bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Replay as context-driven memory reactivation

Zhenglong Zhou*, Michael J. Kahana†, & Anna C. Schapiro†

Department of Psychology, University of Pennsylvania

Abstract

Replay in the brain is not a simple recapitulation of recent experience, with awake replay often unrolling in reverse temporal order upon receipt of reward, in a manner dependent on reward magnitude. These findings have led to the proposal that replay is optimized for learning value-based predictions in accordance with reinforcement learning theories. However, other characteristics of replay are in tension with this proposal, leaving it unclear whether one set of principles governs all replay. We offer a parsimonious memory-focused account, suggesting that the brain associates experiences with the contexts in which they are encoded, at rates modulated by the salience of each experience. During periods of quiescence, replay emerges when contextual cues trigger a cascade of reactivations driven by the reinstatement of each memory's encoding context, which in turn facilitates memory consolidation. Our theory unifies numerous disparate replay phenomena, including findings that existing models fail to account for.

Introduction

Sleep and rest facilitate learning and memory [1, 2, 3]. The replay of neural activity associated with awake behaviors may serve as a mechanism of such offline learning. Over the past several decades, the field of neuroscience has accumulated extensive evidence of replay across animals and humans [4, 5, 6], and there is increasing evidence of its utility to behavior [7, 8, 9]. Replay was initially characterized as the reinstatement of multi-cell spiking patterns of awake experiences. However, subsequent studies revealed that replay does not faithfully obey the statistics and temporal structure of waking experience: Replay over-represents salient experiences [10, 11, 12], unrolls in the reverse order of behavioral sequence upon the receipt of reward [13, 14, 15, 16], and contains never-experienced novel trajectories [17]. These deviations of replay from behavioral sequences call into question the characterization of replay as a simple, direct recapitulation of previous experience. Despite a vast array of observations, it remains unclear how to characterize the nature and purpose of replay.

One influential perspective rooted in the framework of reinforcement learning (RL) characterizes replay as serving value-based offline learning [13, 14, 18, 19, 20]. According to this view, during active interaction with the environment, the animal stores the state transitions and action outcomes it observes as memories. In offline periods, the brain replays these memories to update existing value expectations about outcomes of different actions, in order to improve action policy. RL-based theories posit that replay prioritizes memories of high utility to future behavior, which is consistent with observations that replay over-represents [10, 11] and tends to initiate at rewarded items [21, 15]. The priority of a memory for replay was postulated as its associated reward prediction error [20]. A recent RL model [19] reformulated priority as the expected value of backup (EVB) — the expected change in future reward if a memory is replayed. EVB is the product of a *need* term that prioritizes states that the animal expects to visit soon, and a *gain* term prioritizing states preceding reward prediction errors. By sequentially replaying experiences in order from the highest

*Correspondence: zzhou34@sas.upenn.edu
†Equal contribution

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to the lowest EVB, the model explains a number of replay phenomena, several of which [13, 22, 21, 12, 11] eluded existing accounts of replay. However, the EVB model does not delineate a plausible mechanism of how replay is generated due to its unrealistic assumption that the behavioral consequence of a replaying a memory is known before it is performed. Furthermore, a host of empirical phenomena, such as the decoupling of replay from animals' behavioral preference [23, 24, 17], are not naturally accounted for by this perspective.

Here we offer an alternative theoretical account in which replay reflects memory reactivations guided by associations between contexts and experiences, rather than a memory's utility for learning value predictions. We hypothesize that, during active wakefulness, the animal sequentially associates experiences with the contexts of their occurrence, in a manner modulated by the salience of each experience. During quiescence (both awake rest and sleep), replay arises when spontaneous reactivation of a memory sets off a chain of memory reactivation driven by the reinstatement of memories' associated contexts. In our account, replay does not unroll according to the utility of the memories for learning value predictions. Instead, it arises from a simple memory mechanism operating recursively on established associations between contexts and experiences.

We show that an instantiation of this account — a computational model that extends established context-based memory encoding and retrieval mechanisms [25, 26] unifies numerous replay phenomena including findings that existing models fail to account for. First, in our model, the content and structure of replay sequences vary according to task and behavioral contexts in a manner consistent with previous rodent studies [27, 21, 14, 28]. Second, the model captures prominent effects of valence on properties of replay [13, 10, 11] despite not maintaining nor updating value representations. Third, in line with previous studies [29, 17, 30, 15, 31], replay is not restricted to direct recent experience: The model reactivates non-local and never-experienced novel trajectories. Moreover, our model captures a range of experience-dependent replay phenomena [32, 23, 22, 21, 17, 33, 30, 34], including findings that existing models do not explain. Finally, replay affords additional opportunities for offline learning that benefits memory consolidation in the model, in ways that align with prior observations and theories [35, 36, 37, 18, 16, 38, 39]. As a whole, we outline a general, mechanistic framework to characterize how memories are initially encoded and subsequently reactivated in the service of memory consolidation.

Results

A context model of memory replay

The proposed model of replay builds on retrieved-context theories, as exemplified in the context-maintenance and retrieval model (CMR: [25, 26]). Our implementation, termed CMR-replay (Figure 1b), distinguishes between memory processes operating during interaction with the environment (i.e., active wakefulness) and quiescence. During active wakefulness, our model follows the basic principles of retrieved context theory, storing bidirectional associations between contexts and experiences M^{fc} and M^{cf}, which respectively denote experience-to-context and context-to-experience associations. These associations, in turn, support cue-dependent retrieval, wherein an experience can evoke its past associated contexts via M^{fc} and the present state of context can retrieve its associated experiences via M^{cf}.

Across both wakefulness and quiescence, each experience f_t drives the evolution (or drift) of context according to:

$$c_t = \rho c_{t-1} + \beta c_{f_t} \tag{1}$$

where the evolved context c_t is a weighted combination of the previous context c_{t-1} and the experience's retrieved context c_{f_t}, which f_t retrieves via M^{fc}. Therefore, c_t is a recency-weighted sum of the contexts associated with experiences up to time t. After updating context, CMR-replay strengthens memory associations between the current experience and the updated context using a standard Hebbian learning rule

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

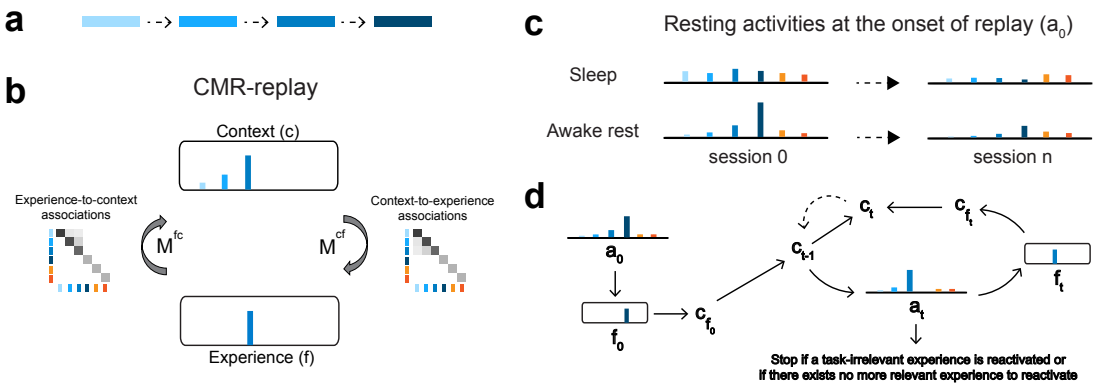

Figure 1: **CMR-replay**. **a**. Consider a task in which the model encodes a sequence consisting of four items, each denoted by a shade of blue. **b**. CMR-replay consists of four components: experience (f), context (c), experience-to-context associations (M^{fc}), and context-to-experience associations (M^{cf}). At each timestep during awake encoding, f represents the current experience and c is a recency-weighted average of contextual features associated with prior experiences. CMR-replay associates f and c at each timestep, updating M^{fc} and M^{cf} according to a Hebbian learning rule (Equations 2-3). M^{fc} and M^{cf} respectively support the retrieval of an experience's associated context and a context's associated experiences. During replay, f represents each reactivated experience and c represents a context that drifts according to reactivated experiences. Here, too, the model updates M^{fc} and M^{cf} according to reactivated f and c. The figure illustrates the representations of f, c, M^{fc}, and M^{cf} as the model encodes the third input during learning. Lengths of color bars in f and c represent relative magnitudes of different features. Shades of grey illustrate the weights in M^{fc} and M^{cf}. Orange features represent task-irrelevant items, which do not appear as inputs during wake learning but compete with encoded items for reactivation during replay. **c**. Consider the activation of items at the onset of sleep and awake rest across sessions of learning. At replay onset, an initial probability distribution across experiences a_0 varies according to the behavioral state (i.e., awake rest or sleep). Compared to sleep, during awake rest, a_0 is strongly biased toward features associated with external inputs during awake rest. For awake rest, the figure shows an example of a_0 when the model receives input features related to the fourth item. Through repeated exposure to the same task sequence across sessions of learning, activities of the four task-related items (i.e., blue items) become inhibited in a_0 relative to task-irrelevant items (i.e., orange items). **d**. Each replay period begins by sampling an experience $f_{t=0}$ according to a_0, where t denotes the current timestep. If $f_{t=0}$ is a task-related item, its associated context $c_{f_{t=0}}$ is reinstated as c_0 to enter a recursive process. During this process, at each timestep $t \geq 1$, c_{t-1} evokes a probability distribution a_t that excludes previously reactivated experiences. Given a_t, the model samples an experience f_t and reinstates f_t's associated context c_{f_t}, which is combined with c_{t-1} to form a new context c_t to guide the ensuing reactivation. Dashed line indicates that c_t becomes c_{t-1} for the next time step. At any t, the replay period ends with a probability of 0.1 or if a task-irrelevant item is reactivated.

according to:

$$\triangle M^{fc} = \gamma_{fc} c_t f_t^T \tag{2}$$

$$\triangle M^{cf} = \gamma_{cf} f_t c_t^T \tag{3}$$

where γ_{fc} and γ_{cf} denote the encoding rates for M^{fc} and M^{cf}. Through this process, CMR-replay associates experiences that appear in similar encoding contexts with similar contexts via M^{fc} and M^{cf}. Due to context drift, the similarity between encoding contexts is the highest between successive experiences.

In CMR-replay, during active wakefulness, context drift is driven by external inputs — a sequence of events that unfolds as the agent interacts with the environment or receives a stream of sensory inputs. Building on prior work [40, 41], CMR-replay embraces the assumption that the salience of each event influences its rate of encoding (i.e., γ_{fc} and γ_{cf}): CMR-replay updates memory associations at higher rates for novel and rewarding experiences than for others.

