Reward revaluation biases hippocampal sequence content away from the preferred outcome

Authors: Alyssa A. Carey[1,2], Youki Tanaka[1], and Matthijs A. A. van der Meer[1*]

[1]Department of Psychological and Brain Sciences, Dartmouth College, Hanover NH 03755. [2]Department of Biology, University of Waterloo, Waterloo ON N2L 3G1, Canada. [*]Correspondence should be addressed to MvdM, Department of Psychological and Brain Sciences, Dartmouth College, 3 Maynard St, Hanover, NH 03755. E-mail: mvdm@dartmouth.edu.

Number of Figures: 6

Abstract Word Count: 80

Main Text Word Count: approx. 3900

Number of Supplementary Figures: 8

Acknowledgments: We thank Nancy Gibson, Martin Ryan and Jean Flanagan for animal care and Min-Ching Kuo, Julia Espinosa and Eric Carmichael for technical assistance. We thank Elyot Grant for developing the SWR detection method used for the main analyses in this paper. This work was supported by the University of Waterloo and Dartmouth College (start-up funds to MvdM), and the Netherlands Organization for Scientific Research (NWO, grant 863.10.013 to MvdM).

Author contributions: AAC performed experiments and pre-processed the data. AAC, YT and MvdM wrote analysis code. AAC and MvdM performed data analysis. MvdM wrote the paper with comments from AAC and YT.

Conflict of Interest: The authors declare no competing financial interests.

Abstract

Internally generated sequences of place cell activity can depict spatial trajectories to reward locations, suggesting a role in model-based behavioral control. A largely separate literature emphasizes reward revaluation as the litmus test for such control, yet the content of hippocampal sequences under revaluation conditions is unknown. We report that the content of sharp wave-ripple hippocampal sequences is biased away from the preferred outcome following reward revaluation, challenging the idea that such sequences reflect recent experience or trajectories toward a goal.

Main text

The spatial navigation literature has identified a role for the hippocampus in flexibly navigating toward goal locations. This behavior is thought to rely on knowledge of the environment (a "cognitive map") to flexibly generate and decide between possible trajectories, an example of model-based behavioral control[1]. Sequences of hippocampal place cell activity, colloquially referred to as "replay" when coincident with sharp wave-ripple complexes (SWRs)[2,3], are a candidate neural underpinning for such planning, as suggested by the bias of such sequences toward goal locations[4] and behavioral impairments resulting from SWR disruption[5]. In contrast, the conditioning literature operationalizes model-based control as sensitivity to reward revaluation[6]. Reward revaluation causes rate remapping of hippocampal place cells[7] and hippocampal damage impairs behaviors that rely on the discrimination of internal motivational states such as hunger and thirst[8,9,10]. However, it is unknown how reward revaluation affects the content of hippocampal sequences. A goal-directed bias in hippocampal sequence content predicts that their content favors the currently highly valued outcome, especially following a recent revaluing. Conversely, if hippocampal sequence content is unaffected by reward revaluation, this would call into question a role for SWR sequences in model-based control.

To determine how hippocampal sequence content is affected by reward revaluation, we recorded neural ensembles from the dorsal CA1 area of the hippocampus as rats (n = 4, male) performed a T-maze task, which offered free choice between food (left arm) and water (right arm) outcomes (Figure 1a). The crucial manipulation in the experimental design was that prior to daily recording sessions, animals were alternately food- or water-restricted. As a result, animals exhibited a clear overall behavioral preference for the restricted outcome (Figure 1b; 89% food choices on food-restricted days, 80% water choices on water-restricted days, $\chi^2_{(df=3)}$ = 188.7, $p < .05$; individual rats all $p < .05$). Across trials within recording sessions, this preference was most clearly apparent from the second trial onward (Figure 1c).

Our analyses focused on SWR-associated ensemble activity patterns ("replay")[2,3]. Specifically, we sought to test how the *content* of such patterns – when corresponding to the "left" or "right" trials on the T-maze – related to the animals' motivational state and behavior. To access SWR content, we applied two different analyses using Bayesian decoding[11] (see Figure S1 for a schematic of the analysis pipelines). The first, *sequenceless decoding* analysis, identified putative SWR events based on spectral properties of the local field potential (LFP) and multi-unit firing activity (MUA). We decoded each putative SWR event as a single time window, and computed the log odds of that event's decoded content corresponding to a left or right trial relative to a shuffled distribution (*Online Methods*; note that left and right trial activity was clearly distinct during running on the track, Figure S2). For the second analysis, *sequence detection*, we first decoded the full session data using a 25 ms moving window, and then applied sequence detection[4] to identify spatiotemporal sequences that decoded to either the left or right trajectory on the track (but not both, *Online Methods*). Only sequences that passed a comparison against a chance distribution based on shuffled neural activity, and that overlapped with a putative SWR event (defined as above), were retained for subsequent analysis.

Sequenceless SWR content is biased away from the preferred outcome. To determine the relationship between motivational state and SWR content, we first quantified SWR content by decoding entire SWR events as a single time window, and z-scoring the log odds of the left vs. right posterior ($log(\frac{p(left)}{p(right)})$) relative to a resampled distribution obtained by randomly permuting left and right arm tuning curves (sequenceless

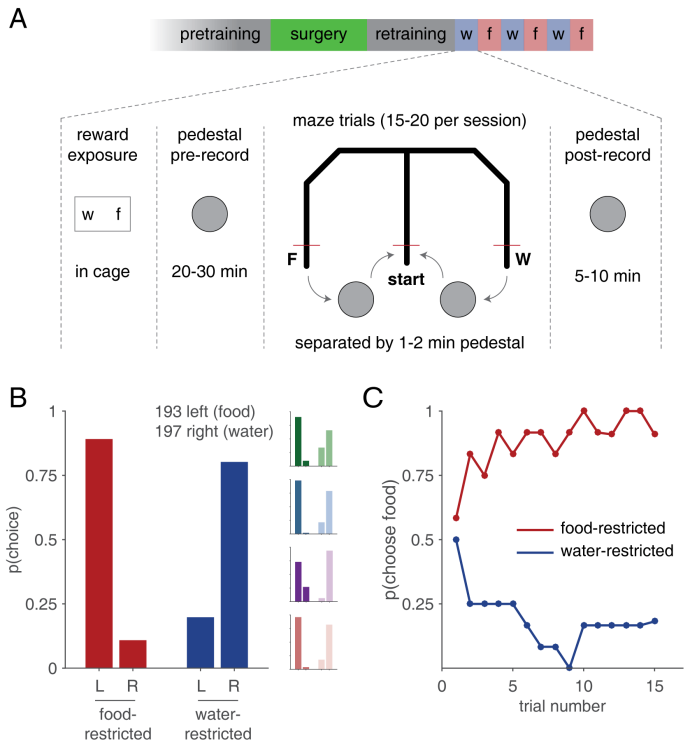

Figure 1: Experimental design, task and behavior. **A**: Rats (n = 4) experienced daily sessions on a T-maze offering them a choice between food (F) at the end of the left arm and water (W) at the end of the right arm. Blocked trials ensured adequate sampling of each outcome (see *Online Methods* for details). Daily recording sessions consisted of exposure to both food and water outcomes in the home cage, followed by a pre-task rest session. 15-20 T-maze trials, separated by a 1-2 minute intertrial interval were performed next, followed by post-task rest. Crucially, across recording sessions, rats were either water-restricted (blue, w) or food-restricted (red, f). **B**: Choice proportions on free-choice trials. Following food restriction (red bars) rats preferred the food (F, left) arm, whereas following water restriction, the water (W, right) arm was preferred. Overall, rats chose both arms a comparable number of times (193 food, 197 water). Small bar graphs show choice proportions for individual subjects. **C**: Within-session choice proportions of choosing the food arm following food restriction (red) and water restriction (blue), averaged across all sessions and subjects. Note that these proportions do not necessarily sum to one because food and water restriction sessions occurred on different days.

decoding; see *Online Methods*). Plotting the resulting log odds z-score across recording sessions revealed a striking pattern: SWR content was systematically biased in the direction *opposite* rats' behavioral preferences (Figure 2). That is, for food restriction sessions (red shading in Figure 2a) rats preferentially chose the food (left) arm, but SWR content was biased toward the water (right) arm. Conversely, for water restriction sessions (blue shading), rats preferred the water (right) arm, but SWR content was biased toward the food (left) arm. More formally, a linear mixed model with the SWR content log odds z-score as the dependent variable was clearly improved by the addition of motivational state (likelihood ratio test, $LRstat_{\Delta df=1}$ =

10.1, $p = .0015$). Similarly, behavioral choice was strongly anticorrelated with SWR content across sessions (Pearson's $r_{(19)} = -.75$, $p < .001$). In a complementary analysis that collapsed across all food restriction sessions and all water restriction sessions respectively (Figure 2b), food restriction (red bar) was associated with a negative z-score indicating a water bias, whereas water restriction (blue bar) was associated with a positive z-score indicating a food bias (food: $-.41 \pm .23$, water $.40 \pm .27$, $p = .004$, Wilcoxon rank sum test). Each individual subject exhibited this pattern (Figure 2b, insets).

