Memory Reconsolidation:

Making Predictions Better

Lynn Nadel and Per B. Sederberg

Abstract

Memory reconsolidation refers to the phenomenon whereby a previously consolidated memory, i.e., one that is resistant to interference or disruption, becomes labile due to reactivation, initiating a short window during which that memory can be modified. With a wide range of potential clinical and educational applications, reconsolidation has been demonstrated across multiple domains and timescales, including fear, motor, and episodic memory. This chapter seeks to clarify the psychological processes involved in reconsolidation, making connections to underlying physiological mechanisms, with the goal of providing a framework for understanding why and when reconsolidation takes place. Drawing on reviews of both human and relevant animal studies, this chapter highlights the importance of both context and predictions for determining whether new experiences give rise to new learning or the updating of previously-learned associations.

Introduction

Memory reconsolidation, and its older relative, memory consolidation, refer to stabilization processes by which memory traces become resistant to interference and disruption. Consolidation refers to the initial fixation of a memory, not simply its strengthening due to item repetition, and reconsolidation refers to the re-fixation of a reactivated memory, possibly updated with some new information. They involve many of the same cellular and molecular mechanisms, including the synthesis of new

proteins, though there are important differences, indicating that reconsolidation is neither an extension nor recapitulation of consolidation.

Inherent in the original notion of consolidation (Müller and Pilzecker, 1900) was the idea that through consolidation a memory becomes increasingly stable until it is permanently fixed and henceforward immutable. Renewed interest in memory and memory consolidation in the 1950s accepted this permanence premise (eg., Hebb, 1949; Gerard, 1963; Glickman, 1961; McGaugh, 1966). Although some questioned the idea (eg., Coons & Miller, 1960; Misanin et al., 1968; see Crowder, 1976; Sara, 2000; Dudai, 2004; Riccio, Millin, & Bogart, 2006 for reviews of the relevant history), the notion of consolidation as a time-limited process, with a permanent end-point, persisted until two critical findings, in separate labs, called it into question.

In the first study (Przybyslawski & Sara, 1997) rats were trained in a spatial maze, and given time to consolidate what they had learned. Memory for the maze was reactivated during a retrieval trial, after which they were immediately injected with an NMDA-receptor antagonist (MK-801) into the stomach where it could affect the entire brain[1]. When tested 24 hours later, the performance of these rats was significantly impaired. Delaying the injection by two or more hours left performance unaffected. Retrieval had induced a time-limited vulnerability in a previously 'consolidated' memory. Przybyslawski and Sara coined the term *memory reconsolidation* to describe their results[2].

Several years later, Nader et al. (2000) injected a protein synthesis inhibitor (anisomycin)[3] directly into basolateral amygdala (BLA) after reactivation of a fear memory that linked a tone stimulus (CS+) with footshock (the US). The next day,

[1] NMDA receptors are a type of glutamate receptor known to be critical in synaptic plasticity.

[2] In hindsight this was an unfortunate choice. The linkage to consolidation this term implies may or may not be warranted. What is more, confusion reigns in the usage of the terms reconsolidation and updating. We will use the term reconsolidation to refer to the notion that an activated memory undergoes some process of re-stabilization. Whether or not a re-stabilized memory **is the same memory** or has **been modified and updated** is an empirical question for any given memory retrieval. The term reconsolidation applies whether or not updating has occurred.

[3] It is widely assumed that protein synthesis is required for stable long-lasting memory, presumably to provide the basis for structural changes, such as the addition of receptors at the synapse.

anisomycin-injected rats showed less fear than control rats (injected with a non-active substance) when exposed to the tone, and also less fear than rats that had not been reminded of the fear learning prior to drug administration. Injection of anisomycin six hours after reactivation of a fear memory did not disrupt memory, confirming the temporal gradient of reconsolidation reported by Przybyslawski and Sara. Reactivation of the fear memory seemed to return it to a labile state requiring re-stabilization. These studies raised profound questions about memory - its fixedness and function.

In parallel with the questions surrounding memory permanence, questions about the organization of memory, prompted largely by brain studies, had been stirring for some time. What had already become clear by the 1980s was that there are multiple systems concerned with memory, and that these were undergirded by separable brain networks (eg., Nadel & O'Keefe, 1974; Cohen & Squire, 1980). Distinctions between explicit and implicit forms of memory are now taken for granted, and further distinctions exist within these types of memory with implications for whether consolidation is the same for all memory types. As we will see, the emergence of multiple memory systems theory affected thinking about memory consolidation, and shed light on what had become a serious problem in the study of consolidation through the 1960s and 1970s.

Many of the early studies focused on trying to determine how long consolidation lasted. Unfortunately, rather different answers emerged, depending on the paradigm used, and the species tested. Much of this early work involved avoidance learning in rats, using post-training manipulations such as electroconvulsive shock, ice-cold water immersion, or drug injections. These studies suggested consolidation times of minutes to hours (Glickman, 1961). In work with humans, direct experimental manipulation was virtually impossible. Estimates of the length of consolidation were instead based on work with patients undergoing ECT for depression (eg., Squire et al. 1976), or suffering from brain damage of one etiology or another. Such estimates were of hours, days, or even months (see Squire, Cohen & Nadel, 1984, for an early review).

It is now the consensus view that consolidation transpires at both the cellular and systems level, the former on relatively short time-scales, the latter more extended in time. Cellular level consolidation processes initiate and sustain changes in synaptic strength, instantiating the gain (or loss) of connectivity within the neural ensembles that capture memories. Systems level consolidation involves apparent shifts in control over memory expression, at the level of brain networks, that occur over weeks and longer (cf., Genzel & Wixted, 2017). Recent memories appear mostly under the control of medial temporal lobe circuits, while more remote memories, in particular those that have lost much of their detail and uniqueness, appear mostly under the control of medial prefrontal circuits (see Moscovitch et al., 2016, for review). Exactly how this happens remains unclear, with hypotheses running the gamut from memories being transferred between systems, to parallel memories being created with shifting strengths and/or impacts over time. Whether traces of remote memories persist in the medial temporal lobe is currently up for debate (Barry & Maguire, 2019; Moscovitch & Nadel, 2019).

Such thinking has implications for the phenomenon of reconsolidation. Does reactivating a memory necessarily destabilize it? And what exactly does it mean to 'destabilize' a memory? If there are parallel memory traces representing an experience, does reactivation destabilize all of them, or only some? Are rules governing the reconsolidation of a destabilized memory the same for all memory systems, and do they apply equally to both recent and remote memories? When does reconsolidation involve merely re-fixing a reactivated memory, leaving it unchanged by current experience (the latter being reflected in a new, distinct memory), and when does it involve adding to, or subtracting from, a reactivated memory in some way? When reconsolidation involves memory "updating", does such updating proceed in the same way in the various memory systems?

These are a lot of questions, only some of which can be addressed in detail here. Before addressing them we will sketch out the broad neurocognitive framework within which we can talk about memory reactivation and reconsolidation. Figure 1 captures the main features of our perspective.

