UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title
A Computational Model of Systems Memory Reconsolidation

Permalink
https://escholarship.org/uc/item/6685t52r

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 35(35)

ISSN
1069-7977

Authors
Helfer, Peter
Shultz, Thomas
Hardt, Oliver
et al.

Publication Date
2013

Peer reviewed

A Computational Model of Systems Memory Reconsolidation

Peter Helfer (peter.helfer@mail.mcgill.ca)
Department of Psychology and Integrated Program in Neuroscience, McGill University, 1205 Penfield Avenue
Montreal QC, Canada H3A 1B1

Thomas R. Shultz (thomas.shultz@mcgill.ca)
Department of Psychology and School of Computer Science, McGill University, 1205 Penfield Avenue
Montreal QC, Canada H3A 1B1

Oliver Hardt (oliver.hardt@me.com)
Department of Psychology, McGill University, 1205 Penfield Avenue
Montreal QC, Canada H3A 1B1

Karim Nader (karim.nader@mcgill.ca)
Department of Psychology, McGill University, 1205 Penfield Avenue
Montreal QC, Canada H3A 1B1

Abstract

Memory reconsolidation, the re-stabilization of consolidated memories after reactivation-induced destabilization, has received considerable attention in recent years. Nevertheless, the neural processes underlying the phenomenon remain elusive. With the aim of contributing to the development of a theory in this area, we here present a computational model of reconsolidation at the "systems" level. The model is an extension of TraceLink, which has previously been used to account for a range of memory phenomena related to consolidation.

Keywords: Memory reconsolidation, neural network, connectionism.

Introduction

The phenomenon of memory reconsolidation, the re-stabilization of consolidated memories after reactivation-induced destabilization, has received considerable attention in recent years with the publication of a series of studies on both animals and human subjects (Nader & Einarsson, 2010; Nader & Hardt, 2009). While several computer simulations have modeled consolidation after initial learning, (McClelland, McNaughton, & O'Reilly, 1995; Murre, 1996), only one model of cellular reconsolidation has been published (Osan, Tort, & Amaral, 2011), and – to our knowledge – no simulation of systems reconsolidation (Debiec, LeDoux, & Nader, 2002). In order to fill this gap, we developed an extended version of a previously published computational model of memory consolidation, TraceLink (Murre, 1996), incorporating features that enable it to also account for reconsolidation.

We begin with a brief introduction to the phenomenon of memory consolidation, followed by a description of the TraceLink model. We then discuss the mechanisms believed to underpin systems memory reconsolidation, describe how we implemented them in the model, and, finally, report our simulation results.

Memory Consolidation

Forgetting and amnesia. The ability to recall acquired memories normally diminishes with time elapsed since learning. Although there is disagreement about the precise shape of the forgetting curve (Anderson & Tweney, 1997), it is often represented as an exponential so-called Ebbinghaus (1885) forgetting curve, as in Figure 1.

Figure 1: Idealized normal forgetting curve.

In contrast with normal forgetting, memory loss after trauma affects recent memories more than remote ones (McClelland et al., 1995; Scoville & Milner, 1957; Squire & Alvarez, 1995), resulting in a curve with the opposite slope, as in Figure 2.
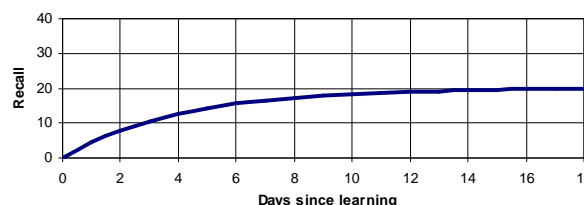

Figure 2: Idealized Ribot gradient.

This graph shows that the ability to recall material learned shortly before onset of amnesia is strongly impaired, whereas older memories are relatively spared. The curve is commonly known as the "Ribot gradient", after the French psychologist Ribot who first postulated it (Ribot, 1882). This temporally graded amnesia gave rise to the idea that a

consolidation process stabilizes newly acquired memories – older memories were less affected in amnesia because they had had more time to stabilize.