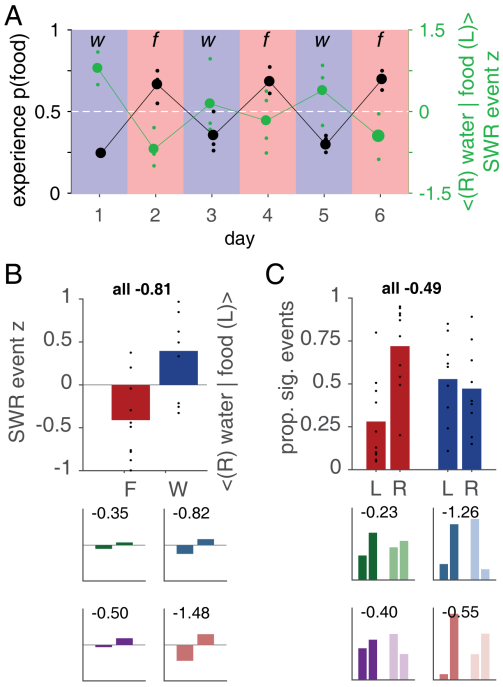

Figure 2: Sequenceless SWR content is biased away from the preferred outcome. **A**: Z-scored log odds (left vs. right) of decoded SWR content (green) plotted across sessions are biased in the opposite direction from behavioral preference (probability of choosing the food arm, black). Positive z-scores correspond to left (food) arm decoding, negative z-scores correspond to right (water) arm decoding. Shaded areas indicate water (blue, *w*) and food-restriction sessions (red, *f*). Large markers show averages across subjects, small markers indicate single recording sessions. **B**: Mean z-scored log odds averaged across food-restriction (left, red) and water-restriction sessions (right, blue). Numerical labels indicate the mean z-score difference between food and water restriction conditions. Markers show individual session data, and insets show individual subject averages. **C**: Proportion of significance-thresholded SWR events that decode to the left (L) and right (R) arms for food-restriction (red) and water-restriction (blue) sessions. Proportions are defined as the number of events for that arm (left, right) divided by the total number of events, yielding one number per session. Numerical labels show *motivational state bias scores* (see main text).

The above analysis included decoding of *all* candidate SWR events, but not all SWR events are thought to map onto any single environment[12]. Thus, it is possible that the above effects are due to an effect of

motivational state on those events that do not in fact clearly decode to one arm or another. To address this issue, we thresholded the z-scored log odds to retain only those events that significantly favored left or right (i.e. that exceeded p = .025 on either side of the shuffled log odds distribution). Figure 2c shows the proportion of these events on the left and right arms depending on motivational state. Consistent with the earlier analysis, food restriction sessions contained a larger proportion of significant events on the water arm (left/food .28, right/water .72) whereas water restriction sessions contained a (slightly) larger proportion of significant events on the food arm (left/food .53, right/water .47, ANOVA interaction of restriction type and arm, $F_{(df=1)}$ = 8.53, p = .006).

An important aspect to appreciate about this analysis is that a number of factors can result in a decoding bias for the left or right maze arm. These factors may be unrelated to the key experimental manipulation (motivational state), such as the number of cells with place fields on left vs. right trials, or idiosyncratic subject-specific behavioral preferences for one arm or another. A test of the hypothesis that motivational state affects SWR content should therefore be based on the *difference*, between motivational states, of left vs. right content. For example, a given rat may show an overall SWR content bias for the right arm on both food- and water-restriction sessions; the key question is whether there is any difference in that bias between motivational states (and which direction that difference is). Is the right (water-containing) arm favored more on water-restriction sessions compared to food-restriction sessions? For the SWR content z-score, this difference is a straightforward subtraction between food-restriction and water-restriction averages (Figure 2b). However, to make this key comparison transparent for SWR proportions (Figure 2c), we defined a *motivational state bias score*, $[p(L)-p(R)]_{food}-[p(L)-p(R)]_{water}$, which is positive for SWR content bias consistent with motivational state, and negative for SWR content bias opposite motivational state, regardless of overall bias. Using this score, it can be seen that rats R042 and R064 both had an overall bias toward the right (water) arm as measured by the proportion of significant events (top left and bottom right insets in Figure 2c). However, this bias was more pronounced during food restriction sessions compared to water restriction sessions, resulting in a negative motivational state bias score, which indicates a shift in SWR content away from the preferred outcome.

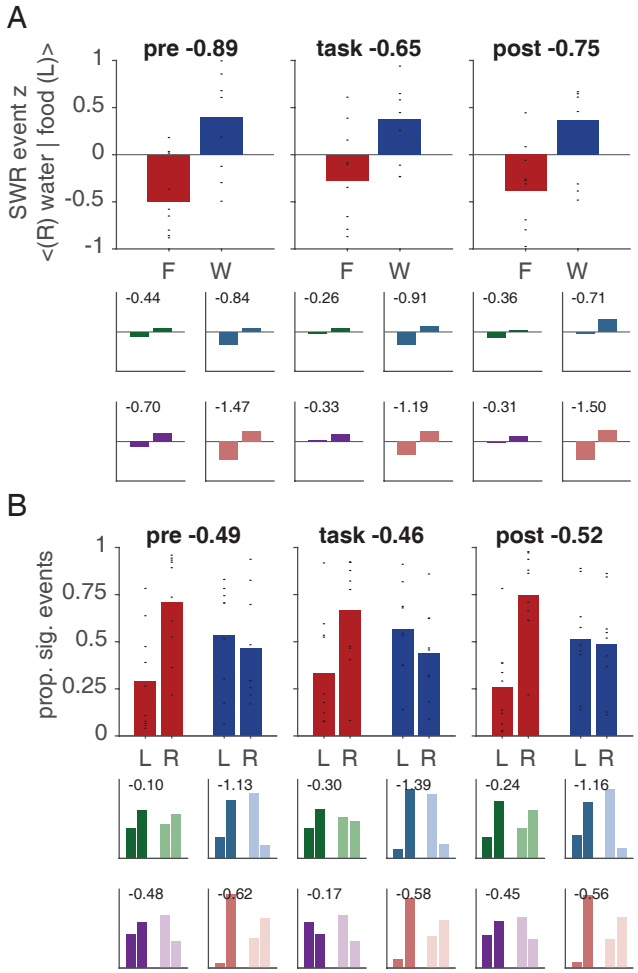

Figure 3: Sequenceless SWR content is biased away from the preferred outcome throughout all epochs within daily recording sessions. **A**: Average z-scored log odds (left arm vs. right arm) of SWR content, broken out by within-session epoch: before experience on the task that day ("pre"), during intertrial intervals ("task") and following experience ("post"). **B**: Proportions of SWR events classified as left and right using the same thresholding procedure as in Figure 2c. Panel layouts are as in Figure 2; red indicates food-restriction sessions, blue indicates water-restriction sessions. Numerical labels indicate the average z-score difference (in **A**) or the motivational state bias score (in **B**).

The analysis so far included *all* SWR events, regardless of when in a session they occurred. However, the observed bias away from the preferred outcome may emerge within a session as a result of experience on the task that day. Alternatively, this bias may be a direct result of the motivational shift alone and therefore *precede* experience that day. To distinguish between these scenarios, we computed separately the proportions of left and right SWR sequences for different task epochs. Confirming the previous analysis, an overall bias away from the preferred outcome was evident, as can be seen from the z-scored left vs. right log odds

(Figure 3a) and the proportion of events that significantly favored left or right (Figure 3b). Interestingly, this bias was clearly present during the pre-task epoch, before experience on the task that day 'pre' SWR log odds z-score food vs. water, –.55 $\pm$ .22 vs. .39 $\pm$ .30, p = .003; 'pre' significant event proportions interaction, $F_{(19,1)}$ = 7.31, p = .011) and persisted through the task and post-task epochs (all other SWR log odds z-score differences $p < .01$, all proportions $p < .035$). Note that each individual subject showed negative across-session bias scores (Figure 3), indicating a bias in SWR content away from the preferred outcome. These results did not depend on the inclusion of putative interneurons in the decoding, because repeating the analysis without these units resulted in comparable effect sizes and preserved the outcomes of all significance tests (Figure S3). Similarly, computing log odds based on the decoded posterior past the choice point (i.e. maze arms only) yielded similar results (z-scored log odds for food restriction, –0.65 $\pm$ 0.16, water restriction 0.49 $\pm$ 0.40, $p = 0.008$; significant-sequences-only interaction $F_{(df=1)} = 8.41$, $p =$ 0.007).

Sequence detection. A strength of the above sequenceless decoding analysis is that it contains no free parameters beyond how SWR event candidates are detected. However, it cannot determine whether SWR event content is organized into a sequence of locations, or the properties of that sequence (forward, backward). Thus, we performed a complementary sequence-based analysis that detects sequences in the full decoded posterior according to a "maximum-jump" criterion and overlap with SWR event candidates[4] (*Online Methods*; see Figure 4 for example sequences).

Using the sequences from the above procedure, we calculated a *single-session bias score* ($\frac{n_{left}}{n_{left}+n_{right}}$) for each session, such that a score of 0.5 indicates an equal number of left and right sequences for that session, 0 indicates all right sequences (water) and 1 indicates all left (food) sequences. We plotted this bias score across the 6 alternating food/water restriction days (Figure 5a). As in the sequenceless analysis above, the bias in sequence content was *opposite* to the animals' behavioral choices. On food-restricted days (red-shaded areas in Figure 5a) animals preferred food (black line above 0.5) but the sequence content was biased toward water (green line below 0.5). Conversely, on water-restricted days (blue-shaded areas) animals preferentially chose

