



Memory's gatekeeper: The role of PFC in the encoding of congruent events

Inês C. Guerreiro^{a,1} and Claudia Clopath^a

Edited by Richard Morris, The University of Edinburgh, Edinburgh, Scotland; received February 21, 2024; accepted June 17, 2024

Theoretical models conventionally portray the consolidation of memories as a slow process that unfolds during sleep. According to the classical Complementary Learning Systems theory, the hippocampus (HPC) rapidly changes its connectivity during wakefulness to encode ongoing events and create memory ensembles that are later transferred to the prefrontal cortex (PFC) during sleep. However, recent experimental studies challenge this notion by showing that new information consistent with prior knowledge can be rapidly consolidated in PFC during wakefulness and that PFC lesions disrupt the encoding of congruent events in the HPC. The contributions of the PFC to memory encoding have therefore largely been overlooked. Moreover, most theoretical frameworks assume random and uncorrelated patterns representing memories, disregarding the correlations between our experiences. To address these shortcomings, we developed a HPC–PFC network model that simulates interactions between the HPC and PFC during the encoding of a memory (awake stage), and subsequent consolidation (sleeping stage) to examine the contributions of each region to the consolidation of novel and congruent memories. Our results show that the PFC network uses stored memory “schemas” consolidated during previous experiences to identify inputs that evoke congruent patterns of activity, quickly integrate it into its network, and gate which components are encoded in the HPC. More specifically, the PFC uses GABAergic long-range projections to inhibit HPC neurons representing input components correlated with a previously stored memory “schema,” eliciting sparse hippocampal activity during exposure to congruent events, as it has been experimentally observed.

memory consolidation | HPC–PFC network | schemas | inhibition | pattern separation

A fundamental question in memory research is how new memories in a labile state are transformed into more robust and permanent memories. A widely accepted framework posits that during learning, the hippocampus (HPC) encodes new information, enabling rapid acquisition of ongoing events without interfering with existing neocortical knowledge. This process is followed by hippocampal replay, during sleep, which effectively “teaches” the recently acquired information to the prefrontal cortex (PFC) (1, 2).

While extensive experimental and theoretical work support this view (2–13), recent experimental studies highlight the involvement of the PFC during the initial stages of learning (14–16), suggesting that it might have a broader role than previously thought, extending beyond offline memory consolidation. Notably, lesions and pharmacological inactivation of PFC disrupts the learning of spatial (17–19), and congruent memory tasks, i.e., tasks that are small variations of previously learned ones (20–22). While the role of PFC in memory encoding (during wakefulness) remains elusive, experimental studies indicate that prior consolidation of a PFC associative memory “schema”—preexisting network of connected neocortical representations (23)—enables rapid learning of congruent information (24, 25). Interestingly, encoding of congruent information has been correlated with a decrease in hippocampal activity (25–28). Furthermore, Guise and Shapiro have shown that mPFC inactivation reduced hippocampal pattern separation of overlapping hippocampal representations (20), a phenomenon typically ascribed to processes supported by the hippocampal neural circuitry, in particular in dentate gyrus (29–34). Taken together, these results emphasize the need to reevaluate the mechanisms by which HPC–PFC interactions support memory processing.

In this study, we propose a computational model of interacting HPC and PFC networks to examine how the PFC modulates hippocampal activity during the encoding of congruent versus novel memories, and how prior knowledge consolidated in PFC influences the learning rate of new information. In addition to excitatory HPC-to-PFC projections considered in previous memory models (for example, refs. 2, 4, 7, and 9), our model includes long-range GABAergic PFC-to-HPC connections, as recently reported

Significance

We propose a theory of memory consolidation that accounts for the role of the prefrontal cortex (PFC) in the rapid consolidation of congruent events. Using a rate-based hippocampus–PFC network model, we show that PFC learning is prior knowledge dependent. New information that is consistent with prior knowledge is learned quickly, in contrast to inconsistent information, which is learned more slowly. Moreover, we show that consolidation of novel information relies on hippocampal replay during sleep, whereas congruent information depends on the strong activation of a consistent PFC schema during wakefulness. In addition, our work highlights the PFC contributions to hippocampal pattern separation, indicating that the PFC plays a more significant role in memory processing than previously understood.

Author affiliations: ^aDepartment of Bioengineering, Imperial College London, London SW7 2AZ, United Kingdom

Author contributions: I.C.G. and C.C. designed research; I.C.G. performed research; and I.C.G. and C.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence may be addressed. Email: ines.completo@gmail.com.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2403648121/-DCSupplemental>.

Published July 17, 2024.

in ref. 35. Our computational model simulates interactions between the hippocampus and prefrontal cortex during the encoding of a memory (Awake stage) and subsequent consolidation (Sleeping stage). We begin by presenting the naive network with a new pattern representing a to-be-learned memory. After confirming that the pattern has been successfully encoded in the HPC and then consolidated into the PFC, we evaluated the responses of both the HPC and PFC when presented with an additional input pattern which is congruent (i.e., overlapping) or novel. We then examined the distinct contributions of HPC-to-PFC and PFC-to-HPC interactions. The simulations capture the rapid encoding of information consistent with prior knowledge described experimentally in ref. 24, and propose a circuit mechanism through which the PFC uses preexisting memory schemas to guide the integration of new information into the HPC–PFC network. Furthermore, our modeling work shows that the PFC creates sparse representations of congruent inputs in the HPC, enabling pattern separation.

Results

We first aim to investigate how the mechanisms underlying the long-term storage of novel and congruent information differ and the specific roles of the hippocampus and PFC in these processes over time. To that end, we begin by storing a pattern A into an untrained (i.e., naive) HPC–PFC network. Once pattern A is consolidated in the PFC, the network receives a pattern B which can either overlap (congruent) or not (novel) with the neural representation of pattern A. In our model, the HPC and PFC are described as recurrent neural networks with plastic Hebbian all-to-all intraregional connections (with a higher learning rate for HPC than for PFC), and fixed one-to-one interregional connections. In addition to excitatory HPC-to-PFC connections and a projection of the external inputs onto the HPC considered in conventional memory models, we implement inhibitory PFC-to-HPC connections, as reported in ref. 35. Based on anatomical studies (36), we also include a projection of the external inputs, i.e., of the pattern to be encoded, onto the PFC.

Encoding and Consolidation of a Memory in a Naive Neural Network. We start by storing a pattern A in the HPC–PFC network that can be used as a reference to compare and classify future incoming inputs as novel or congruent. Storing pattern A involves submitting the network to the awake stage, when it receives pattern A, followed by the sleeping stage.

Starting from a naive state (HPC and PFC connectivity: $W_{\text{HPC}} = W_{\text{PFC}} = 0$; HPC and PFC initial activity: $x_{\text{HPC}}(t_0) = x_{\text{PFC}}(t_0) = 0$), the HPC and PFC receive a pattern A to be encoded (awake stage; Fig. 1a1). At the end of the awake stage, we found strong recurrent connections within the HPC network among neurons activated by pattern A (Fig. 1a2), indicating that a memory trace of pattern A is encoded in HPC. The PFC connectivity, on the other hand, remained unchanged. This is due to the fact that the PFC learning rate is smaller than the one of HPC.

Once the HPC has formed a memory trace of input A, the HPC–PFC network enters the sleeping stage, where it cycles through a rapid eye movement (REM) (uncoupled phase; $W_{\text{PFC-HPC}} = W_{\text{HPC-PFC}} = 0$) and a nonrapid eye movement (NREM) (coupled phase; $W_{\text{PFC-HPC}} = -1$, $W_{\text{HPC-PFC}} = 1$) seven times (Fig. 1b1). Every time the network enters the REM stage, the HPC and PFC neurons are reset to a random noisy state,

and the system evolves autonomously according to its intrinsic dynamics. This enables replay of recently acquired information during sleep, which is believed to facilitate learning (5, 37, 38). In our case, the HPC network will converge to the memory engram A, which means that all neurons encoding memory pattern A will become activated in a similar fashion as during wakefulness. At the end of the sleeping stage, PFC activity reflects the neural representation of pattern A (Fig. 1b1, *Bottom*), and its connectivity resembles the hippocampal memory engram A (Fig. 1b2), although with weaker connectivity weights. Nonetheless, pattern A is consolidated in PFC, as it was confirmed by testing the PFC ability to recall pattern A upon partial activation of its neural ensemble (*SI Appendix*, Fig. S2).