Types of consolidation. Researchers distinguish between two types of memory consolidation, "systems" consolidation and "synaptic" or "cellular" consolidation (Dudai & Morris, 2000). Systems consolidation is a process that transitions initially hippocampus-dependent memories to a hippocampus-independent state. In the mammalian brain, the hippocampal formation is involved with the consolidation of "episodic" memories, explicit memories of experienced events. Animal studies as well as human cases of brain damage have shown that memories initially depend on the hippocampus, but gradually become hippocampus-independent. According to the "standard model of systems consolidation" (McClelland et al., 1995; Squire & Alvarez, 1995), hippocampal memory traces are quickly created but only persist for a limited time, during which they support the more time-consuming construction of neocortical memories. On this view, the temporally graded amnesia observed after hippocampal lesions is due to the fact that older memories have had more time to consolidate in the neocortex, while newer memories are still only weakly represented there (McClelland et al., 1995; Squire & Alvarez, 1995). This process is called "systems consolidation" because it involves interaction between two brain systems, the hippocampus and the neocortex. In contrast, the so-called "cellular" or "synaptic" consolidation process concerns the stabilization of memories within a single system.

The TraceLink Model of Memory Consolidation

TraceLink is a connectionist model of systems memory consolidation (Meeter & Murre, 2005; Murre, 1996). The model has two layers representing hippocampus (HC) and neocortex (NC), respectively. The HC layer has 42 units and the NC layer has 200 units. Each layer is fully connected, i.e. there are independent (asymmetric) connections in both directions between each pair of units, and the two layers are also fully interconnected. Connection weights have values in the range 0.0 to 1.0. The units have discrete activation levels, either 0.0 (inactive) or 1.0 (active), and a stochastic asigmoid activation function:

$$P_j = \frac{1}{1+e^{-\frac{net_j}{temp}}} \qquad [1]$$

where P_j is the probability that unit j will become (or remain) active, net_j is the net input to unit j and *temp* is a parameter that controls the steepness of the asigmoid function, i.e. the amount of randomness in the model. (For small values of *temp*, $P_j(net_i)$ approaches a deterministic step function; for large *temp*, $P_j(net_i)$ is close to 0.5 everywhere, i.e. equal probability of becoming active or inactive regardless of net_i.). A *temp* value of 0.2 was used in all simulations. The net input net_j in equation [1] is calculated according to the following formula:

$$net_j = \sum_i w_{ij} a_i - inhibition_L \qquad [2]$$

where w_{ij} is the weight of the connection from unit i to unit j, and a_i is the activation level of unit i. The term $inhibition_L$ is a layer-specific inhibition quantity that simulates the effect of inhibitory synapses. It is calculated by a feedback algorithm that drives the number of active units in each layer towards a configured equilibrium value, which is also the number of active units in training patterns for the layer. For example, each training pattern for the NC layer has ten active units, and the inhibition mechanism makes the layer preferentially settle into states with that number of active units.

The learning rule is Hebbian with an anti-Hebbian "interference" term that accelerates forgetting of previously learned patterns, especially in the smaller HC layer, where there is more pattern overlap:

$$w_{ij}(t+1) = w_{ij}(t) + \mu_T^+ a_i a_j - \mu_T^- (1-a_i) a_j \qquad [3]$$

where $w_{ij}(t)$ is the connection weight between units i and j at time t, a_i is the activation level of unit i, μ_T^+ is the Hebbian learning rate, and μ_T^- is the interference or "unlearning" rate. The learning rule strengthens connections between units that are both active, and weakens connections from inactive to active units. Learning rates are specified per "tract" (hence the T subscript). A tract is a set of connections with the same source and destination layers: all the connections from HC units to NC units form one tract, all connections internal to the NC layer form another tract, etc. A tract's learning rates (μ_T^+ and μ_T^-) may take on different values during initial acquisition versus consolidation. This simulates the effect of neuromodulation, for example, an increased learning rate in hippocampus in the presence of novel stimuli (Meeter & Murre, 2005; Murre, 1996).