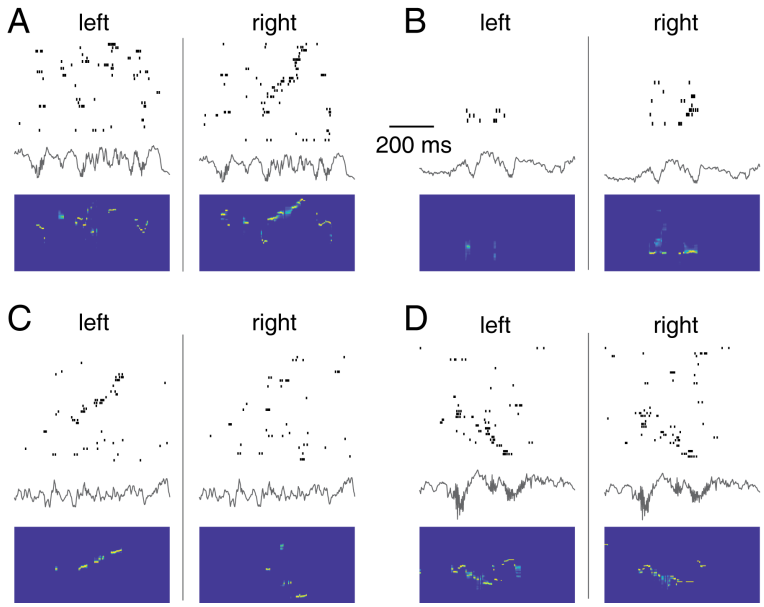

Figure 4: Hippocampal sequences detected during example sharp wave-ripple (SWR) events and their content. Panels **A-D** contain two columns, depicting spiking activity from the same time interval but using tuning curves obtained from "left" and "right" trials respectively. Within each column, a rasterplot (top) contains rows with tickmarks indicating spikes from a single neuron, ordered according to the location of their tuning curve peak firing rate on the track. Below the rasterplot is a local field potential trace (filtered between 1-425 Hz) and the decoded posterior probability distribution (blue: low probability, yellow: high probability). Each panel is an example from a different subject; examples were chosen from representative sessions and include a session with low (B, 53 cells total) and high (D, 162 cells total, note that only cells with place field(s) are shown in the rasterplots) cell counts. The events in panels A-C met the sequence detection criterion for only the right, right and left trajectory respectively and were scored as such; in contrast, the event in panel D met the sequence detection criterion for both left and right trajectories, and was therefore excluded from further analysis. In general, we found forward (A, C), reverse (D) and static sequences (B), which were considered both pooled and separately in further analyses.

water (black line below 0.5) yet sequence content, if anything, was biased toward food (green line above 0.5). A logistic regression model confirmed that motivational state predicted sequence content, changing the mean proportion of left (food) arm sequences from .35 for food-restriction sessions to .57 for water-restriction sessions (likelihood ratio test for models with and without motivational state: $\chi^2_{(df=1)} = 4.27$, $p = .038$). Sequence content was anti-correlated with behavioral choice across sessions ($r_{(n=19)}$ = –0.56, $p = .014$). Thus, overall sequence content tended to be biased in the opposite direction from the behaviorally valued choice.

As with the sequenceless results above, we next computed separately the proportions of left and right SWR

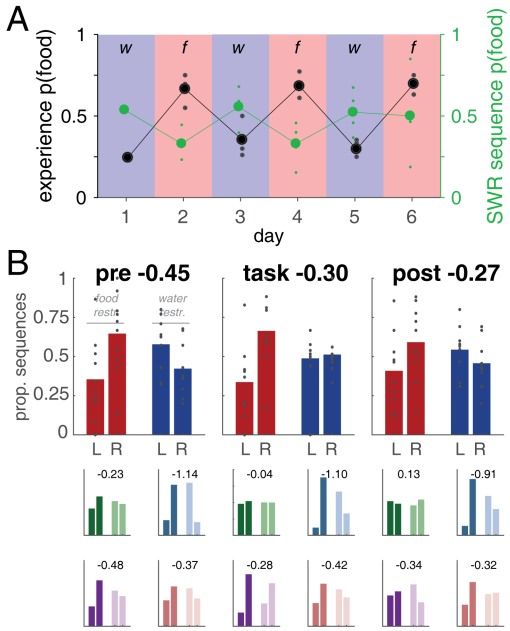

Figure 5: SWR sequence content is biased away from the behaviorally preferred maze arm. **A**: The proportion of detected SWR sequences that decode to the food (left) arm (green data points) is anticorrelated with behavioral preference (black data points). As in Figure 2a, blue shaded areas indicate water-restriction sessions and red shaded areas indicate food restriction sessions. **B**: SWR sequence content averaged across sessions is consistently biased away from the behaviorally preferred maze arm for all task epochs (pre-task, task, and post-task); figure layout as in Figure 3.

sequences for different task epochs (Figure 5b). Confirming the previous analysis, an overall bias away from the preferred outcome was evident, as can be seen from the consistently larger proportion of "water" sequences on food-restriction days (red bars) and a shift toward more "food" sequences on water-restriction days (blue bars; see Figure S4 for raw sequence counts). Again, this bias was clearly present during the pre-task epoch, before experience on the task that day (Figure 5b, left panel; $\chi^2_{df=3}$ = 30.1, $p < .001$). Next, we apply the *motivational state bias score* $[p(L) - p(R)]_{food} - [p(L) - p(R)]_{water}$ used previously for the sequenceless decoding analysis. As in Figure 3, there are some epochs and subjects in which both food and water restriction have more R (water) than L (food) sequences. However, the R (water) preference is larger for food than water restriction, as reflected in a negative bias score. Although motivational state bias scores varied between individual subjects, the motivational shift systematically biased SWR content away from the preferred choice for individual subjects and session epochs, as indicated by consistently negative bias scores (Figure 5b, insets; see Figures S5-S7 for no-interneuron, past-choice point sequences only, and forward vs. backward sequence comparisons). Thus, the bias in SWR sequence content away from the preferred outcome

preceded experience on a given day, and persisted throughout the session.

Relationship with behavior. As shown above, decoded SWR content was anticorrelated with behavioral preference on the task overall when using either the sequenceless or sequence-based measures. However, this overall relationship does not address to what extent trial-by-trial changes in SWR content may relate to behavior. To explore this issue, we constructed logistic regression models that sought to predict trial-by-trial behavioral choice based on a number of regressors, including (1) intertrial interval (ITI) content variables such as the proportion of left vs. right sequences, the most frequent and the most recent sequence, and (2) session-wide variables such as motivational state and session-wide replay content. Evidence of a true trial-by-trial relationship between SWR content and behavior requires that ITI regressors improve model fit relative to session-wide regressors alone. We failed to find any ITI regressors supporting such improvement, suggesting that either trial-by-trial differences in SWR content on this task were unrelated to behavioral choice, or there were insufficient ITI sequences and/or free-choice trials to enable a sufficiently powered test. However, including session-wide SWR content marginally improved the prediction of behavioral choice compared to motivational state alone (motivational state only, AIC 216.91; adding SWR content, AIC 208.99, likelihood ratio test $LRstat_{\Delta df=1} = 9.912$, $p = 0.0016$), suggesting that SWR content bias and behavior are related beyond what is expected from independent effects of motivational state alone.

Null hypothesis and scenarios. A crucial aspect of any analysis examining SWR content across different locations, such as the left and right arms used here, is to compare the observed distribution against an expected distribution that takes decoding accuracy into account. For instance, imagine that for a given recording session, we happened to record many more place fields on the left arm than on the right arm. Similarly, and more insidiously, it is possible that rats' behavior was subtly different on the preferred vs. non-preferred arm, potentially biasing our ability to detect SWR content. The impact of these and other factors can be systematically quantified using a leave-one-trial-out procedure, which computes a cross-validated decoding error (when the animal is running) that can be converted into a null hypothesis of expected SWR content[13]. The null hypothesis quantifies expected differences in *observed* SWR content that arise from

limited and unequal sampling, even though the *true* underlying SWR content is equal between conditions. Using this procedure, we found a small bias of lower error (higher accuracy) for the behaviorally-preferred arm, containing the restricted outcome (Figure 6a, left; note positive motivational state bias score), such that the expected null distribution of SWR sequence content favored the behaviorally-preferred arm (Figure 6a, right; this effect occurs because behavioral trials toward the preferred outcome were less variable[13]). Note that this non-uniform null distribution favors the result that motivational state biases SWR content *toward* the preferred outcome. Importantly, the data show a bias in the opposite direction from the null. Our analyses also controlled for another potential source of bias, the number of trials used to obtain tuning curves for decoding, by subsampling the number of behavioral trials to be equal for left and right trials (*Online Methods*).

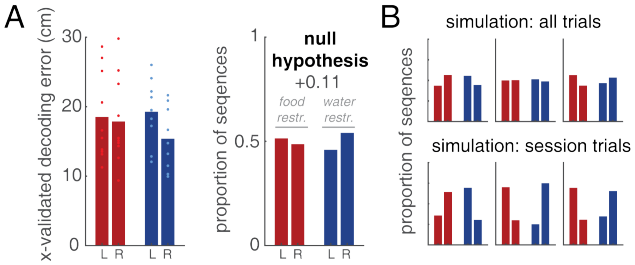

Figure 6: A: Decoding error during trials on the track, as obtained with a leave-one-trial-out cross-validation procedure[13] (25 ms window size, single dots are individual sessions; left panel) provides a null hypothesis of expected SWR sequence content proportions (right panel). The null hypothesis predicts a slight bias toward the arm containing the restricted outcome, whereas the data indicates the opposite (compare Figures 3b and 5b). **B**: Expected SWR sequence content distributions according to simulated scenarios. Top panel: sequence content reflects accumulated experience across all recording days. Bottom panel: sequence content reflects experience within single recording sessions, with pre-task content carried over from the previous day's experience. Note that neither scenario correctly predicts the observed SWR sequence content (compare Figures 3b and 5b).

Given that the observed sequence content cannot be explained by the cross-validated decoding error, and appears robust against specific parameter choices, we next considered potential explanations based on subjects' experience. Following Gupta et al.[14], we simulated scenarios in which each SWR candidate event (detected based on LFP and multi-unit activity properties without reference to actual decoded content) is probabilistically assigned a "left" or "right" label according to the following rules: (1) replay in proportion to experience accumulated over the six recording session days, (2) replay in proportion to within-session experience. For instance, say a SWR event was detected in the inter-trial period following trial 8 of day 2. If all trials that

day were “left”, scenario (2) would assign “left” to that event. If all trials up to that point (including previous sessions) were 55% “right”, then scenario (1) would assign “left” to that event with 45% probability. Since in scenario (2) within-session experience has not occurred yet during the pre-task period, we carried over the post-task distribution from the previous day. The SWR sequence content proportions expected from these scenarios are shown in Figure 6b: although a pre-task bias similar to that in the data was present, the simulations incorrectly predicted a reversal of this bias with experience. Thus, these experience-dependent scenarios could not account for the observed data.

