leads ventral striatum in replay of place-reward information. PLoS Biol 7(8):e1000173.

Lapiz MD, Morilak DA. (2006) Noradrenergic modulation of cognitive function in rat medial prefrontal cortex as measured by attentional set shifting capability. Neurosci 137(3):1039-1049.

Lapiz-Bluhm MD, Soto-Piña AE, Hensler JG, Morilak DA. (2009) Chronic intermittent cold stress and serotonin depletion induce deficits of reversal learning in an attentional set-shifting test in rats. Psychopharmacol (Berl) 202(1-3):329-41. doi: 10.1007/s00213-008-1224-6.

Lenartowicz A, McIntosh AR. (2005) The role of anterior cingulate cortex in working memory is shaped by functional connectivity. J Cogn Neurosci. 17(7):1026-42. doi: 10.1162/0898929054475127.

Li L, Shao J. (1998) Restricted lesions to ventral prefrontal subareas block reversal learning but not visual discrimination learning in rats. Physiol Behav 65:371–379.

Lindsay AJ, Caracheo BF, Grewal JJS, Leibovitz D, Seamans JK. (2018) How much does movement and location encoding impact prefrontal cortex activity? An algorithmic decoding approach in freely moving rats. eNeuro 5(2). pii: ENEURO.0023-18.2018. doi: 10.1523/ENEURO.0023-18.2018

Malagon-Vina H, Ciocchi S, Passecker J, Dorffner G, Klausberger T. (2018) Fluid network dynamics in the prefrontal cortex during multiple strategy switching. Nat Commun 9(1):309.

Malenka RC, Nestler EJ, Hyman SE (2009). "Chapter 13: Higher Cognitive Function and Behavioral Control". In Sydor A, Brown RY (eds.). Molecular Neuropharmacology: A Foundation for Clinical Neuroscience (2nd ed.). New York: McGraw-Hill Medical. pp. 313–321. ISBN 978-0-07-148127-4.

Mansouri FA, Matsumoto K, Tanaka K. (2006) Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin Card Sorting Test analog. J Neurosci 26(10):274556.

Martinet LE, Sheynikhovich D, Benchenane K, Arleo A. (2011) Spatial learning and action planning in a prefrontal cortical network model. PLoS Comp Biol 7(5):e1002045.

Michon F, Sun JJ, Kim CY, Ciliberti D, Kloosterman F. Post-learning hippocampal replay selectively reinforces spatial memory for highly rewarded locations. Curr Biol. 2019;29(9):1436-1444.e5. doi:10.1016/j.cub.2019.03.048

Milner B. (1963) Effects of different brain lesions on card sorting. Arch Neurol 9:90-100.

Miyazaki K, Miyazaki KW, Matsumoto G. (2004) Different representation of forthcoming reward in nucleus accumbens and medial prefrontal cortex. Neuroreport 15(4):721-726.

Mulder AB, Nordquist RE, Orgut O, Pennartz CM. (2003) Learning-related changes in response patterns of prefrontal neurons during instrumental conditioning. Behav Brain Res 146(12):77-88.

O'Keefe J. (1976) Place units in the hippocampus of the freely moving rat. Exp Neurol 51(1):78-109. doi.org/10.1016/0014-4886(76)90055-8.

O'Neill PK, Gordon JA, Sigurdsson T. (2013) Theta oscillations in the medial prefrontal cortex are modulated by spatial working memory and synchronize with the hippocampus through its ventral subregion. J Neurosci 33(35):14211-14224. doi:10.1523/JNEUROSCI.2378-13.2013

Peyrache A, Khamassi M, Benchenane K, Wiener SI, Battaglia FP. (2009) Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. Nature Neurosci 12(7): 919-926.

Poucet B. (1997) Searching for spatial unit firing in the prelimbic area of the rat medial prefrontal cortex. Behav Brain Res 84:151-159.

Powell NJ, Redish AD. (2016) Representational changes of latent strategies in rat medial prefrontal cortex precede changes in behaviour. Nat Commun 7:12830. doi: 10.1038/ncomms12830.

Pratt WE, Mizumori SJ. (2001) Neurons in rat medial prefrontal cortex show anticipatory rate changes to predictable differential rewards in a spatial memory task. Behav Brain Res 123(2):165-183.

Procyk E, Goldman-Rakic PS. (2006) Modulation of dorsolateral prefrontal delay activity during self-organized behavior. J Neurosci 26(44):11313-23.

Ragozzino ME, Detrick S, Kesner RP. (1999a) Involvement of the prelimbic-infralimbic areas of

the rodent prefrontal cortex in behavioral flexibility for place and response learning. J Neurosci 19(11):4585-4594.

Ragozzino ME, Wilcox C, Raso M, Kesner RP. (1999b) Involvement of rodent prefrontal cortex subregions in strategy switching. Behav Neurosci 113(1):32-41.

Ramadan W, Eschenko O, Sara SJ. (2009) Hippocampal sharp wave/ripples during sleep for consolidation of associative memory. PLoS One. 4(8):e6697. doi: 10.1371/journal.pone.0006697.

Raudies F, Hasselmo ME. (2014) A model of hippocampal spiking responses to items during learning of a context-dependent task. Front Syst Neurosci 8:178.

Rich EL, Shapiro M. (2009) Rat prefrontal cortical neurons selectively code strategy switches. J Neurosci 29(22):7208-7219.

Robbins TW. (2007) Shifting and stopping: fronto-striatal substrates, neurochemical modulation and clinical implications. Phil Trans Royal Soc Lond B: Biol Sci 362(1481):917-932.

Rothschild G, Eban E, Frank LM. (2017) A cortical–hippocampal–cortical loop of information processing during memory consolidation. Nat Neurosci 20(2):251-259.

Salazar RF, White W, Lacroix L, Feldon J, White IM. (2004) NMDA lesions in the medial prefrontal cortex impair the ability to inhibit responses during reversal of a simple spatial discrimination. Behav Brain Res 152(2):413-424.

Seo H, Lee D. (2009) Behavioral and neural changes after gains and losses of conditioned reinforcers. J Neurosci 29(11):3627-3641.

Shallice T. (1988) From Neuropsychology to Mental Structure. New York, NY: Cambridge University Press.

Shibata R, Mulder AB, Trullier O, Wiener SI. (2001) Position sensitivity in phasically discharging nucleus accumbens neurons of rats alternating between tasks requiring complementary types of spatial cues. Neurosci 108:391-411.

Siapas AG, Lubenov EV, Wilson MA. (2005) Prefrontal phase locking to hippocampal theta oscillations. Neuron 46:141–151.

Tait DS, Phillips JM, Blackwell AD, Brown VJ. (2017) Effects of lesions of the subthalamic nucleus/zona incerta area and dorsomedial striatum on attentional set-shifting in the rat. Neuroscience. 345:287-296. doi: 10.1016/j.neuroscience.2016.08.008.