By contrast, internally-generated sequential activity drives context drift during quiescence. Here, we extend the CMR framework to distinguish between memory-dynamics during periods of engagement and those of

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

quiescence. For simplicity, we assume that each stretch of quiescence has a fixed number of replay periods for experiences to be reactivated. At the onset of each replay period, an initial probability distribution a_0 (Fig. 1c) represents the likelihood of reactivation across experiences, in which the probability of the i-th unit is:

$$[a_0]_i = \frac{[{a_0}^{spontaneous} + \lambda {a_0}^{retrieved}]_i}{\sum_{j=1}^{n}[{a_0}^{spontaneous} + \lambda {a_0}^{retrieved}]_j} \tag{4}$$

where ${a_0}^{spontaneous}$ represents spontaneous internal activity that we simulated as random noise, ${a_0}^{retrieved}$ is activity that task-related external context cues $c_{external}$ (e.g., the task context of the animal's current location, or task-related cues presented during sleep) evoke according to Eq. 5, and n is the total number of experiences. In our simulations, task-related external context cues are available during local awake rest, or during the re-presentation of learning-related cues in quiescence (Targeted Memory Reactivation, TMR; [42]). In awake rest, the presence of task-related cues biases a_0 to favor their associated experiences (Fig. 1c). As a result, awake replay exhibits an initiation bias — a tendency to initiate at the item associated with $c_{external}$. To account for effects of experience on replay, we incorporated a mechanism that increasingly inhibits task-related items in a_0 through repetition in the same task (Fig. 1c). This mechanism computes the level of inhibition for items in a_0 according to the magnitude of their retrieved contexts in the preceding wake learning session (see Methods for implementation details).

At the onset (i.e., $t = 0$) of each replay period, CMR-replay samples an item $f_{t=0}$ from a_0 and reinstates $f_{t=0}$'s associated context $c_{f_{t=0}}$ as c_0 to enter a series of recursive computations (Fig. 1d). In this recursive process, at each timestep $t \geq 1$, c_{t-1} (i.e., the context at the previous timestep) evokes a probability distribution a_t that excludes items reactivated prior to t. Let U_t denote the set of items that have not been reactivated prior to t. The probability of each item in U_t is:

$$[a_t^{retrieved}]_i = \frac{\exp([M^{cf}c]_i/T_t)}{\sum_{\forall f_j \in U_t}\exp([M^{cf}c]_j/T_t)} \tag{5}$$

where $c = c_{external}$ at $t = 0$ when $c_{external}$ is available, $c = c_{t-1}$ for $t \geq 1$, and T_t is a temperature parameter that scales the relative difference of activities in ${a_t}^{retrieved}$. Based on Eq. 5, all U_t items have non-zero activity in a_t. From a_t, the model samples an experience f_t, whose associated context c_{f_t} enters c_{t-1} to form a new context c_t according to Eq. 1. This process is recursive since c_t guides reactivation at timestep $t + 1$ (Fig. 1d). At each timestep, the replay period ends with a probability of 0.1 or if a task-unrelated experience becomes reactivated. Following prior work [19], we consider replayed sequences (one per replay period) with consecutive segments of length five or greater that preserve the contiguity of wake inputs as replay events. In CMR-replay, replay facilitates a process of gradual offline learning in the absence of external inputs: During replay, the model updates M^{fc} and M^{cf} to strengthen the association between reactivated f_t and c_t at a slower rate compared to wake learning.

Because M^{cf} maps each context to experiences encoded in similar contexts, when c_{t-1} evokes a_t via M^{cf}, the activity of each experience in a_t will reflect the similarity between c_{t-1} and the experience's associated context. Since a key component of c_{t-1} is f_{t-1}'s associated context, this similarity favors awake experiences immediately preceding or following f_{t-1}. Hence, CMR-replay tends to successively reactivate experiences encoded back-to-back during behavior (Fig. 6d left), giving rise to coherent sequences despite the stochasticity of replay (i.e., replay events comprise a small subset of all replay sequences; Fig. 5a left). In the following sections, we will refer to this property as CMR-replay's contiguity bias.

We simulate wake learning in a number of tasks [13, 27, 23, 21, 17, 15, 18, 11] by exposing the model to sequences of experiences that correspond to trajectories of spatial locations or lists of visual stimuli that obey task or environmental constraints. In between sessions of wake learning, we simulate awake rest and/or sleep as periods of autonomous reactivation. In the following sections, we describe sets of empirical phenomena that the model captures, including context-dependent variations of replay, effects of valence on replay, replay

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

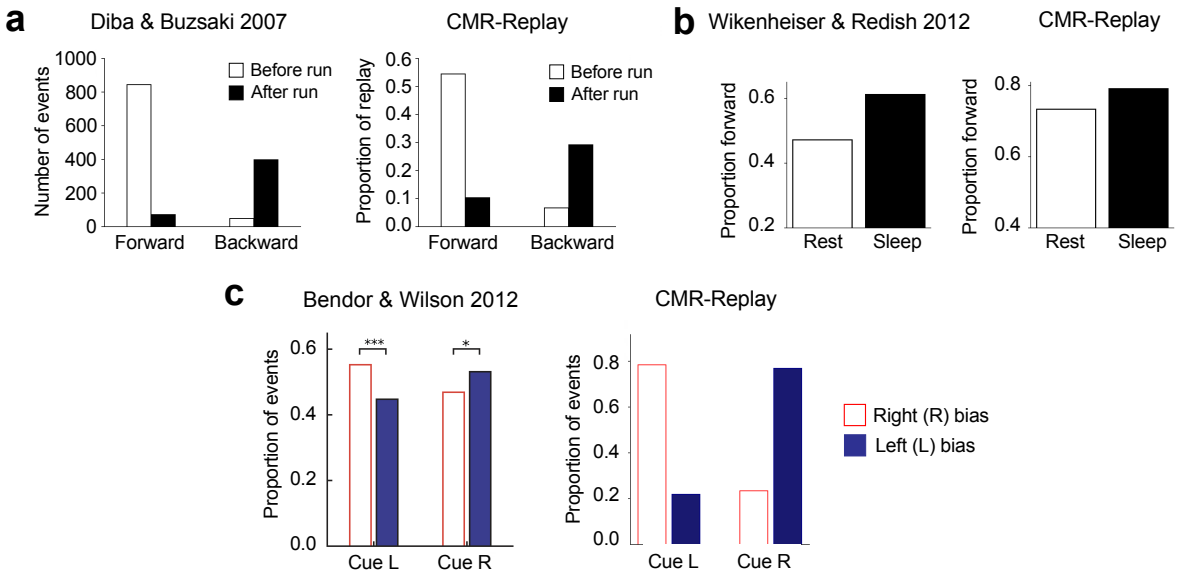

Figure 2: **Context-dependent variations in memory replay**. **a**. As observed in rodents (left), replay in CMR-replay (right) is predominantly forward at the start of a run and backward at the end of a run on a linear track. **b**. Consistent with rodent data (left), in CMR-replay (right), the proportion of forward replay is higher during sleep than during awake rest. **c**. Note that the overall proportion of forward replay is higher in the model than these data, but consistent with that found in Diba and Buzsaki (2007). The presence of external cues during sleep biases replay toward their associated memories both in animals (left) and in CMR-replay (right). *p<0.05; **p<0.01; ***p<0.001.

that goes beyond direct experience, experience-dependent variations of replay, and ways in which replay facilitates memory.

The context-dependency of memory replay

During quiescence, sequential neural firing during sharp-wave ripples (SWRs) recapitulates the temporal pattern of previous waking experience [4]. We distinguish between forward and backward replay, defined as neural activity that either preserves the order of a prior experience (forward replay) or reverses it (backward replay). In animals and humans, the content and directionality of replay systematically vary according to task contexts and behavioral states [27, 21, 15, 28]. For example, animals tend to shift from forward to backward replay between the beginning to the end of a run [21], exhibit more forward replay during sleep [28], and show biased replay of memories associated with external cues during sleep [27]. Some of these observations have led investigators to posit distinct processes underlying forward and backward replay [13, 21, 4, 14, 43, 19, 34], with forward replay supporting planning at choice points [21, 4, 19, 34] and backward replay encoding value expectations from reward outcomes [13, 4, 14]. In line with these observations, the EVB model [19] posits that the agent's estimate of goal-dependent state transitions (i.e., need) drives forward replay whereas reward prediction errors (i.e., gain) drive backward replay. Here we evaluate whether CMR-replay can account for these differential patterns by assuming that replay operates on associations between contexts and experiences.

When animals traverse a linear track to collect a reward, forward replay predominates during pre-run rest [21]. However, during post-run rest when the animal consumes its reward, backward replay predominates (see Fig. 2a left; [21]). We simulate this task by presenting CMR-replay with a sequence of items (Fig. 7a), each representing a distinct location. These item representations can be considered to correspond to place cells in rodents, whose activity is typically used to track replay. During post-run rest, we cue reactivation

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

with the final item's encoding context as $c_{external}$. For pre-run rest, the first item's encoding context serves as $c_{external}$. $c_{external}$ evokes ${a_0}^{retrieved}$ according to Eq. 5, which enters a_0 according to Eq. 4. Through ${a_0}^{retrieved}$, $c_{external}$ biases a_0 so that awake replay initiates disproportionately at the item associated with $c_{external}$ (Fig. 1c), which is consistent with a bias of awake replay to initiate at the resting location [44]. The conjunction of this initiation bias and the model's contiguity bias entails that replay unrolls successively forward from the first item in pre-run rest and backward from the final item in post-run rest. Hence, forward and backward replay respectively predominates during pre-run and post-run rest (Fig. 2a right) as in the data [21]. In contrast to the EVB model [19], CMR-replay captures the graded nature of this phenomenon (Fig. 2a right): Both types of replay appear in both conditions [21].

As with prior retrieved context models [25, 26], CMR-replay encodes stronger forward than backward associations. This asymmetry exists because each item f_t's retrieved context c_{f_t} contribute only to the ensuing items' (e.g., c_{t+1}) encoding contexts. The reinstatement of c_{f_t} thus cues the reactivation of items that followed f_t more than those that preceded f_t, leading to forward asymmetric replay (Fig. 6d left). Absent external cues, sleep replay is less likely to initiate at the final item than rest replay (Fig. 1c), allowing for more forward transitions. This leads to more forward replay during sleep than awake rest (Fig. 2b right), matching empirical observations [4, 45, 28] (Fig. 2b left). In contrast, the EVB model predicts a predominance of reverse sleep replay before behavior stabilizes [19]. Here, and throughout the paper, all differences of interest in the model are highly reliable across the variability present in the replay process.

We next asked whether CMR-replay can simulate Targeted Memory Reactivation (TMR) during sleep. Bendor and Wilson employed the TMR paradigm in rodents, associating distinct auditory cues (L and R) with left and right traversal of a linear track [27]. Playing each auditory cue during sleep elicited replay of place cell activity in the cued direction. We simulate these findings by encoding two sequences that share a start item. To simulate TMR, we present a distinct cue item after each sequence's start item during learning (Fig. 7e), and re-present each cue item (through its associated context) as $c_{external}$ in sleep. Matching Bendor and Wilson [27], CMR-replay preferentially replayed each cue's associated sequence (Fig. 2c right).

Effects of valence

At first glance, our proposal may seem at odds with evidence of valence's influence on replay [13, 14, 33, 15, 18, 16, 10, 11] because CMR-replay neither maintains nor updates value representations during replay. For example, studies suggest that replay over-represents experiences with rewarded or aversive outcomes [33, 10, 46, 12, 11] and awake reverse replay occurs primarily during reward receipt [13, 21, 14]. Moreover, reverse replay's sensitivity to reward [13, 16] argues for a functional distinction between forward and backward replay, with backward replay specialized for learning value-based predictions [13, 14, 18].