A further possibility is that SWR sequence content is experience-dependent, but with a delay (Figure S8a). Such a delay would be necessary to explain the observed opposite-side bias that persists throughout the pre-task, task, and post-task epochs (Figures 3 and 5b). Experience-dependent increases in SWR frequency and content are generally thought to increase sharply immediately following experience, and then gradually decay[15,16,17]. However, there are also observations of replay content that persists for hours and possibly days following experience[12,18]. Thus, we tested the hypothesis that behavioral bias on day *n – 1* should predict the pre-task SWR content bias on day *n*. (Note that behavioral bias, which varies between 0.5 (equally preferred) and 1 (all left, or all right) is different from raw behavior; see Figure S8 and *Online Methods* for details.) For instance, a strong behavioral bias toward the food arm on day *n – 1* would result in a strong SWR content bias toward food the next day (Figure S8b), opposite that day’s restriction type and behavioral preference (now switched to water). However, we found no evidence for correlations between behavioral bias on day *n – 1* and SWR content bias on day *n* (sequence decoding content bias, $r_{(17)} = -0.11$, $p = 0.65$, sequenceless decoding content bias, $r_{(17)} = 0.26$, $p = 0.31$). Moreover, motivational state on day *n* was a numerically better predictor of day *n* SWR content than day *n – 1* behavior (sequence content: behavior AIC 4.15, motiv. state 3.00; non-sequence content: behavior AIC 31.2, motiv state 30.4; lower values are better). Thus, although our experiment was not explicitly designed to rule out the hypothesis that experience on day *n – 1* drives replay content on day *n*, the best statistical account of the data is that motivational state biases SWR sequence content away from the preferred outcome.

Discussion

How do these results inform theories of the content and role of SWR sequences? Two common views of what drives SWR content are that (1) SWRs replay a recency-weighted record of experience, either as part of a dual-store model of memory[19] (usually attributed to SWRs during sleep) or as part of a retrieval process that can support ongoing decisions[20], both "retrospective"; and (2) SWRs depict trajectories toward possible goals, as part of a planning process[4,14] ("prospective"). Our results are not well described by these accounts: simulations show that retrospective biases do not fit the data (Figure 6b and associated analyses). Prospectively simulating a trajectory toward the non-preferred outcome as part of a planning process would certainly be informative in decision making[21]. However, planning content is generally thought of as either agnostically considering options until some termination criterion, and/or containing a bias toward the goal (as in e.g. backward or goal-constrained search)[3,22,23]. These prospective accounts would predict either a bias toward the chosen option, or no bias. Instead, we observed an opposite bias, reminiscent of Gupta et al.[14] who found that during left or right-only blocks on a T-maze, the opposite side was replayed more often than during alternation blocks. Here, we extend that observation of replaying the "other" option to include interoceptive cues (motivational shifts) that precede actual experience as a source of such non-temporal SWR sequence biases.

If replay content on this task is not consistent with a memory trace of recent experience, or with prospective generation of upcoming choice, what could its function be? There is an emerging appreciation that hippocampally-driven consolidation may involve processes beyond the simple offline training of cortical representations based on replay of time-limited information stored in the hippocampus. This broader concept of consolidation includes processes which would benefit from replay content that is not limited to recent experience[24,25]. Examples include the learning and maintenance of predictive "successor representations"[26,27], and updating of decision variables in reinforcement learning (e.g. Dyna[28]) which may prioritize states with high uncertainty[29,30]. When applied to the task used here, we speculate that hippocampal replay enables evaluation and updating of information related to the option not (about to be) taken. These ideas echo the

less specific but related notion that replay could serve to prevent destructive interference of memory traces[14]; alternatively, under certain conditions SWRs may be part of a forgetting mechanism[31,32].

In any case, these possibilities would require that in the task used here, the pattern of SWRs observed reflects a hippocampal mode not necessarily engaged in supporting immediate behavior on the task, but operating in the service of longer-timescale behavior. This conceptualization blurs the distinction between the traditional online/offline modes to more of a continuum from highly task-constrained to more internally driven[25,33]. In this view, it is possible that even on T-maze tasks, highly overtrained animals may not be engaging their hippocampus (including SWRs) to perform the task, freeing it up for (re)consolidation processes. Consistent with this view, Singer et al.[34] found that the animals' choice was predictable from SWR content only during a specific time window during learning, and the animals in Gupta et al.[14] had been extensively trained on the task prior to recording. Thus, an important limitation of these and many other hippocampal sequence recording studies is that it is often unclear to what extent SWRs support behavior on the task; even if hippocampal lesions or inactivations result in an overall impairment, this need not imply that each SWR directly supports performance. Regardless of the role of the hippocampus and SWRs on the present task, the observation that SWR content was modulated by motivational state highlights the nontrivial ways in which SWR content can relate to experience and reward revaluation in particular.

Online Methods

Subjects and timeline. Male Long-Evans rats (n = 4) were trained on a T-maze task, described in detail below, in daily recording sessions until they were running proficiently (range: 4-15 sessions). Subjects were surgically implanted with recording electrodes targeting the dorsal CA1 area of the hippocampus. Following recovery, subjects were re-trained on the task while electrodes were lowered to the CA1 cell layer and subjects were accustomed to running the task while plugged in to the recording system. Upon acquisition of a sufficiently large ensemble of units (range: 6-9 days from start of retraining), a 6-day sequence of daily

recording sessions began, which were analyzed here. Note that this is the same data set used in van der Meer et al.[13] where we established methodological points referenced throughout the analyses reported here. All procedures were approved by the University of Waterloo Animal Care Committee (AUPP 11-06). Animals were kept on a regular day/night cycle (daylight from 7 am to 7 pm) and were run during the light phase.

Surgery and recording probes. Electrode arrays were assembled from a custom 3D-printed base (Shapeways), machined metal parts (00-80 threaded rods and stainless steel cannulas, Component Supply Company; custom fabricated metal nuts and shuttles), and an electrode interface board (EIB-36-16TT, Neuralynx) to contain 15 or 16 independently movable tetrodes (Sandvik 12.7 μm insulated NiCr wire, gold-plated to 300-400 kOhm impedance measured at 1 kHz) and 1-4 reference electrodes. Surgical procedures were as reported previously[35]. Briefly, animals were anesthetized with isoflurane, the dorsal surface of their skull was exposed, and a ground screw was placed through the contralateral parietal bone. Electrode arrays were lowered to the surface of the cortex, and the remaining exposed opening was sealed with a silicone polymer (KwikSil). Then, the arrays were anchored to the skull using small screws and acrylic cement. Rats were given a minimum recovery period of four days, during which antibiotics and analgesics were administered, before retraining began.

Behavioral procedures and apparatus. The apparatus was an elevated T-maze, constructed from wood, painted matte black with white stripes applied to the left arm (Figure 1a) and placed on a metal frame. Rats were trained to run trials on the elevated T-maze in daily sessions, having either been food or water restricted overnight (16 hours). Before beginning the task each day, rats were first given a small quantity of food (90 mg, 2 pellets) and 6% sucrose water (0.1 ml) in their home cage. (This was done because the instrumental conditioning literature shows that such re-exposure following revaluation is required for behavioral change to manifest during a subsequent extinction text, a phenomenon known as incentive learning[36].) Next, they were placed on a movable platform for a 5-20 minute pre-task rest period. Rats were trained over a period of at least 4 days to make a left or right turn at the choice point for food (five TestDiet 5TUL pellets, 45 mg each, dispensed using an automated feeder; Coulbourn) or 0.13 ml 6% sucrose water (dispensed

using tubing equipped with an electrically controlled valve) respectively. Food and water locations remained constant throughout all training and recording sessions used here. A handheld barrier was used to prevent movement in the reverse direction. Rats were allowed to remain at the end of the goal arm until the reward was consumed, or until it was clear that the rat was not interested in the reward, before being transferred to the nearest pedestal where they remained until the beginning of the next trial. Each intertrial interval (on the platform) began at a length of two minutes, and decreased with each trial to as short as one minute or less. A stationary barrier was placed at the choice point during some trials in order to force the rat down the unfavored arm a minimum of five times per session so that place fields could be estimated for both sides of the track. These blocked trials were not included in behavioral choice data (Figure 1b-c) but were included in behavioral experience data (Figures 2a and 5a and associated analyses). Training occurred in a fully lit room during the light phase. Recording sessions proceeded in the same way as training sessions, with minor modifications lengthening the pre-task period to 20-30 minutes, followed by 40 minutes of the task, and finally 15 minutes of post-task recording on the pedestal.

Data acquisition and preprocessing. Neural activity from all tetrodes and references was recorded on a Neuralynx Digital Lynx SX data acquisition system equipped with HS-36 headstages and a motorized commutator. Local field potentials were continuously sampled at 2 kHz, and recorded relative to a reference located in the corpus callosum. Spike waveforms were sampled at 32 kHz for 1 ms when the voltage exceeded an experimenter-set threshold (typically 50 μV) and sorted offline (KlustaKwik, K. D. Harris and MClust 3.5, A. D. Redish). Highly unstable cells were excluded from analysis, as well as cells with fewer than 100 spikes over the entire recording session, or poor isolation statistics. A video tracking system recorded the rat's position based on headstage LEDs picked up by an overhead camera. Signals were preprocessed to remove chewing artifacts and high-amplitude noise transients where necessary.

Session inclusion criteria. As in van der Meer et al.[13], recordings sessions were included for analysis if at least 40 cells total were recorded. This left 19 of 24 total sessions (6 each from R050 and R064, 5 from R042, and 2 from R044).