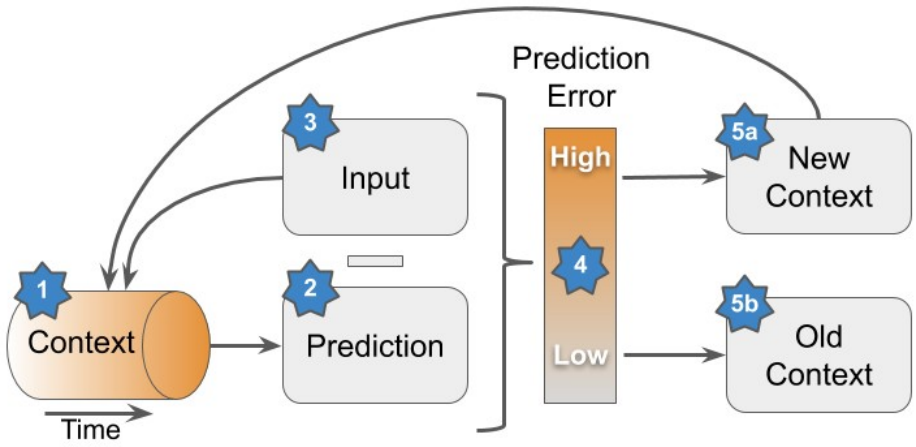

Figure 1: Reconsolidation Framework

1) The brain is constantly active, at any given moment it is in brain context X_t, represented by widely distributed patterns of neural firing. 2) Each cycle generates a prediction from the current state of context about the next input. 3) Brain context 'drifts' in response to modest changes in input.
4)The difference between the new input and the predicted input generates a prediction error that can be either high or low. 5a) If high, the system may shift to a new brain context Y – this is the equivalent of "pattern separation", and gives rise to the formation of a new, distinct, context memory. 5b) If low, the system likely remains within context X – this is the equivalent of "pattern completion", and results in the modification and updating of the reactivated memory as a function of current inputs.

The figure shows that the brain generates a steady flow of predictions, based on its prior state (i.e., mental context; see Chapter on Context) and any inputs it receives, including what we have referred to here as "reminder" cues. The brain states initiated by these cues -- the neural bases of reactivated memories – make up the context at the root of these predictions. When predictions are largely confirmed (i.e., a low prediction error), then any modest changes that are noticed could be incorporated into existing, reactivated, memory traces and reconsolidation of this updated memory can ensue. If the predictions are largely disconfirmed (i.e., a high prediction error), then new, and distinct, memory traces are created. In this case, the reactivated memory trace is left unchanged, and reconsolidated in that unchanged form. It is important to note that the

term reconsolidation applies to what happens to the reactivated trace. Transformed or not, reactivated traces are typically reconsolidated. Unfortunately in much of the literature the term reconsolidation has been reserved for the case where a memory has been transformed after reactivation, ignoring what happens in the absence of transformation. We emphasize here that changed or not, reactivated memories are re-stabilized over time, although the strength of this restabilized trace may vary from its pre-reactivation level.

Hupbach etal. (2013) suggested that a true reconsolidation effect requires: (1) reactivation of a memory; (2) time to manifest in behavior; (3) persistence over time. This was expanded by Elsey etal. (2018), who added that reconsolidation effects had to be specific, affecting only a reactivated memory. They also highlighted the brief "reconsolidation window" after reactivation, during which memories are labile. While these criteria capture the main features of reconsolidation, they cannot readily distinguish reconsolidation from extinction, which is commonly defined as the gradual weakening of a conditioned response until it is no longer observed in the test context. As we outline in more detail below, both start with a reminder, followed by a change in the animal's experience in the test situation. Though both speak to the criteria above, they lead to different outcomes. This distinction is of particular importance in clinical settings, and for that reason has attracted considerable research interest.

In the clinic the distinction between extinction and reconsolidation translates into the difference between relapse and cure, as it is assumed that extinction procedures do <u>not</u> modify the memory traces underlying prior fear learning (cf., Bouton, 2017; Dunsmoor et al., 2015). Rather, they create a new memory of safety in the extinction context, which competes with the fear memory originally trained in that context. Unfortunately, the originally trained fear memory generalizes to new contexts, while the newly-trained extinction memory (of safety) does not. When taken out of this originally dangerous but now safe context, fear typically returns.[4]

[4] Addiction researchers have sought to apply reconsolidation to the problem of altering deeply-ingrained behaviors - addicts too suffer from context-limited 'cures'. See Exton-McGuinness & Milton, 2018, for an up-to-date review of this work.

Extinction, like reconsolidation, is the victim of a misleading name. Standard extinction procedures do not actually extinguish a memory, in any simple sense of the term[5]. Extinction is simply another learning opportunity, where something that once was the case no longer seems to be the case. By definition, in an extinction procedure, the situation reactivates a prior "memory", which predicts something (shock, food, etc). What happens next determines how far off the prediction is, and whether or not the organism treats the situation as a variant of some prior experience, or as an entirely new thing. This, in a nutshell, is the choice between updating/reconsolidating an existing memory, or creating an entirely separate memory. Understanding when you see reconsolidation, and when you don't, could help unravel and illuminate the phenomenon itself. Matters are made more complicated by the fact that predictions come in several varieties. The notion of prediction error first emerged within a reinforcement learning framework, focusing on predictions about reward - either immediate or in the long-term. One of the major concerns within this framework is the 'credit assignment' problem -- how changes in the brain can be restricted to those neural ensembles whose activity increased access to rewards (broadly defined). There is, however, a second kind of prediction, which goes beyond reward value to encompass the "structure" of the organism's world. Any changes in this world, whether or not they involve reward or punishments, can generate prediction errors. Interest in this second kind of prediction, flowing from internal models of the world, has led to concern with another major issue - whether an organism should choose to exploit what it already knows about a situation, or explore what it does not yet know, perhaps to acquire some critical new piece of information[6]. Thus, different kinds of prediction can be generated when one's current situation reactivates traces of prior experience. There are predictions about rewards and punishments, where they might be found, and what kinds of behaviors would get them. And, there are predictions about the general

[5] Extinction procedures were developed within a behaviorist framework, and what was being learned and extinguished were behaviors, not memories. The use of the term has carried over to our more cognitive age, but when one talks about memories rather than behaviors the word extinction fits less well.

[6] The cognitive map theory of hippocampal function put exploration at the center, and discussed how the maps built up during exploration provided the basis for predictions (Nadel & O'Keefe, 1974; O'Keefe & Nadel, 1978). This built on Tolman (1948, 1949), who understood the critical importance of learning that did not involve rewards and punishments.

features of an experience that are crucial to defining the context and hence determining where and when predictions about rewards/punishments apply. Unexpected events, being unpredictable, cannot generate classic prediction error. Instead, surprising events engender "positive prediction error"[7], whose magnitude should be a function of how congruent or incongruent the surprising event is in that context (cf., Long etal., 2016). As such, they undoubtedly exert important influences on what an organism does, and what it learns.

Understanding when ones sees updating and reconsolidation of an existing memory, and when one sees the creation of a new memory (as in extinction) ultimately depends upon understanding the interplay between these various types of prediction, and the neural systems that undergird them. The search for such understanding has mostly involved the use of fear conditioning, primarily in animal models, to which we now turn before we expand our discussion to fear, motor, and episodic memory reconsolidation in humans.

Reconsolidation of Fear Conditioning

Nader et al. (2000) spurred considerable interest in reconsolidation in fear conditioning, not least because of the obvious clinical implications (cf., Lane & Nadel, 2019). The study of fear conditioning, and its consolidation and reconsolidation, has been pursued in both humans and animal models, allowing for exploration of both the behavioral and cellular correlates of the process.

[7] Note that we are using the term positive and negative prediction error differently than Kindt and colleagues. They have focused on reward prediction error, so for them positive PE results from getting more reward than predicted, and negative PE involved getting less reward. Both of these would be considered negative PE in our terminology, as both reflect the disconfirmation of a prediction. For us, positive PE reflects unexpected events that were also unpredicted. In other words, surprise.