Consolidation of Novel Memory Pattern Relies on Hippocampal Replay during Sleep.

Once pattern A has been consolidated in PFC, we set out to examine the HPC–PFC network's behavior when receiving a novel pattern B (0% overlap with pattern A). We consider that a long time has passed since consolidation of pattern A, and that while the PFC network retained the memory trace of A encoded in its connectivity, hippocampal connectivity decayed back to its initial naive state ($W_{\text{HPC}} = 0$). In other words, the HPC does not hold any information about pattern A; it has forgotten it.

When presented with pattern B, the hippocampal and PFC neural ensembles targeted by the novel pattern were strongly activated (Fig. 2a1). At the end of the awake stage, an engram of pattern B was encoded in the HPC connectivity (Fig. 2a2). In contrast, at this stage, the PFC did not incorporate the novel input into its network. Only after going through the sleeping stage was pattern B consolidated in the PFC (Fig. 2b2).

Interestingly, following the reset of the HPC and PFC to a noisy state in the beginning of the REM state, the PFC does not evolve toward the previously consolidated pattern A state of activity (Fig. 2b1, End of REM). This is due to the connectivity of the memory engram of pattern A not being strong enough to drive the PFC network to it and the network decays back to its resting state ($x_{\text{PFC}} = 0$). This suggests that replay during sleep is mainly driven by hippocampal activity, which in this case will evolve toward the novel pattern B.

Our modeling results suggest that incongruent knowledge previously consolidated in the PFC does not influence the mechanisms of long-term storage of novel information—the HPC–PFC network exhibits a similar behavior to the naive case. In other words, consolidation of novel information in PFC relies on offline replay (during sleep) of hippocampal activity learned during wakefulness. These results are consistent with the idea that there is a fast-learning system in HPC that quickly stores information online, which can be replayed to a slower learning system in PFC (1, 2).

Congruent Pattern Is Quickly Stored during Wakefulness. We next sought to examine the effects of previous knowledge (pattern A) on the consolidation of congruent information (pattern B with 90% overlap with A).

Contrarily to what was observed for the case of a novel incongruent pattern, when presented with a congruent input, the PFC quickly integrated its uncorrelated components (highlighted in Fig. 3a2) in its connectivity, suggesting a rapid consolidation of pattern B during the awake stage. The HPC, on the other hand, did not form a memory trace of pattern B (Fig. 3a2, *Left*). During the sleeping stage, there was no replay of pattern B, and

Encoding pattern A

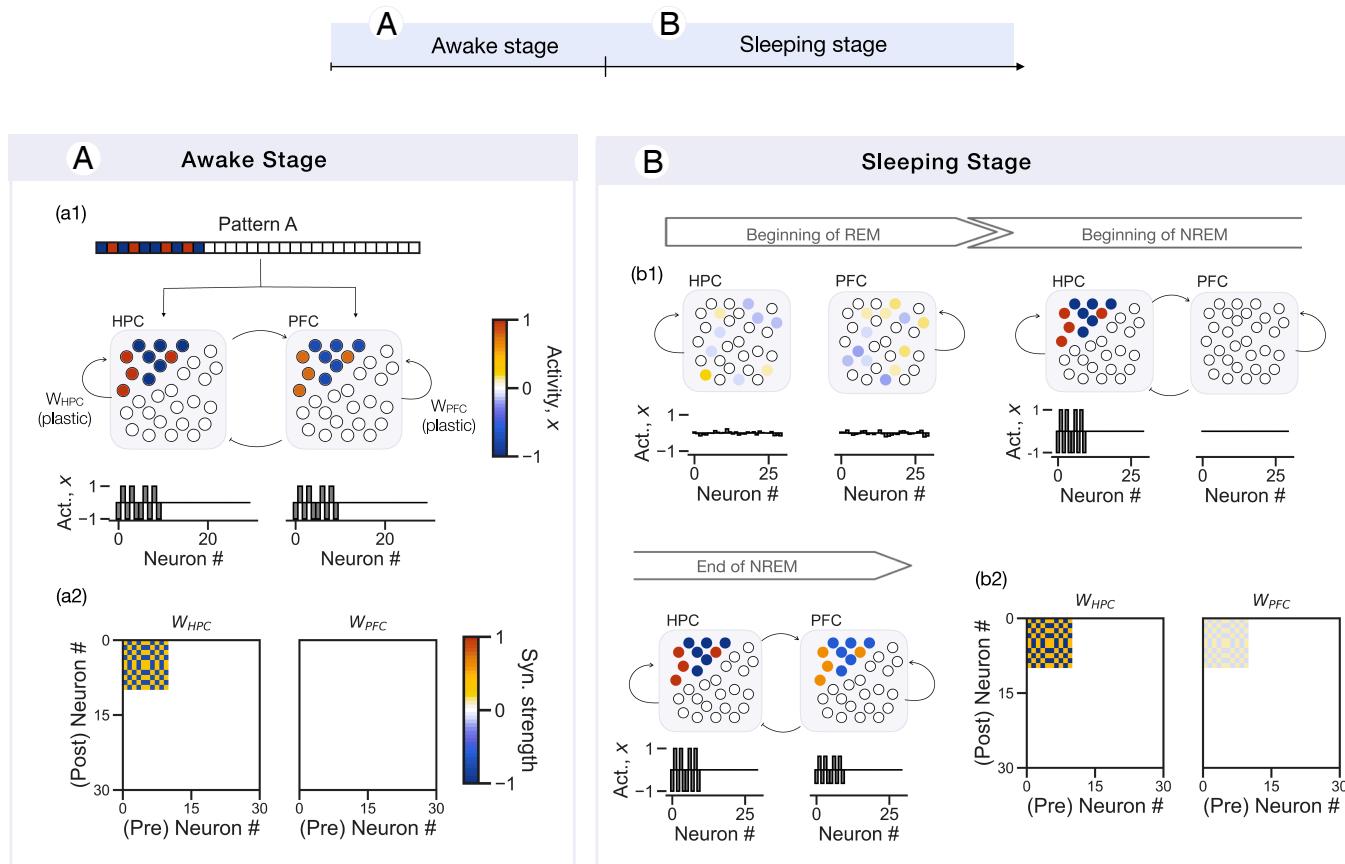


Fig. 1. Encoding and consolidation of a memory in a naive neural network. *Top*: The HPC–PFC network encodes a pattern A. The network goes through two stages: an awake stage, where the network receives pattern A, and a sleeping stage, where the network evolves autonomously according to its intrinsic dynamics. (A) During the awake state, the hippocampus (HPC) and prefrontal cortex (PFC) network receive a pattern A represented by ones (1; red entries) and minus ones (−1; blue entries), targeting the first 10 neurons in the HPC and PFC network. The recurrent connections are plastic. The two regions are coupled through fixed one-to-one HPC-to-PFC excitatory, and PFC-to-HPC inhibitory connections ($W_{HPC-PFC} = 0.5$, $W_{PFC-HPC} = -1$). Each circle represents a neuron of the network, with the color and the height of the corresponding bars representing its activity (*Top* and *Bottom*, respectively). At the end of the awake stage, the HPC and PFC show the same pattern of activation, with the HPC units more strongly activated (a1), and the HPC connectivity has formed an engram of pattern A (a2). (B) The sleeping stage is characterized by a REM phase, when the two regions are uncoupled ($W_{HPC-PFC} = W_{PFC-HPC} = 0$), and a NREM phase, when the two regions are coupled through excitatory HPC-to-PFC and inhibitory PFC-to-HPC connections ($W_{HPC-PFC} = 1$, $W_{PFC-HPC} = -1$). During the sleeping stage, the system cycles through the REM and NREM phases seven times. Every time the network enters the REM phase, the HPC and PFC networks are reset to a noisy random state, from which it evolves according to its intrinsic dynamics. In this case, the HPC converges to memory pattern A and the PFC decays to its naive state (b1; first sleep cycle). At the end of the sleeping stage, the memory engram A is consolidated in the PFC connectivity (b2).

the HPC and PFC connectivity remained the same as at the end of the awake stage (Fig. 3b2).

