Segregation-to-Integration Transformation (SIT) Model of Memory Evolution

Luz Bavassi[1,2], Lluís Fuentemilla[3-5]

[1]Laboratorio de Neurociencias de la Memoria, IFIByNE - UBA, CONICET, Argentina
[2]Departamento de Física, Universidad de Buenos Aires, Buenos Aires, Argentina
[3]Department of Cognition, Development and Education Psychology, University of Barcelona Spain
[4]Institute of Neuroscience (UBNeuro), University of Barcelona, Spain
[5]Bellvitge Institute for Biomedical Research, Hospitalet de Llobregat, Spain

Correspondence: Luz Bavassi (luzbavassi@gmail.com)

Abstract

What dynamics characterize the transformation of memories over time? Here we introduce a neural network formalization that reveals that memory representations undergo a transition from highly segregated to richly integrated network forms, driven by a combination of neural network reactivations, spreading, and synaptic plasticity rules. Modularity, as a fundamental organizing principle, allows information segregation into cohesive modules, preserving specific sets of information while facilitating efficient spread throughout the network. Through our modeling approach, we reveal an optimal window during this transformation where memories are most susceptible to malleability, suggesting a non-linear or inverted U-shaped function in memory evolution. The results of our model integrate a wide range of experimental phenomena along with accounts of memory consolidation and reconsolidation, offering a unique perspective on memory evolution by leveraging simple architectural neural network property rules.

Introduction

Memories, like other systems in nature, have the chance to evolve because if they do, they improve their chance for survival in a behaving organism. Their perpetuity, however, is constrained by the temporally evolving and ever-changing experiential needs of the organism. Therefore, memories, as experience-derived informational entities, are forced to rely on a coding scheme that dynamically adjusts their representational structure maximizing both persistence and utility along their evolutionary path. However, though it is widely accepted that memories evolve at the expense of changes in their architecture, to date, there is still a paucity in understanding how memory persistence and utility are balanced along the way. A major research challenge is to know when and how experience-induced utility drives changes in memory to promote persistence.

Our current understanding of memory evolution suggests that initially, memories are encoded by groups of neurons with synchronized activity, forming stable neural ensembles, also termed engrams, that if reactivated, induce memory retrieval (Josselyn & Tonegawa, 2020). While these neural ensembles are initially identified in the hippocampus, over time, they may undergo structural changes, losing their modular properties (i.e., highly synchronized, and clustered neural ensembles) (Gonzalez et al., 2019), and other neural ensembles beyond the hippocampus, such as the prefrontal cortex, gradually assume their representation (Frankland & Bontempi, 2005). This transition, however, occurs relatively slowly and requires repeated reactivation of the initial neural ensembles to shift the memory representation between networked regions (Frankland & Bontempi, 2005). For a while, the dominating view was that once memories shifted to neocortical neural ensembles, they became consolidated and durable in the long term. More contemporary views advocate for the notion that consolidation and reactivation may act in the service of generalization (Sun et al., 2023). Given that individual memorized experiences rarely repeat exactly, generalization allows us to identify systematic relationships between features of the world, ultimately involving extending learned information to novel contexts. Thus, generalization involves the process of linking and extracting commonalities among various memories. As the brain generalizes, initial memory representations undergo a transformative shift in their structure, becoming intricately connected to related memories (Nadel & Moscovitch, 1997; Winocur et al., 2010). This linkage occurs as the brain identifies patterns, similarities, and overarching principles across different experiences. Through this interconnected web of associations, the initial clustered memory representations become part of a broader network, with each memory influencing and being influenced by others. Together, these processes ensure that the consolidated memories are not isolated but interconnected and applicable to a range of situations, contributing to the flexibility of memories to evolve adaptively.

The concept that memories, once consolidated, are considered immutable has been challenged by research at the cellular level. This research demonstrates that the reactivation of seemingly consolidated memories can render them labile again, making them susceptible

to modification for a brief period. (Dudai & Eisenberg, 2004; Hardt et al., 2010; Nader, 2003; Nader et al., 2000; Sara, 2000). However, memory reconsolidation has boundary conditions (Fernández et al., 2016). For example, the notion that memories can become reconsolidated is time-dependent as reconsolidation dynamics also change with memory age; young memories are susceptible to post-reactivation disruption, while older ones are more resistant (Alberini, 2011; Dudai & Eisenberg, 2004; Frankland & Bontempi, 2005; Milekic & Alberini, 2002; Suzuki et al., 2004). While the mechanistic underpinnings of this issue remain unresolved, a prevailing view suggests the existence of an optimal memory malleability window where the reconsolidation process is triggered. According to this perspective, reconsolidation occurs if new learning, or a specific reminder, re-engages the hippocampal networks that were active during original learning (Debiec et al., 2002; Hupbach et al., 2007; Winocur et al., 2009). On the other hand, memories that have already shifted to a generalized network form would be minimally affected by reactivation (Dudai, 1996; McKenzie & Eichenbaum, 2011). In sum, while much is to be discovered, the fundamental conclusion is that new information consistently integrates, perpetually reshaping memory networks. Nevertheless, deciphering the complexities of this process is still challenging due to a lack of comprehensive understanding of precise mechanisms that support memory transformation at molecular, cellular, and systems levels.

While we wait for further empirical research to advance on the understanding of memory transformation, we here aim to take a step forward from a theoretical perspective. Memories may be represented in sparse, distributed cell ensembles (Roy et al., 2022; Ryan et al., 2021a; Vetere et al., 2017). The connectivity pattern between these engram cells encodes the identity of stored information, and it has been demonstrated that updating the content of memory can be described basically by the synaptic rewiring (Ortega-de San Luis et al., 2023). In this context, Network Science provides a straightforward approach to studying memory evolution as it describes the behavior of complex systems based on the network's wiring diagram (Barabási, 2013). Our framework builds upon the idea that neuronal ensembles can be represented as networks, consisting of neurons (nodes) linked by pairwise connections (edges). Specifically, modularity (or community structure) is a fundamental organizing principle within a network (Girvan & Newman, 2002). These cohesive units preserve specific sets of information, and their interconnected nature interacts with information spread throughout the network, fostering effective communication and exchange within the system (Centola, 2010; Danon et al., 2008; Newman, 2006; Weng et al., 2013). Focusing on a central assumption of memory evolution: we assume that memories are initially stored as segregated forms and over time they tend to an integrated form. Our goal was to develop a simplified model of memory transformation that captures essential properties outlined in existing models of memory consolidation and reconsolidation. However, unlike models that utilize a dual-representational approach involving the hippocampus and neocortex (McClelland et al., 1995; Sun et al., 2023), we opted for a one-layer neural network. The rationale behind this choice was to challenge the notion that memory transformation could be exclusively explained by the reorganization of neural ensembles using a common mechanistic substrate based on principles of information spread (Nematzadeh et al., 2014; Rodriguez et al., 2019).

The central tenet of our proposal is that memory representations tend to shift from highly segregated (highly modular) to richly integrated network forms and that this process is guided by a combination of neural network reactivations, spread of activation, and plasticity rules. A crucial insight from our modeling approach is the finding of an optimal window during this transformation whereby memories are most susceptible to malleability. This implies that the process of memory evolution follows a nonlinear or inverted U-shaped function, thereby highlighting that memories are ever-changing but susceptible to being optimally malleable at specific stages of the transformation process. The resulting model not only unifies various experimental phenomena but also integrates memory consolidation and reconsolidation accounts, offering a perspective by which simple architectural neural network property rules can summarize memory evolution.