We suggest that salience governs encoding rates, which aligns with evidence that salient stimuli bind more strongly to their context [47, 48, 49, 50]. Building on models that adopt this assumption [40, 41], CMR-replay updates M^{fc} and M^{cf} at higher rates for salient experiences, including those with high valence (reward or punishment) or novelty, than for others. In CMR-replay, increasing encoding rates strengthens replay in two distinct ways: Enhancing the M^{cf} encoding rate facilitates the reactivation of an item given its associated context as cue, while enhancing the M^{fc} encoding rate facilitates the faithful retrieval of an item's encoding context. Here, we explore whether these mechanisms allow CMR-replay to account for valence-related phenomena.

After visually exploring a T-maze with one arm with reward, animals preferentially activated sequences representing the rewarded arm during sleep [11] (Fig. 3a left). We simulate this task by presenting CMR-replay with two sequences, one with a rewarded final item and the other with a neutral final item (Fig. 7d). Due to the influence of encoding rates, replay over-represents the rewarded item compared to the matched neutral item (Fig. 3a right) as in empirical observations [33, 10, 11]. CMR-replay exhibits this property without the

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

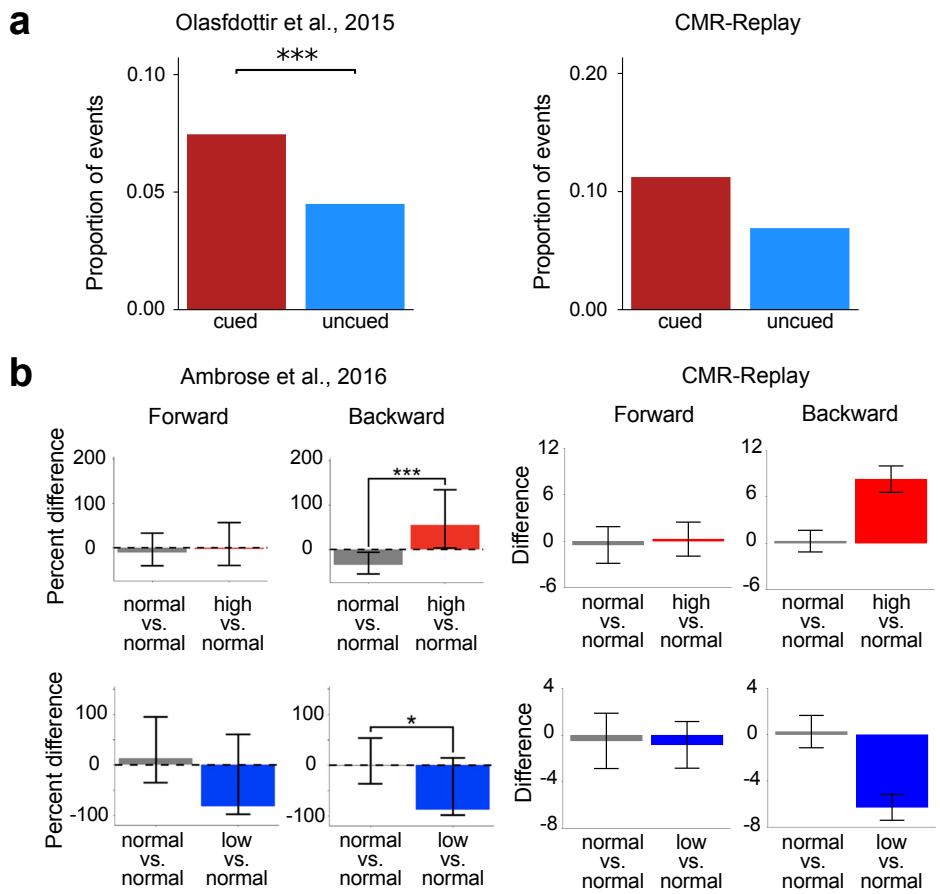

Figure 3: **Reward leads to over-representation in sleep and modulates the rate of backward replay**. **a**. Sleep over-represents experiences associated with reward (i.e., cued) in animals (left) and in CMR-replay (right). **b**. Varying the magnitude of reward outcome leads to differences in the frequency of backward but not forward replay in animals (left) and CMR-replay (right). Error bars show 95% confidence intervals.

assumption that reward-associated items receive more exposure during encoding [51].

Varying the magnitude of reward at the end of a linear track significantly alters the number of backward but not forward replay events [13] (Fig. 3b left). Following Mattar and Daw [19], to disambiguate each location and the direction of a run, we simulated the task with two distinct input sequences, each with a final rewarded item. We manipulated the encoding rate of one reward item so that it's higher (i.e., high reward), lower (i.e., low reward), or identical to that of the reward item in the other sequence (i.e., normal reward). Since the encoding of the reward item primarily influences backward replay in post-run rest, we observed differences in the rate of backward but not forward replay between different reward conditions (Fig. 3b right), matching empirical observations [13, 16].

CMR-replay's ability to account for the effects of reward supports our proposal that valence modulates the initial encoding of memories to shape subsequent replay. After valence exerts its influence during encoding, prioritized replay of rewarded memories can occur even if reward-related activity is absent. Consistent with our proposal that replay itself does not require value-based computations, sleep's preferential consolidation of reward memories does not seem to require dopaminergic activity [52], and the coordination between re-

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

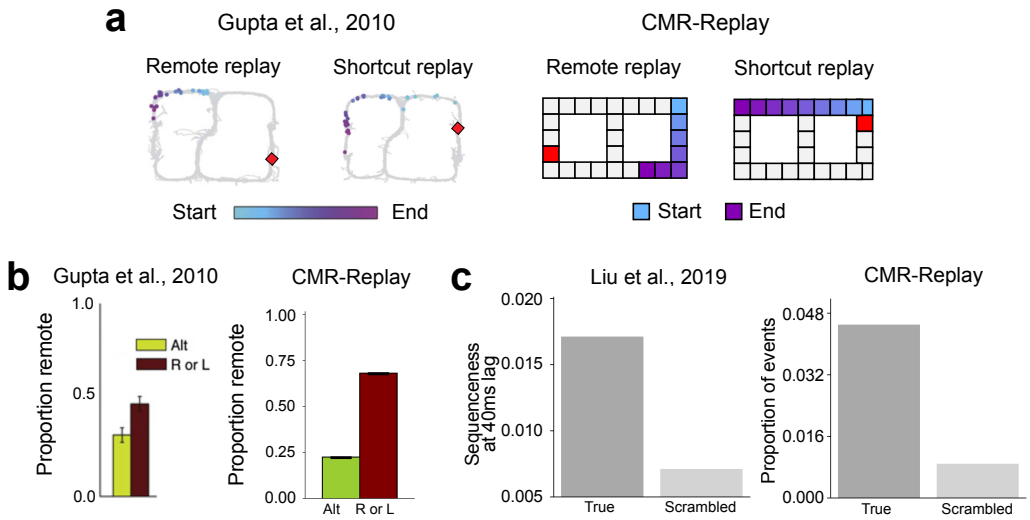

Figure 4: **Replay activates remote experiences and links temporally-separated experiences**. **a**. The left two panels show examples of remote and novel, shortcut replay sequences observed in animals. CMR-replay also generates remote and shortcut rest replay, as illustrated in the two predicted sequences of neural firing in the right two panels. Across all panels, the colored items indicate the temporal order of the sequences (light blue, early; purple, late). The red item denotes the resting position. **b**. Proportion of replay events that contain remote sequences in animals (left) and in CMR-replay (right). Error bars show $+/-$ 1 SEM in the data and model. **c**. In Liu et al. (2019), participants encoded scrambled versions of two true sequences $X_1X_2X_3X_4$ and $Y_1Y_2Y_3Y_4$: $X_1X_2Y_1Y_2$, $X_2X_3Y_2Y_3$, and $X_3X_4Y_3Y_4$ (Fig. 7g) After learning, human spontaneous neural activity showed stronger evidence of sequential reactivation of the true sequences (left). CMR-replay encoded scrambled sequences as in the experiment. Consistent with empirical observation, subsequent replay in CMR-replay over-represents the true sequences (right).

ward responsive neurons and replay-related events is absent in sleep [53]. Our model treats reward as simply a salient experience, generating the prediction that non-reward-related salient experiences that modulate encoding rate should exhibit similar characteristics.

Replay goes beyond direct recent experience

We next asked whether CMR-replay can account for findings in which animals replay sequences learned outside of their present context. Several studies have established this so-called "remote replay" phenomenon [17, 30]. Here we describe one such experiment and show how CMR-replay provides a parsimonious account of its findings. In Gupta et al. [17], animals explored both arms of a T-maze during pre-training. During each subsequent recording session, animals traversed only the left or right arm (L- or R- only conditions) or alternated between them (alternation condition). During reward receipt on the just-explored arm, awake rest exhibited remote replay of its opposite, non-local arm (Fig. 4a left: remote replay)[17]. This observation challenges models that prioritize items near the resting location [19] and recently active neurons [54, 55, 14, 56] throughout replay. To determine whether CMR-replay can reproduce these results, we present the model with sequences that overlap for the first few items (representing the central arm of the T-maze; Fig. 7c). During each of two simulated "pre-training" sessions, the model encodes both sequences. We then run the model through two conditions in an ensuing "experimental" session, where CMR-replay encodes either only one (L or R -only conditions) or both sequences (alternation condition). After encoding the sequences, we simulate reward receipt by presenting CMR-replay with the encoding context of a rewarded item as an external-context cue (i.e., $c_{external}$). As in Gupta et al. [17], CMR-replay is able to generate remote replay of the non-local sequence (Fig. 4a, right; Fig. 4b, right). When CMR-reactivates a

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

non-local item by chance, replay context dramatically shifts by incorporating the non-local item's associated context, thereby triggering a cascade of non-local item reactivation to generate remote replay. Due to its inhibition mechanism, CMR-replay is able to capture the higher prevalence of remote replay in L and R -only conditions (Fig. 4b, right), which we will unpack in a subsequent section. The occurrence of remote replay does not require the inhibition mechanism in CMR-replay, as the model generates remote replay in the alternation condition where inhibition is matched across local and non-local items.

We next examined whether replay in CMR-replay can link temporally-separated experiences to form novel sequences that go beyond direct experience [29, 51, 17, 57, 15]. Gupta et al. [17] showed occurrence of novel replay sequences that link segments of the two arms of a T-maze during rest, even though animals never directly experienced such trajectories [17] (Fig. 4a left: shortcut replay). In our simulation of the study [17], CMR-replay also generates novel rest replay that links segments of the two sequences (Fig. 4a, right): The reactivation of the juncture of the two sequences (the top middle item of Fig. 7c) reinstates context common to the two sequences, allowing replay to stitch together segments of the two sequences. In line with Gupta et al. [17], shortcut replay appeared at very low rates in CMR-replay (the alternation condition: mean proportion of replay events that contain shortcut sequence = 0.0046; L or R conditions: mean proportion = 0.0062). Furthermore, Liu et al. [15] showed that replay in humans reorganizes temporally-separated wake inputs. In their first experiment, participants encoded sequences that scramble pairwise transitions of two true sequences $X_1X_2X_3X_4$ and $Y_1Y_2Y_3Y_4$: $X_1X_2Y_1Y_2$, $X_2X_3Y_2Y_3$, and $X_3X_4Y_3Y_4$. To highlight transitions from the true sequences, the time lag between those transitions (e.g., X_2X_3) was shorter than others (e.g., X_3Y_2) during presentation. Analyses revealed preferential replay of the true as opposed to the scrambled sequences [15] (Fig. 4c, left). We simulated the experiment by presenting CMR-replay with sequences of the same structure (Fig. 7g) and incorporated a distractor item to induce greater context drift for transitions that violate true sequences (see Methods). After learning, CMR-replay performed replay in the absence of external context cues. Consistent with Liu et al. [15], CMR-replay preferentially replayed true sequences (Fig. 4c, right).