Interestingly, we see that HPC neurons targeted by the congruent pattern B were weakly activated, apart from those that represent the input components uncorrelated with pattern A (highlighted units in Fig. 3A), contrasting to what was observed in the case of a novel pattern. We hypothesize that this is directly modulated by inhibition from strongly activated PFC neurons.

Given the substantial overlap (90%) between the representations of patterns A and B, activation of the pattern B neural ensemble in the PFC drives the system into the attractor state formed during the consolidation of pattern A. As a result, PFC neurons that were part of the engram of A and were targeted by B became highly activated, driven by both external inputs and potentiated recurrent intrinsic connections (*SI Appendix, Figs. S4 and S5A*). The hippocampal neurons representing the correlated portion were suppressed by strongly activated PFC cells. This high PFC activity led to rapid plasticity. The strong activation of pattern B components in PFC paired with the activation of uncorrelated components driven by external and hippocampal excitatory inputs (highlighted components in Fig. 3) led to

changes in its connectivity, W_{PFC} , to encode the congruent pattern B. Furthermore, common features between patterns A and B were reinforced in PFC connectivity.

Differential Roles for HPC and PFC in the Encoding of Congruent Inputs.

We next sought to examine the contributions of the interregional connections to the rapid consolidated in PFC and sparse hippocampal activity observed during encoding of a congruent pattern. For that, we repeated the simulations in which we present the HPC–PFC network with the same congruent pattern B but set either the excitatory HPC-to-PFC or the inhibitory PFC-to-HPC to zero. Suppression of HPC-to-PFC excitatory connections ($W_{HPC-PFC} = 0$, $W_{PFC-HPC} = -1$) impaired encoding of the congruent pattern B in PFC during wakefulness (Fig. 4A), while suppressing PFC-to-HPC inhibitory connections ($W_{HPC-PFC} = 0.5$, $W_{PFC-HPC} = 0$) abolished the previously observed sparse hippocampal activity, with the HPC encoding for the full representation of pattern B (Fig. 4B). These results suggest that the bidirectional interregional HPC–PFC connections contribute in distinct ways to the behavior of the HPC–PFC network during encoding of congruent events.

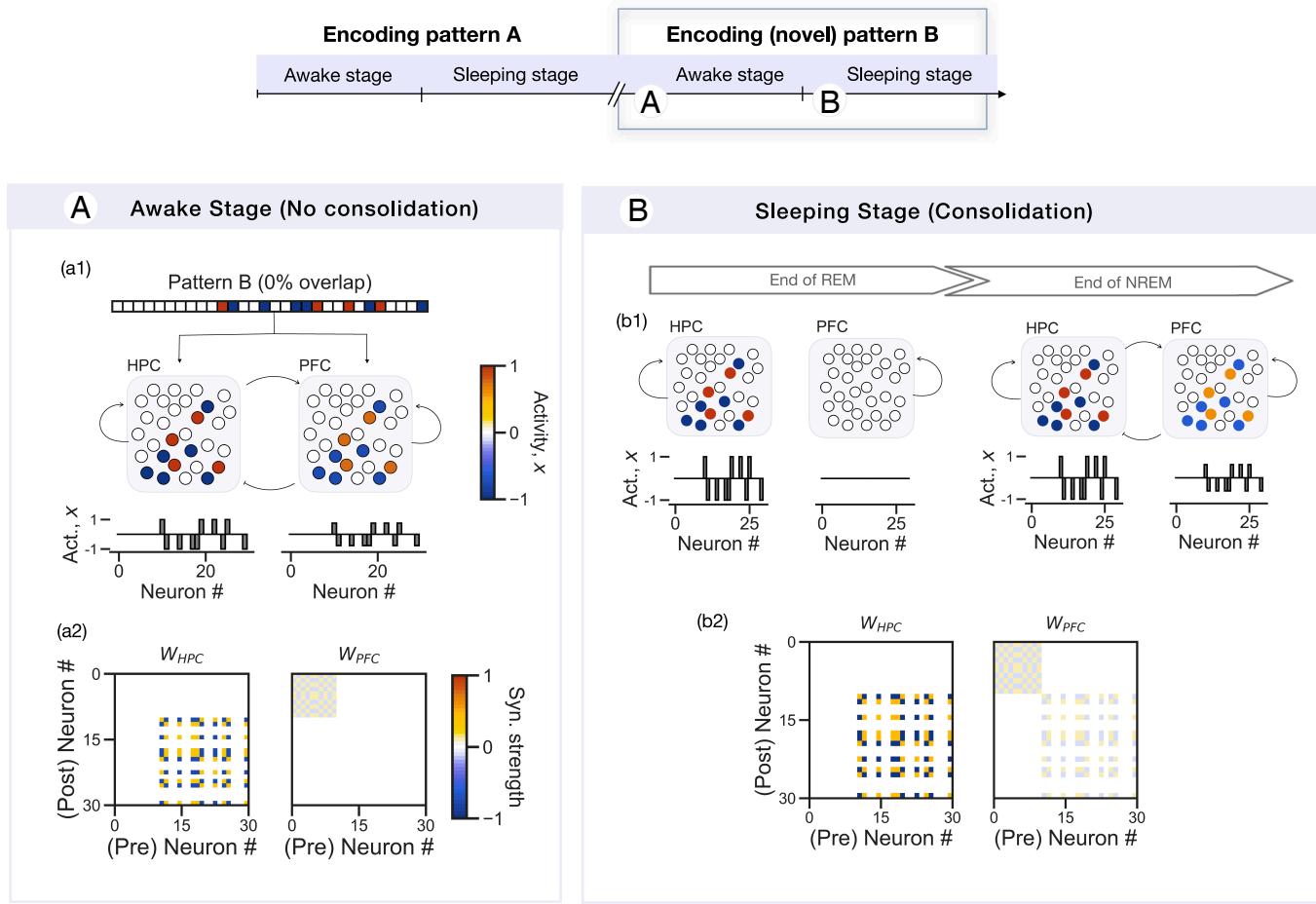


Fig. 2. Consolidation of novel memory pattern relies on hippocampal replay during sleep. *Top:* Hippocampal and prefrontal cortex activity are analyzed during the awake stage, when the network receives a pattern B whose representation does not overlap with the previously consolidated pattern A (overlap 0%), meaning that it targets a different neural ensemble. Encoding of pattern B happens after consolidation of pattern A in the PFC, and decay of its engram in HPC, i.e., when the recurrent hippocampal connectivity is back to its naive state. (A) During the awake state, the hippocampus (HPC) and prefrontal cortex (PFC) network receive a pattern B targeting 10 HPC and PFC units uncorrelated with the units encoding for pattern A (0% overlap). At the end of the awake stage, the HPC and PFC show the same pattern of activation, with the HPC units more strongly activated (a1). The hippocampal network has encoded pattern B in its connectivity, W_{HPC} , forming a memory engram B. The PFC connectivity, W_{PFC} , remains unaltered, i.e., it only encodes the memory engram A (a2). (B) During the sleeping stage, the HPC network converges to the memory pattern B at during the REM stage (b1, first sleep cycle). At the end of the sleeping stage, the memory engram B is consolidated in the PFC connectivity (b2).

In particular, hippocampal excitatory inputs facilitate the rapid encoding of congruent patterns in PFC, while PFC inhibition mediates HPC sparse activity.

Examining Influence of Degree of Congruency of New Information in PFC Plasticity and HPC-PFC Network Activity. We next sought to examine how these results generalize to degrees of overlap between patterns A and B that range from 0 to 90% (instead of considering just these two extreme cases). More specifically, we wanted to know whether there is a well-defined threshold at which the PFC network identifies an incoming pattern as congruent, triggering rapid consolidation and hippocampal sparse activity, or if it is a smooth and graded process where PFC memory traces of pattern B become stronger the greater the degree of overlap and gradually inhibit more HPC neurons. For that, we quantified changes in connectivity of PFC and HPC neurons targeted by pattern B for different degrees of overlap (0 to 90% with increments of 10%) during the awake stage and estimated the mean HPC and PFC activity (Fig. 5 A and B, respectively).