Results

Our goal was to study the dynamics of the transformative process of memories over time. We conceptualized memories for events as neuronal ensembles that can be represented as complex networks across several circuits in the brain that are susceptible to being reactivated spontaneously or cued by elements of the experience that partially overlap with the original event. We hypothesize the repeated reactivation of the neuronal ensemble would promote changes to the stored neuronal ensemble.

To systematically investigate this theoretical model we built an ensemble of networks with four communities with an equal number of nodes and connecting links within communities (**Figure 1**). In our model, each node could achieve one of two possible values: s = {0,1}, where 1 represented the "active" mode and 0 was the "inactive" one. Unweighted undirected edges illustrated the structural pattern of connections between nodes (1 for nodes that were co-active and 0 for nodes that were not co-active). We studied the dynamics of memory transformation by modeling network changes subserved by creating and removing edges following a Hebbian plasticity rule, as connectivity patterns between engram cells have been implicated as a substrate of memory (Ryan et al., 2015, 2021a; Tonegawa et al., 2015). As plasticity extended to neurons not directly activated (McKenzie et al., 2021; Mugnaini et al., 2023), our model incorporated a deterministic spreading rule throughout the network, enabling the influence of activation beyond the initially active neurons.

We conceptualized our model's results to depict various stages in the evolution of memory. We first determine the initial state of an ensemble of nodes (*"Radiant stage": Early memory forms*). Subsequently, we simulated the network-level repercussions following the reactivation of a single memory (*"Celestial stage": memories evolve*). Finally, we delved into the consequences of recurrent reactivations on the continuous development and network reconfiguration of memories over time (*"Enlightened Stage": the final state*).

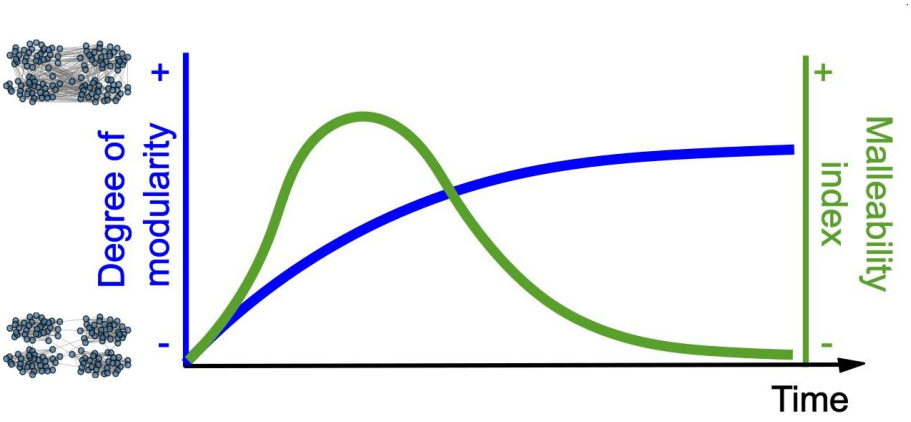

Figure 1 Foundations under the Segregation-to-Integration Transformation (SIT) Model, Memory networks shift from a segregated into an integrated form over time. The model postulates the existence of an optimal memory malleability window.

Memory stages

"Radiant stage": Early memory forms

We aimed to develop a low-dimensional model that captured the fundamental properties described in extant models of memory consolidation and reconsolidation. In these models, memories are initially encoded in a segregated form, described by neuron ensembles with highly synchronized activity. This co-firing pattern of activity describes jointly active neurons organized into motifs, or groups of neurons of high modular activity within a larger network, that underpin spatially selective assemblies representing memories for experienced events (O'Neill et al., 2008). In our model, we defined the first state of memory as a network model with four communities, each of them with the same number of nodes, tuning the degree of modularity. This segregated form means that nodes can be split into internally dense and externally sparse community subnetworks. Highly modular structures serve as a buffer for perturbations, as disturbances are contained within the community where they originated rather than spreading throughout the entire network (Nematzadeh et al., 2014). Additionally, a sparsely segregated network expands the range of possible configurations, thereby enhancing storage capacity (Brunel, 2016). Consequently, a segregated network organization represents an optimized configuration for early memories, effectively balancing the need to shield the network from disruptions or interference while accommodating a substantial amount of information storage.

"Celestial stage": memories evolve

Though highly segregated memory forms are an optimal design, as they provide a pattern-separated code from other memories (Teyler & DiScenna, 1986) and recurrent dynamics can easily be implemented to drive pattern completion (Hopfield, 1982; Rolls, 2013), these memories are very "expensive" in terms of energy consumption (Legenstein,

2018) and lacked the flexibility to ensure memory update and generalization. Indeed, rodent studies showed that while a stable synchronous activity of initially segregated neuronal ensembles is thought to support memory persistence over time (Gonzalez et al., 2019), the mnemonic operation of updating the representation of an encoded event via its reactivation causes the topological reorganization in the co-activity structure of the network (Gava et al., 2021). Therefore, we aimed to study the influence of memory reactivation on the topographical arrangement of the initial memory form. Since, in our model, memories are represented as interconnected nodes, we conducted a series of simulations to investigate alterations in connectivity patterns triggered by the reactivation of the network. To computationally simulate this phenomenon, we utilized a reactivation strategy that entailed activating a random set of nodes and studied the ensuing propagation of this activity in the network. We described this propagated activity with a deterministic rule, under the assumption that the reactivation of a target set of features of a memory might, also, reach more information than the one that is directly represented in the original neural ensemble (Anderson, 1983). We modeled this cascade dynamics in three steps: Turn-on/Activation, Spreading, and Plasticity defined as follows:

i. Turn-on

We represented a memory by a collection of N nodes linked by L undirected edges. A behavioral experience that partially overlaps with the original event activates a collection of nodes of the network, transitioning its mode from an inactive to an active state (from $0 \rightarrow 1$). We referred to "*Intensity*" *(Int)* for how many nodes may become active. We explored the reactivation process covering a rate of nodes turned active between 0.1 and 0.6. Under the idea that the degree of overlap between a current experience and a previously learned event can influence the activation patterns in the network, we assumed that for a higher degree of overlap, it is reasonable to expect a stronger reactivation response, resulting in a higher *Intensity*.

ii. Spreading

We adopted a linear threshold model, commonly applied in studies of information diffusion in social networks (Nematzadeh et al., 2014), to account for activity propagation in the network. As explained before, the condition of node i at time t was expressed through a binary variable $s_i(t) = \{0, 1\}$, with 1 denoting the "active" state and 0 indicating the "inactive" one. At time $t = 0$, a proportion of randomly chosen nodes, referred to as "seeds," are set in the active state. In the subsequent step, the state of each node is simultaneously modified based on the following threshold rule (**Equation 1**):

$$S_n(t = i + 1) = \begin{cases} 1 & if \quad \theta k_n < \sum_{j \in N_m} S_m(t = i) \\ 0 & if \quad \text{otherwise} \end{cases} \qquad (1)$$

where θ was the threshold parameter, k_n was the node degree of node n and $j \in N_m$ represented the set of neighbors of node n. The spreading persisted until the network reached

a stable state ($S_n(t=i+1) = S_n(t=i),\ \forall n$). In our modeling, we set different threshold levels of propagation (θ) in line with the notion that the degree to which a neuron becomes part of a memory engram depends on the intrinsic excitability at encoding (i.e., memory allocation) (Josselyn & Frankland, 2018).

iii. Plasticity

Once the network reached a stable state, a Hebbian plasticity rule was implemented. This step reconfigured the connectivity matrix of the memory network by creating an edge between two nodes $L_{n,m}$ if both connected nodes are in an active mode and removing it if both two connected nodes are in an inactive mode (**Equation 2**):

$$L_{n,m}(t=i+1) = \begin{cases} 1 & if\ \ S_n(t=i) = S_m(t=i) = 1 \\ 0 & if\ \ S_n(t=i) = S_m(t=i) = 0 \\ L_{n,m} & \text{otherwise} \end{cases} \quad (2)$$

To quantify the degree of changes induced in a network (plasticity), we defined network malleability index (ΔL) as the number of created and removed edges relative to the total number of edges of the network.