The influence of experience

Task exposure influences replay, with replay appearing less frequently in familiar as compared with unfamiliar environments [22, 58, 56]. Task repetition similarly reduces replay [21, 34]. After gaining experience along multiple trajectories, animals and humans can exhibit enhanced replay of non-recently explored trajectories [23, 24, 17, 31]. Overall, these findings demonstrate a negative relationship between the degree and recency of experience and the frequency of replay. This pattern challenges models in which experience enhances the reactivation of local items [54, 55, 51, 14, 19, 56].

In the EVB model [19], learning reduces replay frequency. Learning reduces reward prediction errors (i.e., gain), which drive backward replay, and stabilizes the model's estimation of goal-dependent state transitions (i.e., need), which drive forward replay. Since the combination of gain and need determines the scheduling of replay, learning reduces the overall frequency of replay. Due to the differential influence of gain and need on backward and forward replay, these asymmetrical changes in the two quantities entail a reduction in the proportion of replay events that are backward [19]. The prediction that learning reduces the proportion of backward replay events exemplifies RL theories' assumption that value prediction errors, which learning minimizes, govern backward replay. However, recent findings indicate that the proportion of backward replay does not decrease through learning [33, 34].

In CMR-replay, experience shapes replay in two opposing ways. First, repetition strengthens M^{fc} and M^{cf}, allowing replay to better preserve the structure of waking inputs. Second, by enhancing M^{fc}, repetition increases the activity of items' retrieved contexts during learning. Since CMR-replay inhibits items' activity in a_0 as a function their activity during learning, repetition increases the inhibition of task items' activity, reducing their probability of reactivation. Such an inhibitory mechanism may be adaptive in allowing replay

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

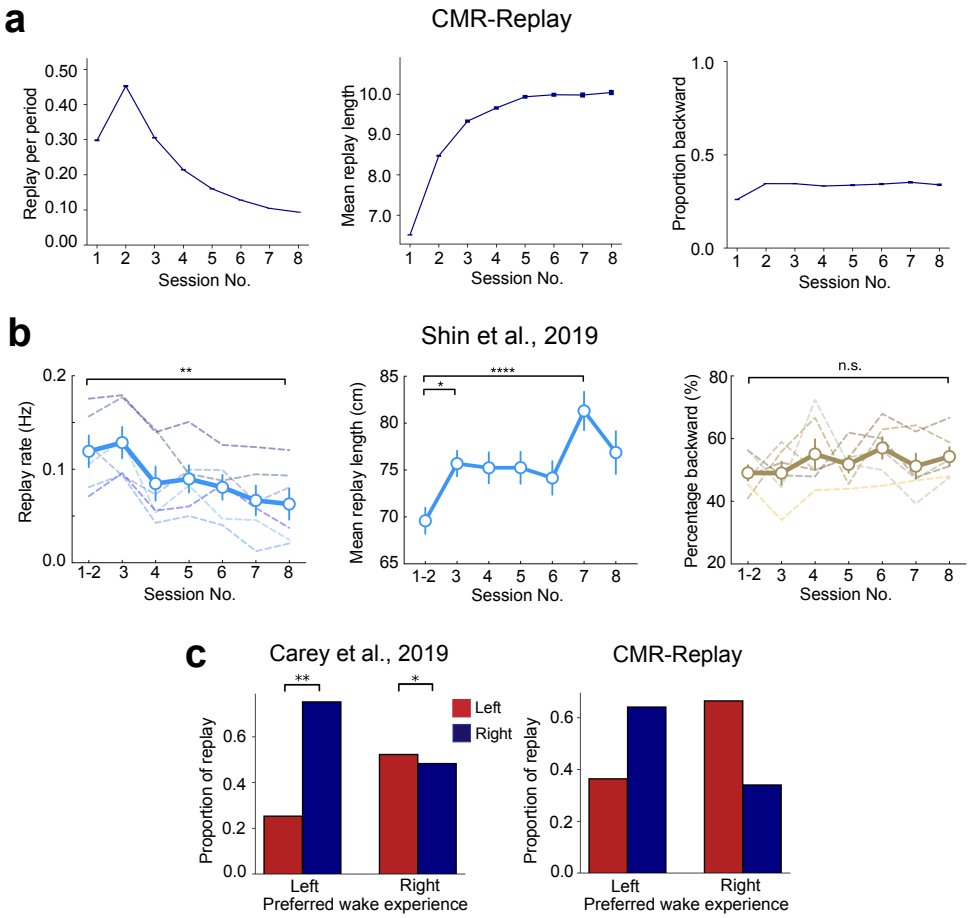

Figure 5: **Variations in replay as a function of experience.** **a**. In CMR-replay, through repeated exposure to the same task, the frequency of replay events decreases (left), the average length of replay events increases (middle), and the proportion of replay events that are backward remains stable (after a slight initial uptick; right). Error bars denote +/− 1 SEM in the model. **b**. With repeated experience in the same task, animals exhibit lower rates of replay (left) and longer replay sequences (middle), while the proportion of replay events that are backward stays relatively stable (right). Error bars denote +/− 1 SEM in the data. **c**. In a T-maze task, where animals display a preference for traversing a particular arm of the maze, replay more frequently reflects the opposite arm [23] (left). CMR-replay preferentially replays the right arm after exposure to the left arm and vice versa (right).

to benefit not just the most recently- and/or strongly- encoded items. Our proposal that inhibition plays a role in replay aligns with models that regulate place cell reactivation via inhibition [59], and empirical observations of increases in hippocampal inhibitory interneuron activity with experience [32].

The next set of simulations illustrates CMR-replay's account of experience-dependent changes in replay [23, 21, 17, 33, 34]. We first examined how replay changes through repeated encoding of the same inputs following our linear track simulation illustrated in Fig. 7a. Here, CMR-replay encodes the same sequence across learning sessions, with awake rest after each session. Initially, experience increases the prevalence of replay (Fig. 5a: left). As repetition enhances the inhibition of task-related items in a_0, replay frequency subsequently decreases in CMR-replay (Fig. 5a: left). Through experience, the average length of replay increases (Fig. 5a: middle), suggesting that repetition strengthens sequence memory in the model. In contrast to the EVB model [19], the proportion of replay events that are backward does not decrease (Fig. 5a right) in CMR-replay. This result highlights that, unlike the EVB model, CMR-replay does not employ distinct

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

variables to drive forward versus backward replay.

In an experiment where animals learned the same task across eight behavioral sessions, Shin et al. [34] observed similar patterns of results. As shown in Figure 5b, animals exhibited lower rates of replay but longer replay sequences in later sessions (left, middle). As in our CMR-replay simulations, the proportion of forward relative to backward replay events remained relatively stable across sessions (right). Furthermore, consistent with reduced reactivation of task-related units in CMR-replay, the study observed decreased reactivation of task-related place cells through experience. In contrast, item reactivation increases monotonically through repetition in alternative models [51]. Several other studies using varied experimental procedures have reported similar effects of repeated experience on replay, including a reduction in the prevalence of replay [22, 21], an increase in replay length [32], and no reduction in the proportion of replay events that are backward [33].

In CMR-replay, items' retrieved context activity in a learning session modulates their inhibition during ensuing quiescence. As a result, items that get more exposure in a session may receive more inhibition than others at the onset of replay, facilitating the reactivation of their competitors. In our simulation of [17] (Fig. 7c), in the L and R -only conditions, since the sequence presented during learning receives more inhibition, remote replay is more prevalent than in the alternation condition, where both sequences appear during learning (Fig. 4b). In the L or R -only conditions, when CMR-replay performs post-learning replay in the absence of external context cues, replay over-represents the alternative sequence (Fig. 5c), which aligns with the observation that replay exhibits a bias away from the arm of a T-maze that animals preferred during behavior [23]. This property is also consonant with recent findings that replay preferentially activates non-recent trajectories [24].

The function of replay

Many have proposed adaptive functions for replay, including for memory consolidation [36, 1, 60], retrieval [36, 9, 61], credit assignment [13, 14, 20], and planning [62, 63, 64]. Growing causal evidence suggests that replay benefits memory: TMR enhances memory [42], and disrupting SWRs impairs memory [7, 8, 9]. Replay facilitates offline learning in our model by updating M^{fc} and M^{cf} according to the internally-reactivated items and contexts during replay. In the following set of simulations, we characterize ways in which replay facilitates memory in the model.

One of the most robust benefits of sleep is on motor sequence memory [37]. To simulate the impacts of sleep replay on sequence memory, we presented CMR-replay with a five-item sequence and examined whether sleep enhanced memory of the sequence. Before and after sleep, we assessed the proportion of replay sequences that matched the input sequence. The assessment occurred in "test" periods, where learning rates were set to zero and $c_{external}$ was absent. In post-sleep test, CMR-replay generated a higher proportion of sequences matching the correct sequence than in pre-sleep test (Fig. 6a), indicating that sleep enhances sequence memory in the model.

Replay preferentially enhances rewarded memories [16], and sleep preferentially consolidates salient experiences [38, 39]. In our simulation of a T-maze with reward in one of the two arms [11], we also included pre- and post- sleep test periods to assess how sleep in CMR-replay shapes rewarded versus non-rewarded memory. Through sleep, CMR-replay exhibited a greater increase in its reactivation of the rewarded item compared to a matched neutral item (Fig. 6b), suggesting that sleep preferentially enhances memory associations for rewarded items in CMR-replay.