We found that while there is a tendency to have a stronger memory trace of pattern B in PFC during wakefulness (i.e., bigger

changes in PFC connectivity encoding for pattern B) the bigger the overlap with the stored pattern A, we observe a decrease in connectivity changes when pattern B overlaps with A by 40% and 50% compared with the case of 30%. The same propensity appears in the mean PFC activity, which shows a slight increase with the degree of overlap, except for an overlap of 40 and 50% (Fig. 5 B, *Inset* and *SI Appendix, Fig. S5B*). This indicates that, in the framework here considered, the PFC is able to adopt a fast (slow) speed of consolidation when the incoming information is clearly congruent (novel), i.e., it overlaps by 90% (0%) with previous knowledge. However, if an incoming pattern overlaps by 40 to 50%, the network shows an ambiguous behavior. These results align with previous indications that memories tend to be stronger when the encoded information either aligns with our previous knowledge or is completely novel (39).

Surprisingly, we found that there is a clear threshold for which hippocampal activity and plasticity dramatically decreases. This result can be explained if we consider a form of a race between the HPC and PFC regions to form a memory trace of an incoming pattern B. During the awake stage, both the HPC and PFC receive the pattern to be encoded. As they do, the activity of the HPC and PFC neurons encoding the pattern starts to increase

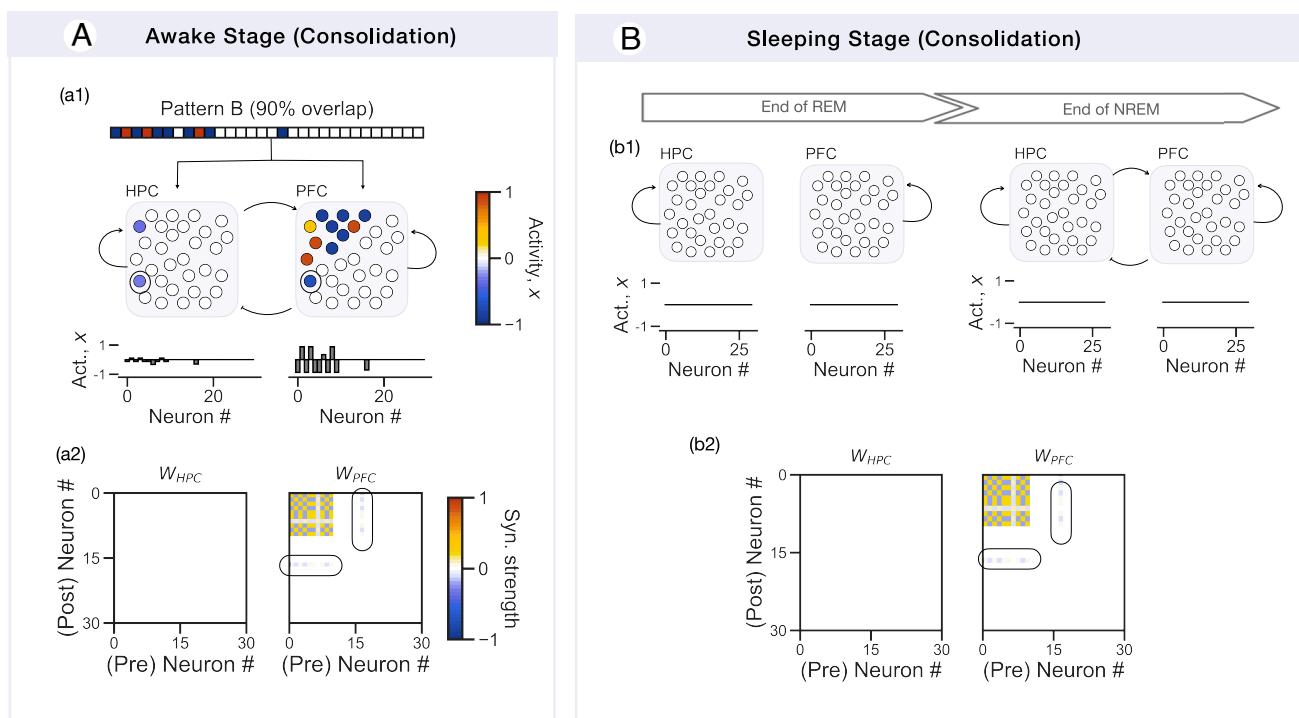
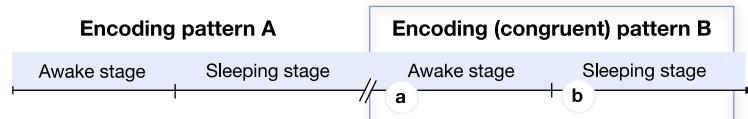


Fig. 3. Congruent pattern is quickly stored during wakefulness. *Top*: Hippocampal and prefrontal cortex activity are analyzed during the awake stage, when the network receives a pattern B whose representation overlaps with previously consolidated pattern A by 90%. Encoding of pattern B happens after consolidation of pattern A in the PFC, and decay of its engram in HPC. (A) The HPC network shows sparse activity, while the PFC units targeted by input B are strongly activated (a1). At the end of the awake stage, the HPC network connectivity remains unaltered (i.e., in its naive state). The PFC network, on the other hand, has integrated the uncorrelated components of pattern B in its connectivity with the memory engram A (a2). The circle highlights the nonoverlapping, i.e., uncorrelated, components of pattern B. (B) During the sleeping stage, the HPC network converges to its naive state (b1). At the end of the sleeping stage, the memory engram B is consolidated in the PFC connectivity but not in HPC (b2).

until it reaches a point ($x_i = 0.4$) where they start to strengthen their intraregional connections with coactivated neurons. If the PFC neurons reach a level of activation comparable to the level of activation of HPC neurons they target before HPC has the chance to significantly change its connections with other neurons representing pattern B, then their activity is suppressed by PFC (SI Appendix, Fig. S6).

Congruent Inputs Are Linked in PFC, Whereas Novel Stimuli Exhibit Pattern Separation. To examine how congruent inputs are integrated with consistent knowledge in PFC, we tested the ability of PFC to recall memory engram A and B (memory linking), or just memory engram B (pattern separation) at the end of the awake stage, when congruent inputs are consolidated, and at the end of the sleeping stage, when the HPC–PFC network replays the patterns encoded during wakefulness (Fig. 6). To that end, we examined the response of engram A and B neurons to activation of a subset of engram B cells. If all neurons of engram B were activated, but not engram A, this means that the two memories are stored independently (pattern separation). If both engrams A and B were activated, then the two memories were linked together (memory linking). For simplicity, we start by analyzing the two extreme cases: when pattern B overlaps by 10% (novel) and 90% (congruent) with A. If pattern B is novel (only overlaps by 10% with the representation of A) activating

90% of engram B cells at the end of the awake stage (Before sleep test) will not prompt the recall of either the pattern B or A in PFC (Fig. 6a1). However, if pattern B is congruent (overlaps by 90% with A) activation of solely 30% engram B units recalls not only the full pattern B, but also pattern A, indicating that at the end of the awake stage, the two memories are linked in PFC (Fig. 6a2). When performing the same test at the end of the sleeping stage, we now get that activation of 90% of engram cells of a novel pattern B causes recall of B, without recall of A (Fig. 6b1), indicating pattern separation of the two representation. For a congruent pattern B, the network shows the same performance as after the awake stage, i.e., activation of 30% engram B cells recalls both engrams A and B (Fig. 6b2).