Tanaka S (2007) Stable and unstable activation of the prefrontal cortex with dopaminergic modulation. Pps. 235-236 in: Tseng KY, Atzori M (eds) *Monoaminergic Modulation of Cortical Excitability*. Springer, Boston, MA.

Tang W, Shin JD, Frank LM, Jadhav SP. (2017) Hippocampal-prefrontal reactivation during learning is stronger in awake compared with sleep states. J Neurosci 37(49):11789-11805.

Tatsuno M, Malek S, Kalvi L, Ponce-Alvarez A, Ali K, Euston DR, Grün S, McNaughton BL. (2020) Memory reactivation in rat medial prefrontal cortex occurs in a subtype of cortical UP state during slow-wave sleep. Philos Trans R Soc Lond B Biol Sci. 375(1799):20190227. doi: 10.1098/rstb.2019.0227.

1234 1. Thierry AM, Gioanni Y, Dégénétais E, Glowinski J (2000) Hippocampo-prefrontal cortex pathway: anatomical and electrophysiological characteristics. Hippocampus 10:411–419. doi.org/10.1002/1098-1063(2000)10:4<411::AID-HIPO7>3.0.CO;2-A

Trouche S, Koren V, Doig NM, Ellender TJ, El-Gaby M, Lopes Dos-Santos V, et al. (2019) A hippocampus-accumbens tripartite neuronal motif guides appetitive memory in space. Cell 176(6):1393-1406.e16.

van Heukelum S, Mars RB, Guthrie M, Buitelaar JK, Beckmann CF, Tiesinga PHE, Vogt BA, Glennon JC, Havenith MN. (2020) Where is cingulate cortex? A cross-species view. Trends Neurosci. 43(5):285-299. doi: 10.1016/j.tins.2020.03.007

Vinogradova OS. (2001) Hippocampus as comparator: role of the two input and two output systems of the hippocampus in selection and registration of information. Hippocampus 11:578–598.

Wagatsuma H, Yamaguchi Y. (2004) Cognitive map formation through sequence encoding by theta phase precession. Neural Comput 16**:** 2665–2697.

Wierzynski CM, Lubenov EV, Gu M, Siapas AG. (2009) State-dependent spike-timing relationships between hippocampal and prefrontal circuits during sleep. Neuron 61(4):587-96.

Yu JY, Frank LM. (2015) Hippocampal-cortical interaction in decision making. Neurobiol Learn Mem. 117:34-41. doi: 10.1016/j.nlm.2014.02.002.

Yu JY, Kay K, Liu DF, Grossrubatscher I, Loback A, Sosa M, Chung JE, Karlsson MP, Larkin MC, Frank LM. (2017) Distinct hippocampal-cortical memory representations for experiences associated with movement versus immobility. Elife 6:e27621. doi: 10.7554/eLife.27621.

Yu JY, Liu DF, Loback A, Grossrubatscher I, Frank LM. (2018) Specific hippocampal representations are linked to generalized cortical representations in memory. Nat Commun. 9(1):2209.

Supplementary Data

In 341 neurons (83% of the 411 selected neurons; 36% of the total pool of 944 responsive neurons) there were SDF differences on all trials taken together and also on either left choice trials alone or right choice trials alone. Among the 331 neurons that could be analyzed in the 18 RC+SC sessions, 140 (42.3%) showed a significant shift in activity (Monte Carlo, $p<0.05$, two-tailed). Among these, 121 (121/140: 86.4%; or 36.6% of the total of 331 neurons) retained these differences on either left or right choice trials alone. These neural activity transitions in 83 neurons occurred fewer trials from an SC (and thus will be considered "SC-responsive"), while 90 were fewer trials from an RC (thus considered "RC-responsive"), while in 9 neurons this could not be determined unequivocally. Here individual neurons could show two activity transitions in close proximity to a RC and a SC respectively (62/418, or 15%, of SC-responsive neurons and 62/145, or 43%, of RC-responsive neurons). Because of this ambiguity, analyses were further pursued for SC-only and RC-only sessions only.

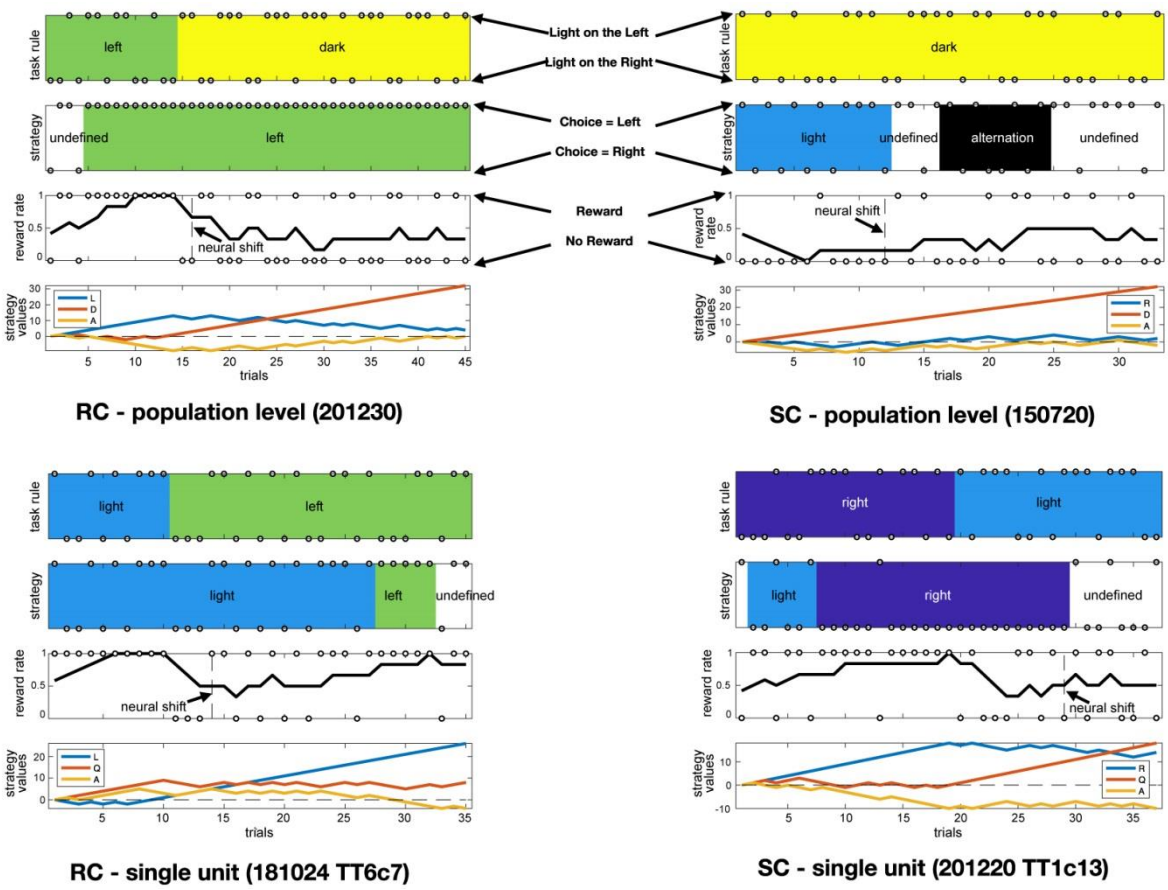
For each session with a RC or a SC, the significant SDF values of trial-by-trial bootstrap Monte Carlo analyses of the neurons were plotted as distributions across trials for different spatial positions on the maze, as shown in the second column of **Fig. 5**. This generally revealed smooth progressions of values with maximal or minimal values. However, the trial with the peak SDF value (e.g., trial 8 in **Fig. 5A**) did not necessarily represent the most marked inter-