Initial acquisition. The TraceLink system is trained by presenting a training pattern to both layers[1] and applying the learning rule to adjust connection weights. The intra-HC and NC-HC tracts have high learning rates and learn patterns well in a single presentation. The intra-NC tract has a much lower learning rate, and as a result a single training cycle only creates a weak trace there.

Recall. To test recall of a training pattern, a subset of the pattern's active NC units (a "cue pattern") are held ("clamped") in the "on" state, and the rest of the units in both layers are randomly set to either the active or inactive state, with equal probability. The whole system is then repeatedly cycled by executing the activation function for all the unclamped units in random order and updating their activation levels accordingly. At the end of each cycle, the inhibition algorithm adjusts the inhibition coefficients of both layers. After a configurable number of such cycles (we

---

[1] It would be more realistic to present only the NC pattern, and let the model discover an HC representation autonomously. This is the subject of a planned enhancement of the present model.

used 70 in all simulations), the activation pattern into which the system has settled is compared to the original training pattern. Recall accuracy is measured as the percentage of non-cued NC units in the training pattern that have been successfully turned on.

Lesioning. Hippocampal lesion is simulated by simply disconnecting the HC layer (setting all inter-layer connections weights to zero). After initial training, the intact system can normally recall patterns quite well, because the NC-HC and HC-HC connections provide linkage between the pattern's NC units, but after virtual lesioning recall is poor, because the NC-NC connections are not strong enough to independently enable the system to complete the pattern correctly.

Consolidation. Memory consolidation is simulated by randomly setting each unit's activation level to either 0.0 or 1.0, letting the system "settle" in the same manner as for recall (but without any cue pattern), and reinforcing whatever state it settles into by applying the learning rule in the NC layer. Because the system is more likely to settle into trained patterns (Hopfield, 1982), this procedure gradually strengthens those patterns in the NC layer. After a pattern has been reinforced in this manner a sufficient number of times, its NC connections become strong enough that the pattern can be recalled even after HC lesioning.

Simulations. In a typical TraceLink simulation, a series of training patterns are presented, one per simulated "day", each followed by a number of consolidation cycles (Meeter & Murre, 2005; Murre, 1996). Because of interference, especially in the smaller HC layer where patterns overlap more, earlier patterns are gradually overwritten by newer ones. When recall is tested after training a number of patterns, a forgetting curve can be observed: older patterns are recalled less successfully than newer ones. The model is thus able to account for normal forgetting (the idea that interference plays a major role in hippocampal forgetting may be debatable (Hardt, Nader, & Nadel, 2013)).

While patterns are slowly forgotten in the HC layer, they are gradually strengthened in the NC layer due to consolidation. If the HC layer is "lesioned" after a number of days, the earlier training patterns, which have had more time to consolidate and therefore have a stronger NC representation, are recalled more successfully than the newer ones. The model is thus also able to account for the Ribot gradient observed after hippocampal lesion. See Meeter & Murre (2005), for more details about the TraceLink model, including accounts of simulations that reproduce a range of human memory phenomena.

Memory Reconsolidation

It has been shown that reactivating a consolidated memory can return it to a labile state, from which it needs to reconsolidate in order to persist (Nader & Hardt, 2009). During the period of instability, the so-called "reconsolidation window", memory impairments may be produced by the same types of intervention that can interfere with initial consolidation, such as lesions and protein synthesis inhibition (Debiec et al., 2002; Nader, Schafe, & Le Doux, 2000). Some have suggested that that post-reactivation plasticity allows knowledge to be modified when new information is acquired (Hardt, Einarsson, & Nader, 2010; Lee, 2009). As is the case with memory consolidation, memory reconsolidation has been documented at both the systems and cellular level. The former type, systems reconsolidation, is "the demonstration that reactivation of a remote memory returns the trace to being hippocampus dependent again for a period of time before once again becoming independent of hippocampus" (Debiec et al., 2002).