Fear conditioning is deceptively simple. Typically, mice or rats are placed in an enclosed apparatus, and after some time, given one or more electric shocks through a grid floor. This procedure generates fear of that experimental context in the animal, expressed as increased 'freezing', or immobility, when the trained animal is placed back in that context in an hour, or the next day, or week, or even month. Sometimes a cue (a tone, or a flashing light) is used to signal an upcoming shock during training, in which case the animals learn about relations between cue, context, and shock. Cue and context fear learning engage only partially overlapping neural systems; cue-fear associations require the amygdala, whereas context-fear associations also engage the hippocampus (Phillips & LeDoux, 1992).

Failures to observe reconsolidation of fear conditioning typically involve memories that are very strong, very old, or both (cf., Alberini & LeDoux, 2013). A focus on these variables brought attention to the phenomenon of *metaplasticity*, or the control of plasticity. Reconsolidation that gives rise to modification of the original memory trace, it is assumed, requires reactivation of memory circuits and something that amounts to their "destabilization", which opens them up to that change. One current notion (cf., Zhang et al., 2018) is that susceptibility to destabilization is controlled by the ratio of two NMDA receptor subunits (GluN2A/GluN2B). As this ratio increases, it becomes harder for reactivations to initiate destabilization, so it is of considerable interest that both memory strength and memory age can influence, and generally will increase, this ratio. One implication of this boundary condition is that highly ingrained habits are likely to be hard to change. Another is that memory reactivation does not always lead to memory destabilization. Old, and strong, memories appear resistant to this effect.

Cue Specificity and Context

Given that reconsolidation starts with memory reactivation, one critical question is how similar a situation has to be in order for it to reactivate a particular memory. In regard to the fear conditioning literature, this question was first asked by varying the contexts within which fear was established and retrieved. Matters became complicated when it

was shown that context specificity, itself, changes with the passage of time. When fear of a context (X) is established, for a number of days after initial learning fear seems relatively restricted to the training context (X) alone. However, a month later, fear is evinced in other contexts that bear some, but perhaps only scant, resemblance to the training context. This shift reflects systems-level consolidation, which stretches over weeks or months in the case of fear conditioning, and even years in other cases. There remains considerable debate about what happens during systems consolidation, but for present purposes we can focus on the fact that in rodent fear conditioning it appears to involve: (1) the parallel creation of memory traces in the hippocampal formation and prefrontal cortex; and (2) the initial dominance of the hippocampal trace, but with time a shift towards dominance of the cortical trace, which is accompanied by generalization of fear to (even vaguely) similar contexts. Given this, one can question exactly what happens during context fear memory reconsolidation?

Winocur et. al. (2009) used hippocampal lesions to explore reconsolidation of context fear. In a first experiment, rats were trained to have a context fear and subjected to hippocampal or control lesions either 24 hours or 28 days later, followed by retention tests. Hippocampal lesions made a day after training eliminated fear, but the same lesions made at 28 days had no effect. In this study they also manipulated context, testing for fear either in the training context or a novel, safe, context. They found, as others had suggested (Anagnostaras et. al., 1999), that fear is context-specific early on (day 1), but generalizes to other, ostensibly similar, contexts over time (by day 28). This result reflects a more generalized memory for context fear established outside the hippocampus. This memory trace, localized by some to the prefrontal cortex, is assumed to be more schematic and hence applicable to contexts beyond the originally trained one (eg., Frankland, 2004).

In a second experiment, Winocur et al. (2009) went on to probe the nature of the 28 day old memory trace that survives hippocampal damage, using what is essentially a reconsolidation design. Rats were, as before, trained to have a context fear. Four weeks later they were “reminded” by being put into either the training context (X) or a

different context (Y) that only resembled context X with regard to general features. Then, a day later, they were subjected to either hippocampal or control lesions. Retention tests followed, in either X or Y. One major result was that it mattered greatly whether the reminder was given in context X or Y. When rats were reminded in the same context in which they were trained, hippocampal lesions pretty much eliminated fear - whether tested in X or Y. When rats were reminded in a novel context, hippocampal lesions had no effect on fear levels compared to controls.

These and other results (Land et al., 2000; Debiec et al., 2002) support the view that there are multiple traces capable of supporting context fear, but disagreement persists as to the fate of these traces over time. Some argue that the more specific hippocampally-based trace disappears over time (eg., Barry & Maguire, 2019), whereas others argue that a trace, however faint, remains in hippocampus (eg., Moscovitch & Nadel, 2019). The fact that a reminder can render a 28 day-old context memory susceptible to hippocampal lesions supports the latter view. Further support is offered by recent optogenetic studies, which show that memory traces (of different specificity) are created in parallel from the very start of context fear conditioning (eg., Kitamura et al., 2017). What seems to change over time is which trace governs behavior, and this can be affected by reactivation and reconsolidation, as the Winocur et al. study shows.

The persistence of fear after extinction training, and the inconsistent impact of pharmacological interventions in human studies (see below) led Monfils et al. (2009) to try a novel behavioral intervention: an isolated reminder trial was used to open the 'reconsolidation window', and this was followed by the standard extinction protocol. This procedure, referred to as the Reminder + Extinction (R+E) protocol, can lead to real weakening of the fear memory. When it works, this protocol does not result in classical 'extinction' of the sort that suffers from spontaneous recovery or renewal. Instead, it is said to update and reconsolidate the existing fear memory. Schiller et al. (2010) reported a similar effect in humans, but failures to find the effect have been reported in both animal models and human studies (cf. Goode et al., 2017; Monfils & Holmes, 2018; Cahill et al., 2018). A recent meta-analysis (Kredlow et al., 2016)

considering results from more than 30 studies in animals and humans suggests that post-reactivation extinction can sometimes yield long-lasting behavioral change, but the effects when present are generally small. We return briefly to this protocol when we consider models of reconsolidation below.

Reconsolidation in Humans

Shortly after reconsolidation was reported in maze learning and fear conditioning in rats, a similar result emerged in the domain of motor learning in humans (Walker et al., 2003). The paradigm was simple, involving sequences of finger-tapping. Right-handed participants learned to tap one of four buttons, using the four fingers on their left hand. They were trained on a particular sequence (eg., 4-1-3-2-4) and the impact of both sleep and the learning of a second, different, sequence on retention of the first sequence were assessed. Results showed that reactivating a human motor sequence memory before learning a different one interfered with the retention of the first, reactivated, sequence.

For several years this result was the most prominent example of reconsolidation in humans, and its apparent non-replication (Hardwicke, Taqi & Shanks, 2016) drew lots of attention, especially as the authors generalized their results beyond the domain of motor memory. While it appears that there are conditions under which reconsolidation is not observed in the procedural learning domain, most often it is (Walker & Stickgold, 2018; Silva & Soares, 2018). And, its occurrence can be mobilized in the service of improving motor skills (Wymbs et al., 2016).