trial activity transition (e.g., between trials 22 and 23 in **Fig. 5A**), because of non-monotonic trial-by-trial progression of SDF values.

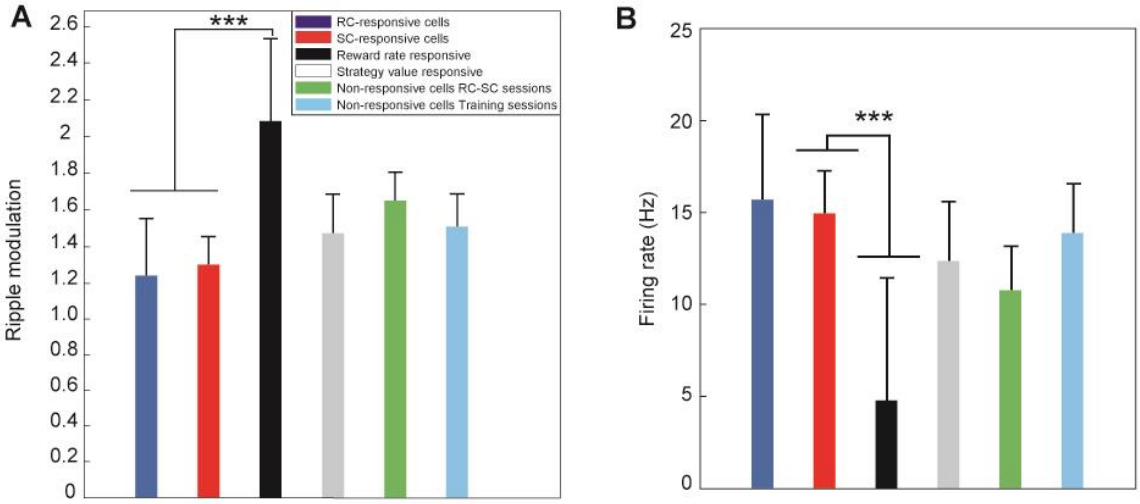
To verify that the latter results are not confounded by other behavioral or task related parameters, we re-examined the beta distributions for RC-responsive and SC-responsive neurons, excluding those that also showed behaviorally correlated activity including left vs right reward arm choice, correct vs error trial, and light vs dark reward arm choice. In RC-only sessions, the beta distribution means of SDF values was 4.7±3.1 trials after the first unrewarded trial providing evidence for the RC. For SC-only sessions the mean was 2.3±1.6 trials before the SC. In RC+SC sessions, the mean was 4.1±2.9 trials after the RC while the mean was 1.8±1.4 trials before the SC. All of these results are consistent with the above analyses.

The difference in the proportion of theta-modulated cells in RC-only vs SC-only sessions was not due to differences in running speed (which affects theta). The mean velocity of animals during RC-only sessions (36.0±3.8cm/s), SC-only sessions (35.9±1.7cm/s), RC+SC sessions (38.4±2.3 cm/s) and Training sessions (31.1±2.7cm/s) were not significantly different (Kruskal-Wallis test, χ^2=2.3, df=3, p=0.5125). There also was not a significant difference between RC-only and SC-only sessions in the change in running speed between the 5 trials before and the 5 trials after the RC or SC event (RC-only sessions: -8.0±4.7 cm/s; SC-only sessions: -3.8±1.8 cm/s; Kruskal-Wallis test, χ^2 = 2.78, df=2, p=0.25). Thus the greater incidence of theta modulation in RC-responsive neurons is not likely to be confounded with differences in speed changes between RC and SC. Furthermore, note that a recent study showed very minor impact of movement parameters on the activity of individual PFC neurons (Lindsay, et al., 2018; see also Powell and Redish, 2016).

Supplementary Figures



Supplementary Figure 1. Examples of neural shift events illustrated relative to the rat's behavior, and other potentially relevant information such as reward rate and strategy values. Top, neural shifts detected at the population level (shifts between blocks within population activity matrix) in two rats, Rat 20 and Rat 15. Bottom, neural shifts in single neurons in two rats: Rat 18 and Rat 20. R=Right strategy, L=Left strategy, Q=Light strategy, D=Dark strategy, A=Alternation strategy.



Supplementary Figure 2. A) Amplitude of ripple modulation among the cell response types before and after sessions. (B) Control for firing rates ***- p<0.001 (Chi-square test)

Supplementary Annex

Preliminary statements

The goal of this work is to provide a descriptive analysis of animals behavior in a binary rewarded decision task. Descriptive means that we don't try to explain why or how the animals perform the task, though we have to specify several explicit hypotheses, as follows, which are all of course debatable:

(H1) We assume that each animal decision is based on a finite number of information, which are all available at the decision time (i.e. the behavior is causal)

(H2)We assume that all the cues are binary variables (or can be reformulated as a set of binary variables). As a counterexample of this assumption, the animal decision might be instead based on the accumulated reward obtained from the beginning of the experimental session.

(H3) We suppose that all these cues are also known to the experimenter, and correctly recorded. Note: this hypothesis is quite strong; animal decision might depend on uncontrolled or unrecorded cues. This issue could be eventually handled in the permissive analysis (see below)

(H4) We suppose that at each decision time, the animal will apply a deterministic rule (*i.e.* a Boolean function of the set of binary cues). Again, this is quite a strong hypothesis, which will be relaxed in the permissive analysis.

(H5)We admit that at some trials (the transition trials), the animal switches to a different rule, drawn from a prespecified set of rules. We will not try to explain why the transition occurs, but only when it is likely to occur. That's why the analysis is in some sense only descriptive and not subjective.

(H6) Obviously, the number of transitions has to be restricted in some way. If not, the mere sequence of observed animal decisions would be the best result of the analysis. In the following, we will assume a *constant transition probability* $\varepsilon < 0{:}5$. We will show that, according to the non permissive analysis, maximizing the posterior probability of sequence of rules and transition (given the sequence of observed decisions) is equivalent to minimizing the number of transitions.