A recent study [18] presented evidence that replay facilitates nonlocal value learning. Human participants encoded six sequences, each of which links one of three start items to one of two end items. When they received reward at the end of a sequence, awake replay reflected the just-visited (local) sequence as well as nonlocal sequences that shared the same end item (i.e., sequences that do not begin with the current

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

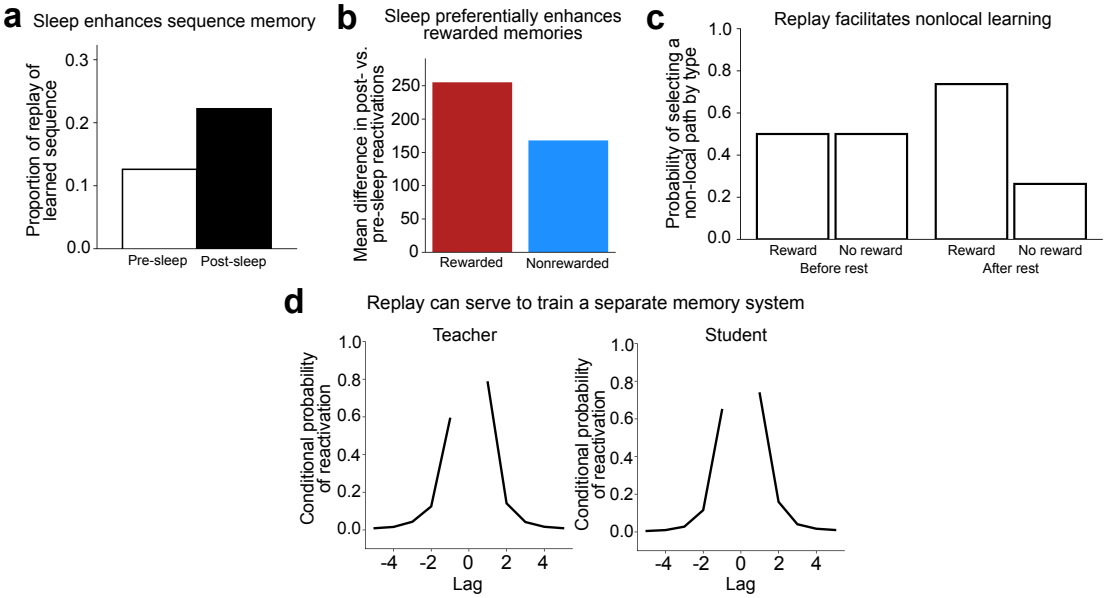

Figure 6: **Learning from replay**. **a**. Sleep increased the likelihood of reactivating the learned sequence in the correct temporal order in CMR-replay, as seen in an increase in the proportion of replay for learned sequences post-sleep. **b**. Sleep leads to greater reactivation of rewarded than non-rewarded experiences, indicating that sleep preferentially strengthens rewarded memories in CMR-replay. **c**. CMR-replay encoded six sequences, each of which transitioned from one of three start items to one of two end items. After receiving a reward outcome for the end item of a sequence, we simulated a period of rest. After but not before rest, CMR-replay exhibited a preference for non-local sequences that led to the rewarded item. This preference emerged through rest despite the fact that the model never observed reward in conjunction with those non-local sequences, suggesting that rest replay facilitates non-local learning in the model. **d**. We trained a "teacher" CMR-replay model on a sequence of items. After encoding the sequence, the teacher generated replay sequences during sleep. We then trained a separate blank-slate "student" CMR-replay model exclusively on the teacher's sleep replay sequences. To assess knowledge of the original sequence, we collected sleep replay sequences from both models, and assessed the probability that each model reactivates the item at position i + lag of the sequence immediately following the reactivation of the i-th item of the sequence, conditioned on the availability of the i-th item for reactivation. Both models demonstrated a tendency to reactivate the item that immediately follows or precedes the just-reactivated item on the original sequence. This result suggests that the student acquired knowledge of the temporal structure of original sequence by encoding only the teacher's replay sequences.

sequence's start item but terminate with the same end item). Participants then exhibited a behavioral preference for sequences that terminated in the rewarded item, despite no direct recent experience with reward in that sequence. The authors suggested that replay propagated value to associated items, allowing participants to select nonlocal sequences associated with reward without direct observation. In our simulation of this paradigm, CMR-replay encoded six sequences of the same structure (Fig. 7b), with increased encoding rates to simulate reward receipt, as in the simulations above. During rest after reward receipt, we presented the encoding context of the rewarded item as $c_{external}$. Before and after rest, we assessed the activity that each non-local start state's associated context evokes considering only its two ensuing items according to Eq. 5. After but not before rest, CMR-replay preferentially activated the item that leads to the rewarded end item (Fig. 6c). In CMR-replay, this preference emerged without value updates during replay, suggesting that replay can facilitate nonlocal learning by re-organizing memory associations.

There has been much interest in the memory literature in the possibility that hippocampal replay serves to train neocortical systems to represent recent memories [65, 35, 54, 36, 60, 66, 67]. We explored whether replay in CMR-replay can serve to transfer one model's knowledge to another. After a "teacher" CMR-replay encodes a sequence, we collected its sleep replay sequences to train a blank-slate "student" CMR-replay at

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

replay's learning rates. Through this process, the student inherited the contiguity bias of the teacher (Fig. 6d), suggesting it acquired knowledge of the structure of the teacher's training sequence. This simulation provides a proof of concept that replay in CMR-replay can serve to facilitate memory transfer across systems, in addition to promoting local learning.

Discussion

What is the nature and function of neural replay? We suggest a simple memory-focused framework that explains a wide array of replay phenomena. First, the brain associates experiences with their encoding contexts in accordance with the salience of each experience. Then, in quiescence, the brain replays by spontaneously reactivating a memory and retrieving its associated context to guide subsequent reactivation. Learning continues to occur from these endogenously reactivated sequences. A model embodying these ideas – CMR-replay – accounts for many empirical characteristics of replay and its impacts, including evidence previously interpreted as diagnostic of reinforcement learning accounts.

First, CMR-replay demonstrates properties of replay that other models exhibit (or could easily accommodate) [51, 43, 57, 19, 68], including replay's recapitulation of the temporal pattern of past experience during rest and sleep [21, 28], bias toward external cues' associated memories [27], and ability to stitch together temporally-separated experiences to form novel sequences [17, 15]. Second, CMR-replay captures findings that have been interpreted as evidence that replay serves value-based reinforcement learning, including over-representation of memories associated with reward [11], reverse replay upon reward receipt [21, 14], and the unique sensitivity of reverse replay to reward magnitude [13]. Third, CMR-replay accounts for observations that are not naturally accounted for with prior models, including a stable proportion of backward replay through learning [34], reduced item reactivation and sequential replay through experience [34, 22], increased prevalence of forward replay in sleep [28], enhanced replay outside of the current context [17], and a tendency for replay to cover non-behaviorally-preferred experiences [23]. Finally, replay facilitates memory in CMR-replay in ways that align with empirical findings [37, 18, 16, 39, 38], including that it improves sequence memory, preferentially strengthens rewarded memories, facilitates nonlocal learning, and can serve to train a separate memory system in the absence of external inputs.

Many memory consolidation theories are aligned with CMR-replay in suggesting that replay actively strengthens and re-organizes memories [69, 35, 54, 70, 1, 60, 66]. Contextual binding theory [71], however, takes a different approach, suggesting that residual encoding-related activity may elicit merely epiphenomenal replay as context drifts during quiescence. Our theory echoes this perspective in characterizing replay as an outcome of context-guided processes. However, we diverge in suggesting that the emergent replay does significantly benefit memory by strengthening learned associations. Moreover, our model captures observations of enhanced replay of infrequent and remote experiences, which are in tension with the perspective that replay is primarily guided by recent activity.

An ongoing debate concerns to what extent awake replay reflects a process of planning that simulates future scenarios to support immediate decision-making [64, 62, 63], versus to what extent replay serves to store, update, and maintain memory without directly guiding behavior [72, 8, 73]. Evidence supporting the planning hypothesis comes from studies that demonstrate enhanced replay of upcoming behavioral trajectories [64, 74]. However, in tasks that track representations of multiple, temporally- and spatially- separated experiences, animals exhibit replay that appears to be decoupled from their behavioral preference [23, 24, 17]. Our model aligns more with the memory perspective, as it is able to capture existing findings without positing that replay serves to optimize behavioral outcome. However, replay of this kind could at times be read out and used by downstream decision-making systems. Recent work argues that the dynamics of the retrieval processes in this class of models could support adaptive choice in sequential decision tasks [75]. Overall, our framework argues that replay characteristics are primarily driven by memory principles, and that replay

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

serves to strengthen and reorganize memories, which benefits subsequent — but not necessarily immediate — behavior [24, 73].

Our model has mainly considered replay occurring during sharp wave ripples. During active behavior in rodents, ordered place cell sequences also activate during the theta oscillation (theta sequences) [76]. Similar to ripple-based replay, theta sequences manifest in both forward- and reverse-order [77], initiate at the animal's location, extend further into upcoming locations through experience [78, 79, 80, 81], cluster around behaviorally-relevant items [82], and have been proposed to correspond to cued memory retrieval [83]. These parallels lead us to speculate that the context-driven mechanisms we have laid out for findings of replay mainly during sharp wave ripples may also be relevant in understanding theta sequences.

An important area for future work is to investigate the mapping between the components of CMR-replay and neural circuitry. Our model employs a series of bidirectional operations between context and item representations to generate replay. These operations might be implemented within the recurrent connections of CA3 in the case of temporally-compressed sharp wave ripple replay. It is possible that these interactions could also play out across the "big loop" of the hippocampus [57] or within cortical circuits [84, 85, 86, 87], which could correspond to slower forms of replay [18, 88]. In quiescence, we posit that the hippocampus can serve as a "teacher" that endogenously samples memory sequences to help establish these associations in neocortical areas, with local context-item loops within the teacher and student areas. In alignment with the observation that disrupting entorhinal cortex input to the hippocampus affects only awake replay whereas manipulating hippocampal subfield CA3 activity affects both awake and sleep replay [89], in CMR-replay, the principal distinction between awake rest and sleep is whether external inputs bias replay. There are likely other variables, such as task engagement [90], that modulate the influence of external inputs on replay.

There exists a range of computational models that simulate replay at different levels of biological detail [91, 92, 19, 93, 94], account for different features of replay [51, 43, 57, 63, 19, 68], and posit distinct functions for replay [95, 96, 63, 19, 60, 97, 66]. Our theory follows in a lineage of memory-focused replay models, showing the power of this perspective in accounting for data that have been assumed to require optimization of value-based predictions. As CMR-replay builds on existing theories of memory recall, our account is in line with recent proposals that reactivation and recall may have similar underlying mechanisms and utility for behavior [98]. Our theory unifies a wealth of phenomena, offering a parsimonious and mechanistic framework characterizing how the brain initially encodes and subsequently replays memories to facilitate behavior.

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Methods

Representation and Initialization

CMR-replay comprises four components: experience (f), context (c), experience-to-context associations (M^{fc}), and context-to-experience associations (M^{cf}). At each moment in time, f is the current experience (i.e., an external input presented during awake learning or a reactivated experience during replay) and c is a recency-weighted sum of features associated with recent experiences. For notation, we refer to the experience associated with the i-th feature in f as f_i. By contrast, we refer to the experience at the i-th timestep of a process as $f_{t=i}$ or generically as f_t. During both awake learning and replay, at each timestep, the model updates M^{fc} and M^{cf}, which respectively support the retrieval of an experience's associated context and a context's associated experiences.

In our simulations, CMR-replay employs a one-hot representation of f (i.e., a localist item representation): Each experience is represented by a one-hot vector of length n in which only the unit representing the experience is on (i.e., has an activity of 1) and all other units are off. As illustrated in Fig. 1, in addition to task-related items shown as inputs during learning, we include task-irrelevant items that do not appear during learning but compete with encoded items for reactivation during replay. We use n_{task}, $n_{non-task}$, and n to respectively denote the number of task-related items, the number of task-irrelevant items, and the total number of items (i.e., the sum of n_{task} and $n_{non-task}$). To allow for sufficient competition between task-related and task-irrelevant items, we set $n_{non-task}$ to be roughly one half of n_{task} (i.e., rounded up when n_{task} is odd). We note that the particular ratio of $n_{non-task}$ to n_{task} is not critical to the pattern of results in our simulations.

In each simulation, M^{fc} and M^{cf} are initialized as identity matrices of rank n, which are scaled respectively by 1.0 and 0.7. These scaling factors were chosen to qualitatively match the empirically observed proportions of forward and backward replay in different conditions (though the forward/backward asymmetry is always observed in the model). Our initialization of these two matrices as identity matrices differs from the initialization strategy in prior work [40, 25, 99, 26], where M^{fc} and M^{cf} are initialized to reflect pre-experimental similarity among items. Thus, prior to learning, M^{fc} maps distinct experiences onto orthogonal context features, and M^{cf} maps each context feature to one experience. Before the model encodes each input sequence, c is reset as a zero vector of length n. Resetting contexts in between sequence presentations demarcates boundaries between discrete events as in prior work [100].