Data analysis overview. All analyses were performed using MATLAB 2017a and can be reproduced using code available on our public GitHub repository (http://github.com/vandermeerlab/papers). Data files including metadata are publicly available on DataLad (http://datasets.datalad.org, 'MotivationalT' data set). We first provide an overview of the major data analysis steps, and then discuss each step in more detail.

- **Estimating encoding models (tuning curves).** From the behavioral data on the task (position tracking) and spiking data, we estimate the relationship between location on the track and unit firing rates. We estimate these tuning curves in three conceptually and procedurally distinct sets. The first set separates left and right trials and is used to evaluate decoding accuracy when animals are running on the task using leave-one-out cross-validation[13]. A similar set of tuning curves, without leave-one-out, is used for the sequence-based decoding analysis. The third set concatenates left and right trials in a single tuning curve and is used for sequenceless decoding. A key step used for all sets is equalizing the number of trials used for "left" and "right" tuning curves.

- **Decoding.** Using the above tuning curves, we decode (1) running data for evaluating decoding accuracy to obtain a null hypothesis for decoded sequence content, and (2) SWR-associated activity only for sequenceless decoding, and (3) all activity for subsequent SWR-associated sequence detection. Key parameters of these steps include the kernel width used to obtain spike density functions, time step and window size, and the minimum number of cells active during each decoding window.

- **Sequenceless decoding content.** Because sequenceless decoding is based on tuning curves that include both left and right trials in parallel, the decoded posterior probabilities for both can be compared, and converted into a z-scored log odds relative to a shuffled distribution, reflecting SWR content.

- **Sequence candidate detection and selection.** Using the full-session decoded posterior, sequences are detected using the following key parameters: maximum jump distance, time gap allowed before sequence is broken, and minimum sequence length. A subsequent selection step only retains those

sequences that meet certain criteria, such as coinciding with a SWR candidate and passing a shuffle comparison, as described below. Further analyses apply additional criteria such as only analyzing sequences that occurred during a specific task epoch, that have certain properties such as depicting a forward or reverse trajectory or specific part of the track, et cetera. Only sequences for either left or right trials, but not both, are retained.

- **SWR detection.** From local field potential and multi-unit activity, detect candidate events. Key parameters: features used for detection, detection threshold(s).

Estimating encoding models (tuning curves). Position data was first linearized by mapping position samples to the nearest location on experimenter-drawn idealized trajectories for left and right trials separately. Linearized trajectories were binned using approx. 3 cm per bin, such that the 334 cm track length resulted in 111 bins (except for one subject, R042, who used a track length of 257 cm with the same number of bins). Next, "run" epochs were identified when the rat's speed (estimated using an adaptive filtering procedure) exceeded a fixed threshold (approx. 4 cm/s). A matched set of trials was identified such that (1) the number of left and right trials was equal, and (2) the time difference between the left and right trials was minimized. This matching procedure was necessary to prevent biases in decoding accuracy for left and right sequences due to differences in how tuning curves for left and right trials were estimated. Using these matched trials, raw average firing rates for left and right trials were estimated separately from run epochs only by dividing spike counts by occupancy for each linearized bin. Raw firing rates were smoothed with a $\sigma = 1$ bin kernel[13]. Bins at the start and end of the track were excluded from analysis. As explained in the next section, tuning curves were estimated from different sets of trials on the track, as required for analysis of SWR sequences and analysis of decoding accuracy.

Decoding. We ran three separate decoding analyses. The first analysis decodes each SWR as a single time window and determines its content (sequenceless decoding). The second analysis decodes all the data and then runs a sequence detection algorithm, and the final analysis decodes only running data (not SWRs) to

establish decoding accuracy for left and right trajectories (a control analysis which provides a null hypothesis for SWR content across conditions). These analyses differ in (a) which trials were used to estimate tuning curves, (b) what data was decoded, and (c) what further analyses were ran on the output of the decoder (all described in detail in the next sections). All decoding analyses used the standard memoryless Bayesian decoding procedure[11,13,37].

For *SWR sequenceless decoding*, we first constructed tuning curves for left and right trials as described above, and then concatenated them into a single $[nCells \text{ x } 2 * nLocationBins]$ matrix, keeping track of which locations corresponded to left and right trials respectively. Next, we used the putative SWR event intervals found by the SWR detector (described below) to decode each event as a single time bin. The decoded posterior was summed for all left and right bins separately and converted into a log odds ($log(\frac{p(L)}{p(R)})$). Repeating this procedure 1000 times for tuning curves in which random subsets of spatial locations had their left and right tuning curves swapped yielded a shuffled distribution which was used to obtain two SWR content measures: a z-score (averaged over all events in a session, yielding a single number) and a significant event count. The latter was obtained by including only those events that exceeded either the left- or rightmost 2.5% of the shuffled distribution, yielding left and right counts for each session. Only SWR events with at least 4 neurons active were included. To compare SWR content across multiple sessions, we defined a *motivational state bias score* $[p(L) - p(R)]_{food} - [p(L) - p(R)]_{water}$, where $p(L, R)$ are the proportions of sequences depicting left and right trajectories on the track respectively, and the food and water subscripts indicate the session restriction type.

For *SWR sequence decoding*, we first obtained "left" and "right" tuning curves with matched trials as described above, and used these to obtain left and right decoded posteriors by decoding spiking data from the entire session. Spike trains were converted to spike density functions by convolving with a $\sigma = 2$ ms Gaussian kernel[13]. We used the canonical one-step Bayesian decoding procedure with a uniform spatial prior and a 25 ms moving window, which was stepped in 5 ms increments. At least 4 cells had to be active in each window for decoded output to be produced. The output of this decoding procedure was then used for

sequence detection analysis, described below.

For the analysis of decoding accuracy, we decoded each trial separately, using tuning curves obtained from all other trials of the same type (left or right). This leave-one-out procedure tests how well tuning curves generalize to the decoding of withheld trials, a situation that mimics some aspects of generalizing tuning curves to decoding of SWR sequences[13]. The output from this decoding procedure was used to compute decoding error as the mean distance between the maximum likelihood (or equivalently, maximum a posteriori, since we used a uniform prior) decoded location and the true, observed location. Decoding parameters were the same as above (25 ms window, $\sigma = 2$ ms, minimum 4 cells active per bin). To derive the null hypothesis of sequence content, decoding errors were converted into proportions as $(1 - p_l) = \frac{err_l}{err_l + err_r}$ and $(1 - p_r) = \frac{err_r}{err_l + err_r}$.

Sequence detection and selection. To detect candidate sequences from the decoded output, we applied the procedure introduced by Pfeiffer and Foster[4]. A candidate sequence starts at the first available sample and extends to the next sample, unless the distance between decoded locations (defined as the bin with the maximum likelihood) in two adjacent time steps exceeds a threshold (40 cm unless described otherwise) or no decoded location was available (due to insufficient number of cells being active at that time step). These conditions break the candidate sequence, and a new one is started at the next available decoded sample. The resulting pool of candidate sequences is then subjected to further criteria, starting with a minimum length (50 ms), a minimum number of neurons participating in the event (4), overlap with a putative SWR for any non-zero amount of time (see the next section for how these are detected). Next, for each time bin of each remaining candidate sequence, we computed in what fraction of identity-permuted shuffles (n = 500) that bin was also included in a sequence. Only candidate sequences in which the average proportion of shuffle-included bins was 0.05 or less were retained for further analysis. Finally, because this analysis was performed separately for “left” and “right” trajectories, based on their respective tuning curves, the final step only retained those sequences that were classified as either left or right, but not both. More precisely, any ambiguous sequence that overlapped for any amount of time with any opposite-side sequence resulted in

removal of both sequences.

The results reported in the main text used our default analysis parameters, which allowed no "jumps" or "skips" in the sequence detection and included all sequences, regardless of whether they were forward, reverse, or neither. In the supplementary figures we report results with different parameter settings and selection steps, including for sequences that are forward, reverse, and other restrictions. To determine if a sequence was forward or reverse, the beta coefficient for the time regressor needed to be significantly different from zero in predicting location in a linear regression (t-statistic, $p < .05$); because animals ran the track in one direction only, positive betas indicate forward sequences and negative betas indicate reverse sequences.

Sequence content analysis. The output of the above procedure is, for each session, a set of detected sequences, which are intervals (start and end times) with a label indicating which trajectory on the track they depict (left/right). Because the raw number of sequences detected will vary substantially due to factors of no relevance to the main questions of interest, such as the number of cells recorded or the number of SWRs detected on any given day, the main analyses use *proportions* of sequences. As above, we defined a *single session bias score*, which is simply the fraction of sequences for each trajectory and the total number of sequences, e.g. $\frac{n_{left}}{n_{left}+n_{right}}$ for left sequences. To compare sequence content across multiple sessions, we used the *motivational state bias score* $[p(L)-p(R)]_{food}-[p(L)-p(R)]_{water}$ as in the sequenceless analysis above.

SWR event detection. For the results in the main text, we used a thresholding procedure based on spectral content of the local field potential, combined with adaptive thresholding of multi-unit activity, to identify putative sharp wave-ripple complexes. Briefly, for each recording session, a noise-corrected SWR Fourier spectrum is obtained by manual identification of example SWR events, and subtracting a non-SWR "noise" Fourier spectrum. The dot product between this noise-corrected frequency spectrum and the frequency spectrum of the LFP is then obtained using a sliding 60-ms window to obtain a "SWR similarity score". Separately, a

multi-unit activity detector was defined, starting from a spike density function obtained by convolving the spiking activity of all recorded units with a σ = 20 ms kernel. An adaptive contrast enhancement step was applied to the raw MUA to emphasize transient changes in firing. Finally, a composite SWR detector was obtained by taking the geometric mean of the noise-corrected frequency spectrum and the modified MUA. The resulting unitless quantity was thresholded at 4 to yield candidate events, which were further restricted by the following requirements: (1) subjects could not be moving faster than 4 cm/s, (2) theta power could not be more than 2 SDs above the mean, (3) minimum duration of 20 ms, and (4) minimum 5 units active. We believe this method performs the best of those we have tested in identifying true SWR events while minimizing false positives. Rats appeared to be mostly awake during recording sessions, although it is likely some SWRs in the pre- and post-task rest periods occurred during early sleep stages.