1 The non permissive analysis

Let M^t the observed animal decision at trial t with 1≤ t ≤T , R^t the unknown rule (or strategy), one element of the set of allowed rules $\mathcal{R}$, applied by the animal at trial t, and C^t the set of binary cues available to the animal (and recorded by the experimenter) at trial t. In the non permissive analysis, we apply a Dirac likelihood distribution, as follows:

$$P(M^t = m \,|\, R^t = r; C^t = c) = m \wedge r(c)$$

This equation means that the likelihood is equal to 1 when the animal decision is exactly predicted by the application of the rule r to the set of cues c, otherwise the likelihood is equal to 0.

According to (H6), the transition probability between rules is specified by:

$$P(R^{t+1} = r' \neq r \,|R^t = r) = \varepsilon/(n-1)$$

$$P(R^{t+1} = r \,|\, R^t = r) = 1 - \varepsilon$$

where n is the number of allowed rules and ε is the (unique) constant parameter used in this method. We also assume that the prior distribution on R is uniform (no rule is "privileged"), so that:

$$P(R^1 = r) = 1/n$$

The posterior distribution over the whole sequence of rules is given by:

$$P(R^{1:T} \,|\, M^{1:T}, C^{1:T}) \propto P(M^1 \,|\, R^1, C^1) \prod_{t=2:T} P(R^t \,|\, R^{t-1}) P(M^t \,|\, R^t, C^t)$$

The problem is to find the sequence of rules which maximizes the posterior probability, given the observed decisions and the recorded cues, *i.e.*:

$$r_{opt}^{1:T} \in Argmax\, P(R^{1:T} \,|\, m^{1:T}, c^{1:T})$$

Note that several sequences could have the same probability, therefore the analysis might eventually provide several optimal solutions.

To solve this problem, we first remark that if for any single trial the rule is not compatible with the observed decision, *i.e.*$\exists t : m^t \neq r^t(c^t)$, the whole sequence has a null probability, hence the term "non permissive". We then constrain the rules to belong to the series of subsets S^t defined as the subsets of compatible rules at trial t:

$$r \in S^t \Leftrightarrow m^t = r(c^t)$$

We also remark that the transition probability at each trial can have only two values: either $\varepsilon/(n-1)$ if there is a rule change, or $1-\varepsilon$ if not. It follows that

the Log posterior probability of the whole sequence of rules (LP(R)) depends only on the number n_T of transition trials (providing that all rules belongs to the compatible subsets):

$$LP(R) = LogP(R^{1:T} \,|\, m^{1:T},\, c^{1:T}) = cste\text{-}n_T Log(\frac{(1\text{-}\varepsilon)(n-1)}{\varepsilon})$$

Providing that $\varepsilon < 0.5 \leq \frac{n-1}{n}$, an optimal sequence of rules is therefore a sequence of compatible rules (*i.e.* $r^t \in S^t$) with a minimal number of transition n_T. To compute the minimal n_T and the sequence of optimal rules, we compute the sequence of rule subsets Q^t as follows:

(1) Initialize $n^T = 0$ and $Q^1 = S^1$

(2) For t = 2 to t = T compute $Q^t = Q^{t-1} \cap S^t$ and continue while $Q^t \neq \emptyset$

(3) If $Q^t = \emptyset$, t is a transition trial, then increment n_T and reset $Q^t = S^t$

(4) Between two transition trials $[t_0,\, t_1 - 1]$ (or between the first trial and the first transition $[t_0 = 1,\, t_1 - 1]$; or between the last transition and the last trial), we have $R^t = r$ where r can be freely chosen in the subset $Q^{t_1 - 1}$.

Interestingly the non permissive analysis, though formalized in probabilistic terms, reduced to a logic algorithm, and the resulting optimal solutions do not depend on the parameter ε. However, it clearly depends on the set of allowed rules $\mathcal{R}$ (see results). One has also to keep in mind that the output of this algorithm is one optimal solution; one cannot exclude that there are other optimal solutions, with different transition trials. Suppose for instance that the algorithm output is:

$$aaaaaaaaaabbbbb$$

The rule a is valid for the first 10 trials, then one has to switch to rule b for the remaining 5 trials. Let's suppose that the rule b is not valid for the first 5 trials, but thereafter becomes valid for the remaining 10 trials. It follows that the sequence:

$$aaaaabbbbbbbbbb$$

is as optimal as the first one. It would have been the output of the same algorithm working backward from trial T to trial 1.

The Reduced Rule Set Problem

Suppose that we get an optimal solution $r_{opt}^{1:T}$ by taking rules from a set $\mathcal{R}$. The question is: is there a minimal set of rules $\mathcal{R}_{min} \subset \mathcal{R}$ such that, but constrained rules to belong to the reduced set of allowed rules, we obtain exactly the same number of transitions n_T than with a full set of rules ? Unfortunately, this problem is NP complete, and therefore one cannot build the true $\mathcal{R}_{min}$ in a reasonable time. Fortunately, there exists an approximate algorithm (a "greedy" algorithm) which provides a reduced rule set (RRS) which is generally a good approximate of the minimal set.

The principle of the greedy algorithm is the following:

(1) one builds the list of $n_T + 1$ subsets of valid rules with the forward algorithm described before:

$$\{Q^{t_1-1}, Q^{t_2-1}, ..., Q^{t_{n_T}-1}, Q^T\}$$

One initializes $\mathcal{R}_{min} = \emptyset$ and $k = card(\mathcal{R}_{min}) = 0$

(2) one identifies the rule r which belongs to the greatest number of subsets from that list. One takes r as a new element of $\mathcal{R}_{min}$ and k is incremented

(3) one removes from the list all the subsets which contain r. If the list is now empty, the algorithm is finished, otherwise one goes back to step (2).

Results for rat behavioral experiments

We applied the non permissive analysis and the RRS algorithm to the behavioral data recorded from 5 rats and a total of 3322 trials. The following table gives the minimal number of transitions, and the minimal (or approximately minimal) number of rules in RRS, starting with different sets of allowed rules $\mathcal{R}$. We tested 10 sets of rules. The first two sets contain the two cue-independent rules (Left or Right, L/R) plus the two single cue dependent rules L_0 (rules based on the current light cue) and M_1 (rules based on the previous movement). The following seven sets are defined as the full set of 256 rules dependent on the indicated group of three cues. The last set is the union of the seven full sets of 256 rules. Note that this big set contains several instances of the same rule (*e.g.* the Left rule appears in each of the full sets of 256 rules, therefore, it appears seven times in the so-called All set).