Context Drift

In both active wakefulness and quiescence, each time an experience is presented or reactivates, c drifts by incorporating contextual features associated with that experience. Concretely, at each timestep t, CMR-replay reinstates the current experience f_t's associated context c_{f_t} via experience-to-context matrix M^{fc}_{t-1} according to:

$$c_{f_t} = \frac{M^{fc}_{t-1} f_t}{\parallel M^{fc}_{t-1} f_t \parallel} \tag{6}$$

Given c_{f_t}, c_{t-1} drifts and forms a new context c_t according to Eq. 1, in which ρ and β determine the relative contribution of c_{t-1} and c_{f_t} to c_t. To ensure that c_t has a unit length, ρ is computed according to:

$$\rho = \sqrt{1 + \beta^2[(c_{t-1}c_{f_t})^2 - 1]} - \beta(c_{t-1}c_{f_t}) \tag{7}$$

Operations that drive context drift in our simulations, including those specified by Eqs. 1, 6, and 7, are identical to those in prior work [40, 26, 99]. In all simulations, β is 0.75 (similar to drifts rates for temporal context features reported in Polyn et al. [26]), except when distractors cause context drift in the simulation of Liu et al [15].

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Updating M^{fc} and M^{cf}

Each time the context drifts, CMR-replay updates M^{fc} and M^{cf} to strengthen associations between the current c_t and f_t. The model updates M^{fc} and M^{cf} using a standard Hebbian learning rule according to Eq. 2 and Eq. 3, in which γ_{fc} and γ_{cf} control the rates at which M^{fc} and M^{cf} are updated.

During active wakefulness, the rate of updating M^{fc} and M^{cf} is higher for novel and rewarding experiences than for other experiences. For all simulations, the base learning rates γ_{fc} and γ_{cf} are 1.0. For rewarded experiences, learning rates vary according to the magnitude of reward: Learning rates $\gamma_{fc_{\text{low}}}$ and $\gamma_{cf_{\text{low}}}$ are 1.0 for experiences associated with low reward, $\gamma_{fc_{\text{normal}}}$ and $\gamma_{cf_{\text{normal}}}$ are 1.5 for those with standard reward, and $\gamma_{fc_{\text{high}}}$ and $\gamma_{cf_{\text{high}}}$ are 2.0 for those with high reward. These values are consistent with prior work [40, 41], in which these scaling factors are 1.0 for items that evoke no arousal [40] and greater than 1.0 for those assumed to evoke emotional arousal [40, 41]. When the same input is repeated across sessions, its learning rate is $\frac{\gamma}{i}$, where i is the index of the current session and γ is its initial learning rate.

Replay

After each session of awake learning, CMR-replay autonomously reactivates experiences in a number of replay periods. For simplicity, we assumed that the number of replay periods is fixed, rather than determined by task-related variables.

In each replay period, at each timestep t, the model samples an experience from a distribution of activities across experiences a_t according to Eq. 5, which follows previous work [25]. At $t = 0$, as shown in Eq. 4, a_0 is a normalized combination of ${a_0}^{spontaneous}$, which simulates spontaneous internal activity, and ${a_0}^{retrieved}$, which simulates activity evoked by external context cues in awake rest and sleep TMR. ${a_0}^{spontaneous}$ is a vector of size n whose elements are independently and randomly drawn from the interval $[0, 0.001]$. In awake rest and sleep, a task-related external context cue $c_{external}$ that represents the animal's current state (i.e., the context at the end of a sequence presentation or the context elicited by an external cue presented during sleep) evokes ${a_0}^{retrieved}$. By contrast, at $t \geq 1$, c_{t-1} — a recency-weighted average of contexts evoked by experiences reactivated at previous timesteps, evokes ${a_t}^{retrieved}$ which becomes a_t. Therefore, after time $t = 0$, reactivated experiences' associated contexts determine subsequent a_t.

At each timestep t, the model samples an experience $f_t \in U_t$ from a_t according to Eq. 5, where T_0 is 0.1 and T_t is 0.14 for all $t \geq 1$. If f_t is a task-relevant experience, the model reinstates its associated context c_{f_t} according to Eq. 6. When $t = 0$, c_{f_t} becomes c_0. When $t \geq 1$, c_{f_t} drifts c_{t-1} to form c_t according to Eq. 1. As in awake learning, the model updates the association between f_t and the drifted context c_t in M^{fc} and M^{cf} according to Eq. 2 and Eq. 3, but at a slower learning rate γ_{replay} of 0.001. At each timestep, replay terminates with a probability of 0.1 or if f_t is a task-irrelevant experience.

Inhibition

To model how repeated exposure influences replay, CMR-replay incorporates a mechanism that inhibits the activity of experiences in a_0 according the magnitude of context activity in the preceding active wakeful period. This mechanism differs from the use of inhibition in prior work [99, 26], which scales the degree of competition among items during recall. For each sequence item presented in a wake learning session, its activity in a_0 is multiplied by:

$$\omega = exp(-C) \tag{8}$$

where C is the L2 norm of the item's retrieved context vector in the recent wake learning session. For experiences not shown in the wake session, C is 0.0 and thus ω is 1.0.

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

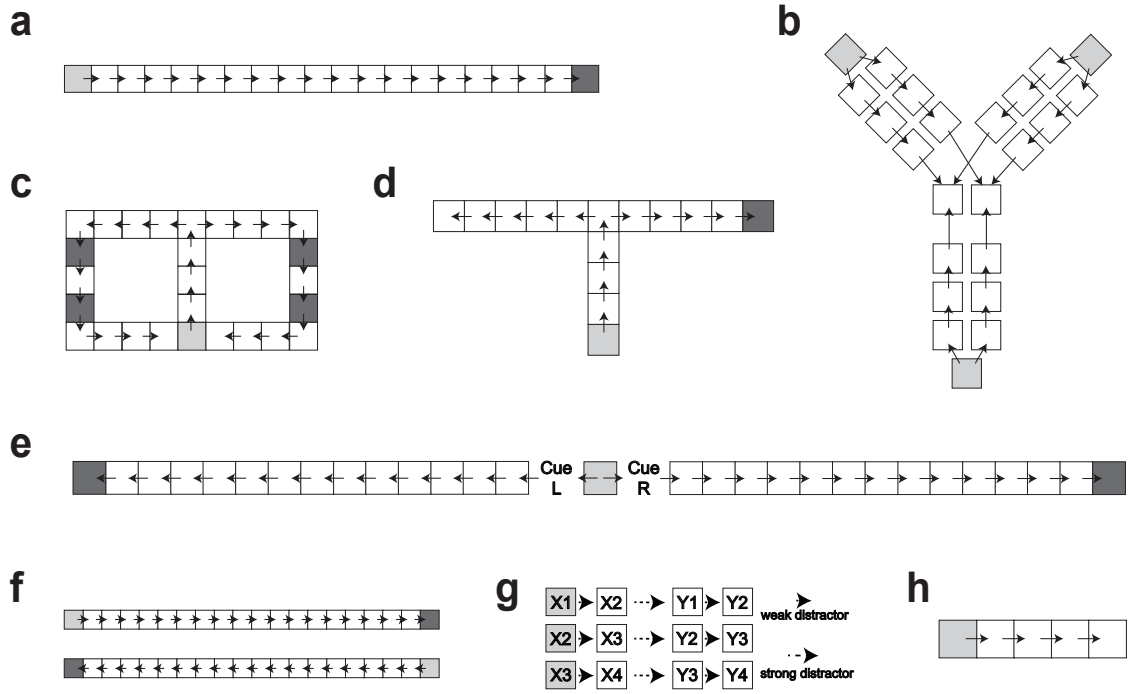

Figure 7: **Task simulations**. Each enclosed box corresponds to a unique stimulus. Each arrow represents a valid transition between two items. Each dashed arrow represents a distractor that causes a drift in context in between two items. Dark grey items represent salient items in each task. Task sequences initiate at light grey items. For tasks with multiple valid sequences, the order in which sequences are presented is randomized. **a**. Simulation of a linear Track. **b**. Simulation of the task in Liu et al. [18]. **c**. Simulation of a two-choice T-maze. **d**. Simulation of a T-maze. **e**. Simulation of the task in Bendor and Wilson [27]. **f**. Simulation of a linear track task with distinct directions of travel. **g**. Simulation of input sequences in Liu et al. [15]. **h**. Simulation of a fix-item sequence.

Task Simulations

During awake learning, CMR-replay encodes sequences of experiences, representing spatial trajectories or other stimulus sequences (Fig. 7). After wake learning, the model participates in a number of awake rest or sleep replay periods. In each simulation, for each condition, we ran 100 instantiations of the model with the same initialization. Variability in replay sequences across models arises from the stochastic nature of the replay process. Due to the variability in replay sequences, different models develop distinct M^{fc} and M^{cf} as they learn from replay. Unlike prior work that identified the best-fitting parameters for each simulation [26, 99, 40], CMR-replay employs the same set of model parameters across simulations with varying input structures.

In the simulation that examines context-dependent forward and backward replay through experience (Figs. 2a and 5a), across a total of 8 sessions of awake learning, CMR-replay encodes an input sequence shown in Fig. 7a, which simulates a linear track run with no ambiguity in the direction of inputs. In this simulation, learning rates for the rewarded item are $\gamma_{fc_{\text{normal}}}$ and $\gamma_{cf_{\text{normal}}}$. After each wake learning session, we simulate 500 awake rest replay periods at the end of a run followed by another 500 periods at the start of a run. For rest at the end of a run, $c_{external}$ is the context associated with the final item in the sequence. For rest at the end of a run, $c_{external}$ is the context associated with the start item.

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

In the simulation that contrasts forward and backward replay between rest and sleep (Fig. 2b), the model encodes the input sequence shown in Fig. 7a for a single session. After encoding, each model either participates in 500 awake rest or sleep replay periods, with 100 models in each condition (i.e., awake rest or sleep).

In the simulation of TMR (Fig. 2c), each model encodes two sequences shown in Fig. 7e in a randomized order in a session of wake learning. During input presentation, for each sequence, a separate cue item (i.e., cue L or cue R) is presented immediately after the start item. The models encode the goal item at rates $\gamma_{fc_{\text{normal}}}$ and $\gamma_{cf_{\text{normal}}}$. After wake learning, each model engages in 500 sleep replay periods. In each replay period, the context associated with cue L or cue R is randomly presented as $c_{external}$.

In the simulation that contrasts replay of rewarded versus non-rewarded experiences (Fig. 3a and 6b), each model encodes two sequences shown in Fig. 7d in a randomized order in a session of wake learning. The models encode the goal item at rates $\gamma_{fc_{\text{normal}}}$ and $\gamma_{cf_{\text{normal}}}$. After wake learning, each model engages in an extended phase with 5000 sleep replay periods. To quantify changes in memory through sleep, in each model, we additionally simulated 5000 replay periods before and after extended sleep with no learning (i.e., M^{fc} and M^{cf} are not updated) and no $c_{external}$.