Degree of modularity (Z)

We hypothesized that memories would tend to shift from highly segregated to richly integrated state forms. To formalize the state of the network, we defined Z to quantified the degree of modularity following **Equation 3** :

$$Z = \frac{\sum_{i,j \notin \mathcal{C}} 1/2 L_{i,j}^{ext}}{L} \quad (3)$$

where L is the total number of edges and $L_{i,j}^{ext}$ are edges that linked node i, a member of the initial community $\mathcal{C}$, with node j that belongs to another community. In this context, a network highly segregated has a $Z \sim 0$. As the number of edges between communities increases, the degree of modularity r also increases, eventually reaching a value of $Z \sim 1$ (**Figure 2**).

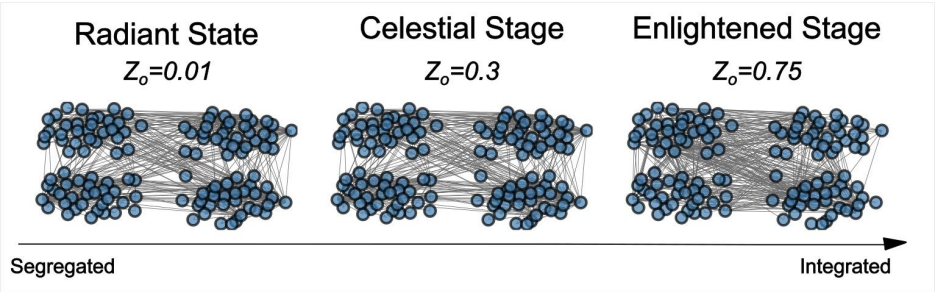

Figure 2. Examples of three networks with different degrees of modularity (Z). All networks have the same number of nodes (N = 128) and they were built by joining four similar communities of 32 nodes each.

In our model, reactivation is primarily influenced by two factors: *Intensity* (Int) and the threshold of propagation (θ). However, the degree of modularity within the network also plays a crucial role in determining the outcomes of reactivation. To offer a comprehensive understanding of the impact of these factors, we described the results of a single reactivation in terms of variations in the degree of modularity (Z) and the malleability index (ΔL).The results of these analyses are summarized in **Figure 3**.

First, we explored shifts in the degree of modularity (ΔZ) by considering various degrees of modularity within previous networks(Z_o) and for reactivations with different Intensities (Int) (**Figure 3A**). We found that ΔZ depended differentially on the modularity degree of the early state of the network. Specifically, highly initially segregated networks (lower Z_o) presented higher shifts in their modularity than initially integrated networks (higher Z_o). Interestingly, all values of ΔZ were higher than 0, indicating that a single reactivation transformed the original network state into a more integrated form after a reactivation. This point holds significant importance because, although the SIT model suggested a unidirectional shift, it was not inherently imposed in the reactivation rule itself; rather, it emerged because of the spreading of activation and the application of the Hebbian plasticity rule.