Given its clear translational implications, most research on human memory reconsolidation has used fear or threat conditioning. In one of the first studies of this type (Kindt et al. 2009), spider pictures were paired with unpleasant electric shocks. Learning was expressed both in terms of conscious shock-expectancy ratings and enhanced startle responses. A day after training three groups were composed: two

groups of participants received a single CS reminder cue, the third did not. One of the two reminded groups, along with the non-reminded group, were given injections of propranolol[8] 90 minutes before memory reactivation. The other reminded group received placebo injections. On the third day, both conscious expectancy of shock and unconsciously-driven enhanced startle response were assessed. The participants that had been reminded, and had received propanolol beforehand, no longer showed enhanced startle responses, though their conscious expectations were not changed. This reconsolidation procedure had affected one aspect of fear learning but not the other. A series of follow-up studies demonstrated that this loss of enhanced startle responding was likely not the result of extinction, as classically defined, since manipulations that usually lead to the recovery of an "extinguished" memory did not do so in this case (cf., Elsey & Kindt, 2017, for further discussion).[9]

Attempts to replicate the reconsolidation-blocking effects of propanolol yielded mixed results, propelling Schiller et al. (2010) to use the reminding behavioral manipulation described above. As already noted, Schiller and her colleagues reported successful reconsolidation blockage using the R+E protocol, but failures are common. Agren et al. (2017) report some success with a different behavioral technique, involving "imaginal" extinction. Others have tried different physiological interventions: electroconvulsive shock was reported to impact episodic memory reconsolidation (Kroes et al., 2014) and a recent study reported disruption of memory for a reactivated emotional slideshow story followed by propofol-induced sedation (Vallejo etal., 2019).

Results from the study of fear conditioning in humans tell us several important things. First, they confirm that reconsolidation can be observed in humans. Second, they suggest some potential avenues to persistent behavior change. Third, by showing that reconsolidation plays out differently in implicit and explicit memory systems they connect to work with rodents indicating that parallel memory traces are engaged in fear

[8] Propanolol is a so-called beta-blocker that interferes with norepinephrine and through that, synaptic plasticity.

[9] This study shows the awkwardness of the term 'extinction'. In this case, the one method that actually leads to 'extinguishing' a response cannot be called "extinction". Because 'true' extinction is subject to reinstatement and renewal!

conditioning. Is it possible that the distinction between “reconsolidation” and “new memory formation” is entangled with these parallel memory systems? We return to this question after considering reconsolidation in human episodic memory - perhaps the locus of the state-based prediction system.

Two studies reporting reconsolidation-like effects in human episodic memory were published independently in 2007. Both used variations of standard human verbal learning paradigms, and both verified the existence of reconsolidation in human episodic memory. Forcato et al. (2007) trained their participants on two paired-associates tasks, and demonstrated the existence of reconsolidation-like effects of second list learning on retention of the first list. Forcato et al. (2010) went on to show that updating and reconsolidation in paired-associates learning only happens when the participants are appropriately reminded, and/or given instructions to integrate the two lists, thereby creating a shared list context. Absent these conditions they instead kept the two lists separate, forming a distinctly independent memory, and distinct context, for the second list.

Hupbach and colleagues (2007, 2008, 2009) exposed their participants to a set of 20 objects, presented in physical form, one at a time. They were then asked to recall as many as they could. Training proceeded until a criterion of 17/20 items recalled was reached, to a maximum of 4 trials. Two days later participants were assigned either to a reminder group or a no-reminder group. The reminder group returned to the same room, met with the same experimenter and was asked a leading question about what had happened two days earlier. Together, these reminders served to reinstate the context of the initial learning episode. They then learned a second set of objects, presented in a different way. The no-reminder group came to a new room, was met by a new experimenter, and was taught the second set straightaway, without an explicit reinstatement of the context from the first list presentation. Two days later all participants returned to the original training room and were asked to recall as many items from Set 1 (or Set 2) as they could. In other experiments a recognition paradigm was used instead, in which participants first indicated whether or not they recognized an

item, and if they said yes, which set it was in. Both procedures yielded the same result: participants in the reminder group attributed a substantial number of items from Set 2 to Set 1. The authors interpreted these "intrusions" as reflections of a reconsolidation and updating process. Their result satisfied the criteria noted earlier: it took time to emerge (immediate tests after learning Set 2 showed few intrusions), and it persisted for at least several weeks. What is more, it could not be readily explained by general source memory confusion - items from Set 1 were rarely attributed to Set 2.

Recently, Sinclair and Barense (2018) used naturalistic stimuli (event videos), and showed that incomplete reminders from interrupting the video just prior to the expected outcome rendered memory for the video susceptible to interference from newly presented videos, replicating the effect of partial reminding observed by Hupbach and her colleagues. The greater the surprise, or prediction error, occasioned by the video interruption during the reminder, the more false memories (intrusions) were reported. These effects, crucially, were time-dependent, suggesting that reconsolidation processes were engaged.

Initial attempts have been made to explore the brain correlates of the effects observed in several of these paradigms. The first such study, using a variant of the Hupbach paradigm (Gershman et al. 2013), showed a positive relation between intrusions and neural context reinstatement from the first set study period, providing an intriguing parallel to the behavioral work demonstrating the importance of context (Hupbach et al., 2007). The second study used the paradigm in a different way, focusing on the fate of items from Set 2, and the brain activity associated with correctly attributing an item to Set 2, or incorrectly attributing it to Set 1. Simon et al. (2017) showed that the best predictor of this distinction was activity in the temporoparietal junction (TPJ). They argued that this activity indirectly reflected the extent of prediction error, and that when prediction error is high enough the system separated ongoing experience from activity associated with any reactivated memory. This would account for why increases in TPJ activity were associated with fewer intrusions on a participant by participant basis.

Bavassi et al. (2019) explored the impact of reconsolidation on neural activity in a paradigm similar to the one used by Forcato et al. (2007, 2010). The goal of this study was to compare two reactivation procedures, one involving direct retrieval of the trained list of paired associates (picture - unrelated word), the other merely reactivating the memory of the training without eliciting the pairs themselves. Bavassi et al. showed that the latter kind of reactivation elicits reconsolidation and memory updating, while direct retrieval does not. While they observed different neural signatures associated with the two conditions, the only brain region that reliably distinguished between the two was the posterior cingulate cortex. This brain area, generally thought to be a hub in several core brain networks, often shows activity closely linked to the TPJ.

In sum, initial studies of the neural correlates of memory reconsolidation have identified several brain regions of interest, and have added to the growing consensus that processes concerned with the representation of contexts, and the generation of prediction and prediction errors, are central to this phenomenon. These considerations, amongst others, have helped shape theoretical and computational approaches to memory reconsolidation, to which we now turn.

Models of Reconsolidation

All forms of reconsolidation involve memory, though potentially at widely different temporal scales, relevant brain regions, and processes through which the memories are acquired. As such, any mechanistic theory of consolidation and reconsolidation processes must involve the formation and modification of connections between representations. The key to explaining reconsolidation thus entails characterizing the nature of the representations involved and the processes that bind them.

Regardless of the level of explanation, all theories of reconsolidation have one thing in common: they place a heavy emphasis on the role of reactivating the original memory trace. It is only in this context that one can even talk about reconsolidation, however, it

is important to define what it means to create, maintain, reactivate, and modify a memory, as opposed to forming a new one. Most models of memory and representation begin with the assumption that information is represented in the brain as patterns of firing across neurons or ensembles of neurons with some degree of region specificity (i.e., sensory information is represented in regions of cortex that are largely distinct from regions representing motor information, while other regions integrate information from multiple regions.) Behavior, for example due to some stimulus, manifests as a transformation of patterns of activation as information flows through the brain via synaptic connections between the ensembles of neurons. Thus, memory can be assessed in multiple ways: as mappings between a stimulus and either a conscious (explicit) behavioral response, a subconscious (implicit) physiological response, or an elicited pattern of brain activity.