$\mathcal{R}$	$card(\mathcal{R})$	$min\, n_T$	$card(RRS)$
$L_0 + L/R$	4	874	4
$L_0 + M_1 + L/R$	6	739	6
$L_0 M_1 R_1$	256	498	43
$L_0 L_1 R_1$	256	476	49
$L_0 L_1 M_1$	256	481	45
$L_0 M_1 M_2$	256	489	45
$L_0 R_1 R_2$	256	512	60
$M_1 M_2 R_1$	256	578	37
$M_1 R_1 R_2$	256	606	43
All	1792	372	97

Table 1: Non permissive analysis and number of rules in the reduced set for various rule sets.

As a general tendency, increasing the size of the rules set yields fewer transitions.

2 The permissive analysis

In this analysis, the Dirac likelihood distribution is replaced by:

$$P(M^t = m \,|\, R^t = r; C^t = c) = \alpha + (1\text{-}2\alpha).(m \wedge r(c))$$

This form of likelihood allows the animal to sometimes perform a movement opposite to the rule he is presumed to apply. The probability to perform such an error is quantified by the parameter $\alpha \ll 1$. One simple way to interpret the parameter α is to relax the deterministic hypothesis (H4): applying a rule r consists to perform the movement $r(c)$ with the probability $1-\alpha$ and to perform the opposite movement with the probability α. Another way to interpret the parameter is to relax the well controlled cues hypothesis (H3): the rules are deterministic functions of the known cues plus some unknown cue (or unknown rule), which is generally constant, but can change at some unpredictable trials with a probability α so that the animal change its decision. Anyway, the mathematical formulation remains the same.

The permissive analysis is specified by two parameters (ε and α). At each trial, the transition probability can take one among two values as before; the likelihood can also take one among two values: either α if it is an error trial or $1-\alpha$ if not. Let n_T the number of transitions and n_E the number of errors, the Log posterior probability of the whole sequences of rules is :

$$LP(R) = cste\text{-}n_T Log(\frac{(1\text{-}\varepsilon)(n\text{-}1)}{\varepsilon}) - n_E Log(\frac{1\text{-}\varepsilon}{\varepsilon})$$

The issue is then to find the best compromise between minimizing n_T and minimizing n_E. The critical parameter is the *permissive factor* p:

$$p = Log(\frac{(1-\varepsilon)(n-1)}{\varepsilon})/Log(\frac{1-\alpha)}{\alpha})$$

The cost J to be minimized is simply defined as:

$$J = n_T - n_E/p$$

with:

$$LP(R) = cste\text{-}J.Log(\frac{(1\text{-}\varepsilon)(n\text{-}1)}{\varepsilon})$$

We first remark that if $p < 0.5$ the permissive analysis turns out to be equivalent to the non permissive analysis, *i.e.* the permissive factor is too low. In fact, let's consider the following sequence or rules given by the non permissive model:

$$aa...aabaa...aa$$

There is two transitions and no error, so that the cost is equal to 2. Consider now the following sequence where the unique b is replaced by an error trial. The new sequence:

$$aa...aaaaa...aa$$

contains no transition and one single error. Its cost is equal to $1/p$. Therefore if $p < 0.5$, the permissive analysis cannot give rise to better sequences than the one produced by the non permissive analysis.

Suppose now that $p > 1$. One can see that a full sequence of a interrupted by two$b$:

$$aa...aabbaa...aa$$

has a higher cost that a sequence containing only "a" (two errors are better than two transitions). More generally, any pairs of errors in a sequence would be suppressed. We consider (but this is debatable) that $p > 1$ is too permissive, and would allow the suppression of transitions which are in fact significant. Therefore, we will restrain ourselves to the weakly permissive analysis where $0.5 < p < 1$. In practice, we choose $p = 2/3$ (*i.e.* the cost of two transitions is equal to the cost of three errors).

As far as we know, there is no algorithm - running in a reasonable time - which is able to build the optimal sequence of rules according to the weakly permissive analysis. We then adopt a greedy forward algorithm as followed

(1) one builds the list of T subsets of compatible rules$\{S^1; ..., S^T\}$, *i.e.*:

$$r \in S^t \Leftrightarrow m^t = r(c^t)$$

(2) Initialize $n_T = n_E = 0$ and $Q^1 = S^1$

(3) For $t = 2$ to $t = T$ compute $Q^t = Q^{t-1} \cap S^t$ and continue while $Q^t \neq \emptyset$

(4) If $Q^t = \emptyset$, t is either a transition trial or an error trial. It is an error trial if and only if $Q^{t-1} \cap S^{t+1} \neq \emptyset$ (that is the trial t can be bypassed). In that case, increment n_E and set $Q^t = Q^{t-1} \cap S^{t+1}$. Otherwise, increment n_T and reset $Q^t = S^t$.

Results for rat behavioral experiments

We applied the weakly permissive analysis (WPA) to the same data and same set of rules used with the non permissive analysis (NPA). The cost for NPA is always equal to the number of transitions. The cost for WPA depends on the number of transitions and errors ($WPAn_T$ and $WPAn_E$) with a permissive factor $p = 2/3$.

$\mathcal{R}$	$NPA Cost$	$WPA cost$	$WPA n_T$	$WPA n_E$
$L_0 + L/R$	874	656.0	534	153
$L_0 + M_1 + L/R$	739	694.66	662	49
$L_0 M_1 R_1$	498	480.66	472	16
$L_0 L_1 R_1$	476	466.66	458	10
$L_0 L_1 M_1$	481	471	465	9
$L_0 M_1 M_2$	489	470.33	453	26
$L_0 R_1 R_2$	512	479	447	48
$M_1 M_2 R_1$	578	560.66	546	22
$M_1 R_1 R_2$	606	562.66	524	58
All	372	369.66	367	4

Table 2: Comparison of the weakly permissive and non permissive analysis for various rule sets.

We remark that the permissive factor benefits mainly to the smallest set of allowed rules. Still, the lowest cost and the minimal number of transitions are obtained with the big All set of rules (which, as we have seen before can be reduced to 97 different rules). In second position, the groups specified by the cues$L_0 L_1 R_1$,$L_0 M_1 M_2$ and $L_0 L_1 M_1$ make quite a good job with only 49 and 45 rules.

Finally, if we look at the most frequently used rule (by counting the number of valid subsets Q^t containing a given rule), we observe that the "persistence rule" (which consists to reproduce the previous movement) is generally ahead, of course providing that it is included in the allowed set $\mathcal{R}$. Within the biggest rule set, the most frequently used rule consists to reproduce the previous movement if it has been rewarded, otherwise to reproduce the movement performed two trials before (the penultimate movement).

Incremental Analysis

In this section, we start with only two complementary rules (e.g. Left and Right or Light and Dark or Altern and Persist). Then, we look for the best additional rule which minimizes the NPA cost (i.e. the number of transitions without errors). The process is repeated until no further improvement can be achieved. Because the process is incremental (and not decremental as in the search for Reduced Rule Set) we expect to obtain at the end a NPA cost slightly higher than in the previous analysis. The advantage is to get a (monotonically decreasing) curve for NPA cost as a function of the number of rules.