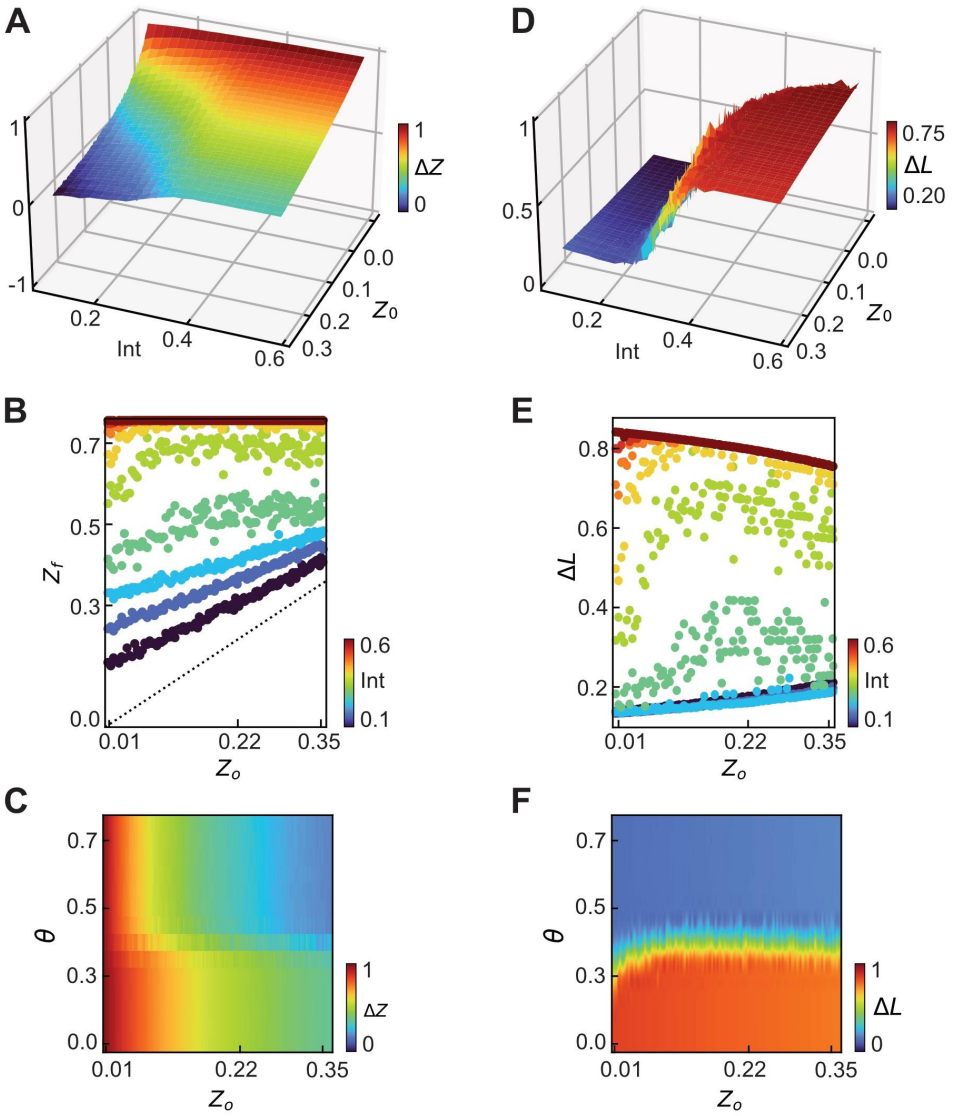

Figure 3. Description of reactivation concerning *Intensity*, the degree of modularity, and the threshold parameter. A. Surface of ΔZ (shifts of Z) as a function of both the *intensity* and degree of segregation of the former network ($\theta = 0.4$). **B.** Transition from $Z_o \rightarrow Z_f$ after a reactivation with $\theta = 0.4$. Dots depict the relation between the degree of segregation pre (Z_o) and post (Z_f) reactivation. Cold colors represent reactivations with low intensities while warm colors represent higher intensities. **C.** Heatmap of ΔZ as a function of the degree of modularity of the former network and threshold of propagation. Here, *intensity* is fixed ($Int = 0.3$). **D.** Surface of the malleability index (ΔL) as a function of *intensity* and degree of modularity of the former network ($\theta = 0.4$). **E.** Relation between malleability index and the degree of modularity of the former network, $\theta = 0.4$. Colors represent the reactivation *intensity*. **F.** Heatmap of the malleability index as a function of the degree of modularity of the former network and excitability threshold. *Intensity* is fixed ($Int = 0.3$). The reported data is an average of multiple experimental runs (25) where the activation of nodes (turn-on) was randomized in each run.

We also found that shifts in modularity were affected by the degree of Int of the reactivation. **Figure 3B** depicts the transition of the modularity degree after the reactivation (

Z_f) as a function of the initial on (Z_o), colored by Int. The increase in Int revealed staggered jumps in the transition of Z until a limit value was reached (low *intensity*: dark blue dots, higher intensities: red dots). Notably, the resulting network was less segregated compared to its predecessor, as indicated by data points located above the identity (black line). In other words, a single reactivation led to a shift in memory representations towards a more integrated form, independently of the initial community configuration. However, when the initial forms of memory were highly integrated, the changes in its structure were more subtle. For more segregated configurations, the size of the transition depends on the *intensity* (more jump at higher *intensity*). The same general trend of transitioning from a less segregated to an integrated form was found when simulating reactivations with different degrees of activity threshold (θ) (**Figure 3C**). Surprisingly, θ did not yield an effect on ΔZ for a fixed Int. Instead, we observed that the initial configuration has a stronger impact on changes in the memory configuration than the intrinsic excitability of neurons.

Next, we computed the malleability index (ΔL: number of connections that were created or removed after a reactivation) through adjustments to degree of modularity (Z_o) of the initial memories and the range of activation intensities ($Int.$). An outline of these results can be seen in **Figure 3D**. This analysis showed that the malleability surface exhibited a pronounced dependence on *intensity*, adhering to a sigmoid profile. Lower Int of the reactivated memory led to low malleability of the network, whereas high Int resulted in an increased degree of malleability. Essentially, reactivations with low activation Int induced subtle changes in the number of edges, suggesting a nuanced modification in the connectivity pattern. Conversely, reactivations with high Int prompted the creation and removal of a large number of edges in the initial memory network, essentially rewiring the entire network. Intriguingly, while the network's malleability raised with Int, we observed a nonlinear response - a distinctive inverted-U shape - in malleability. This phenomenon was particularly notable when the original network occupied an intermediate modular configuration, situated between highly segregated and highly integrated forms, and when approached $Z \sim 0.2$ (**Figure 3E**). This effect was particularly noticeable for intermediate values of Int, where malleability showed the greatest sudden increase. Finally, we inspected the influence of the threshold of propagation (θ) on malleability (**Figure 3F**). From this analysis, it became apparent that the threshold delineated the phase space into two halves, with more stringent thresholds constraining the spread of activation and consequently diminishing memory malleability.

In summary, we presented a deterministic formalism to elucidate the influence of a single reactivation on the existing memory network. These mathematical rules unveiled that initial memory stages tend to transition towards a more integrated configuration following a single reactivation. Our model posited that this shift is more pronounced with higher reactivation intensities. Conversely, the degree of malleability is contingent on the interplay between *intensity* and the modularity degree of the original memory. The spreading parameter facilitates alterations in the community configuration but imposes constraints on the creation or removal of edges.

"Enlightened Stage": the final state

After describing the results of a single reactivation, we addressed the effects of repeated reactivation interventions to account for the evolution of memory over time. Our goal was to test the hypothesis that networks undergo a gradual transition from a segregated to an integrated network configuration. Additionally, we aimed to explore whether there exists a specific window during this transition where memory is most susceptible to malleabilityTherefore, as in our previous analyses, we quantified the changes in its degree of modularity (Z) and malleability (ΔL), both extracted from the connectivity pattern of the network, as a function of different activation thresholds (θ), intensities (Int) and the modular configuration of the initial network (Z_o).

We first focused on assessing how repeated reactivations elicited changes in the degree of modularity (Z) in the network. More specifically, we simulated 5 reactivations with varying thresholds of propagation (θ) and *Intensity* (Int) and studied the evolution of Z for a memory network that originally is very segregated ($Z = 0.01$) and gradually shifted to an integrated form ($Z = 0.76$) (**Figure 4A**). The results of these simulations revealed that the network transformation occurred at different speeds for various parameter values. However, in all cases, the networks ultimately achieved a fully integrated form. The Int parameter was shown to induce a uniform effect on the network when the θ was set to its lowest value, leading to an immediate shift towards an integrated form. Conversely, for higher θ, the *intensity* parameter elicited a differential effect. The transformation took longer to occur for reactivations with low Int. This result aligned with our previous findings, which showed that higher Int resulted in larger shifts of Z (i.e., **Figure 3B**). **Figure 4B** and **4C** illustrate some examples of the dynamics of Z for a subset of selected Int and θ, respectively. Our results showed that as the Int increases, Z reached a ceiling state more rapidly, accomplishing this state in a single reactivation at the highest Int. A similar scenario was found with θ, which revealed that Z tended to reach a fully integrated form at different speeds with repeated reactivations as θ increased. However, compared to Int, reaching a fully integrated form took longer as θ grew, indicating that θ was more restrictive in eliciting a transformation shift memory networks from a segregated to an integrated form.