As noted above, many of the observed reconsolidation effects are *context dependent*, implying that the neural representation of the stimulus that elicits a learned response contains more than just the most-recent sensory input, but in fact involves a more-nuanced brain state integrating multiple dimensions of experience through space and time. Consequently, many theories of memory formation and retrieval define context as a brain state comprising a recency-weighted running average of experience in ensembles of neurons distributed in many areas throughout the brain. For example, one prominent theory of episodic memory formation and retrieval, the temporal context model (TCM), posits that we form associations between the features of experience and the spatiotemporal context in which we experienced them (Howard & Kahana, 2002; Sederberg et al., 2008; Polyn et al., 2009; Lohnas et al., 2015). Consequently, the best cue for retrieving a memory is a recapitulation of the distributed patterns of neural firing (the ‘context’) prevailing at the time the memory was formed, providing an explanation for the extensive context effects observed in memory retrieval (Smith & Vela, 2001). Importantly, this theory also posits that retrieving a memory can facilitate reinstatement of the context bound to the content of that memory, giving rise to a sense of mental time travel (Tulving, 1985), but also providing a potential mechanism by which previous memories can be updated. If a reminder serves to reinstate the context of an original

memory, then new information could be bound to that context, thereby changing any context-dependent behavior.

This is precisely the approach taken by Sederberg and colleagues (2011) when attempting to explain the pattern of episodic memory reconsolidation results reported by Hupbach and colleagues outlined in detail above (Hupbach et al., 2007, 2008, 2009). In these studies the reminder on day two served to reinstate the physical and mental context that was present on day one when the first list of items was studied. In the Sederberg et al. model, context was implemented as a vector of units representing the studied items, with recently experienced items more active than older items, but also including units for the features of each study period, including the task, room, and experimenter. As in standard TCM approaches, the representation of context "drifts" with each new item presentation, with old items becoming weaker as new items enter into context. In this application, however, the additional features representing the task, room, and experimenter were maintained throughout the entire study session in a manner similar to the context maintenance and retrieval (CMR) model, which is built on top of the TCM framework (Polyn et al, 2009). As each item was studied, it was bound via a standard Hebbian association to the current context, which, in the reminder condition on day two, would also include the reinstated Set 1 context, in addition to the current Set 2. Mathematically, a simple Hebbian association involves increasing the connection strength between units of the item and context vectors, proportional to the amount they are both active and, in the approach taken here, it does not weaken connections between units that are not coactive. As we will revisit below, this assumption makes sense for many instances of episodic memory where you do not unlearn one episode just because you experienced another in a similar context, however, there are other forms of memory where such a simple learning mechanism will not work.

At test on day three, context either provides the cue for recall (in the case of free recall) or source memory strength is extracted from the context retrieved from an item cue (in the case of a source memory test). Thus, memory strength for an item will depend on

the extent to which either the Set 1 or Set 2 context was bound to that item and the cue used for memory retrieval. This simple model based on context reinstatement was able to account for the full range of recall intrusions (the tendency to recall Set 2 items as being part of Set 1 on the third day only in the reminder condition) and source recognition errors (an increased probability of misattributing the source of Set 2 items to Set 1 in the reminder condition) across the Hupbach et al. (2007, 2009) studies (Sederberg et al., 2011). It is important to note that this TCM-based account does not need to invoke a notion of memory destabilization. Instead, all memories are potentially malleable if the correct context is reinstated and then bound to new information[10].

While this context-based theory with Hebbian associations can capture reconsolidation effects that involve adding new information to an existing memory, other instances of reconsolidation described below give rise to some degree of unlearning or weakening, not just competition-based masking, of previously-learned associations. This is one area that sets reconsolidation apart from extinction, which can be explained via competition between newly learned associations and largely-unmodified previous associations (Bouton, 2002). Memory modification by means of unlearning is especially prevalent in conditioning experiments, for which the reinforcement learning framework has seen wide success in capturing a range of observed effects (Sutton & Barto, 2018). At the core of reinforcement learning is the assumption that we use memory in service of our goals, i.e., to gain rewards and avoid punishments. As agents, the term used to identify the decision-making individual in reinforcement learning, we extract regularities from our experiences in our environment and use these to make predictions from our current context to help us decide how to behave to maximize future reward. Therefore, learning (and unlearning) are determined by predictions, and the processes governing these predictions must be able to resolve prediction errors that arise from comparing observed outcomes to predictions.

Prediction-error learning further underlines the importance of context: one **must** have a current mental state from which to make predictions and one common assumption is

[10] Much the same assumption was made by multiple trace theory (MTT; Nadel & Moscovitch, 1997), which proposed that every retrieval of a memory led to a new re-encoding.

that your mental representation of the current spatiotemporal context is this state. In fact, it has been shown that temporal context in TCM is similar to the eligibility trace, a decaying memory trace that improves learning efficiency in many reinforcement learning models, and is mathematically equivalent to one widely successful reinforcement learning approach, called the successor representation, under certain conditions (Gershman et al., 2012). This hints at similar underlying mechanisms to explain findings in both reinforcement learning and episodic memory, each of which seem to rely on context and predictions.

The actual learning that takes place due to an event depends on the type of prediction error. Positive prediction errors, generated by events that are not predicted, give rise to new learning, whereas negative prediction errors, generated by the non-occurrence of events that were predicted, give rise to unlearning[11]. If there is no error in prediction there is nothing to learn, though it remains possible that such confirmation strengthens the circuits generating the correct predictions (e.g., Tay et al., 2019; Smith et al., 2013). Interestingly, this implies that if there are no predictions, or the predictions are ignored, there should be no unlearning regardless of the observed outcome because it is not possible to have negative prediction errors without predictions. When predictions are not borne out, there are two potential ways for the agent to deal with the resultant prediction errors. One way involves unlearning the previous association, which may have been learned over many trials (potentially over many days) and, hence, may be difficult to unlearn. Alternatively, the agent could assume that the current context is different from that driving the original association and simply stop using the latter to make predictions, or discount the predictions. In this second scenario there is no actual modification of the original associative memory, however, it is possible that the agent will update it's context representation so that it does not repeat the predictions that gave rise to the original prediction error. Importantly, these two scenarios are not mutually

[11]Whereas there is much debate over the need to evoke the notion of unlearning to account for behavioral patterns in many episodic memory studies (e.g., Postman and Underwood, 1973), if we accept that synaptic plasticity is one key mechanism for learning, then there is ample evidence that synaptic connections can be both strengthened and weakened, which provides the foundation for discussing unlearning and what role it might play in reconsolidation.

exclusive and may occur together along a continuum, helping to explain the range of observed extinction and reconsolidation phenomena (Clem & Schiller, 2016).

Whether or not an agent continues to use a particular context representation to make predictions in the face of prediction errors will therefore determine if an existing association between the context and the current outcome is updated, or whether an entirely new memory is formed. For example, after an agent has learned to associate a conditioned stimulus and an aversive shock, the experimenter may attempt to extinguish that association by presenting the stimulus without the shock. As long as the agent assumes that the current context is the same as the one in which the original association was formed, then it will continue to make predictions of a shock (though potentially with slowly decreasing magnitudes) giving rise to negative prediction errors when a shock does not occur. This will lead to some degree of unlearning with each trial. If, however, upon experiencing strong negative prediction error early in the extinction process, the agent quickly stops making (or using) the prediction of a shock, there will be little or no prediction error to correct, and there will be little to no weakening or unlearning of the original CS-US association, even though the lack of a response linked to an expected shock might indicate some learning has taken place. Consequently, when tested sometime later, the agent will exhibit spontaneous recovery of the original learned response.