Overall, our results show that there is a congruency threshold (40% overlap) above which the PFC network rapidly integrates incoming input with previously stored congruent patterns, linking both representations. On the other hand, inconsistent information (less than 30% overlap) relies on HPC–PFC activity replay during sleep to be consolidated in the PFC network without interfering with previously consolidated knowledge (Fig. 6C and SI Appendix, Fig. S7). Note that, according to our model, there is an intermediate result (overlap between 30 and 40%) for which incoming inputs are congruent enough to evoke recurrent activation of a correlated schema and rapidly consolidate this information with the uncorrelated components

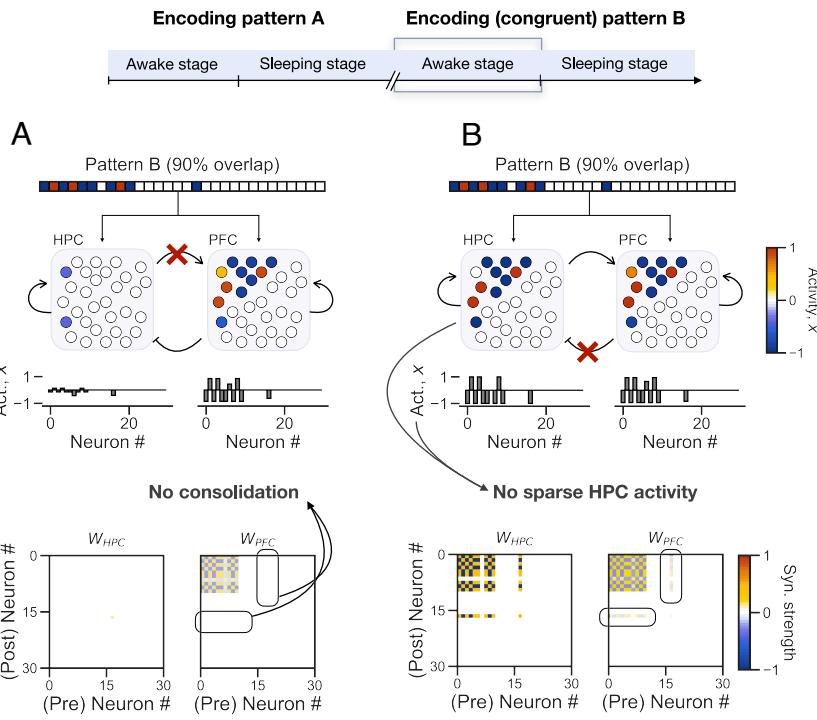


Fig. 4. Differential roles for HPC and PFC in the encoding of congruent inputs. (A) The HPC-PFC network receives an input B overlapping with memory A by 90%. Coupling between the two regions is only mediated by inhibitory PFC projections ($W_{HPC-PFC} = 0$, $W_{PFC-HPC} = -1$). At the end of the awake stage, the PFC network did not encode for the uncorrelated components of input B in its connectivity, indicating that the input was not consolidated. (B) The HPC-PFC network receives an input B that overlaps with memory A by 90%. Coupling between the two regions is only mediated by excitatory HPC projections ($W_{HPC-PFC} = 0.5$, $W_{PFC-HPC} = 0$). At the end of the awake stage, both the hippocampal and PFC network have encoded input B in its connectivity. However, we no longer have the sparse hippocampal activity observed during the encoding of congruent memories (26–28).

of the input, but the overlapping schema is not big enough to recruit the full engram of the previously stored consistent pattern. This means that such inputs are rapidly consolidated during wakefulness, but are not linked to consistent patterns.

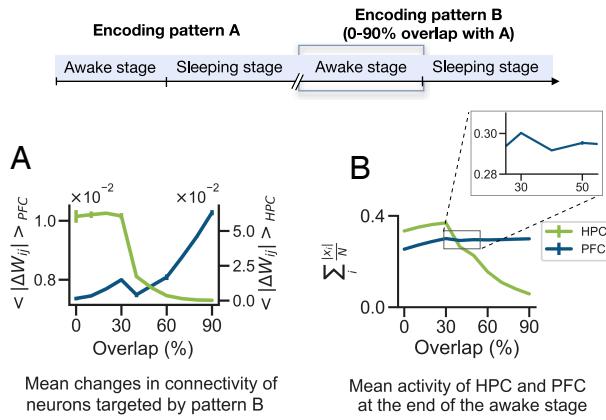


Fig. 5. Examining influence of degree of congruency of new information in plasticity and HPC-PFC network activity. Top: HPC and PFC activity and changes in connectivity are analyzed during the awake stage, when the network receives a pattern B whose representation overlaps with the previously consolidated pattern A by 0 to 90%. (A) Mean absolute changes of the PFC (blue line) and HPC (green line) connections between neurons encoding pattern B, estimated at the end of the awake stage, for patterns B overlapping by 0 to 90% with A. For each degree of overlap, we considered 10 randomly generated patterns B, and the average of the mean connectivity changes obtained for each pattern. (B) Mean absolute activity of all the HPC and PFC neurons (green and blue line, respectively). Once more, for each degree of overlap, we considered 10 randomly generated patterns B.

These results were consistent regardless of the specific pattern A considered, as we obtained the same outcomes with a different pattern A that had a different ratio of positive to negative entries (SI Appendix, Fig. S8).

We also note that the more congruent an incoming input is (the bigger the overlap with stored information), the easier it is to recall it. For example, considering the cases where pattern B overlaps by 50 and 90% with A, both patterns are consolidated in PFC during wakefulness and linked to pattern A. However, with a 90% overlap, we only need to activate 30% of engram B cells to recall it, while recall for the case of a 50% overlap requires the activation of 90% engram B cells (SI Appendix, Fig. S7). In other words, the linking between patterns is stronger the bigger the overlap. This will impact how new patterns are encoded. For instance, a new pattern C that partially overlaps with patterns A and B may or may not become linked to the existing patterns, depending on the strength of the linkage between A and B (SI Appendix, Fig. S9).

Considering that, for a certain degree of overlap, the neurons representing the correlated components of a pattern B are randomly selected, different patterns are going to have the same degree of overlap with pattern A. While examining how different patterns with a same degree of overlap affect the results (we considered 10 randomly generated patterns), we found that we have approximately consistent results (same as for Fig. 6C) except for pattern with a degree of overlap of 40% with stored information. When testing recall of pattern B with a 40% overlap with A at the end of the awake stage, for certain patterns the PFC could not recall it (see SI Appendix, Fig. S10 for examples of a pattern B with 40% overlap resulting in a successful and unsuccessful recall). This aligns with previous