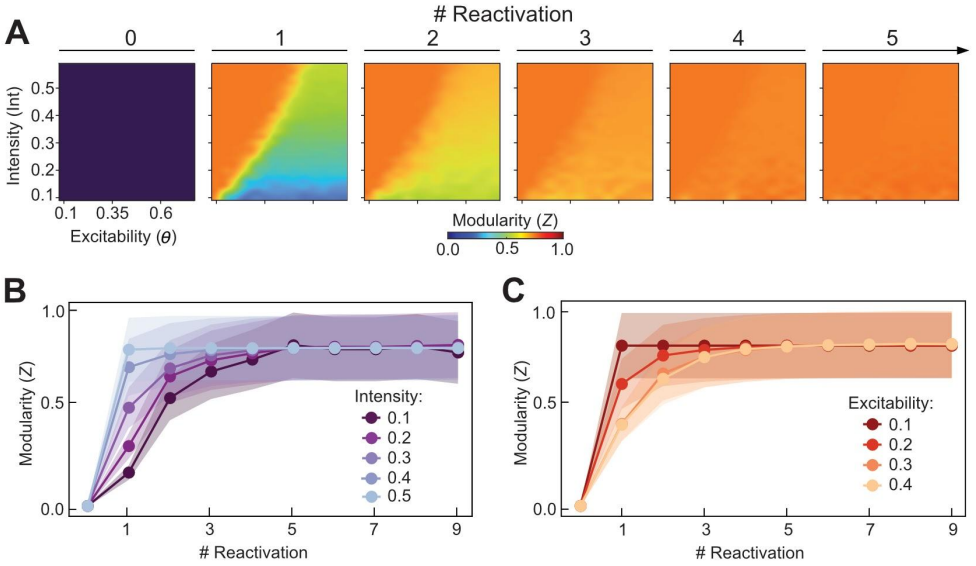

Figure 4. Changes in the degree of modularity of the network for repeated reactivations. A. Transformation of Z over 5 repeated reactivations. Each panel is a heatmap of Z as a function of *Intensity* and the threshold of propagation. The initial state (0) has $Z_o = 0.01$. After 5 reactivations, the network is completely integrated with $Z = 0.76$. **B.** Curves of Z for different intensities and fixed threshold of propagation ($\theta = 0.4$, mean +- SE). **C.** Z transformation for different thresholds propagation and fixed *intensity* ($Int = 0.3$, mean +- SE). The data reported is the outcome of 25 runs with the same initial state with randomized activated nodes.The initial network has 128 nodes, 4 similar communities of 32 nodes each and $Z_o = 0.01$.

Finally, we aimed to investigate if the degree of malleability of memories changed for repeated reactivations. To address this issue, we quantified the malleability of an initial segregated network form ($Z = 0.01$) as a function of repeated reactivations with varying degrees of Int and θ. This simulation revealed that, for repeated reactivations, the peak of malleability was relocated within the parameter space (**Figure 5A**). Concretely, under low θ, the malleability peak was observed during the initial reactivation, with subsequent reactivations exhibiting minimal changes in their structural configuration. Conversely, more restrictive θ led to almost null malleability throughout reactivations, indicating marginal alterations in the number of edges of the network. Interestingly, intermediate levels of θ elicited clear malleability peaks were observed at the variable reactivation stage, depending on the Int of the reactivation. To exemplify the behavior of malleability, we plotted the dynamics of this measure as a function of Int for a selected θ value ($\theta = 0.4$; **Figure 5B**). The curves displayed in **Figure 5B** revealed important distinctions, particularly in the first reactivation, where the curve corresponding to the highest Int exhibited a pronounced maximum malleability. This maximum gradually diminished with increasing Int, in line with the observations in **Figure 3E**. The other curves also indicated peaks of malleability for early reactivations, albeit with lower amplitudes. These findings substantiate previous results when investigating the consequences of a single reactivation, underscoring that memory malleability conforms to an inverted-U shape function, whereby specific windows of

opportunity may be optimal for instigating changes in the structure of a previously encoded memory through reactivation.

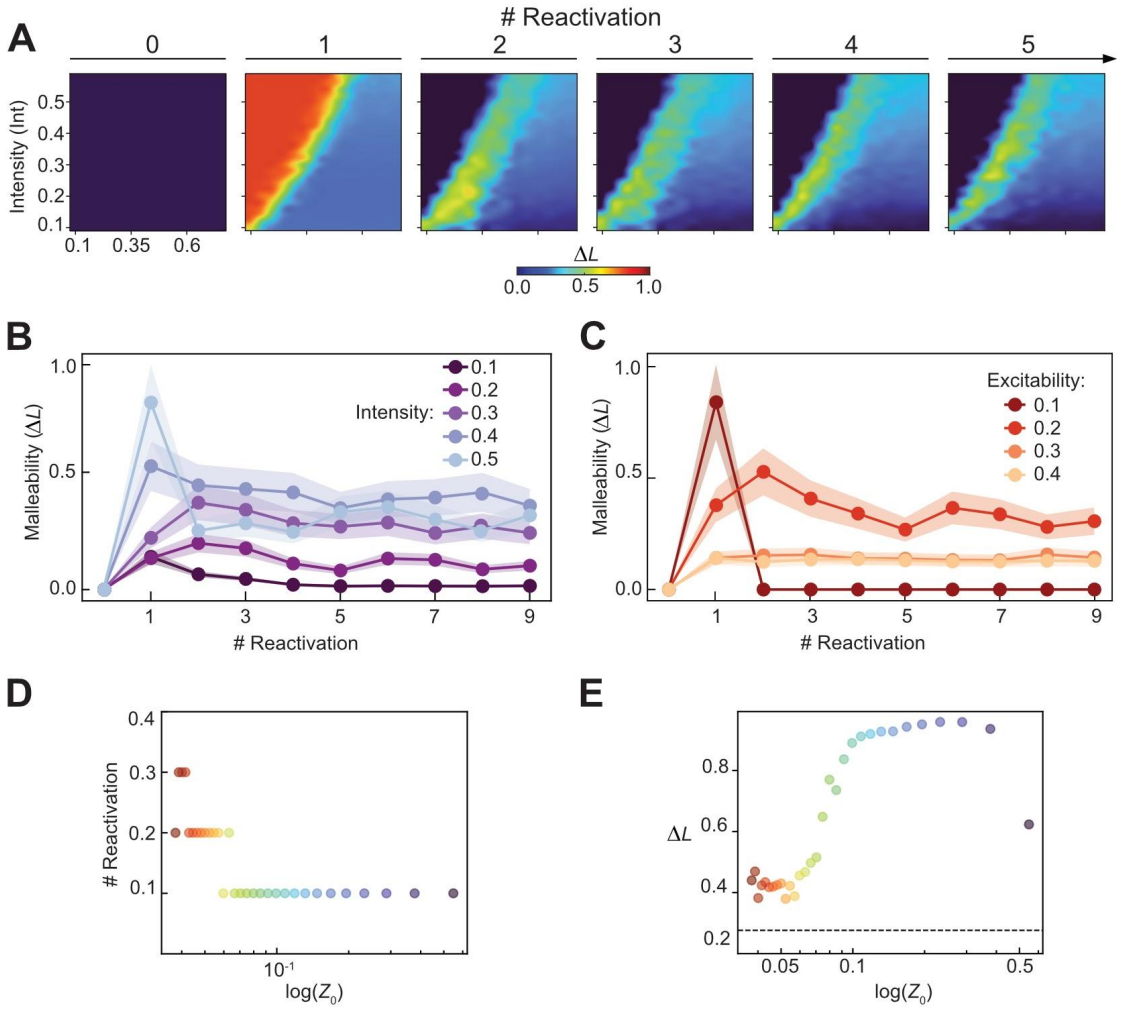

Figure 5. Changes in the Malleability index over repeated reactivations. A. Transformation of ΔL over 5 repeated reactivations. Each panel is a heatmap ΔLas a function of both *intensity* and the threshold of propagation. The initial state has $Z_o = 0.01$. **B.** Curves of $\Delta L$for different intensities and fixed threshold of propagation ($\theta = 0.4$, mean +- SE). **C.** ΔL transformation for different thresholds of propagation and fixed *intensity* ($Int = 0.3$, mean +- SE). **E**. Position (in terms of the number of reactivations) of the maximum of ΔL as a function of the degree of modularity of the initial network configuration ($\theta = 0.4, Int = 0.3$). **F.** Amplitude of the maximum of ΔL as a function of the degree of modularity of the initial network ($\theta = 0.4$, $Int = 0.3$). The data reported is the outcome of 25 runs that have the same initial configuration with randomized activated nodes. The initial network has 128 nodes, 4 similar communities of 32 nodes each.