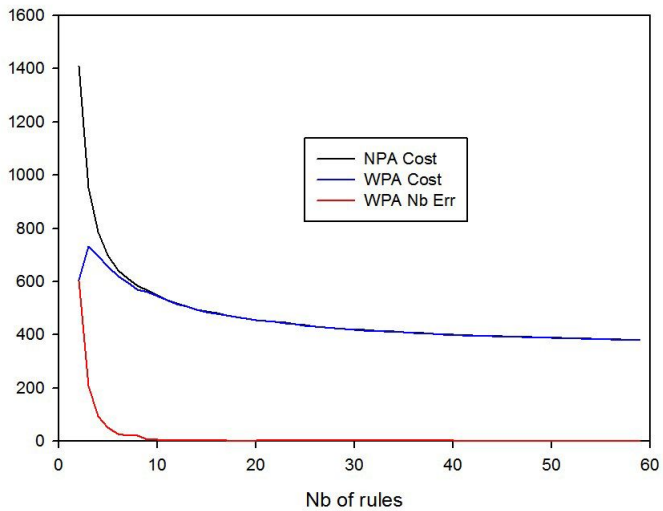


Fig. 1: Minimal cost as a function of the number of rules, obtained by the incremental analysis based on minimal NPA cost, and starting with the two complementary rules Left and Right.

The results depend on the choice of the pair of complementary rules, as indicating in the following table. The influence of the starting pair is mostly at the beginning of the process (*i.e.* for a relatively small number of rules). The best choice appears to be (Altern, Persist).

Start	2	4	6	8	16	End	Final Nb
L + R	1409	786	642	586	481	380	59
A + P	1082	742	638	582	478	384	58
L + D	1535	789	672	602	484	383	57

Table 3: Minimal NPA cost obtained by the incremental analysis starting with the three complementary pairs: L+R = (Left, Right); A+P = (Altern, Persist) L+D = (Light, Dark). The last two columns give the final number of transition (End) and the final number of rules (Final Nb) when the algorithm stops, i.e. when no decreasing cost can be obtained.

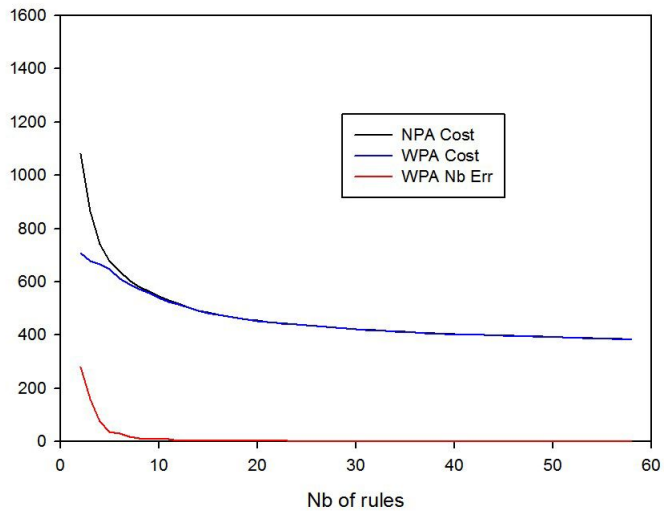


Figure 2: Results of the incremental analysis based on the minimal NPA cost, and starting with the (Altern, Persist) pair.

The same incremental analysis can be used with the minimal WPA cost, in order to obtain this time a monotonically decreasing curve for WPA. As before, one can compare the resulting costs for the three starting pairs of complementary rules (see Table 4).

Start	Value	2	4	6	8	16	End	Final Nb
L + R	Cost	605	584	575.7	572	562.7	562.7	16
L + R	Errors	603	474	472	471	448	448	16
A + P	Cost	708.7	655.7	588	556.3	476.7	383.7	52
A + P	Errors	280	157	99	59	10	4	52
L + D	Cost	824.3	632.7	599.7	568.7	492.33	393.33	60
L + D	Errors	533	193	115	94	35	8	60

Table 4: Minimal WPA cost and number of errors obtained by the incremental analysis (based on the WPA criterion) starting with the three complementary pairs.

Note that the Left/Right pair obtains the best initial cost, at the price of a very large number of errors. The cost and the number of errors can be only slightly improved by increasing the number of rules. Again the winner starting pair appears to be Altern/Persist.

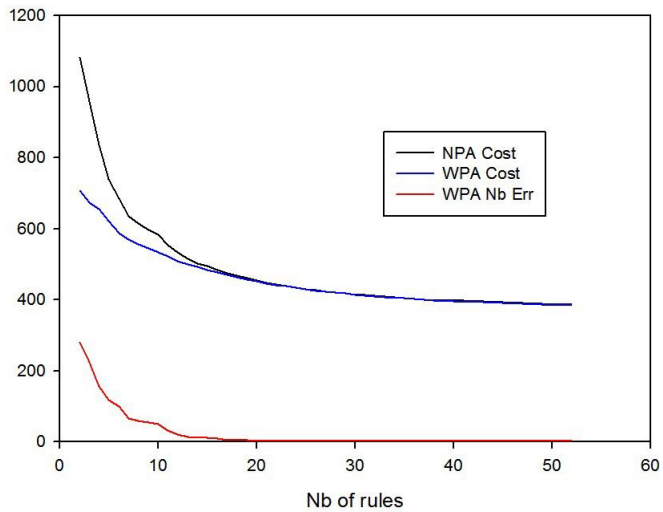


Fig. 3: Results of the incremental analysis based on the WPA cost and starting with the (Altern, Persist) pair.

In general, it is quite hard to understand the "logic" of the first rules which are added to the initial pair by the incremental analysis. To give one single example, starting with the Altern / Persist pair, and looking for the lowest WPA cost, the first two new rules that one of the rat (R20) seems to apply are defined as follows:

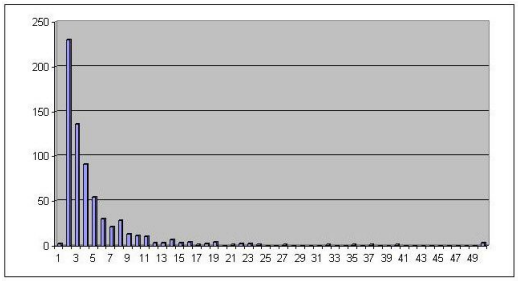
- if the previous trial was rewarded, move to the dark arm (*i.e.* $Not(L_0)$); if not previously rewarded, move to the arm which was previously lightened (*i.e.* L_1)