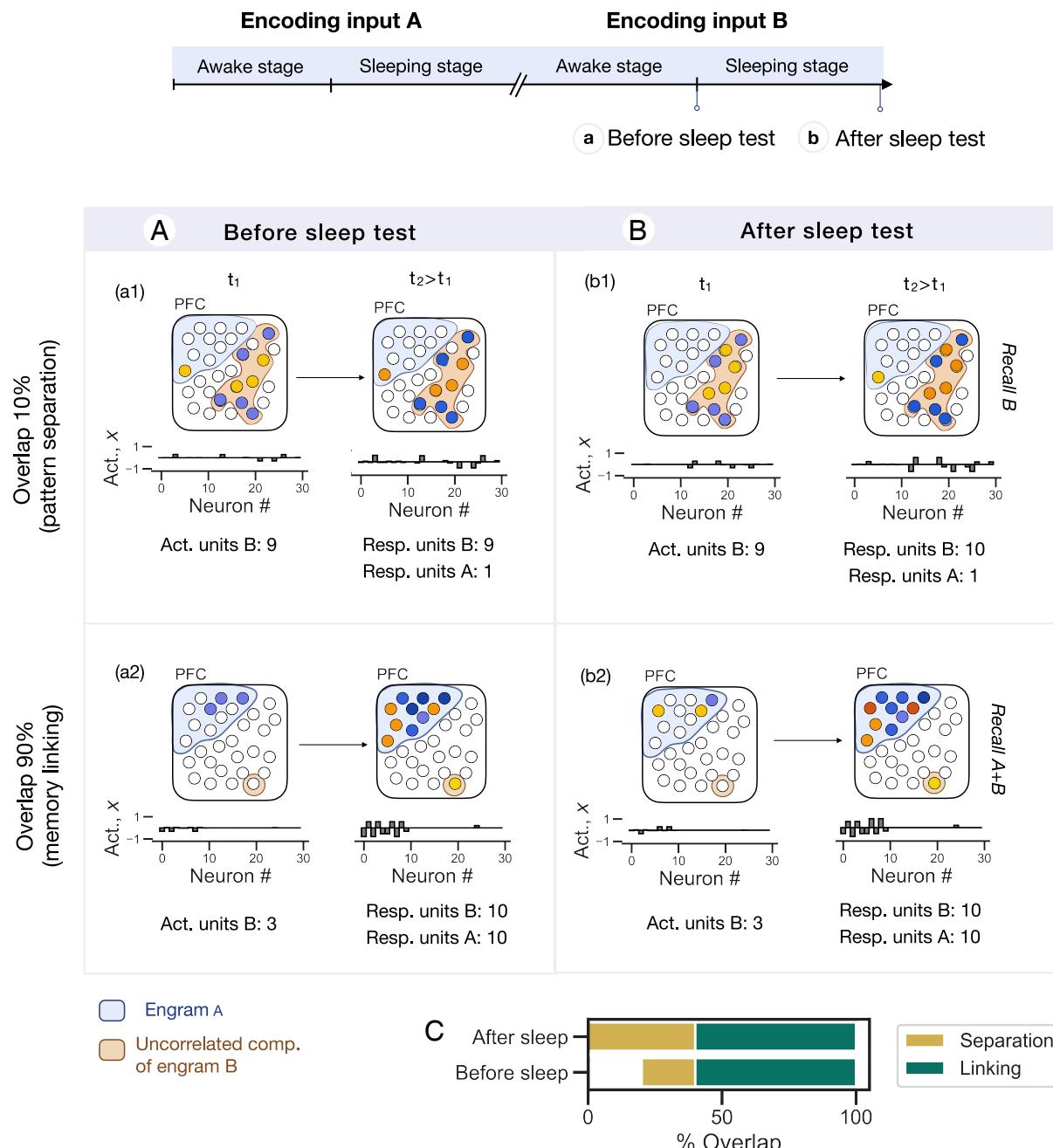


Fig. 6. Congruent inputs are linked in PFC whereas uncongruent stimuli exhibit pattern separation. *Top:* Testing PFC ability to perform pattern separation and memory linking at the end of the awake stage and at the end of the sleeping stage. (*A*) Testing the ability of PFC to recall pattern B and pattern A at the end of the awake stage. If pattern B overlaps by 10% with pattern A, the PFC will not be able to recall engram B or engram A (*a1*). The PFC pattern of activation 90 time steps (t_1) and 7,600 time steps (t_2) after activating 9 out of 10 engram B units is the same (*Left* and *Right*, respectively). If pattern B overlaps by 90% with pattern A, activating 3 engram B units results in the recall of both engram A and engram B, indicating that the two patterns are linked (i.e., activation of engram A plus the uncorrelated components of engram B; *a2*). (*B*) Testing the PFC ability to recall pattern B and pattern A at the end of the sleeping stage. If pattern B overlaps by 10% with pattern A, activating a subset of engram B units results in recall of engram B but not engram A, indicating pattern separation (*b1*). If pattern B overlaps by 90%, activation of a subset of engram B units recalls engram A and B, similar to what was observed at the end of the awake stage (*b2*). (*C*) Classifying pattern separation and memory linking at the end of the awake stage (Before sleep) and at the end of the sleeping stage (After sleep) for patterns B overlapping by 0 to 90% with pattern A. Pattern separation is defined as recall of engram B without recall of A. Congruent inputs (overlap >40%) are encoded in PFC during awake stage and are linked to previously consolidated overlapping representations. Novel inputs are separated and consolidated during sleep.

results showing that there is a break in the general tendency of increased changes in the PFC connectivity during wakefulness with the degree of overlap at 40% (Fig. 5*A*), reinforcing the

idea that the ease in which “in-between” events (in this case, with a congruence degree of 40%) are recalled is inconsistent. This ambiguous behavior is particularly accentuated during

the awake stage of encoding, with results stabilizing during sleep.

Altogether, our model proposes that congruent inputs are linked in PFC to congruent information during wakefulness, while inconsistent information relies on HPC and PFC replay during sleep to be consolidated without interfering with stored knowledge.

Discussion

In the wild, adult animals rarely encounter new information in isolation. Their experiences are usually linked to what they have encountered before. Previous research shows that past experiences affect how new memories are processed (40, 41). For instance, consider how easily you grasp new information related to your field compared to unrelated information. However, most memory studies overlook the impact of previously acquired knowledge in experiments. In recent work, Tse et al. (22, 24) showed that new associations consistent with a previously consolidated PFC schema quickly become HPC-independent, suggesting rapid PFC consolidation. Consequent theoretical work by McClelland has shown that a Rumelhart network is able to integrate new information into existing consistent knowledge (42). However, such networks use a backpropagation learning rule, which is nonlocal, and it requires the propagation of error signals backward through symmetric feedback connections, making it biologically implausible. Thus, the neural and circuit mechanisms underlying the rapid consolidation of congruent information remain elusive.

In this work, we hypothesize that schemas stored in PFC promote rapid integration of congruent events without sleep by enabling strong activation of engram cells representing the overlapping representation of the stored and new events. According to our model, rapid consolidation relies on activation of PFC neurons by an external and hippocampal input. By adopting a Hebbian learning rule, where changes in synaptic strength depend on the level of activation of the pre- and postsynaptic units, we can overcome the slow learning rates characteristic of the PFC by strongly activating the interacting neurons that form a memory engram. A congruent external input drives the PFC network to the closest attractor state, previously formed during the consolidation of a similar pattern and characterized by strong recurrent connections. Neurons activated by both the external input and strong recurrent connections, i.e., neurons representing the correlated components of the congruent input, will be activated strongly enough to quickly strengthen their connections with neurons representing the uncorrelated components. In other words, the “new part” of a congruent input will readily be incorporated with the stored schema. If, on the other hand, the circuit receives a novel input, the PFC network will not converge to an attractor state, and activation of the targeted neurons will be solely due to the action of the external input. Neurons representing the novel input will be weakly activated, and will not be able to strengthen their synapses and form an engram. In this case, the HPC–PFC network needs to go through the sleeping stage, where the HPC repeatedly reactivates the neural ensembles representing the novel input, in order to consolidate it. It is important to note that, in this case, our model is in line with conventional accounts of memory consolidation mechanisms (for example, ref. 1).

Our model captures several important experimental findings; namely, the quick consolidation of events consistent with prior knowledge, that is disrupted by removal of the hippocampus

(22, 24). Our results grant a broader role to the PFC in the encoding of events than classically considered. Besides hypothesizing the PFC potential to consolidate memory events without hippocampal replay during sleep, we also propose its potential role in supporting HPC by modulating its activity during the encoding phase. When incoming information is congruent with previously acquired knowledge, the PFC quickly incorporates its uncorrelated components with a preexisting overlapping schema in its network through strengthening of its cortico-cortical functional connections and inhibits the hippocampal activity encoding for the correlated components of the new input.

Many studies have suggested that PFC exerts top-down control over information processing in the HPC (35, 43–46). Here, motivated by recent anatomical studies reporting long-range GABAergic projections from the PFC to HPC (35), we propose a network mechanism through which the PFC exerts top-down inhibition over hippocampal activity during the encoding of congruent events. We propose that this mechanism may be responsible for the sparse activity observed in the hippocampus during the encoding of congruent events. A large body of work has focused on hippocampal contributions to pattern separation (29–34). However, converging evidence suggests that pattern separation is supported by a network of brain regions (47). Here, we predict that PFC can contribute to hippocampal pattern separation by suppressing hippocampal activity encoding information already consolidated in PFC. Such a mechanism could also promote increased memory capacity of the HPC (SI Appendix, Fig. S11).

In summary, our modeling work suggests the following: 1) rapid encoding of consistent information in PFC during wakefulness is mediated by hippocampal inputs and strong activation of a congruent PFC schema; 2) inconsistent information relies on hippocampal replay during sleep to be consolidated; 3) congruent information is integrated in PFC into a congruent schema whereas novel information undergoes pattern separation, producing little interference with memories already stored in PFC; 4) GABAergic PFC-to-HPC projections induce sparse hippocampal activity during the encoding of congruent events, enabling pattern separation.

Methods

Multiregion Recurrent Neural Network. The hippocampus and prefrontal cortex regions are modeled as rate-based recurrent networks. Each region is composed of $N = 30$ units with all-to-all connections. The dynamics of each unit x_i is described by the following equation:

$$\frac{dx_i}{dt} = -x_i + \phi \left(\sum_j W_{ij} x_j + I_i \right), \quad [1]$$

where ϕ is a nonlinearity applied to the total input each unit receives. It is modeled as $\phi(x) = \tanh(x)$ for the hippocampal units and $\phi(x) = \tanh(0.5x)$ for the prefrontal cortex units, reflecting differences in the responsiveness of neurons in both regions (48). More specifically, PFC units exhibit a more gradual response compared to HPC. W_{ij} is the synaptic weight between the pre- and postsynaptic units, j and i , respectively, and I_i the external input. We considered both regions to receive identical external inputs. Patterns of ones (1) and minus ones (−1) describe the pattern we want to store into the network. We consider inputs represented by neural ensembles with 10 units. In other words, an external input is composed of 10 entries of values 1 or −1, and the remaining 20 entries are zero. We assume that different patterns are represented by different neural ensembles of the same size. We will use the term “congruent input” to refer

to an external input that is represented by a subset of neurons that overlaps significantly with the representation of a previously stored input.