In the simulation of forward and backward replay with different levels of reward (Fig. 3b), the model encodes two sequences (Fig. 7f) in a randomized order in a single session. The inclusion of two disjoint sequences follows the approach in [19], which simulates different directions of travel to distinguish place cells with directional preference for replay decoding in animal studies. The simulation consists of three conditions: normal vs. normal reward, low vs. normal reward, and high vs. normal reward. In the normal vs. normal condition, each model encodes goal locations in both sequences at rates $\gamma_{fc_{\text{normal}}}$ and $\gamma_{cf_{\text{normal}}}$. In the low vs. normal condition, each model encodes the goal location at rates $\gamma_{fc_{\text{low}}}$ and $\gamma_{cf_{\text{low}}}$ for one sequence and at rates $\gamma_{fc_{\text{normal}}}$ and $\gamma_{cf_{\text{normal}}}$ for the other. Finally, in the high vs. normal condition, each model encodes the goal location at rates $\gamma_{fc_{\text{high}}}$ and $\gamma_{cf_{\text{high}}}$ for one sequence and at rates $\gamma_{fc_{\text{normal}}}$ and $\gamma_{cf_{\text{normal}}}$ for the other. After encoding a sequence, we simulate 500 awake rest replay periods at the end of a run followed by another 500 at the start of a run.

In the simulation of remote replay, shortcut replay, and the over-representation of non-behaviorally-preferred experiences in replay (Figs. 4a, 4b, and 5c), each model encodes two sequences (Fig. 7c) in a randomized order in a total of three sessions. Learning rates for each goal location are $\gamma_{fc_{\text{normal}}}$ and $\gamma_{cf_{\text{normal}}}$. In these simulations, we treat the first two sessions as the period in which an animal is pre-trained extensively on the task. After wake learning in the third session, for the results shown in Figs. 4a and b, each model engages in 500 awake rest replay periods at each of the four goal locations in a randomized order. For the results shown in Figs. 5c, to simulate replay away from the task environment, each model engages in 500 replay periods with no external context cue.

In the simulation of Liu et al. [15] (Fig. 4c), each model encodes three sequences (Fig. 7g) shown in a randomized order. These three sequences $X_1X_2Y_1Y_2$, $X_2X_3Y_2Y_3$, and $X_3X_4Y_3Y_4$ are scrambled versions of pairwise transitions from true sequences $X_1X_2X_3X_4$ and $Y_1Y_2Y_3Y_4$. A distractor item, which is a distinct item that does not participate in replay, induces context drift between successive items. The item induces context drift at a β of 0.99 for transitions that do not exist in the true sequences and at a β of 0.3 for transitions that exist in the true sequences.

In the simulation that examines sequence memory through sleep (Fig. 6a), each model encodes a five-item sequence (Fig. 7h) in a session. After wake learning, each model participates in an extended period of sleep with 5000 replay periods. As in (iv), in each model, we additionally simulated 5000 replay periods before and after extended sleep with no learning (i.e., M^{fc} and M^{cf} are not updated) and no $c_{external}$.

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

CMR-Replay

In the simulation that examines replay's role in non-local learning (Fig. 6c), each model encodes six sequences (Fig. 7b) in a randomized order in a session. Sequences in this simulation consist of three start states and two end states. Each start state has a unique sequence that connects it to each of the two end states. The model encodes the final item in the final sequence at rates $\gamma_{fc_{\text{high}}}$ and $\gamma_{cf_{\text{high}}}$ and encodes all other items at base learning rates γ_{fc} and γ_{cf}. After the encoding of all six sequences, each model participates in 5000 awake rest replay periods with the final item's associated context as $c_{external}$.

In the simulation of "teacher" and "student" CMR-replay (Fig. 6d), each "teacher" model encodes a sequence (Fig. 7a) in a session. Each teacher model encodes the goal location at learning rates $\gamma_{fc_{\text{normal}}}$ and $\gamma_{cf_{\text{normal}}}$. After wake learning, we simulate an extended period of sleep with 5000 replay periods in each model. We then present each teacher model's 5000 replayed sequences as inputs to train a different blank-slate "student" model with learning rates γ_{replay}.

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

1. Klinzing, J. G., Niethard, N., and Born, J. (2019). Mechanisms of systems memory consolidation during sleep. *Nature neuroscience*, 22(10):1598–1610.

2. Walker, M. P. and Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron*, 44(1):121–133.

3. Wamsley, E. J. (2019). Memory consolidation during waking rest. *Trends in cognitive sciences*, 23(3):171–173.

4. Foster, D. J. (2017). Replay comes of age. *Annual review of neuroscience*, 40:581–602.

5. Liu, Y., Nour, M. M., Schuck, N. W., Behrens, T. E., and Dolan, R. J. (2022). Decoding cognition from spontaneous neural activity. *Nature Reviews Neuroscience*, 23(4):204–214.

6. Ólafsdóttir, H. F., Bush, D., and Barry, C. (2018). The role of hippocampal replay in memory and planning. *Current Biology*, 28(1):R37–R50.

7. Ego-Stengel, V. and Wilson, M. A. (2010). Disruption of ripple-associated hippocampal activity during rest impairs spatial learning in the rat. *Hippocampus*, 20(1):1–10.

8. Girardeau, G., Benchenane, K., Wiener, S. I., Buzsáki, G., and Zugaro, M. B. (2009). Selective suppression of hippocampal ripples impairs spatial memory. *Nature neuroscience*, 12(10):1222–1223.

9. Jadhav, S. P., Kemere, C., German, P. W., and Frank, L. M. (2012). Awake hippocampal sharp-wave ripples support spatial memory. *Science*, 336(6087):1454–1458.

10. Singer, A. C. and Frank, L. M. (2009). Rewarded outcomes enhance reactivation of experience in the hippocampus. *Neuron*, 64(6):910–921.

11. Ólafsdóttir, H. F., Barry, C., Saleem, A. B., Hassabis, D., and Spiers, H. J. (2015). Hippocampal place cells construct reward related sequences through unexplored space. *Elife*, 4:e06063.

12. Wu, C.-T., Haggerty, D., Kemere, C., and Ji, D. (2017). Hippocampal awake replay in fear memory retrieval. *Nature neuroscience*, 20(4):571–580.

13. Ambrose, R. E., Pfeiffer, B. E., and Foster, D. J. (2016). Reverse replay of hippocampal place cells is uniquely modulated by changing reward. *Neuron*, 91(5):1124–1136.

14. Foster, D. J. and Wilson, M. A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, 440(7084):680–683.

15. Liu, Y., Dolan, R. J., Kurth-Nelson, Z., and Behrens, T. E. (2019). Human replay spontaneously reorganizes experience. *Cell*, 178(3):640–652.

16. Michon, F., Sun, J.-J., Kim, C. Y., Ciliberti, D., and Kloosterman, F. (2019). Post-learning hippocampal replay selectively reinforces spatial memory for highly rewarded locations. *Current Biology*, 29(9):1436–1444.

17. Gupta, A. S., van der Meer, M. A., Touretzky, D. S., and Redish, A. D. (2010). Hippocampal replay is not a simple function of experience. *Neuron*, 65(5):695–705.

18. Liu, Y., Mattar, M. G., Behrens, T. E., Daw, N. D., and Dolan, R. J. (2021). Experience replay is associated with efficient nonlocal learning. *Science*, 372(6544):eabf1357.

19. Mattar, M. G. and Daw, N. D. (2018). Prioritized memory access explains planning and hippocampal replay. *Nature neuroscience*, 21(11):1609–1617.

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

20. Momennejad, I., Otto, A. R., Daw, N. D., and Norman, K. A. (2018). Offline replay supports planning in human reinforcement learning. *Elife*, 7:e32548.

21. Diba, K. and Buzsáki, G. (2007). Forward and reverse hippocampal place-cell sequences during ripples. *Nature neuroscience*, 10(10):1241–1242.

22. Cheng, S. and Frank, L. M. (2008). New experiences enhance coordinated neural activity in the hippocampus. *Neuron*, 57(2):303–313.

23. Carey, A. A., Tanaka, Y., and van Der Meer, M. A. (2019). Reward revaluation biases hippocampal replay content away from the preferred outcome. *Nature neuroscience*, 22(9):1450–1459.

24. Gillespie, A. K., Maya, D. A. A., Denovellis, E. L., Liu, D. F., Kastner, D. B., Coulter, M. E., Roumis, D. K., Eden, U. T., and Frank, L. M. (2021). Hippocampal replay reflects specific past experiences rather than a plan for subsequent choice. *Neuron*, 109(19):3149–3163.

25. Howard, M. W. and Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of mathematical psychology*, 46(3):269–299.

26. Polyn, S. M., Norman, K. A., and Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological review*, 116(1):129.

27. Bendor, D. and Wilson, M. A. (2012). Biasing the content of hippocampal replay during sleep. *Nature neuroscience*, 15(10):1439–1444.

28. Wikenheiser, A. M. and Redish, A. D. (2013). The balance of forward and backward hippocampal sequences shifts across behavioral states. *Hippocampus*, 23(1):22–29.

29. Barron, H. C., Reeve, H. M., Koolschijn, R. S., Perestenko, P. V., Shpektor, A., Nili, H., Rothaermel, R., Campo-Urriza, N., O'Reilly, J. X., Bannerman, D. M., Behrens, T. E., and Dupret, D. (2020). Neuronal computation underlying inferential reasoning in humans and mice. *Cell*, 183(1):228–243.

30. Karlsson, M. P. and Frank, L. M. (2009). Awake replay of remote experiences in the hippocampus. *Nature neuroscience*, 12(7):913–918.

31. Wimmer, G. E., Liu, Y., McNamee, D., and Dolan, R. J. (2023). Distinct replay signatures for prospective decision-making and memory preservation. *Proceedings of the National Academy of Sciences*, 120(6).

32. Berners-Lee, A., Feng, T., Silva, D., Wu, X., Ambrose, E. R., Pfeiffer, B. E., and Foster, D. J. (2022). Hippocampal replays appear after a single experience and incorporate greater detail with more experience. *Neuron*, 110(11):1829–1842.

33. Igata, H., Ikegaya, Y., and Sasaki, T. (2021). Prioritized experience replays on a hippocampal predictive map for learning. *Proceedings of the National Academy of Sciences*, 118(1):e2011266118.

34. Shin, J. D., Tang, W., and Jadhav, S. P. (2019). Dynamics of awake hippocampal-prefrontal replay for spatial learning and memory-guided decision making. *Neuron*, 104(6):1110–1125.

35. Born, J. and Wilhelm, I. (2012). System consolidation of memory during sleep. *Psychological research*, 76(2):192–203.

36. Carr, M. F., Jadhav, S. P., and Frank, L. M. (2011). Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nature neuroscience*, 14(2):147–153.

37. King, B. R., Hoedlmoser, K., Hirschauer, F., Dolfen, N., and Albouy, G. (2017). Sleeping on the motor engram: the multifaceted nature of sleep-related motor memory consolidation. *Neuroscience & Biobehavioral Reviews*, 80:1–22.

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

38. Payne, J. D. and Kensinger, E. A. (2010). Sleep's role in the consolidation of emotional episodic memories. *Current Directions in Psychological Science*, 19(5):290–295.

39. Payne, J. D., Stickgold, R., Swanberg, K., and Kensinger, E. A. (2008). Sleep preferentially enhances memory for emotional components of scenes. *Psychological science*, 19(8):781–788.

40. Cohen, R. T. and Kahana, M. J. (2022). A memory-based theory of emotional disorders. *Psychological Review*, 129(4):742.