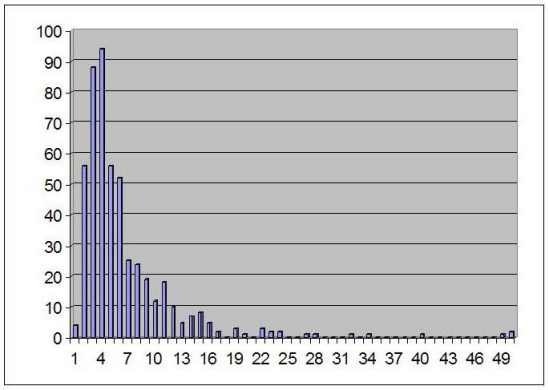
- if the previous trial was rewarded, move to the lightened arm (L_0); if not previously rewarded, move to the arm which was previously in the dark ($Not(L_1)$)

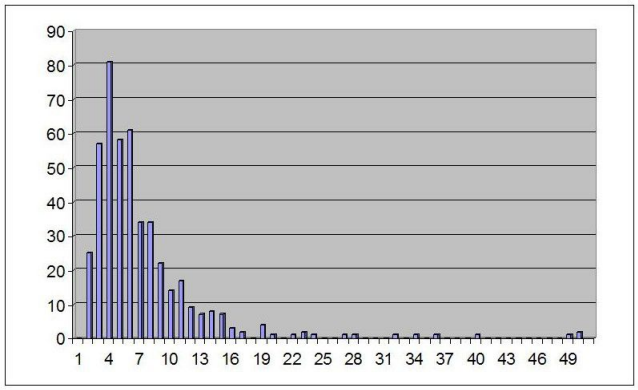
These two rules are in this case complementary, but this is exceptional. Note that these two rules can explain why the rat seems to have learned the light and dark experimenter's rules.

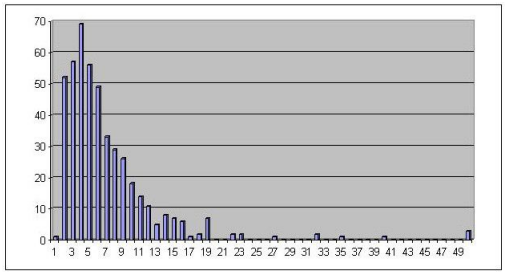
Histograms of sequence lengths

In this analysis, we compute the number of sequences of a given length where the same rule (or the same subset of rules) applies to consecutive trials, for a given set of rules $\mathcal{R}$ and for both NPA and WPA. Below are shown the results for the group of 6 rules (left, right, light, dark, altern & persist), for the group of 256 rules based on the cues $L_0L_1R_1$ and the full set of 1792 rules (All set), according to the WP analysis.


250
200
150
100
50
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49


100
90
80
70
60
50
40
30
20
10
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49


90
80
70
60
50
40
30
20
10
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49


70
60
50
40
30
20
10
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49

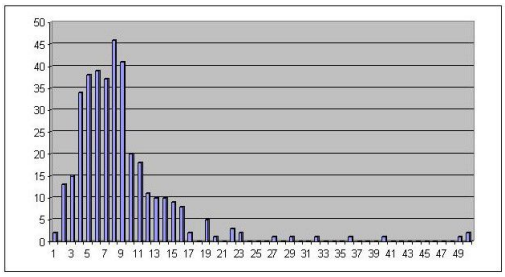


Fig. 4: Histograms of sequence length with WPA applied to various sets of rules. From top to bottom: group of 6 basic rules (left, right, light, dark, altern, persist) with a total of 663 sequences, group of 12 best rules with a total of 496 sequences, group of 20 best rules with a total of 451 sequences, group of 256 rules based on the cues $L_0L_1R_1$ with a total of 459 sequences and finally for the group of 1792 rules with a total of 368 sequences. Note that the peak shifts from 3 to 8, as the number of rules increases.

Significance

A quite difficult (because it's ill posed) issue is to evaluate the significance of the NPA or WPA results. One simple way is to compute the probability of any sub-optimal sequence of rules within a given frame, *i.e* within NPA or WPA for a given set of rules. Within NPA, the probability of a sequence of rules is uniquely determined by the number of transitions n_T. Within WPA, the probability of a sequence of rules is uniquely determined by the number of transitions n_T and the number of errors n_E, though the key number is the cost $J = n_T + n_E/p$. The cost is proportional to $-Log(P)$ with a proportionality coefficient equal to $Log(\frac{(1-\varepsilon)(n-1)}{\varepsilon}) = Log(n-1) + Log(\frac{(1-\varepsilon)}{\varepsilon})$. The parameter ε has no influence on the selection of the optimal sequence, but it has a key impact on the probability of sub-optimal sequence. In the worst case, we admit $\varepsilon = 0.5$ (which means that the probability to change the rule from trial to trial is 50%). In this worst case, within NPA, a sub-optimal sequence which consists to replace a given sequence $a...a$ by the sequence $a..aba...a$ has the following relative probability (with respect to the optimal sequence $a...a$) function of the number of rules n:

n	p
4	0.111
6	0.04
12	0.00826
20	0.00277
40	0.00066
90	0.000126

According to the scientific standard in biology, it's significant when $n \geq 6$. Within WPA, one can consider the sub-optimal sequence $a....a$ (no transition,

two errors, cost=3) with respect to the optimal sequence $a..abba...a$ (two transitions, no error, cost = 2). The relative probability of this sub-optimal sequence is:

n	p
4	0.333
6	0.2
12	0.0909
20	0.0526
40	0.0256
90	0.0112

Again, according to the scientific standard, a sub-optimal sequence of rules is significantly sub-optimal for n $>$ 20. One can also compute the relative probability of a sub-optimal sequence $a....a$ (no transitions, 3 errors, cost = 9/2) with respect to the optimal sequence $a..abbba...a$ (two transitions, no error, cost = 2):

n	p
4	0.06415
6	0.0179
12	0.00249
20	0.000635
40	0.000105
90	0.0000133

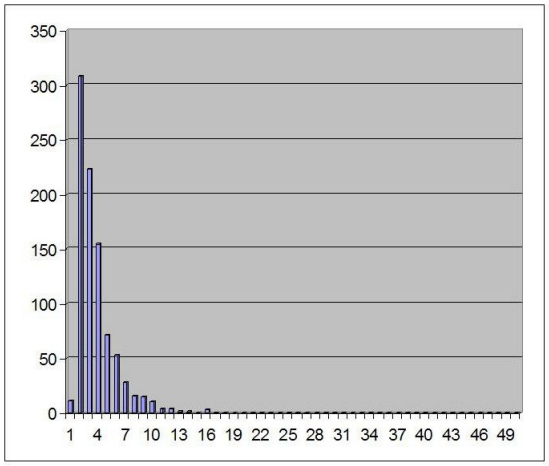
Clearly, the sub-optimal sequence is significantly (according to scientific standard) less probable as soon as n $>$ 4.