Method

Although the physiological events underlying systems memory reconsolidation are not known, researchers have proposed hypothetical mechanisms that could explain the observed phenomena. The present work is a neural-network model of such a hypothesis (Debiec et al., 2002; Hardt et al., 2010; Nadel & Hardt, 2010; Nader et al., 2000). According to this hypothesis, (1) consolidation renders remote memories hippocampus-independent; (2) reactivation of a consolidated neocortical memory creates a temporary hippocampal trace (or strengthens the existing but decaying trace); (3) the hippocampal trace stimulates the neocortical trace through back-projections; (4) this stimulation has the effect of initially destabilizing the neocortical synapses, making them susceptible to decay and/or modification; (5) continued hippocampal reinforcement prevents decay of (or even strengthens) the neocortical trace while it restabilizes. The model thus provides an explanation for the observed fact that reactivation followed by hippocampal lesion produces amnesia, but neither reactivation nor lesion alone causes memory loss.

Implementation

In order to model this hypothesis, we implemented a two-layer network along the lines of TraceLink, but with a few additional features: (a) connections have a plasticity attribute; (b) connection weights are subject to time-based decay (Hardt et al., 2013); and (c) the simulation now includes a "reactivation" phase to trigger memory reconsolidation.

Plasticity. The plasticity attribute has a value between 0.0 and 1.0, representing minimum and maximum plasticity, respectively. Our new learning rule takes plasticity into account:

$$w_{ij}(t+1) = w_{ij}(t) + p_{ij}(\mu_T^+ a_i a_j - \mu_T^-(1-a_i)a_j) \quad [4]$$

where p_{ij} is the plasticity of the connection from unit i to unit j. Thus the plasticity affects a connection's sensitivity to training and also its susceptibility to interference.

Connections are created with a p_{ij} value of 1.0 (fully plastic), which subsequently decreases exponentially over simulated time, as expressed by the following formula:

$$p_{ij}(t+1) = p_{ij}(t) \cdot (1 - pdr_T) \qquad [5]$$

where pdr_T is a plasticity decay rate specific to the tract to which connection ij belongs. In the simulations reported here, the pdr_T value was 0.1 for the NC-NC tract, and 0.0 for the other tracts, i.e. plasticity variations in hippocampus were not simulated.

Decay. Connection weights are subject to exponential decay at a rate that is configurable on a per-tract basis. A connection's weight decays by its decay rate modulated by its plasticity, according to the following formula:

$$w_{ij}(t+1) = w_{ij}(t) \cdot (1 - p_{ij} wdr_T) \qquad [6]$$

where wdr_T is the weight decay rate specified for the tract to which the connection belongs. Thus, as a connection becomes less plastic, it becomes more resistant to decay (Hardt et al., 2013).

Reactivation. In addition to TraceLink's "Acquisition" and "Consolidation" phases, our model has a "Reactivation" phase, during which one or more previously trained patterns are activated, the learning rule [4] is applied, and the plasticity between active units is restored to its maximum value 1.0. Following reactivation, a number of consolidation periods may be executed, as after initial learning.

Simulations

The following simulations were carried out:
A. *Consolidation*
    1. Train a single pattern.
    2. Execute 40 consolidation periods (simulated "days"). At each day, test recall in the intact system and with "lesioned" (deactivated) HC layer.
B. *Reactivation/Reconsolidation*
    Same procedures for training, consolidation and testing as in simulation A, but on day 20, reactivate the trained pattern, then continue daily consolidation and testing.
C. *Reactivation and HC lesion*
    Same procedure as in simulation B, but on day 21, permanently lesion the HC layer.

The same parameter settings were used in all three simulations, as indicated in Table 1.

An explanatory note about the daily recall tests with intact and "temporarily lesioned" HC: these tests are performed without affecting the continued evolution of the system. No learning or (re)consolidation takes place, and HC is turned back on after testing. The simulation then continues as if the tests had not taken place. Researchers with live subjects, of course, do not have this luxury; in an analogous experiment, they would only be able to get one data point from each subject.