If unlearning is the desired outcome, and if active predictions are key to unlearning, then it is crucial that the agent not stop making predictions based on the reactivated context. That is, it is important for the agent to act on the assumption that it is in that context, until proven otherwise. Given this, it stands to reason that gradual extinction will give rise to unlearning of the original association that is resistant to spontaneous recovery. To test this hypothesis, Gershman and colleagues slowly decreased the proportion of trials in which a previously conditioned rat received a shock during extinction, thereby coaxing them into still making and utilizing predictions that the shock would appear. Consistent with the hypothesis, they found unlearning that was resistant to spontaneous recovery, but only in the gradual extinction group and not in rats that experienced

standard non-reinforced extinction (Gershman, Jones, Norman, Monfils, & Niv, 2013). Auchter etal. (2017) provided further support for the importance of maintaining predictions in the face of uncertainty by showing that varying the inter-trial interval (ITI) during extinction induces greater unlearning than fixed ITIs. Higher freezing by rats during the variable ITI condition suggests they were continuing to make predictions that they would receive shocks.

Given the range of timescales memory updating may involve, this prediction-error account could vary as a function of the memory system involved. For quickly-formed associations, perhaps those mediated preferentially by the hippocampus, we would expect to see fast unlearning in the cases where there was active prediction. Latent extinction provides an example of this kind of fast unlearning (cf. Goodman & Packard, 2019). For memories that take longer to form, such as in conditioning experiments or motor sequence learning, the agent may instead stop making predictions from the current context, occasioned by the assumption of a new context, and little to no unlearning will take place[12]. Importantly, even when no unlearning has taken place, behavior during the extinction period may mimic unlearning because the agent stops reacting on the basis of predicting a negative outcome. Consequently, any measured fear response will decline.

This brings us directly to the core of reconsolidation -- the necessary role of the reminder. In our framework (see Figure 2), reminders serve to reinstate the previous context, which allows the agent to make predictions of what to expect going forward. If this contextual reinstatement is followed by new information that involves only positive prediction errors because no specific items were predicted, such as those in the Hupbach studies where participants studied a new set of objects following a reminder (2007, 2008, 2009), negative prediction errors need not be resolved and new learning will involve binding new items to the reinstated context, creating a kind of hybrid

[12] It is also relevant that human learners require the aversive event to occur probabilistically with the CS. This implies that, just as in extinction, the learning process becomes more stable in conditions with active prediction. If the learner is actively retrieving and making predictions, this is akin to a testing effect whereby item retrieval becomes facilitated (Roediger and Karpicke, 2006). Learning can also take place from context-based predictions, themselves, such as in Smith et al. (2013).

between the reinstated and present context. If, instead, what follows the reminder primarily gives rise to negative prediction errors because the reminder also gave rise to specific item-level predictions, the agent will either modify or unlearn the disconfirmed associations to the current context (e.g., Kim et al., 2014) or, if prediction error levels are sufficiently high, shift their mental state to reflect and represent a new context that does not make the same error-laden predictions, and form a new association in that new context, leaving the initial association untouched. Figure 2 updates the original initial reconsolidation framework with the basic flow of this prediction-error and context update model of reconsolidation.

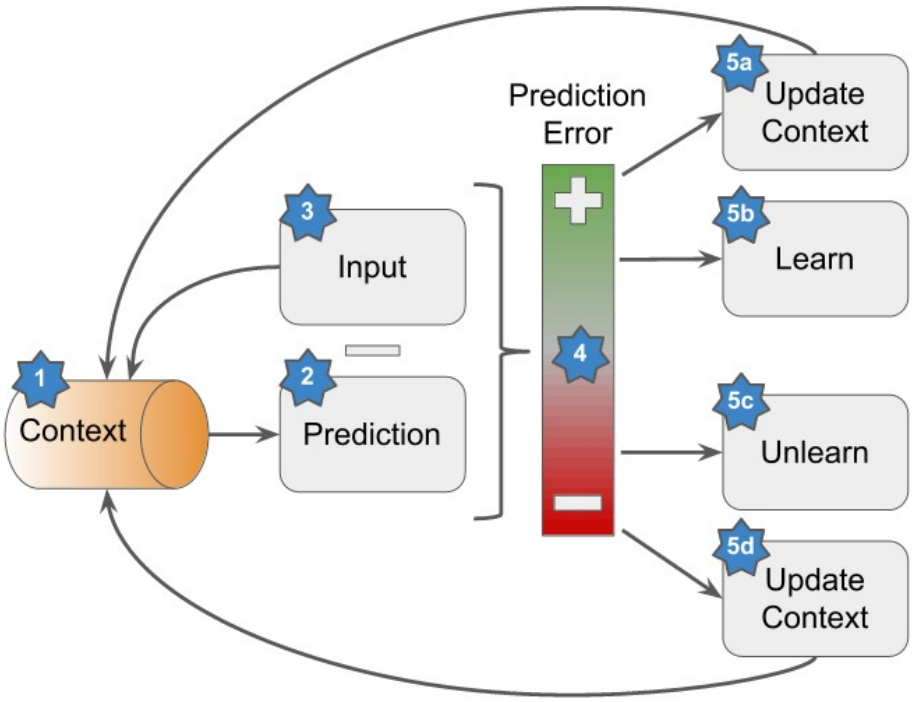

Figure 2: Model of Prediction-Error and Context Update in Reconsolidation.

1) The brain maintains a recency-weighted running average of experience, referred to as its "temporal context". 2) This context generates predictions via the previously-learned associations with that context. 3) New events provide input to the current context and 4) Any difference between the context-generated

prediction and the new input determines the prediction error (PE). 5) Depending on PE magnitude and direction, different actions could resolve it going forward. If PE is large (especially if large and negative), the choice may be to update to a new context that will no longer generate error-prone predictions (5a and 5d). If the PE is smaller, the best choice may be to maintain the current context, but modify the relevant context-based predictions by either learning (5b) or unlearning (5c) context--item associations.

With context reset in mind, we can briefly reconsider the reminder plus extinction (R+E) protocol (e.g., Schiller et al., 2010), which might diminish fear in a way that is resistant to spontaneous recovery. In these paradigms, the participant typically receives an isolated extinction trial, a CS without the US reinforcement, followed by a block of extinction trials between 10 minutes and a few hours later. Participants in the control group simply receive the block of extinction trials without the reminder. According to the context-based prediction error hypothesis, the negative prediction error instigated by this initial extinction trial cannot be resolved by a single reminder[13]. Under these conditions of uncertainty, hippocampal engagement, and hence context reactivation, increases (Rigoli et al., 2019), leading to stronger predictions, and the possibility of more unlearning during the subsequent extinction trials.

Whether or not the R+E protocol consistently works, any protocol that increases uncertainty, such as increased intertrial variability during extinction, should give rise to more reliance on retrieved rather than novel contexts (e.g., Auchter et al., 2017). Conversely, a reminder that reduces uncertainty, such as a set of extinction trials lasting more than 30 seconds, will bias participants into forming a novel context and evincing less robust extinction following the delay (Hu et al., 2018).