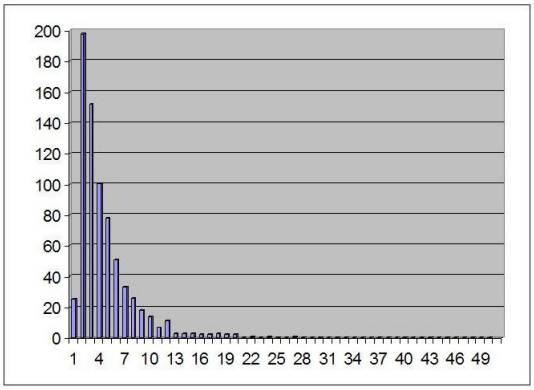
Analysis of random movements

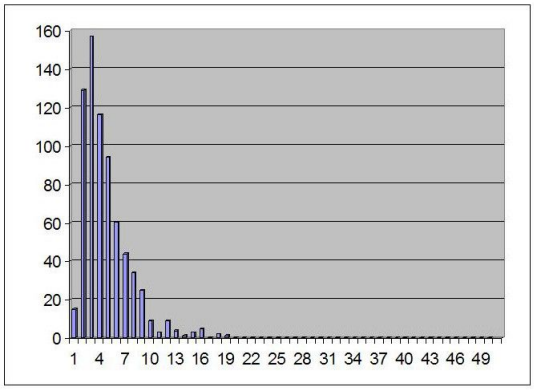
Another way to assess the significance of NPA and WPA is to run these methods on a set of (pseudo) random movements. For this, we create pseudo data files with the same number of trials and same cues except for the movement which was randomly drawn with the Mersenne-Twister random generator.

$\mathcal{R}$	NPA Cost	WPA Cost	WPA n_T	WPA n_E
$L_0 + L/R$	1206	977.7	819	238
$L_0 + M_1 + L/R$	1061	975.0	903	108
$L_0M_1R_1$	678	659.7	643	25
$L_0L_1R_1$	697	672.3	651	32
$L_0L_1M_1$	620	613.0	603	15
$L_0M_1M_2$	636	619.7	605	22
$L_0R_1R_2$	674	636.7	608	43
$M_1M_2R_1$	706	680.66	658	34
$M_1R_1R_2$	720	674.0	632	63
All	448	447.3	446	2

Table 5: Comparison of the weakly permissive and non permissive analysis for various rule sets, applied to the random data. There is obviously a dramatic increase of the cost, for both NPA and WPA, and all sets of rules.







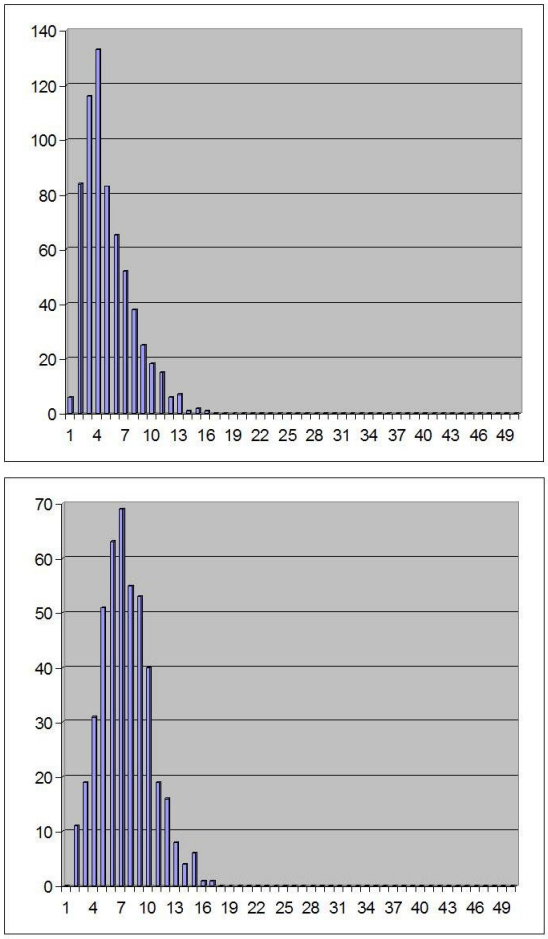


Fig. 5: Histograms of sequence length with WPA applied to the random behavior. The sets of rules are the same than in Fig. 4, *i.e.* from top to bottom: 6, 12, 20, 256 and 1792 rules; the total numbers of sequences are, respectively, 904, 615, 669, 604 and 447. The numbers of errors are also very high, except for the larger sets: 108, 326, 158, 32 and 2. As compared to real data, the peak is sharper and shifts less, as the number of rules increases.

Consider for instance the group of 12 rules (a reasonable size). Compared to random data analysis, the real data analysis provides significantly less short sequences (length < 4), roughly similar number of medium size sequences (length between 4 and 12) and a higher number of long sequences (length > 12). Similar conclusion can drawn for the sets of 6 and 20 rules. For 6 rules, the real rat results exceeds the random results for sequences longer than 10 (though 3 sequences of length 16 are found in the random data, against 4 for the real rat...). For 20 rules, real rat results exceeds random results for sequences longer

than 12, though one sequence of length 19 is found in the random data). As a consequence, one cannot exclude that randomness can account for the existence of one long sequence of consecutive trials (up to about 20 trials). It is the number of long sequences that matters. If we look at the subset of rules which are compatible with sequences of length > 12, most rules of the reduced rule set are observed. For instance, in the ascending method 11 among the first 12 best rules, and 15 among the first 20 best rules are observed during long sequences.

Outside NPA & WPA: the unknown rule ...

We consider in this section the hypothesis that the rat can use sometimes an "unknown rule" u in addition to the regular rules $r \in \mathcal{R}$. As this unknown rule explains any movement, it should be penalized in some way. Otherwise, the whole data would be perfectly explained by a unique sequence of u, without transition and without error. The weak permissive analysis was a first way to introduce this idea; the penalization was quantified by the permissiveness factor. However, this factor was set to a minimal value of 2/3, in order to prevent the possibility of two consecutive errors. Another way to formalize this idea is to penalize u through the specification of transition probability. We need now three parameters in order to define:

- as before the transition probability between two different regular rules $\varepsilon/(n-1)$

- the transition probability from regular to unknown rule: ε_1

- the transition probability from unknown to regular rule: ε_2

The difficulty of this approach results from the arbitrary choice of these parameters. As in section 2, the parameters can be translated into cost *i.e.* the change in Log P resulting from a regular transition c_t, a transition towards u C_u, a return transition from u C_r and finally the cost to maintain u between two consecutive trials C_m. Simple reasoning shows that we must have :

(1) $C_u + C_r < 2C_t$ otherwise u will never be used to account for the data

(2) $C_u + C_r + kC_m > 2C_t$ for some maximal length k otherwise infinite sequence of u will be used instead of regular rules. These considerations lead to some restriction for the parameters, but they cannot be precisely specified without some arbitrary choice.