Statistical analyses. For the regression analyses linking behavior and SWR content, we fit linear mixed models in MATLAB (R2017a, *fitglme()* function). Results reported are for best model fit as determined by the lowest AIC, among alternative models fit with and without subject-specific intercepts and slopes. Formal model comparisons were performed with a likelihood ratio test (Williams, 2001; MATLAB *compare()* function). Correlations (Pearson's r) and associated p-values were computed with the MATLAB *corrcoef()* function.

Supplementary Figures

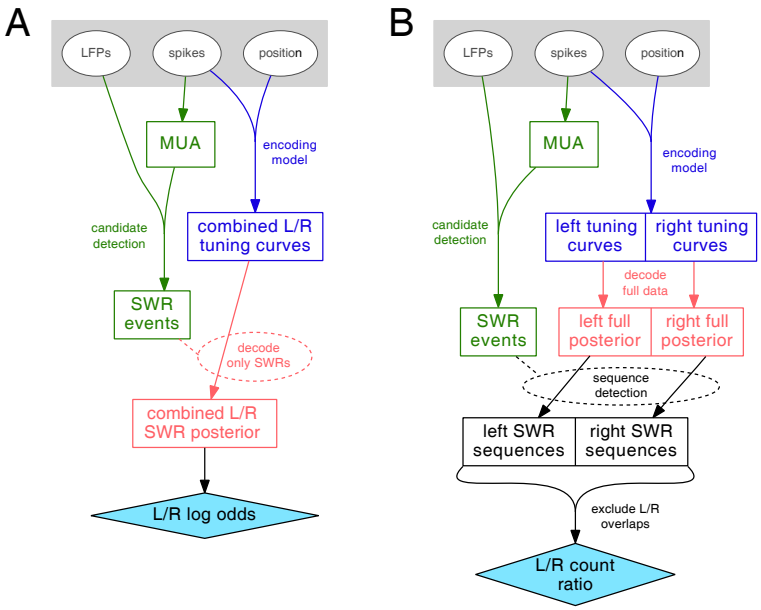

Figure S1: Schematic of SWR content analysis using sequenceless decoding (**A**) and sequence-based decoding (**B**). In the sequenceless decoding analysis, each candidate SWR event is decoded as a single time bin, using joint tuning curves containing left and right trial data. In contrast, sequence-based decoding first decodes all data in a given session for left and right trial tuning curves separately, using a 25 ms moving window (step size, 5 ms). Next, sequences are detected in the left and right posteriors separately, before removing sequences that (1) did not overlap with a candidate SWR event, (2) were not sufficiently distinct from the number of sequences obtained from a random resampling procedure, or (3) were sequences for *both* left and right.

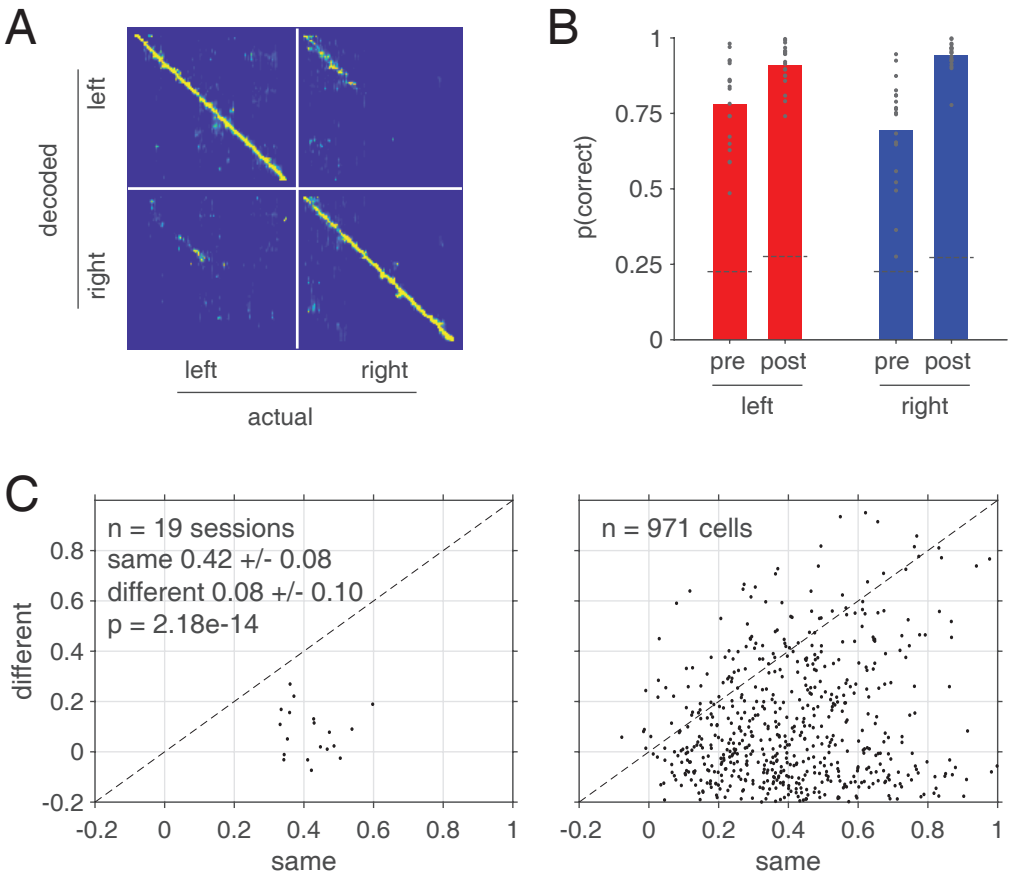

Figure S2: Left and right trials are clearly distinguishable during running on the track. **A**: Decoding confusion matrix for a single example session. Each decoded time bin (bin size = 100 ms) is assigned to the corresponding actual trial type (left, right) and position of the animal (horizontal axis). Decoded posteriors for all bins at that location are averaged to obtain the values in each column of the matrix. Perfect decoding would result in all diagonal elements being 1. Note the clear diagonal indicating non-random decoding overall, and the fainter, off-diagonal elements corresponding to confusion of left and right trials. **B**: Average classification performance across all subjects and sessions, based on the trial type (left, right) and location (before the choice point, 'pre'; or after, 'post') of the maximum a posteriori decoding. Note that for all trial types and locations, classification performance is clearly above chance (all $p < 0.001$). **C**: Single-trial tuning curve correlations between trials of the same type (left-left, right-right) are systematically higher than correlations between trials of different type (left-right), both when averaged across all cells within a session (left panel) and on a cell-by-cell basis (right panel). Only positions on the common, central stem of the maze were included in this analysis. The higher correlations between trials of the same type, compared to the correlations across types, indicate left and right trials are distinct. Thus, as measured by two different approaches, left and right trials are clearly distinguishable during behavior on the track, even on the common portion of the maze.

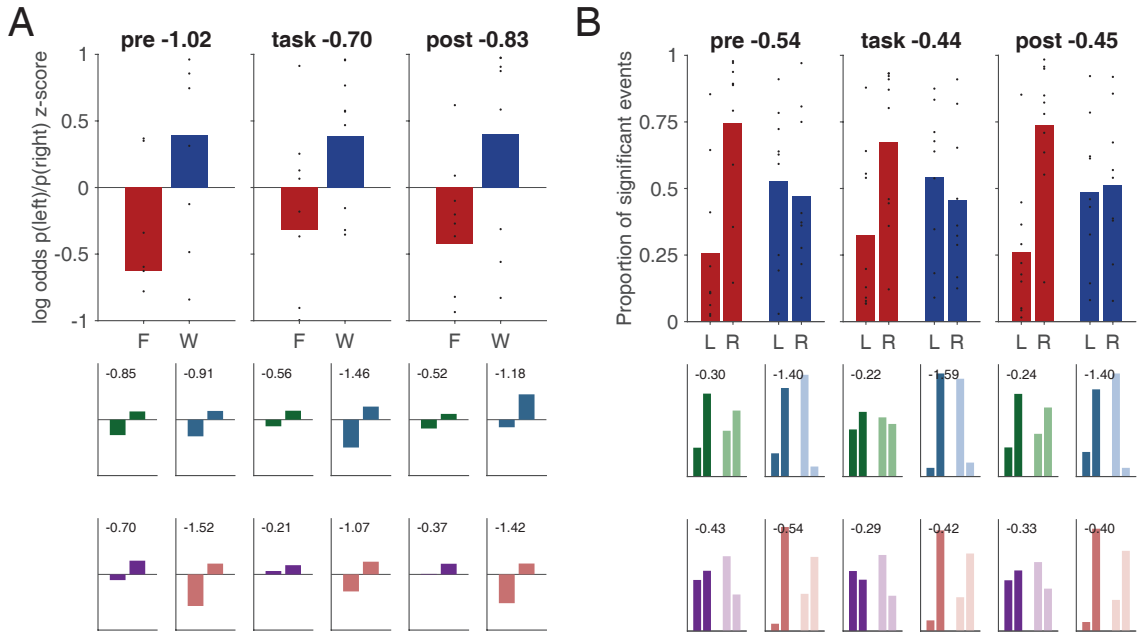

Figure S3: Sequenceless decoding results with putative interneurons excluded from the analysis.