[13] This makes intuitive sense. Memory provides one piece of data (CS (or context) > US), the current situation provides another (CS (or context) > no US). Which to believe? This is the “reconsolidation window” during which training trials should strengthen the CS/Context-US link, and extinction trials should weaken it. Merlo et al. (2014) provide behavioral and molecular evidence for what they call a “limbo” state after fear memory reactivation where neither reconsolidation/strengthening nor extinction/weakening is yet engaged. Jezek et al. (2011) used teleportation in rats to create instantaneous context shifts and observed something like this limbo state directly - competing hippocampal place representations flickered back and forth for several theta cycles before ultimately settling on one “context”.

What remains unclear is how an agent determines when negative prediction error resulting from the omission of something expected, or positive prediction error resulting from the presence of something unexpected, is sufficient to signal that it is in a distinctly different context. Or, as in one recent computational account of reconsolidation, how does the agent infer the *latent cause* for the observed event (Gershman et al., 2017)? This recognition of a boundary between one context and another, or one event and another, is one of the core functions of the hippocampus (cf. Maurer & Nadel, in press). Recent work shows that as the agent's uncertainty about its situation increases, greater reliance is placed on the hippocampus (Rigoli etal. 2019). While we cannot go into detail here, there is considerable evidence suggesting that the locus coeruleus (LC) plays a critical role in the agent's recognition of and reaction to such boundaries. It has been suggested that the LC is the source of a "reset" signal that shifts the agent's internal context, making it possible to create a new contextual representation when sufficient PE indicates this is necessary (Bouret & Sara, 2005; Grella et al., 2019). It is presently not known where the LC gets its information about prediction error(s). Importantly, we cannot say whether what we have called positive and negative PE derive from different processes, and have a different impact on LC. Whatever the underlying neural story, this reset signal has many effects on the organism - within the domain of memory it initiates the context shift that actively prevents reconsolidation by changing brain state. At the same time, it facilitates spatial learning (Kempadoo et al., 2016) and the rapid formation of a new context representation (Wagatsuma etal., 2018).

The prediction-error theory outlined above mostly discusses reconsolidation at the systems level, without explicitly referencing the underlying neurobiology and the cellular and molecular processes that play a role in reconsolidation. Much could be said about the role of various neurotransmitters, such as dopamine and norepinephrine, or certain synaptic receptors, such as AMPA and NMDA, but that is beyond our scope. Nor have we discussed a potential role of sleep in reconsolidation, which is rendered plausible by the time dependence of the reconsolidation effect (see Simon etal. 2020, for a recent review). Such considerations might help us understand how synaptic tagging (Frey & Morris, 1997) modulates the effectiveness of the R+E reconsolidation protocol, and why

full effectiveness might depend on an optimal delay following the reminder. To our knowledge, no computational model has attempted to incorporate both systems and cellular/molecular mechanisms in the same theory, basically leaving the field with largely non-overlapping accounts at present. Bridging this gap remains an avenue for future theoretical work (Helfer et al., 2013).

CONCLUDING COMMENTS

In this chapter we hope to have provided both an overview of the domain of memory reconsolidation and a particular conceptual approach to it that we believe best captures its nuances. Those seeking more details are encouraged to read one or more of the numerous reviews that have been published within the past few years (eg., Kredlow et al., 2016; Gershman et al., 2017; Lee et al., 2017; Elsey et al., 2018; Wideman et al., 2018; Gisquet-Verrier & Riccio, 2018; Orederu & Schiller, 2019).

Our approach emphasizes the following points:

1. Memory reconsolidation is an integral part of the dynamics by which neural processes engage adaptively with the environment, allowing organisms to continuously update their models of the world.
2. It starts with the reactivation, by some cue(s), of a prior state of the brain - this is the “context” that existed when a particular event took place and a memory of that event formed.
3. This context-based memory provides the basis for predictions via learned associations from that context about what should happen next.
4. When something predicted fails to occur, or occurs in changed form, negative prediction error results.
5. When something unpredicted occurs, positive prediction error results.
6. The extent (and probably type) of prediction error determines which of two, possibly non-exclusive, outcomes will ensue: (a) the reactivated memory will be

adjusted to incorporate the change; or (b) a new event memory will be created, separate from the reactivated one, which will return, with possibly some minor loss of strength, to its pre-reactivated, condition.

These fundamental tenets link the approach to memory reconsolidation spelled out here to a number of other, related, literatures[14]. The fact is, any procedure that involves the reactivation of a memory necessarily engages the dynamics we discussed here. Which, of course, means reconsolidation is relevant to just about any procedure that claims to initiate and test for memory-formation, and/or its retrieval. The implications of this fact for the field of memory research remain to be fully appreciated.

[14] Some obvious ones: the testing effect, the retrieval interference effect, encoding specificity effects, the think/nothink paradigm and any study of retrieval specificity. See also Antony et al., (2017).

Bibliography

Agren, T., Björkstrand, J., & Fredrikson, M. (2017). Disruption of human fear reconsolidation using imaginal and in vivo extinction. *Behavioural brain research*, *319*, 9-15.

Alberini, C. M., & LeDoux, J. E. (2013). Memory reconsolidation. *Current Biology*, *23*(17), R746-R750.

Anagnostaras, S. G., Maren, S., & Fanselow, M. S. (1999). Temporally graded retrograde amnesia of contextual fear after hippocampal damage in rats: within-subjects examination. *Journal of Neuroscience*, 19(3), 1106-1114.

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

Auchter, A., Cormack, L. K., Niv, Y., Gonzalez-Lima, F., & Monfils, M. H. (2017). Reconsolidation-Extinction interactions in fear memory attenuation: the role of Inter-Trial interval variability. *Frontiers in behavioral neuroscience*, *11*, 2.

Barry, D. N., & Maguire, E. A. (2018). Remote memory and the hippocampus: a constructive critique. *Trends in cognitive sciences*.

Bavassi, L., Forcato, C., Fernández, R. S., De Pino, G., Pedreira, M. E., & Villarreal, M. F. (2019). Retrieval of retrained and reconsolidated memories are associated with a distinct neural network. *Scientific Reports*, 9(1), 784.

Bouret, S., & Sara, S. J. (2005). Network reset: a simplified overarching theory of locus coeruleus noradrenaline function. *Trends in neurosciences*, *28*(11), 574-582.

Bouton, M. E. (2002). Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biological psychiatry*, 52(10), 976-986.

Bouton, M.E. (2017) Extinction: Behavioral Mechanisms and Their Implications. In: Menzel, R. (ed.), Learning Theory and Behavior, Vol . 1 of Learning and Memory: A Comprehensive Reference, 2nd edition , Byrne, J.H. (ed.). pp. 61–83. Oxford: Academic Press.

Cahill, E. N., Wood, M. A., Everitt, B. J., & Milton, A. L. (2018). The role of prediction error and memory destabilization in extinction of cued-fear within the reconsolidation window. *Neuropsychopharmacology*, 1.

Clem, R. L., & Schiller, D. (2016). New learning and unlearning: strangers or accomplices in threat memory attenuation?. *Trends in Neurosciences*, 39(5), 340-351.

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, *210*(4466), 207-210.

Coons, E. E., & Miller, N. E. (1960). Conflict versus consolidation of memory traces to explain" retrograde amnesia" produced by ECS. *Journal of comparative and physiological Psychology*, *53*(6), 524.

Crowder, R. G. (1976). Principles of learning and memory, Lawrence Erlbaum Associates, New Jersey.