Our previous analysis when studying the effects of a single reactivation in the memory network revealed that memory malleability was more sensitive to θ than Int (e.g., **Figure 3B**). To further study this issue in the context of repeated reactivations, we quantified

the degree of malleability throughout repeated reactivations with varying degrees of θ. The results of these simulations are displayed in **Figure 5C**. We found that under more permissive θ (i.e., $\theta < 0.3$), a peak of memory malleability was evident in the initial reactivations, gradually diminishing subsequently. Interestingly, following the attainment of the maximum, malleability persisted, as ΔL remained above 0, but at a much lower level, indicating that memories were permeable to structural changes over the entire course of repeated reactivations. In contrast, under more restrictive θ conditions (i.e., $\theta > 0.2$), malleability remained very low throughout all reactivations, indicating that, under these conditions, memories are minimally susceptible to changing their structure over time.

In summary, we found that malleability exhibited a nonlinearity in the context of memory transformation, characterized by an inverted U-shape pattern. This implies that memories within intermediate modular configurations manifest the highest level of malleability (with *intensity* and the excitability threshold fixed, $\theta = 0.4, Int = 0.3$). To determine the location and magnitude of the malleability peak, we analyzed the position and amplitude of the nonlinearity for initial networks with varying degrees of segregation and *Intensity* and excitability threshold, as illustrated in **Figures 5D** and **5E,** respectively. Networks with greater segregation peaked later and with a smaller magnitude, while the magnitude of the peak demonstrates a linear increase as the initial network loses its compartmentalization.

Discussion

The theoretical model presented here provides a normative and quantitative framework for assessing the conditions under which memories evolve. The central premise of this model is that, over time, memories shift from highly segregated (i.e., highly modular) to integrated (i.e., less modular) network forms, guided by neural network reactivations, activation spread, and plasticity rules. Our modeling identifies an optimal window during this transformation, revealing a nonlinear or inverted U-shaped function for memory evolution. Memories are most malleable at early stages and gradually become less susceptible to changes over time, unifying experimental phenomena and integrating consolidation and reconsolidation accounts. This model simplifies memory evolution, summarizing it through architectural neural network property rules, and emphasizing the dynamic and optimal malleability of memories throughout the transformation process.

Our model proposes that the transformation of memory relies on modifications in the connectivity patterns among nodes within the memory network. This perspective aligns with the idea that learning involves modifications to the wiring diagram of a neural ensemble, where previously unconnected units establish connections, and vice versa. However, while alterations in the wiring engram offer a potential substrate for encoding more extensive information in sparse coding models (Chklovskii et al., 2004; Knoblauch et al., 2010), it also

underscores that other forms of plasticity based on changes in synaptic weights, the strength of connections between cells, are crucial for comprehending memory evolution, as they necessitate less complex biological machinery and entails a more rapid learning process (Bonhoeffer & Yuste, 2002; Chklovskii et al., 2004). While both mechanisms are likely involved in engram cell formation and function (Poo et al., 2016), the investigation into network properties in hippocampal code representation has shown that the wiring diagrams effectively encode specific experiences (Ortega-de San Luis et al., 2023; Ryan et al., 2021b) and shifts in these connectivity patterns provide a more robust explanation for memory transformation than alterations in the individual firing properties of neurons (Gava et al., 2021). In subsequent work, the simplicity of the model will enable us to readily incorporate new variables that account for changes in synaptic weights, encompassing both modifications in the wiring diagram and the strength of the connections.

Perhaps the most notorious finding in the SIT model is that a simple property such as the degree of modularity in a neural network reconciles the theoretical findings within the memory consolidation and reconsolidation literature. These models commonly argue that memories are transformed over time but while consolidation views offer a one-shot direction of the effects, the reconsolidation view highlights that memories could undergo continuous changes over time (Nadel et al., 2012). The results of our modeling approach offer a reconciling framework by showing that while memories can evolve perpetually there may exist an optimal window of malleability during this course. These optimal memory malleability windows appeared at early stages since memory formation, in line with the notion that reconsolidation is a time-dependent phenomenon, as young memories are susceptible to disruption, while older ones would be more resistant to changes (Alberini, 2011; Dudai & Eisenberg, 2004; Frankland & Bontempi, 2005; Milekic & Alberini, 2002; Suzuki et al., 2004). Moreover, our findings align better with the idea that once this malleability opportunity window has passed. Memories may still change, albeit at a minimal level.

We conceptualized that memory changes were driven by reactivating structural properties of a memory network and that the effects of this reactivation would be modulated by state-dependent parameters of the existing network such as the threshold of propagation and *Intensity*. These two parameters were included in our model to accommodate findings from rodent studies that showed, on the one hand, that the recruitment of cells to code memory for a specific event, or engram, relied on the intrinsic excitability state of the cell, a phenomenon termed memory allocation (Josselyn & Frankland, 2018). The basic premise of these findings is that eligible neurons compete for allocation to a given engram, with more excitable neurons winning this competition. More importantly, though, in the context of our modeling work, several lines of evidence in both experimental animals and humans suggest that one of the key functions of allocation is to direct engrams underlying different experiences related by contextual variables to become linked and unrelated memories to become disambiguated (Eichenbaum, 2000; Schlichting & Frankland, 2017). The emerging principle from these findings is that fluctuations in neuronal excitability determine how engrams interact, promoting either memory integration (via coallocation to overlapping

engrams) or separation (via disallocation to non-overlapping engrams) (Josselyn & Frankland, 2018). When applying this parameter to our modeling approach, we found that, indeed, the degree of intrinsic excitability of the network state was important in determining the rate at which memories would shift to an integrated form. More permissive excitability thresholds, allowing more nodes to be "active", promote memory malleability whereas more stringent excitability thresholds (i.e., a lesser number of nodes to be "active") restrict the spreading of activation in the network, and decreasing memory malleability. However, intermediate levels of excitability thresholds induced distinct malleability peaks at variable reactivation stages, dependent on reactivation *intensity*. These memory configurations necessitated a greater number of "active" nodes (i.e., higher *Intensity*) during early reactivations to promote structural changes in the memory (**Figure 5A**). While the relevance of cell excitability in engram formation and modification has gathered substantial evidence, the reactivation *Intensity* and its functional role are still not well understood. Future studies are needed to unravel the existent relation between the behavioral experience and the *Intensity* of reactivation.

Our memory model aligns with the idea that a modular architecture, allowing independent adjustments within memory networks, is beneficial for maintaining enriched detailed memories. However, our framework proposes that the breakdown of modularity occurs as memories evolve and that this occurs by repeated reactivation, perhaps cued by overlapping novel experiences. We propose that the shift of memories from segregated to integrated forms responds to the memory system's adaptation to highly dynamic environments, where novel and past experiences may only partially overlap, necessitating a balanced interplay between stored information and accessibility. In such a context, highly segregated memories may function suboptimally due to limited accessibility (Hintze & Adami, 2008). On the contrary, highly integrated memory network forms prioritize accessibility over storage quality, offering an adaptive structural configuration in ever-changing contexts. Network science has demonstrated that at the intersection of these two extreme network configurations, there exists an optimal modularity configuration that effectively balances memory storage and information diffusion (Rodriguez et al., 2019). Similarly, we propose that such an optimal modularity structural state of a memory form is ideal for inducing changes, as it provides a well-balanced trade-off between the probability of being reactivated and the diffusion of activation throughout the network.