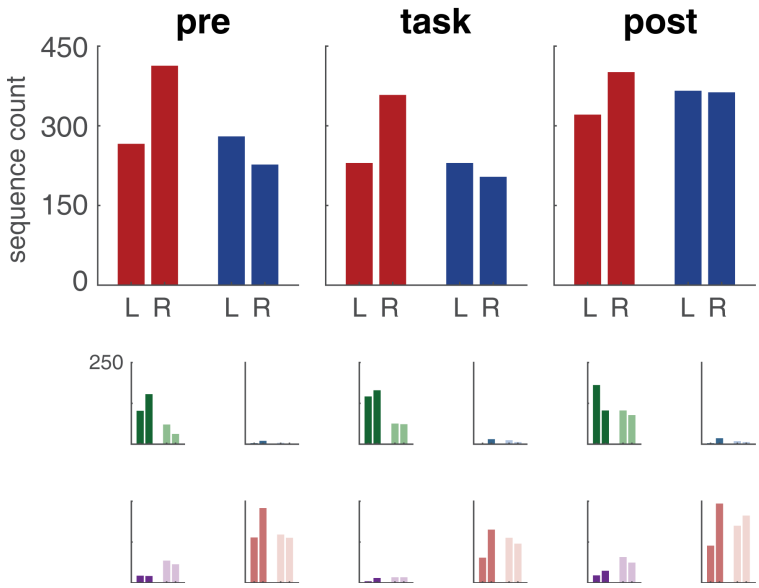

Figure S4: Raw counts of sequences detected.

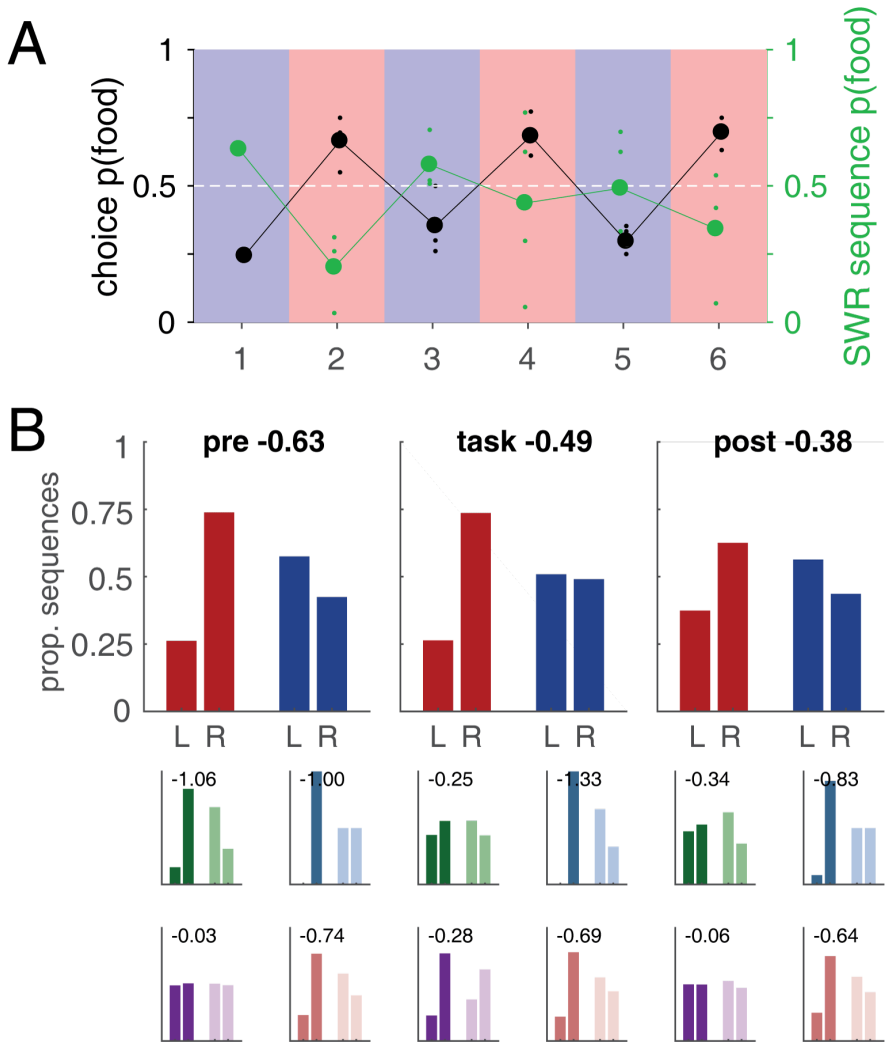

Figure S5: Sequence decoding results with putative interneurons excluded from the analysis.

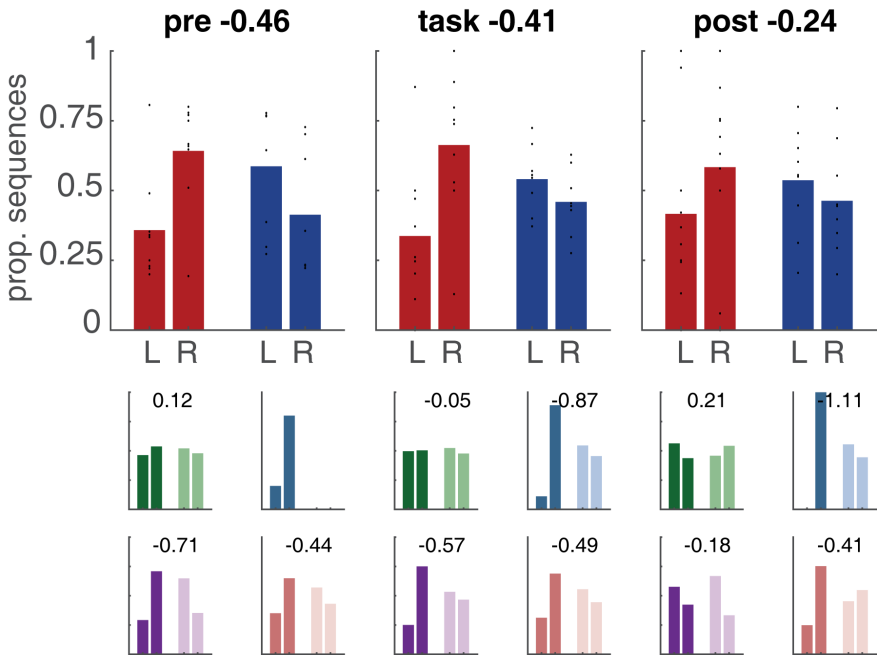

Figure S6: Sequence decoding results when including only those sequences with a start or end location past the choice point.

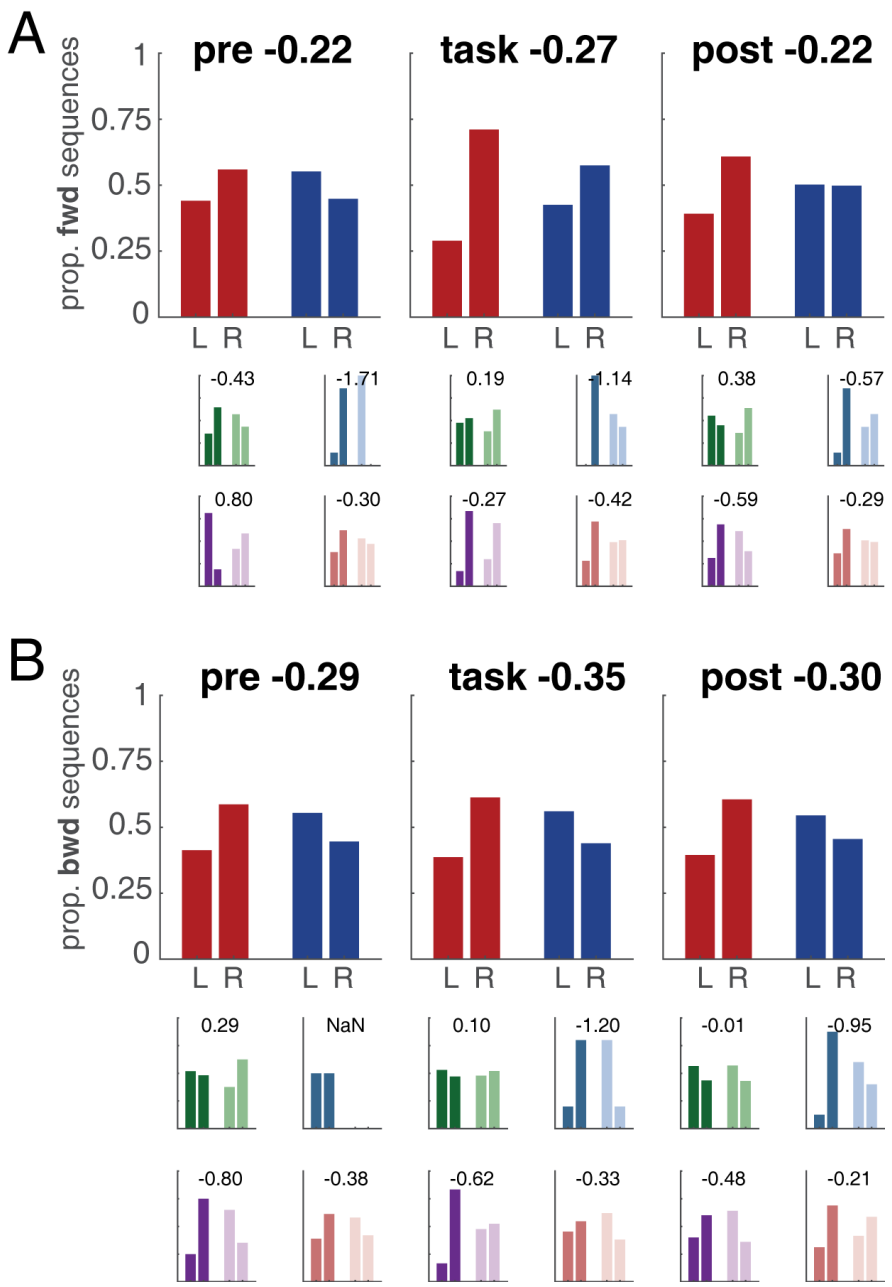

Figure S7: Sequence decoding results shown separately for forward (fwd, **A**) and backward (bwd, **B**) sequences.