Table 1: Parameter values used in the simulations

Parameter	Values	
	NC	HC
Learning rate during initial acquisition	0.06	0.4
Learning rate during consolidation	0.02	0.0
Learning rate during reactivation	0.0	0.2
Unlearning rate	75% of learning rate	
Weight decay rate	0.1	0.1
Plasticity decay rate	0.1	0.0
Number of units	200	42
Active units at equilibrium (=pattern size)	10	7
Cue pattern size (units)	5	0

The values in the "NC" column apply to the NC layer and intra-NC tract. The values in the "HC" column apply to the HC layer, intra-HC tract and inter-layer tracts.

Results

A. Consolidation

Figures 3a and 3b show the weight and plasticity of a representative individual connection in the HC-HC and NC-NC tracts, respectively, during the consolidation simulation. Each of the two monitored connections joined two units that were simultaneously active in the training pattern, i.e. they were connections where significant Hebbian learning took place.
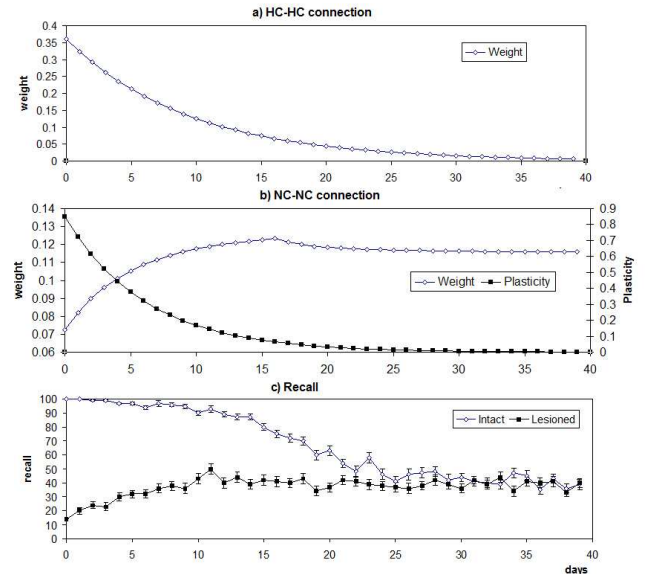

Figure 3: Consolidation. a) Connection weight of a hippocampal connection. b) Weight and plasticity of a neocortical connection. c) Recall performance (averaged results from fifty simulations). Each point on the "lesioned" curve shows the performance with deactivated HC, i.e. as if HC had been lesioned on that day. Vertical bars show standard error.

As expected, HC connections quickly learn the presented pattern, and then decay exponentially. NC connections, on the other hand, quickly become very plastic, but learn only gradually. Around day 17 the HC trace has become too faint for any further consolidation to take place, and the NC trace starts to decay somewhat, but the decay slows down as the plasticity diminishes further and the trace becomes stabilized.

Figure 3c shows the recall performance during the simulation. The upper curve, representing recall in the intact system, shows normal forgetting. The lower curve, recall performance with disabled HC layer, shows a gradient during the consolidation "window", followed by constant performance. These results are similar to those obtained with the original TraceLink model (Meeter & Murre, 2005); the difference is that forgetting there was purely interference-based, whereas in this simulation it is caused by a combination of interference and decay. (Interference plays a role even though only a single pattern is trained, because the patterns reinforced during (re)consolidation may differ from the trained pattern.)