While SIT provides a theoretical solution for many defining characteristics in memory consolidation and reconsolidation literature, it remains indifferent to several pertinent aspects. For example, an open question concerns how network configuration and its transformation are distributed in the brain. Our model posited that these transformations may occur throughout various brain regions, yet existing models suggest that changes in this network may coincide with localized alterations in specific brain regions (Squire, 1992). Others, however, emphasized the distributed nature of memory formation and transformation (Tonegawa et al., 2018). While we await a better characterization of this issue at the cellular and structural level, our model offers the simplest proposed principle based on the topological structure of neural networks by which memories transform over time. We anticipate the

current model would be of interest to psychologists and neurobiologists to find motivation in testing and challenging the SIT model.

Methods

Simulations were done using homemade Python codes mainly using Networkx (https://networkx.org/, a package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks).

Building networks

To analyze the Segregated-to-Integrated Model in detail, we studied the evolution of a set of unweighted and undirected networks. We only studied the simplest case and analyzed simulations where the initial condition was a network of 128 equal nodes and four equal communities of 32 nodes. We adopted this simplistic approach to thoroughly explore the roles of different parameters. The degree of the initial modularity of these networks was tuned by varying the mean node degree within the community (k_{int}: from 1 to 31) and the number of edges that connected different communities (L^{ext}: from 1 to 128). To quantify the degree of modularity, we defined the parameter Z (**Equation 3**). We used the block model approach to build each community (Holland et al., 1983). We then linked the communities with a number of edges (L^{ext}) that were assigned randomly to pairs of nodes in different communities. This arrangement ensured that every node and community was not isolated and that they played a similar role in the network.

Reactivation

To shift the nodes from inactive to active mode (turn-on step), we introduced the reactivation intensity ($Int.$) as a parameter ranging between 0 and 1. The number of activated nodes per community was determined by sampling from a normal probability distribution with a mean of $Int.$ and standard deviation of 0.05. The value of $Int.$ ranged from 0.1 and 0.6, with increments of 0.02. To determine which nodes became active, we utilized a uniform random number generator that selected values between 1 and 32.

The primary parameter of propagation is θ. Throughout the simulations θ varied between 0.1 and 0.8, with increments of 0.05. Spreading was modeled using a deterministic linear threshold rule (**Equation 1**), which implies that once a node becomes active, it will remain forever, and the propagation will continue until the system reaches a steady state or exceeds 50 iterations (Nematzadeh et al., 2014).

Finally, we applied the plasticity rule (**Equation 2**), where edges were rewired. If two nodes achieve an active mode, an edge is created between them. If two nodes remained inactive, the edges between them were removed. Note that the Plasticity rule can be expanded

to a weighted network by adjusting the edge weights with a fixed positive (negative) parameter between two active (inactive) nodes. This adjustment indicates an increase (decrease) in strength.

Memory evolution

We simulated the evolution of memory networks by applying 10 consecutive re-activations. The intensity ($Int.$) remained constant throughout each re-activation. The nodes activated in each iteration were randomly assigned. We repeated the entire network evolution with the same set of parameters 25 times to characterize the main transformation behavior, independently of the random assignment of nodes.

To analyze the outcome of only one reactivation (**Figure 3**), we focused on the first iteration of the sequence of 10 repeated reactivations.

Quantification of memory transformation

We studied the changes in reactivation printed in the connectivity patterns of the networks. First, we analyzed the transformations in the modular structure of the network. We inspected the dynamics of Z over a set of repeated reactivations. Also, we figured out shifts of Z as a consequence of one reactivation (**Equation 4**, $Z_o \rightarrow Z_i$),

$$\Delta Z = \frac{Z_i - Z_o}{Z_i + Z_o} \quad (4)$$

ΔZ ranging between -1 and 1; 1 meaning transformation from network to segregated to integrated and -1 the other way around.

Finally, we studied changes in the number of edges. malleability index (ΔL) was defined as the number of created and removed edges relative to the total number of edges before reactivation. This is a positive parameter, 0 value means no change in the number of edges.

Author contributions: L.B. and L.F. conceptualized the model. L.B. ran the simulations. L.B. and L.F. wrote the manuscript.

Acknowledgments: This work was supported by the Spanish Ministerio de Ciencia, Innovación y Universidades, which is part of Agencia Estatal de Investigación (AEI), through the project PID2019-111199GB-I00 (Co-funded by European Regional Development Fund. ERDF, a way to build Europe), to L.F. We thank CERCA Programme/Generalitat de

Catalunya for institutional support. This work was supported by Agencia Nacional de Promoción Científica y Tecnológica PICT 2020 - 00956, to L.B.

Competing Interests: The authors declare no competing financial interests.

Code Availability: Code will be provided upon manuscript publication.

References

Alberini, C. M. (2011). The role of reconsolidation and the dynamic process of long-term memory formation and storage. *Frontiers in Behavioral Neuroscience*, *5*, 12.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning & Verbal Behavior*, *22*(3), 261–295. https://doi.org/10.1016/S0022-5371(83)90201-3

Barabási, A.-L. (2013). Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *371*(1987), 20120375.

Bonhoeffer, T., & Yuste, R. (2002). Spine motility: Phenomenology, mechanisms, and function. *Neuron*, *35*(6), 1019–1027.

Brunel, N. (2016). Is cortical connectivity optimized for storing information? *Nature Neuroscience*, *19*(5), Article 5. https://doi.org/10.1038/nn.4286

Centola, D. (2010). The spread of behavior in an online social network experiment. *Science*, *329*(5996), 1194–1197.

Chklovskii, D. B., Mel, B., & Svoboda, K. (2004). Cortical rewiring and information storage. *Nature*, *431*(7010), 782–788.

Danon, L., Arenas, A., & Diaz-Guilera, A. (2008). Impact of community structure on information transfer. *Physical Review E*, *77*(3), 036103.

Debiec, J., LeDoux, J. E., & Nader, K. (2002). Cellular and systems reconsolidation in the hippocampus. *Neuron*, *36*(3), 527–538.

Dudai, Y. (1996). Consolidation: Fragility on the road to the engram. *Neuron*, *17*(3), 367–370.

Dudai, Y., & Eisenberg, M. (2004). Rites of passage of the engram: Reconsolidation and the

lingering consolidation hypothesis. *Neuron*, *44*(1), 93–100.

Eichenbaum, H. (2000). Hippocampus: Mapping or memory? *Current Biology*, *10*(21), R785–R787.

Fernández, R. S., Boccia, M. M., & Pedreira, M. E. (2016). The fate of memory: Reconsolidation and the case of Prediction Error. *Neuroscience & Biobehavioral Reviews*, *68*, 423–441.

Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience*, *6*(2), 119–130. https://doi.org/10.1038/nrn1607

Gava, G. P., McHugh, S. B., Lefèvre, L., Lopes-dos-Santos, V., Trouche, S., El-Gaby, M., Schultz, S. R., & Dupret, D. (2021). Integrating new memories into the hippocampal network activity space. *Nature Neuroscience*, *24*(3), 326–330.

Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, *99*(12), 7821–7826.