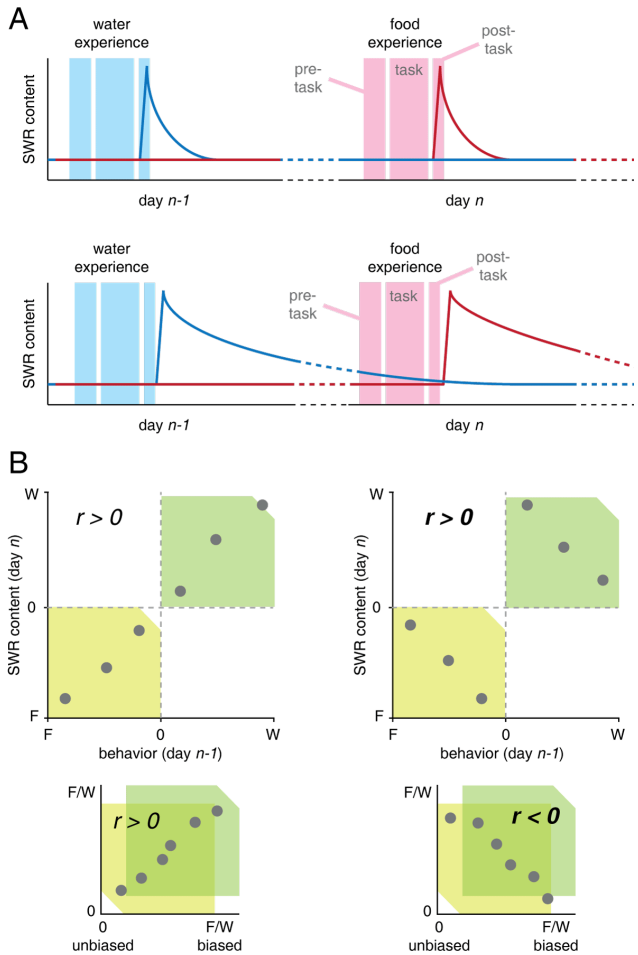

Figure S8: Schematic depicting the hypothesis that SWR content on the task is a delayed reflection of recent experience (**A**) and its predictions (**B**). Prevailing views of the effects of experience on replay hold that SWR content should be biased toward recent experience, which predicts that in the post-task epoch following water choices, SWR content should be biased toward water (and the opposite for food, **A**, top) before returning to baseline. However, it is possible that these effects of experience are sufficiently delayed, as well as sufficiently persistent, to be absent during the post-task epoch and carry over into the next day's recording session (**A**, bottom). This hypothesis predicts a specific relationship between behavior on day $n-1$ and SWR content on day n (**B**, top left). It is tempting to use a direct correlation between these variables as a test of this hypothesis; however, as illustrated in panel **B** (top right) this test lacks specificity because even the opposite relationship would result in an overall positive correlation! The correct test, illustrated on the bottom row, uses $bias$, defined as $max(p_{left}, 1 - p_{left})$, for both behavior and SWR content. Bias varies between 0.5 for no preference (equal left and right) and 1 for all left or all right. Geometrically, the bias aligns the food and water quadrants to a common axis (**B**, bottom row). Note that now the prediction of a positive correlation is specific to the hypothesis tested.

References

[1] van der Meer, M. A. A., Kurth-Nelson, Z. & Redish, A. D. Information processing in decision-making systems. *The Neuroscientist* **18**, 342–59 (2012).

[2] Buzsáki, G. Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. *Hippocampus* **25**, 1073–188 (2015).

[3] Foster, D. J. Replay Comes of Age. *Annual Review of Neuroscience* **40**, 581–602 (2017).

[4] Pfeiffer, B. E. & Foster, D. J. Hippocampal place-cell sequences depict future paths to remembered goals. *Nature* **497**, 74–9 (2013).

[5] Jadhav, S. P., Kemere, C., German, P. W. & Frank, L. M. Awake hippocampal sharp-wave ripples support spatial memory. *Science* **336**, 1454–8 (2012).

[6] Dolan, R. J. & Dayan, P. Goals and habits in the brain. *Neuron* **80**, 312–325 (2013).

[7] Kennedy, P. J. & Shapiro, M. L. Motivational states activate distinct hippocampal representations to guide goal-directed behaviors. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 10805–10 (2009).

[8] Hebben, N., Corkin, S., Eichenbaum, H. & Shedlack, K. Diminished ability to interpret and report internal states after bilateral medial temporal resection: Case H.M. *Behavioral Neuroscience* **99**, 1031–1039 (1985).

[9] Davidson, T. & Jarrard, L. E. A role for hippocampus in the utilization of hunger signals. *Behavioral and Neural Biology* **59**, 167–171 (1993).

[10] Kennedy, P. J. & Shapiro, M. L. Retrieving Memories via Internal Context Requires the Hippocampus. *Journal of Neuroscience* **24** (2004).

[11] Zhang, K., Ginzburg, I., McNaughton, B. L. & Sejnowski, T. J. Interpreting Neuronal Population Activity by Reconstruction: Unified Framework With Application to Hippocampal Place Cells. *Journal of Neurophysiology* **79**, 1017–1044 (1998).

[12] Karlsson, M. P. & Frank, L. M. Awake replay of remote experiences in the hippocampus. *Nature Neuroscience* **12**, 913–8 (2009).

[13] van der Meer, M. A. A., Carey, A. A. & Tanaka, Y. Optimizing for generalization in the decoding of internally generated activity in the hippocampus. *Hippocampus* **27**, 580–595 (2017).

[14] Gupta, A. S., van der Meer, M. A. A., Touretzky, D. S. & Redish, A. D. Hippocampal replay is not a simple function of experience. *Neuron* **65**, 695–705 (2010).

[15] Kudrimoti, H. S., Barnes, C. A. & McNaughton, B. L. Reactivation of hippocampal cell assemblies: effects of behavioral state, experience, and EEG dynamics. *Journal of Neuroscience* **19**, 4090–101 (1999).

[16] Foster, D. & Wilson, M. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* **440**, 680–683 (2006).

[17] Eschenko, O., Ramadan, W., Mölle, M., Born, J. & Sara, S. J. Sustained increase in hippocampal sharp-wave ripple activity during slow-wave sleep after learning. *Learning & Memory* **15**, 222–228 (2008).

[18] Louie, K. & Wilson, M. A. Temporally Structured Replay of Awake Hippocampal Ensemble Activity during Rapid Eye Movement Sleep. *Neuron* **29**, 145–156 (2001).

[19] Wang, S.-H. & Morris, R. G. M. Hippocampal-neocortical interactions in memory formation, consolidation, and reconsolidation. *Annual Review of Psychology* **61**, 49–79, C1–C4 (2010).

[20] Carr, M. F., Jadhav, S. P. & Frank, L. M. Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nature Neuroscience* **14**, 147–53 (2011).

[21] Wu, C.-T., Haggerty, D., Kemere, C. & Ji, D. Hippocampal awake replay in fear memory retrieval. *Nature neuroscience* **20**, 571 (2017).

[22] Ólafsdóttir, H. F., Bush, D. & Barry, C. The Role of Hippocampal Replay in Memory and Planning. *Current Biology* **28**, R37—-R50 (2018).

[23] Pfeiffer, B. E. The content of hippocampal ”replay”. *Hippocampus* (2017).

[24] Javadi, A.-H., Tolat, A. & Spiers, H. J. Sleep enhances a spatially mediated generalization of learned values. *Learning & memory* **22**, 532–6 (2015).

[25] Pezzulo, G., Kemere, C. & van der Meer, M. A. A. Internally generated hippocampal sequences as a vantage point to probe future-oriented forms of cognition. *Annals of the New York Academy of Sciences* (2017).

[26] Stachenfeld, K. L., Botvinick, M. M. & Gershman, S. J. The hippocampus as a predictive map. *Nature neuroscience* **20**, 1643 (2017).

[27] Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J. & Daw, N. D. Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS computational biology* **13**, e1005768 (2017).

[28] Sutton, R. S. First results with Dyna, an integrated architecture for learning, planning and reacting. *Neural Networks for Control* 179 (1990).

[29] Momennejad, I., Otto, A. R., Daw, N. D. & Norman, K. A. Offline Replay Supports Planning: fMRI Evidence from Reward Revaluation. *bioRxiv* 196758 (2017).

[30] Mattar, M. G. & Daw, N. D. Prioritized memory access explains planning and hippocampal replay. *bioRxiv* 225664 (2018).

[31] Colgin, L. L., Kubota, D., Jia, Y., Rex, C. S. & Lynch, G. Long-term potentiation is impaired in rat hippocampal slices that produce spontaneous sharp waves. *The Journal of physiology* **558**, 953–961 (2004).

[32] Mehta, M. R. Cortico-hippocampal interaction during up-down states and memory consolidation. *Nature Neuroscience* **10**, 13–15 (2007).

[33] Olafsdottir, H. F., Carpenter, F. & Barry, C. Task Demands Predict a Dynamic Switch in the Content of Awake Hippocampal Replay. *bioRxiv* (2017).

[34] Singer, A. C., Carr, M. F., Karlsson, M. P. & Frank, L. M. Hippocampal SWR activity predicts correct decisions during the initial learning of an alternation task. *Neuron* **77**, 1163–73 (2013).

[35] Malhotra, S., Cross, R. W., Zhang, A. & Van Der Meer, M. A. A. Ventral striatal gamma oscillations are highly variable from trial to trial, and are dominated by behavioural state, and only weakly influenced by outcome value. *European Journal of Neuroscience* **42**, 2818–2832 (2015).

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

[37] Davidson, T. J., Kloosterman, F. & Wilson, M. A. Hippocampal replay of extended experience. *Neuron* **63**, 497–507 (2009).