B. Reconsolidation

As shown in Figure 4, if the pattern is reactivated on day 20, then (a) the hippocampal trace is rapidly strengthened, (b) the necocortical trace is quickly destabilized and then gradually strengthened and restabilized in a round of reconsolidation, and (c) the recall performance is somewhat improved after the reminder.
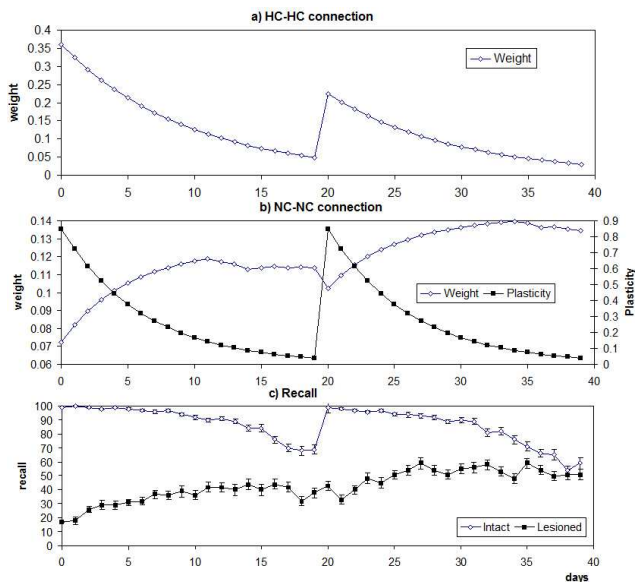

Figure 4: Reconsolidation. a) Connection weight of a hippocampal connection. b) Weight and plasticity of a neocortical connection. c) Recall performance (averaged results from fifty simulations).

C. Reactivation followed by HC lesion

When the HC layer is permanently lesioned after memory reactivation, the results are as illustrated in Figure 5: (a) The

hippocampal trace decays after initial training as in the previous simulation and is boosted by the reactivation
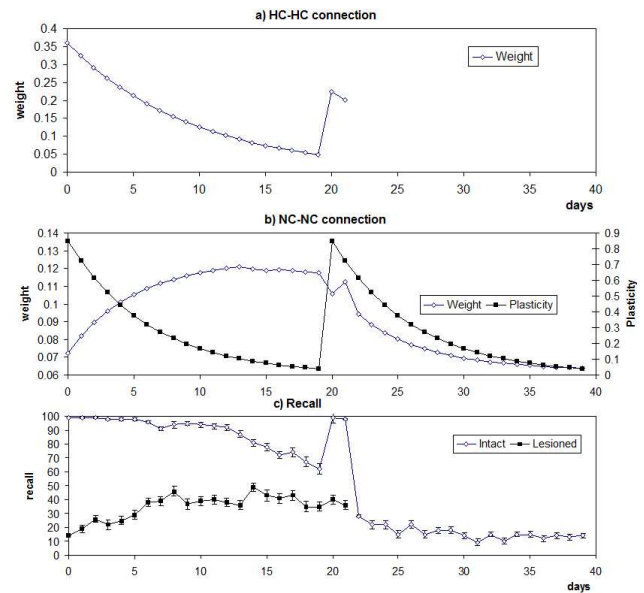

Figure 5: HC lesioning following reactivation. a) Connection weight of a hippocampal connection. b) Weight and plasticity of a neocortical connection. c) Recall performance (averaged results from fifty simulations). The points on the "intact" curve after day 21 show the performance of the lesioned system.

on day 20. The plot ends at the hippocampal lesion on day 21. (b) The neocortical trace evolves as in experiment B until day 20, the day of reactivation. Following the HC lesion on day 21, instead of being strengthened by reconsolidation, the destabilized NC trace rapidly decays. (c) The recall performance shows rapid onset of amnesia after the hippocampal lesion.

Discussion

In spite of a growing number of studies on both humans and animals, the neural mechanisms underlying memory reconsolidation are not well understood. The present paper seeks to contribute to the development of a theory by introducing a computational model of reconsolidation.

The key finding in system memory reconsolidation studies is that lesioning after reactivation produces amnesia, whereas neither reactivation alone nor lesioning alone causes memory impairment (Debiec et al., 2002; Nader & Hardt, 2009). With this in mind, it is interesting to compare Figures 3-5. Figure 3c shows that, once a memory is consolidated in the model, hippocampal lesions without preceding memory reactivation have little effect on it, whereas Figure 5c illustrates that post-reactivation lesions lead to a dramatic drop in recall performance. The cause of this difference is that, after reactivation, the plasticity of the neocortical trace is high, allowing for rapid decay. In Figure 4c, on the other hand, where hippocampus is left intact after

reactivation, reconsolidation more than compensates for the decay, resulting in moderate strengthening of the memory trace after reactivation.