Gonzalez, W. G., Zhang, H., Harutyunyan, A., & Lois, C. (2019). Persistence of neuronal representations through time and damage in the hippocampus. *Science*, *365*(6455), 821–825.

Hardt, O., Einarsson, E. Ö., & Nader, K. (2010). A bridge over troubled water: Reconsolidation as a link between cognitive and neuroscientific memory research traditions. *Annual Review of Psychology*, *61*, 141–167.

Hintze, A., & Adami, C. (2008). Evolution of complex modular biological networks. *PLoS Computational Biology*, *4*(2), e23.

Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, *5*(2), 109–137. https://doi.org/10.1016/0378-8733(83)90021-7

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, *79*(8), 2554–2558.

Hupbach, A., Gomez, R., Hardt, O., & Nadel, L. (2007). Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learning &*

Memory, *14*(1–2), 47–53.

Josselyn, S. A., & Frankland, P. W. (2018). Memory allocation: Mechanisms and function. *Annual Review of Neuroscience*, *41*, 389–413.

Josselyn, S. A., & Tonegawa, S. (2020). Memory engrams: Recalling the past and imagining the future. *Science*, *367*(6473), eaaw4325.

Knoblauch, A., Palm, G., & Sommer, F. T. (2010). Memory capacities for synaptic and structural plasticity. *Neural Computation*, *22*(2), 289–341.

Legenstein, R. A. (2018). Long term memory and the densest k-subgraph problem. *9th Innovations in Theoretical Computer Science Conference*.

McClelland, J. L., McNaughton, B. L., & 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.

McKenzie, S., & Eichenbaum, H. (2011). Consolidation and reconsolidation: Two lives of memories? *Neuron*, *71*(2), 224–233.

McKenzie, S., Huszár, R., English, D. F., Kim, K., Christensen, F., Yoon, E., & Buzsáki, G. (2021). Preexisting hippocampal network dynamics constrain optogenetically induced place fields. *Neuron*, *109*(6), 1040–1054.

Milekic, M. H., & Alberini, C. M. (2002). Temporally graded requirement for protein synthesis following memory reactivation. *Neuron*, *36*(3), 521–525.

Mugnaini, M., Trinchero, M. F., Schinder, A. F., Piatti, V. C., & Kropff, E. (2023). Unique potential of immature adult-born neurons for the remodeling of CA3 spatial maps. *Cell Reports*, *42*(9).

Nadel, L., Hupbach, A., Gomez, R., & Newman-Smith, K. (2012). Memory formation, consolidation and transformation. *Memory Formation*, *36*(7), 1640–1645. https://doi.org/10.1016/j.neubiorev.2012.03.001

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, *7*(2), 217–227.

Nader, K. (2003). Memory traces unbound. *Trends in Neurosciences*, *26*(2), 65–72.

Nader, K., Schafe, G. E., & Le Doux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, *406*(6797), 722–726.

Nematzadeh, A., Ferrara, E., Flammini, A., & Ahn, Y.-Y. (2014). Optimal network modularity for information diffusion. *Physical Review Letters*, *113*(8), 088701.

Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, *103*(23), 8577–8582.

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

Ortega-de San Luis, C., Pezzoli, M., Urrieta, E., & Ryan, T. J. (2023). Engram cell connectivity as a mechanism for information encoding and memory function. *Current Biology*.

Poo, M., Pignatelli, M., Ryan, T. J., Tonegawa, S., Bonhoeffer, T., Martin, K. C., Rudenko, A., Tsai, L.-H., Tsien, R. W., Fishell, G., & others. (2016). What is memory? The present state of the engram. *BMC Biology*, *14*, 1–18.

Rodriguez, N., Izquierdo, E., & Ahn, Y.-Y. (2019). Optimal modularity and memory capacity of neural reservoirs. *Network Neuroscience*, *3*(2), 551–566.

Rolls, E. T. (2013). The mechanisms for pattern completion and pattern separation in the hippocampus. *Frontiers in Systems Neuroscience*, *7*, 74.

Roy, D. S., Park, Y.-G., Kim, M. E., Zhang, Y., Ogawa, S. K., DiNapoli, N., Gu, X., Cho, J. H., Choi, H., Kamentsky, L., & others. (2022). Brain-wide mapping reveals that engrams for a single memory are distributed across multiple brain regions. *Nature Communications*, *13*(1), 1799.

Ryan, T. J., Ortega-de San Luis, C., Pezzoli, M., & Sen, S. (2021a). Engram cell connectivity: An evolving substrate for information storage. *Current Opinion in Neurobiology*, *67*, 215–225.

Ryan, T. J., Ortega-de San Luis, C., Pezzoli, M., & Sen, S. (2021b). Engram cell

connectivity: An evolving substrate for information storage. *Current Opinion in Neurobiology*, *67*, 215–225.

Ryan, T. J., Roy, D. S., Pignatelli, M., Arons, A., & Tonegawa, S. (2015). Engram cells retain memory under retrograde amnesia. *Science*, *348*(6238), 1007–1013.

Sara, S. J. (2000). Retrieval and reconsolidation: Toward a neurobiology of remembering. *Learning & Memory*, *7*(2), 73–84.

Schlichting, M. L., & Frankland, P. W. (2017). Memory allocation and integration in rodents and humans. *Current Opinion in Behavioral Sciences*, *17*, 90–98.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*(2), 195–231. https://doi.org/10.1037/0033-295X.99.2.195

Sun, W., Advani, M., Spruston, N., Saxe, A., & Fitzgerald, J. E. (2023). Organizing memories for generalization in complementary learning systems. *Nature Neuroscience*, *26*(8), 1438–1448.

Suzuki, A., Josselyn, S. A., Frankland, P. W., Masushige, S., Silva, A. J., & Kida, S. (2004). Memory reconsolidation and extinction have distinct temporal and biochemical signatures. *Journal of Neuroscience*, *24*(20), 4787–4795.

Teyler, T. J., & DiScenna, P. (1986). The hippocampal memory indexing theory. *Behavioral Neuroscience*, *100*(2), 147.

Tonegawa, S., Morrissey, M. D., & Kitamura, T. (2018). The role of engram cells in the systems consolidation of memory. *Nature Reviews Neuroscience*, *19*(8), 485–498. https://doi.org/10.1038/s41583-018-0031-2

Tonegawa, S., Pignatelli, M., Roy, D. S., & Ryan, T. J. (2015). Memory engram storage and retrieval. *Current Opinion in Neurobiology*, *35*, 101–109.

Vetere, G., Kenney, J. W., Tran, L. M., Xia, F., Steadman, P. E., Parkinson, J., Josselyn, S. A., & Frankland, P. W. (2017). Chemogenetic Interrogation of a Brain-wide Fear Memory Network in Mice. *Neuron*, *94*(2), 363-374.e4. https://doi.org/10.1016/j.neuron.2017.03.037

Weng, L., Menczer, F., & Ahn, Y.-Y. (2013). Virality prediction and community structure in social networks. *Scientific Reports*, *3*(1), 1–6.

Winocur, G., Frankland, P. W., Sekeres, M., Fogel, S., & Moscovitch, M. (2009). Changes in context-specificity during memory reconsolidation: Selective effects of hippocampal lesions. *Learning & Memory*, *16*(11), 722–729.

Winocur, G., Moscovitch, M., & Bontempi, B. (2010). Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal–neocortical interactions. *Neuropsychologia*, *48*(8), 2339–2356.