41. Talmi, D., Lohnas, L. J., and Daw, N. D. (2019). A retrieved context model of the emotional modulation of memory. *Psychological Review*, 126(4):455.

42. Oudiette, D. and Paller, K. A. (2013). Upgrading the sleeping brain with targeted memory reactivation. *Trends in cognitive sciences*, 17(3):142–149.

43. Khamassi, M. and Girard, B. (2020). Modeling awake hippocampal reactivations with model-based bidirectional search. *Biological Cybernetics*, 114(2):231–248.

44. Davidson, T. J., Kloosterman, F., and Wilson, M. A. (2009). Hippocampal replay of extended experience. *Neuron*, 63(4):497–507.

45. Lee, A. K. and Wilson, M. A. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron*, 36(6):1183–1194.

46. Sterpenich, V., van Schie, M. K. M., Catsiyannis, M., Ramyead, A., Perrig, S., Yang, H.-D., Ville, D. V. D., and Schwartz, S. (2021). Reward biases spontaneous neural reactivation during sleep. *Nature communications*, 12(1):4162.

47. Anderson, A. K. (2005). Affective influences on the attentional dynamics supporting awareness. *Journal of experimental psychology: General*, 134(2):258.

48. MacKay, D. G., Hadley, C. B., and Schwartz, J. H. (2005). Relations between emotion, illusory word perception, and orthographic repetition blindness: Tests of binding theory. *The Quarterly Journal of Experimental Psychology Section A*, 58(8):1514–1533.

49. MacKay, D. G., Shafto, M., Taylor, J. K., Marian, D. E., Abrams, L., and Dyer, J. R. (2004). Relations between emotion, memory, and attention: Evidence from taboo Stroop, lexical decision, and immediate memory tasks. *Memory & cognition*, 32(3):474–488.

50. Mather, M. (2007). Emotional arousal and memory binding: An object-based framework. *Perspectives on Psychological Science*, 2(1):33–52.

51. Diekmann, N. and Cheng, S. (2023). A model of hippocampal replay driven by experience and environmental structure facilitates spatial learning. *bioRxiv*.

52. Asfestani, M. A., Brechtmann, V., Santiago, J., Peter, A., Born, J., and Feld, G. B. (2020). Consolidation of reward memory during sleep does not require dopaminergic activation. *Journal of Cognitive Neuroscience*, 32(9):1688–1703.

53. Gomperts, S. N., Kloosterman, F., and Wilson, M. A. (2015). VTA neurons coordinate with the hippocampal reactivation of spatial experience. *Elife*, 4:e05360.

54. Buzsáki, G. (1989). Two-stage model of memory trace formation: a role for "noisy" brain states. *Neuroscience*, 31(3):551–570.

55. Csicsvari, J., O'Neill, J., Allen, K., and Senior, T. (2007). Place-selective firing contributes to the reverse-order reactivation of CA1 pyramidal cells during sharp waves in open-field exploration. *European Journal of Neuroscience*, 26(3):704–716.

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

56. O'Neill, J., Senior, T. J., Allen, K., Huxter, J. R., and Csicsvari, J. (2008). Reactivation of experience-dependent cell assembly patterns in the hippocampus. *Nature neuroscience*, 11(2):209–215.

57. Kumaran, D. and McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: a model of the hippocampal system. *Psychological review*, 119(3):573.

58. Giri, B., Miyawaki, H., Mizuseki, K., Cheng, S., and Diba, K. (2019). Hippocampal reactivation extends for several hours following novel experience. *Journal of Neuroscience*, 39(5):866–875.

59. Malerba, P., Krishnan, G. P., Fellous, J.-M., and Bazhenov, M. (2016). Hippocampal CA1 ripples as inhibitory transients. *PLoS computational biology*, 12(4):e1004880.

60. McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, 102(3):419.

61. Wimmer, G. E., Liu, Y., Vehar, N., Behrens, T. E., and Dolan, R. J. (2020). Episodic memory retrieval success is associated with rapid replay of episode content. *Nature neuroscience*, 23(8):1025–1033.

62. Mattar, M. G. and Lengyel, M. (2022). Planning in the brain. *Neuron*.

63. Jensen, K. T., Hennequin, G., and Mattar, M. G. (2023). A recurrent network model of planning explains hippocampal replay and human behavior. *bioRxiv*.

64. Pfeiffer, B. E. and Foster, D. J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, 497(7447):74–79.

65. Alvarez, P. and Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: a simple network model. *Proceedings of the national academy of sciences*, 91(15):7041–7045.

66. Singh, D., Norman, K. A., , and Schapiro, A. C. (2022). A model of autonomous interactions between hippocampus and neocortex driving sleep-dependent memory consolidation. *Proceedings of the National Academy of Sciences*, 119(44):e2123432119.

67. Sun, W., Advani, M., Spruston, N., Saxe, A., and Fitzgerald, J. E. (2021). Organizing memories for generalization in complementary learning systems. *BioRxiv*.

68. McNamee, D. C., Stachenfeld, K. L., Botvinick, M. M., and Gershman, S. J. (2021). Flexible modulation of sequence generation in the entorhinal–hippocampal system. *Nature neuroscience*, 24(6):851–862.

69. Antony, J. W. and Schapiro, A. C. (2019). Active and effective replay: systems consolidation reconsidered again. *Nature Reviews Neuroscience*, 20(8):506–507.

70. Cowan, E. T., Schapiro, A. C., Dunsmoor, J. E., and Murty, V. P. (2021). Memory consolidation as an adaptive process. *Psychonomic Bulletin & Review*, 28:1–15.

71. Yonelinas, A. P., Ranganath, C., Ekstrom, A. D., and Wiltgen, B. J. (2019). A contextual binding theory of episodic memory: systems consolidation reconsidered. *Nature Reviews Neuroscience*, 20(6):364–375.

72. Dupret, D., O'Neill, J., Pleydell-Bouverie, B., and Csicsvari, J. (2010). The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nature neuroscience*, 13(8):995–1002.

73. Joo, H. R. and Frank, L. M. (2018). The hippocampal sharp wave–ripple in memory retrieval for immediate use and consolidation. *Nature Reviews Neuroscience*, 19(12):744–757.

74. Xu, H., Baracskay, P., O'Neill, J., and Csicsvari, J. (2019). Assembly responses of hippocampal CA1 place cells predict learned behavior in goal-directed spatial tasks on the radial eight-arm maze. *Neuron*, 101(1):119–132.

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

75. Zhou, C. Y., Talmi, D., Daw, N. D., and Mattar, M. G. (2023). Episodic retrieval for model-based evaluation in sequential decision tasks. *PsyArXiv*.

76. Johnson, A. and Redish, A. D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *Journal of Neuroscience*, 27(45):12176–12189.

77. Wang, M., Foster, D. J., and Pfeiffer, B. E. (2020). Alternating sequences of future and past behavior encoded within hippocampal theta oscillations. *Science*, 370(6513):247–250.

78. Blum, K. I. and Abbott, L. F. (1996). A model of spatial map formation in the hippocampus of the rat. *Neural computation*, 8(1):85–93.

79. Jensen, O. and Lisman, J. E. (2005). Hippocampal sequence-encoding driven by a cortical multi-item working memory buffer. *Trends in neurosciences*, 28(2):67–72.

80. Mehta, M. R., Barnes, C. A., and McNaughton, B. L. (1997). Experience-dependent, asymmetric expansion of hippocampal place fields. *Proceedings of the National Academy of Sciences*, 94(16):8918–8921.

81. Redish, A. D. and Touretzky, D. S. (1998). The role of the hippocampus in solving the Morris water maze. *Neural computation*, 10(1):73–111.

82. Wikenheiser, A. M. and Redish, A. D. (2015). Hippocampal theta sequences reflect current goals. *Nature neuroscience*, 18(2):289–294.

83. Lisman, J. and Redish, A. D. (2009). Prediction, sequences and the hippocampus. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521):1193–1201.

84. Euston, D. R., Tatsuno, M., and McNaughton, B. L. (2007). Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. *Science*, 318(5853):1147–1150.

85. O'Neill, J., Boccara, C., Stella, F., Schoenenberger, P., and Csicsvari, J. (2017). Superficial layers of the medial entorhinal cortex replay independently of the hippocampus. *Science*, 355(6321):184–188.

86. Vaz, A. P., Wittig Jr, J. H., Inati, S. K., and Zaghloul, K. A. (2020). Replay of cortical spiking sequences during human memory retrieval. *Science*, 367(6482):1131–1134.

87. Wittkuhn, L. and Schuck, N. W. (2021). Dynamics of fmri patterns reflect sub-second activation sequences and reveal replay in human visual cortex. *Nature communications*, 12(1):1795.

88. Denovellis, E. L., Gillespie, A. K., Coulter, M. E., Sosa, M., Chung, J. E., Eden, U. T., and Frank, L. M. (2021). Hippocampal replay of experience at real-world speeds. *Elife*, 10:e64505.

89. Yamamoto, J. and Tonegawa, S. (2017). Direct medial entorhinal cortex input to hippocampal CA1 is crucial for extended quiet awake replay. *Neuron*, 96(1):217–227.

90. Ólafsdóttir, H. F., Carpenter, F., and Barry, C. (2017). Task demands predict a dynamic switch in the content of awake hippocampal replay. *Neuron*, 96(4):925–935.

91. Ecker, A., Bagi, B., Vértes, E., Steinbach-Németh, O., Karlócai, M. R., Papp, O. I., Miklós, I., Norbert Hájos, T. F. F., Gulyás, A. I., and Káli, S. (2022). Hippocampal sharp wave-ripples and the associated sequence replay emerge from structured synaptic interactions in a network model of area ca3. *Elife*, 11:e71850.

92. Hasselmo, M. E. (2008). Temporally structured replay of neural activity in a model of entorhinal cortex, hippocampus and postsubiculum. *European Journal of Neuroscience*, 28(7):1301–1315.

bioRxiv preprint doi: https://doi.org/10.1101/2023.03.22.533833; this version posted March 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

93. Shen, B. and McNaughton, B. L. (1996). Modeling the spontaneous reactivation of experience-specific hippocampal cell assembles during sleep. *Hippocampus*, 6(6):685–692.

94. Spens, E. and Burgess, N. (2023). A generative model of memory construction and consolidation. *bioRxiv*.

95. Barry, D. N. and Love, B. C. (2023). A neural network account of memory replay and knowledge consolidation. *Cerebral Cortex*, 33(1):83–95.

96. Káli, S. and Dayan, P. (2004). Off-line replay maintains declarative memories in a model of hippocampal-neocortical interactions. *Nature neuroscience*, 7(3):286–294.

97. Santoro, A., Frankland, P. W., and Richards, B. A. (2016). Memory transformation enhances reinforcement learning in dynamic environments. *Journal of Neuroscience*, 36(48):12228–12242.

98. Antony, J. W., Ferreira, C. S., Norman, K. A., and Wimber, M. (2017). Retrieval as a fast route to memory consolidation. *Trends in cognitive sciences*, 21(8):573–576.

99. Lohnas, L. J., Polyn, S. M., and Kahana, M. J. (2015). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall. *Psychological review*, 122(2):337.

100. Pu, Y., Kong, X.-Z., Ranganath, C., and Melloni, L. (2022). Event boundaries shape temporal organization of memory by resetting temporal context. *Nature communications*, 13(1):622.