The neural network model presented here is able to reproduce the empirical results by simulating micro-processes that have been hypothesized to underlie memory reconsolidation - controlled variability in synaptic plasticity and plasticity-dependent synaptic decay rates – and thus demonstrates that these mechanisms in fact can account for the observed effects.

An interesting aspect of this model is that it introduces decay-driven forgetting, in contrast with the TraceLink simulations, where all forgetting was due to interference (Meeter & Murre, 2005). It is likely that both types of mechanism play important roles in the consolidation and maintenance of memories (Hardt et al., 2013), and we are planning to apply the model to further investigate the relationship between the two. In particular, work in progress includes simulations with multiple training patterns, which will allow us to study the combined effects of decay and even greater interference.

Another direction in which we are planning to extend this work is to apply the model to manifestations of reconsolidation other than amnesia after hippocampal lesions. These include the effects of protein synthesis inhibition (Debiec et al., 2002; Nader et al., 2000) and interference training in the reconsolidation window (Hupbach, Gomez, Hardt, & Nadel, 2007; Hupbach, Gomez, & Nadel, 2009; Walker, Brakefield, Hobson, & Stickgold, 2003).

Acknowledgments

References

Anderson, R. B., & Tweney, R. D. (1997). Artifactual power curves in forgetting. *Memory & Cognition*, *25*(5), 724–730.

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

Dudai, Y., & Morris, R. (2000). To consolidate or not to consolidate: what are the questions? In *Brain, Perception, Memory. Advances in Cognitive Sciences (Bolhuis J.J. Ed)*. Oxford University Press.

Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*. Duncker & Humblot.

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.

Hardt, Oliver, Nader, K., & Nadel, L. (2013). Decay happens: the role of active forgetting in memory. *Trends in Cognitive Sciences*.

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.

Hupbach, A., Gomez, R., & Nadel, L. (2009). Episodic memory reconsolidation: updating or source confusion? *Memory*, *17*(5), 502–510.

Lee, J. L. C. (2009). Reconsolidation: maintaining memory relevance. *Trends in Neurosciences*, *32*(8), 413–420.

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*, 419–457.

Meeter, M., & Murre, J. M. J. (2005). Tracelink: A model of consolidation and amnesia. *Cognitive Neuropsychology*, *22*(5), 559–587.

Murre, J. M. J. (1996). TraceLink: a model of amnesia and consolidation of memory. *Hippocampus*, *6*(6), 675–684.

Nadel, L., & Hardt, O. (2010). Update on memory systems and processes. *Neuropsychopharmacology*, *36*(1), 251–273.

Nader, K., & Einarsson, E. O. (2010). Memory reconsolidation: an update. *Annals of the New York Academy of Sciences*, *1191*, 27–41.

Nader, K., & Hardt, O. (2009). A single standard for memory: the case for reconsolidation. *Nature Reviews Neuroscience*, *10*, 224–234.

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.

Osan, R., Tort, A. B. L., & Amaral, O. B. (2011). A Mismatch-Based Model for Memory Reconsolidation and Extinction in Attractor Networks. *PLoS ONE*, *6*(8), e23113.

Ribot, T. A. (1882). *Diseases of Memory, an Essay in the Positive Psychology*. New York, NY: D. Appleton and Company.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, *20*(1), 11–21.

Squire, L., & Alvarez, P. (1995). Retrograde-amnesia and memory consolidation - a neurobiological perspective. *Current Opinion in Neurobiology*, *5*(2), 169–177.

Walker, M. P., Brakefield, T., Hobson, J. A., & Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature*, *425*(6958), 616–620.