Intraregional connections are dynamic and change according to the following standard Hebbian learning rule:

$$\frac{dW_{ij}}{dt} = \frac{\lambda_j}{N} (1 - W_{ij}) \Theta(|x_j| - 0.4) \Theta(|x_i| - 0.4) x_j x_i - \frac{\lambda_d}{N} W_{ij}, \quad [2]$$

where Θ is the step function $\Theta(x) = 0$ for $x < 0$ and is 1 otherwise, λ_j is the learning rate ($\lambda_j(HPC) = 0.45$, $\lambda_j(PFC) = 0.06$), and λ_d the decay rate ($\lambda_d(HPC) = 0.55$, $\lambda_d(PFC) = 0$). A gating mechanism has been introduced to the learning rule, contingent on the activity of the neurons involved. When the absolute activities of the pre- and postsynaptic neurons, x_j and x_i , exceed a threshold of 0.4, the standard Hebbian term contributes to the adjustment of synaptic strengths, aiming to reinforce connections that are correlated with neuronal activity. This gating mechanism enables the learning rule to account for situations where neuronal activity is insufficient to drive synaptic potentiation, allowing for a more biologically realistic representation of learning dynamics in the network.

Interregional connections are mediated through one-to-one excitatory HPC-to-PFC connections and reciprocate inhibitory PFC-to-HPC connections. Note that, these connections are not of the same nature. Excitatory functional connections from HPC to PFC are mainly modulated by intermediate regions [for example, the thalamus (49)]. In this case, the strength of the connections $W_{HPC-PFC}$ reflects the coupling strength of hippocampal and PFC oscillatory activity. On the other hand, inhibitory PFC-HPC connections represent long-range GABAergic projections, i.e., direct connections from the PFC to HPC (35). The value of $W_{PFC-HPC}$ is representative of the conductance of GABA receptors in HPC neurons.

For simulations, the differential equations were solved using Euler's method with a time step $\Delta t = 0.01$ (a.u.).

Simulating Awake and Sleep. During the awake stage, the HPC and PFC receive an external input, I . The two regions are coupled through one-to-one excitatory HPC-to-PFC connections ($W_{HPC-PFC} = 0.5$), and one-to-one inhibitory PFC-to-HPC projections ($W_{PFC-HPC} = -1$). The awake stage has a duration of 7,600 time steps.

Following the awake stage, the model enters the sleep period. The model has a rapid eye movement (REM) sleep stage, where the two regions are uncoupled and evolve autonomously according to their intrinsic dynamics, and a non-rapid eye movement (NREM) sleep stage, where dynamics between the hippocampus and PFC are tightly coupled (4). During sleep, coupling is modeled by setting $W_{HPC-PFC} = 1$ and $W_{PFC-HPC} = -1$. This setup is inspired by previous studies suggesting that during NREM sleep, the coupling between HPC and PFC tends to be relatively strong, while in REM sleep, the coupling between the HPC and PFC is generally weaker (5, 37). We find that alternating between the REM and NREM sleep stages 7 times facilitates the consolidation of hippocampal-dependent memories into the PFC network, which has a small learning rate. Each time the model enters the REM stage, the hippocampal and PFC neurons are reset to a noisy state (38). The state of each neuron is randomly drawn from a normal distribution with SD 0.1 and mean 0.

The REM and NREM stages have a duration of 9,000 and 900 time steps, respectively.

Model Simulations. Starting the network from a naive state, i.e., with hippocampal and prefrontal cortex connectivities set to zero ($W_{HPC} = 0$, $W_{PFC} = 0$), the network receives a pattern A to encode. Pattern A is represented by 10 HPC and PFC units; in other words, it targets the first 10 neurons of each region (Pattern A = $[-1 \ 1 \ -1 \ 1 \ -1 \ -1 \ 1 \ -1 \ 0 \ \dots \ 0]$). Storing pattern A involves submitting the network to the awake stage, when it receives pattern A, followed by the sleeping stage, instead of being set up directly in the PFC network. This is done to ensure that the strength which pattern A is encoded in the PFC connectivity is a result of the natural awake-sleep cycle to avoid introducing a bias in simulations that follow.

After going through the awake and sleeping stage, we verify that pattern A was successfully consolidated in the PFC network. We note that although in our simulations pattern A is encoded in PFC after one awake-sleep cycle, which

corresponds to one day, and under experimental conditions learning a new task requires several days, replicating awake-sleeping repetitions during encoding of pattern A does not change our results (SI Appendix, Fig. S1). Similarly, setting the HPC network connectivity W_{HPC} to a random state (each entry drawn from a normal distribution with mean 0 and SD 0.03) at the beginning of the awake and sleeping stage, instead of resetting to a naive state, won't change the results (SI Appendix, Fig. S3).

We then assume that a long time has passed (say months) and that while the PFC network retained the memory trace of A in its connectivity, the hippocampal network decayed back to its naive state. At this point, the network receives an input B whose representation overlaps with pattern A.

Pattern B is generated by randomly selecting a fraction f of the first 10 entries of pattern A to remain unaltered (i.e., the entries with values 1 and -1) and the rest of the initial 10 entries ($10-f$) are set to zero. The parameter f reflects the extent of overlap between patterns A and B. For instance, to generate a pattern B with an 80% overlap with A, 8 out of the 10 first entries in pattern A remain the same. To ensure that any disparities in activity observed in the HPC and PFC result exclusively from their interactions, and not because input B targets a different number of units, out of the 20 zero entries of pattern A, we randomly select ($10-f$) entries to change to 1 or -1 (also in a random independent way).

Testing Pattern Completion, Pattern Separation, and Memory Linking. To evaluate the PFC network's ability to perform pattern completion, we assessed the responsiveness of neurons to the activation of a subset of engram cells. A neuron, denoted i , was deemed responsive if after 7,600 time steps $|x_i| > 0.2$, while activity below this threshold was considered noise. Successful pattern completion refers to the network's capacity to reactivate all cells within a memory engram when a subset is activated. This implies the consolidation of the memory within the network. The degree of consolidation is determined by the size of the subset required for pattern completion, which we refer to as the "Recall threshold." The smaller the subset of cells needed to achieve pattern completion, the stronger the consolidation of the memory. To examine pattern separation and memory linking, we tested the network's ability to perform pattern completion of engram B when a subset of engram B neurons was activated, as well as the pattern completion of engrams A and B when a subset of engram B cells were activated. The successful pattern completion of both engrams when only a subset of engram B is activated indicates the linking of both memories. On the other hand, pattern completion of engram B without engram A indicates pattern separation. To ensure the accuracy of our findings, we assessed the responsiveness of the total number of cells forming an engram to the activation of subsets of different sizes. The subset of cells to be activated was randomly selected. For each subset size, we repeated the analysis 50 times, activating a different subset of neurons each time.

Sparsity Index. In this study, two distinct approaches are utilized to quantify and analyze sparsity. First, we investigate the amplitude and position of density distribution peaks in the HPC and PFC activity. A higher peak centered around zero indicates sparser activity, suggesting that a majority of neurons within the network exhibit zero activity. To visualize these distributions, we employ kernel density estimation (KDE) plots. Second, we calculate a sparsity index. This index is determined by computing the average absolute activity across all neurons in the network, represented as $\sum_i \frac{|x_i|}{N}$. A lower value of the sparsity index suggests sparser activity. To examine the sparsity dynamics during the encoding of various types of information, we conduct the analysis for a range of inputs that have different degrees of overlap (0% to 90%) with a previously stored memory.

Data, Materials, and Software Availability. Code data have been deposited in (https://github.com/inesCompleto/role_PFC_consolidation) (50).

ACKNOWLEDGMENTS. This work was supported by Biotechnology and Biological Sciences Research Council BB/N013956/1, BB/N019008/1, Wellcome Trust200790/Z/16/Z, Simons Foundation 564408, and Engineering and Physical Sciences Research Council EP/R035806/1.

1. J. McClelland, B. L. McNaughton, R. C. O'Reilly, Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* **102**, 419–457 (1995).
2. D. Marr, Simple memory: A theory for archicortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **262**, 23–81 (1971).
3. B. Milner, W. Penfield, The effect of hippocampal lesions on recent memory. *Trans. Am. Neurol. Assoc.* **80**, 42–48 (1955).
4. D. Singh, K. A. Norman, A. C. Schapiro, A model of autonomous interactions between hippocampus and neocortex driving sleep-dependent memory consolidation. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2123432119 (2022).
5. S. Diekelmann, J. Born, The memory function of sleep. *Nat. Rev. Neurosci.* **11**, 114–126 (2010).
6. J. G. Klinzing, N. Niethard, J. Born, Mechanisms of systems memory consolidation during sleep. *Nat. Neurosci.* **22**, 1598–1610 (2019).
7. L. R. Squire, Mechanisms of memory. *Science* **232**, 1612–1619 (1986).
8. T. J. Teyler, P. DiScenna, The hippocampal memory indexing theory. *Behav. Neurosci.* **100**, 147–154 (1986).
9. P. Alvarez, L. R. Squire, Memory consolidation and the medial temporal lobe: A simple network model. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7041–7045 (1994).
10. E. T. Rolls, A theory of hippocampal function in memory. *Hippocampus* **6**, 601–620 (1996).
11. T. D. Goode, K. X. Tanaka, A. Sahay, T. J. McHugh, An integrated index: Engrams, place cells, and hippocampal memory. *Neuron* **107**, 805–820 (2020).
12. L. Nadel, G. Winocur, L. Ryan, M. Moscovitch, Systems consolidation and hippocampus: Two views. *Debates Neurosci.* **1**, 55–66 (2007).
13. M. J. Sekeres, G. Winocur, M. Moscovitch, The hippocampus and related neocortical structures in memory transformation. *Neurosci. Lett.* **680**, 39–53 (2018).
14. P. K. Dash, A. E. Hebert, J. D. Runyan, A unified theory for systems and cellular memory consolidation. *Brain Res. Rev.* **45**, 30–37 (2004).
15. H. Miyawaki, K. Mizuseki, De novo inter-regional coactivations of preconfigured local ensembles support memory. *Nat. Commun.* **13**, 1272 (2022).
16. T. Kitamura *et al.*, Engrams and circuits crucial for systems consolidation of a memory. *Science* **356**, 73–78 (2017).
17. R. J. Kyd, D. K. Bilkey, Prefrontal cortex lesions modify the spatial properties of hippocampal place cells. *Cereb. Cortex* **13**, 444–451 (2003).
18. L. M. DeVito, H. Eichenbaum, Distinct contributions of the hippocampus and medial prefrontal cortex to the "what-where-when" components of episodic-like memory in mice. *Behav. Brain Res.* **215**, 318–325 (2010).
19. J. C. Churchwell, A. M. Morris, N. D. Musso, R. P. Kesner, Prefrontal and hippocampal contributions to encoding and retrieval of spatial memory. *Neurobiol. Learn. Memory* **93**, 415–421 (2010).
20. K. G. Guise, Medial prefrontal cortex reduces memory interference by modifying hippocampal encoding. *Neuron* **94**, 183–192.e8 (2017).
21. L. M. DeVito, C. Lykken, B. R. Kanter, H. Eichenbaum, Prefrontal cortex: Role in acquisition of overlapping associations and transitive inference. *Learn. Memory* **17**, 161–167 (2010).
22. D. Tse *et al.*, Schema-dependent gene activation and memory encoding in neocortex. *Science* **333**, 891–895 (2011).
23. M. T. R. van Kesteren, T. I. Brown, A. D. Wagner, Learned spatial schemas and prospective hippocampal activity support navigation after one-shot learning. *Front. Hum. Neurosci.* **12**, 486 (2018).
24. D. Tse *et al.*, Schemas and memory consolidation. *Science* **316**, 76–82 (2007).
25. M. T. R. van Kesteren *et al.*, Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: From congruent to incongruent. *Neuropsychologia* **51**, 2352–2359 (2013).
26. M. P. Karlsson, L. M. Frank, Network dynamics underlying the formation of sparse, informative representations in the hippocampus. *J. Neurosci.* **28**, 14271–14281 (2008).
27. J. D. Cohen, M. Bolstad, A. K. Lee, Experience-dependent shaping of hippocampal CA1 intracellular activity in novel and familiar environments. *eLife* **6**, e23040 (2017).
28. J. S. Lee, J. J. Briguglio, J. D. Cohen, S. Romani, A. K. Lee, The statistical structure of the hippocampal code for space as a function of time, context, and value. *Cell* **183**, 620–635.e22 (2020).
29. R. C. O'Reilly, J. L. McClelland, Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off. *Hippocampus* **4**, 661–682 (1994).
30. A. Treves, E. T. Rolls, Computational analysis of the role of the hippocampus in memory. *Hippocampus* **4**, 374–391 (1994).
31. J. K. Leutgeb, S. Leutgeb, M. B. Moser, E. I. Moser, Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science* **315**, 961–966 (2007).
32. M. A. Yassa, C. E. Stark, Pattern separation in the hippocampus. *Trend. Neurosci.* **34**, 515–525 (2011).
33. J. P. Neunuebel, J. J. Knierim, CA3 retrieves coherent representations from degraded input: Direct evidence for CA3 pattern completion and dentate gyrus pattern separation. *Neuron* **81**, 416–427 (2014).
34. J. J. Sakon, W. A. Suzuki, A neural signature of pattern separation in the monkey hippocampus. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 9634–9643 (2019).
35. R. Malik, Y. Li, S. Schamiloglu, V. S. Sohal, Top-down control of hippocampal signal-to-noise by prefrontal long-range inhibition. *Cell* **185**, 1620–1617.e17 (2022).
36. L. Golmaya, A. Nunez, L. Zaborszky, Electrophysiological evidence for the existence of a posterior cortical-prefrontal basal forebrain circuitry in modulating sensory responses in visual and somatosensory rat cortical areas. *Neuroscience* **119**, 597–609 (2003).
37. S. A. Cairney, S. J. Durrant, R. Power, P. A. Lewis, Complementary roles of slow-wave sleep and rapid eye movement sleep in emotional memory consolidation. *Cereb. Cortex* **25**, 1565–1575 (2015).
38. B. Rasch, J. Born, About sleep's role in memory. *Physiol. Rev.* **93**, 681–766 (2013).
39. A. Alonso, J. van der Meij, D. Tse, L. Genzel, Naïve to expert: Considering the role of previous knowledge in memory. *Brain Neurosci. Adv.* **4**, 2398212820948686 (2020).
40. F. C. Bartlett, *Remembering: A Study in Experimental and Social Psychology* (Cambridge University Press, 1932).
41. H. F. Harlow, The formation of learning sets. *Psychol. Rev.* **56**, 51–65 (1949).
42. J. L. McClelland, Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *J. Exp. Psychol. Gen.* **142**, 1190–1210 (2013).
43. H. Eichenbaum, Prefrontal-hippocampal interactions in episodic memory. *Nat. Rev. Neurosci.* **18**, 547–558 (2017).
44. R. G. Benoit, J. C. Hulbert, E. Huddleston, M. C. Anderson, Adaptive top-down suppression of hippocampal activity and the purging of intrusive memories from consciousness. *J. Cognit. Neurosci.* **27**, 96–111 (2015).
45. P. Rajasethupathy *et al.*, Projections from neocortex mediate top-down control of memory retrieval. *Nature* **526**, 653–659 (2015).
46. A. R. Preston, H. Eichenbaum, Interplay of hippocampus and prefrontal cortex in memory. *Curr. Biol.* **23**, R764–R773 (2013).
47. T. Amer, L. Davachi, Extra-hippocampal contributions to pattern separation. *eLife* **12**, e82250 (2023).
48. R. N. Ruggiero *et al.*, Neuromodulation of hippocampal-prefrontal cortical synaptic plasticity and functional connectivity: Implications for neuropsychiatric disorders. *Front. Cell. Neurosci.* **15**, 732360 (2021).
49. D. F. Tomé, S. Sadeh, C. Clopath, Coordinated hippocampal-thalamic-cortical communication crucial for engram dynamics underneath systems consolidation. *Nat. Commun.* **13**, 840 (2022).
50. I. C. Guerreiro, Code for simulations from "Memory's gatekeeper: the role of PFC in the encoding of congruent events". Github. https://github.com/inesCompleto/role_PFC_consolidation. Deposited 19 March 2024.