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

Dębiec, J., & LeDoux, J. E. (2004). Disruption of reconsolidation but not consolidation of auditory fear conditioning by noradrenergic blockade in the amygdala. *Neuroscience*, *129*(2), 267-272.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram?. *Annu. Rev. Psychol.*, *55*, 51-86.

Dunsmoor, J.E. & Kroes, M.C.W. Emotion-memory interactions: implications for the reconsolidation of negative memories. In “The Neuroscience of Enduring Change, R.Lane & L. Nadel (eds), in press.

Dunsmoor, J.E., Niv, Y., Daw, N. and Phelps, E.A. (2015). Rethinking extinction. Neuron, **88**, 47- 63.

Dunsmoor, J. E., Kroes, M. C., Li, J., Daw, N. D., Simpson, H. B., & Phelps, E. A. (2019). Role of human ventromedial prefrontal cortex in learning and recall of enhanced extinction. *Journal of Neuroscience*, *39*(17), 3264-3276.

Elsey, J. W., & Kindt, M. (2017). Breaking boundaries: optimizing reconsolidation-based interventions for strong and old memories. *Learning & Memory*, 24(9), 472-479.

Elsey, J. W., Van Ast, V. A., & Kindt, M. (2018). Human memory reconsolidation: A guiding framework and critical review of the evidence. *Psychological bulletin*.

Exton-McGuinness, M.T.J. and Milton, A.L. (2018) Reconsolidation blockade for the treatment of addiction: challenges, new targets, and opportunities. Learning and Memory, 25, 492-500.

Forcato, C., Burgos, V. L., Argibay, P. F., Molina, V. A., Pedreira, M. E., & Maldonado, H. (2007). Reconsolidation of declarative memory in humans. *Learning & Memory*, 14(4), 295-303.

Forcato, C., Rodríguez, M. L., Pedreira, M. E., & Maldonado, H. (2010). Reconsolidation in humans opens up declarative memory to the entrance of new information. *Neurobiology of learning and memory*, 93(1), 77-84.

Frankland, P. W., Bontempi, B., Talton, L. E., Kaczmarek, L., & Silva, A. J. (2004). The involvement of the anterior cingulate cortex in remote contextual fear memory. *Science*, *304*(5672), 881-883.

Frey, U., & Morris, R. G. (1997). Synaptic tagging and long-term potentiation. *Nature*, 385(6616), 533-536.

Genzel, L., & Wixted, J. T. (2017). Cellular and systems consolidation of declarative memory. In *Cognitive neuroscience of memory consolidation* (pp. 3-16). Springer, Cham.

Gerard, R. W. (1963). The material basis of memory. Journal of Verbal Learning and Verbal Behavior, 2, 22-33.

Gershman, S. J., Moore, C. D., Todd, M. T., Norman, K. A., & Sederberg, P. B. (2012). The successor representation and temporal context. *Neural Computation*, 24(6), 1553-1568.

Gershman, S. J., Jones, C. E., Norman, K. A., Monfils, M. H., & Niv, Y. (2013). Gradual extinction prevents the return of fear: implications for the discovery of state. *Frontiers in behavioral neuroscience*, 7, 164.

Gershman, S. J., Schapiro, A. C., Hupbach, A., & Norman, K. A. (2013). Neural context reinstatement predicts memory misattribution. *Journal of Neuroscience*, *33*(20), 8590-8595.

Gershman, S. J., Monfils, M. H., Norman, K. A., & Niv, Y. (2017). The computational nature of memory modification. *Elife*, *6*, e23763.

Gisquet-Verrier, P., & Riccio, D. C. (2018). Memory integration: an alternative to the consolidation/reconsolidation hypothesis. *Progress in neurobiology*.

Glickman, S. E. (1961). Perseverative neural processes and consolidation of the memory trace. *Psychological Bulletin*, *58*(3), 218.

Goode, T. D., Holloway-Erickson, C. M., & Maren, S. (2017). Extinction after fear memory reactivation fails to eliminate renewal in rats. *Neurobiology of learning and memory*, *142*, 41-47.

Goodman, J., & Packard, M. G. (2019). There is more than one kind of extinction learning. Frontiers in systems neuroscience, 13.

Grella, S. L., Neil, J. M., Edison, H. T., Strong, V. D., Odintsova, I. V., Walling, S. G., ... & Harley, C. W. (2019). Locus coeruleus phasic, but not tonic, activation initiates global remapping in a familiar environment. *Journal of Neuroscience*, *39*(3), 445-455.

Hardwicke, T. E., Taqi, M., & Shanks, D. R. (2016). Postretrieval new learning does not reliably induce human memory updating via reconsolidation. *Proceedings of the National Academy of Sciences*, *113*(19), 5206-5211.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Psychology Press.

Helfer, P., Shultz, T. R., Hardt, O., & Nader, K. (2013). A computational model of systems memory reconsolidation. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, 2512-2517. Austin, TX: Cognitive Science Society.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46(3), 269-299.

Hu, J., Wang, W., Homan, P., Wang, P., Zheng, X., & Schiller, D. (2018). Reminder duration determines threat memory modification in humans. *Scientific reports*, 8(1), 8848.

Hupbach, A., Gomez, R., Hardt, O., & Nadel, L. (2007). Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learning & Memory*, 14, 47–53.

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

Hupbach, A., Gomez, R., & Nadel, L. (2013). Episodic memory reconsolidation: An update. In *Memory reconsolidation* (pp. 233-247). Academic Press.

Hupbach, A., Hardt, O., Gomez, R., & Nadel, L. (2008). The dynamics of memory: Context-dependent updating. *Learning & Memory*, 15, 574–579.

Jezek, K., Henriksen, E. J., Treves, A., Moser, E. I., & Moser, M. B. (2011). Theta-paced flickering between place-cell maps in the hippocampus. *Nature*, *478*(7368), 246.

Kempadoo, K. A., Mosharov, E. V., Choi, S. J., Sulzer, D., & Kandel, E. R. (2016). Dopamine release from the locus coeruleus to the dorsal hippocampus promotes spatial learning and memory. *Proceedings of the National Academy of Sciences*, *113*(51), 14835-14840.

Kim, J. J., & Fanselow, M. S. (1992). Modality-specific retrograde amnesia of fear. *Science*, *256*(5057), 675-677.

Kim, G., Lewis-Peacock, J. A., Norman, K. A., & Turk-Browne, N. B. (2014). Pruning of memories by context-based prediction error. *Proceedings of the National Academy of Sciences*, 111(24), 8997-9002.

Kitamura, T., Ogawa, S.K., Roy, D.S., Okuyama, T., Morrissey, M.D., Smith, L.M. et al. (2017). Engrams and circuits crucial for systems consolidation of a memory, Science. 356, 73–78.

Kredlow, M. A., Unger, L. D., & Otto, M. W. (2016). Harnessing reconsolidation to weaken fear and appetitive memories: A meta-analysis of post-retrieval extinction effects. *Psychological Bulletin*, *142*(3), 314.

Kroes, M. C., Tendolkar, I., Van Wingen, G. A., Van Waarde, J. A., Strange, B. A., & Fernández, G. (2014). An electroconvulsive therapy procedure impairs reconsolidation of episodic memories in humans. *Nature Neuroscience*, *17*(2), 204.

Land, C., Bunsey, M., & Riccio, D. C. (2000). Anomalous properties of hippocampal lesion-induced retrograde amnesia. *Psychobiology*, *28*(4), 476-485.

Lane, R. and Nadel, L. Eds., The Neuroscience of Enduring Change, Oxford University Press, 2019 (in press).

Lee, J. L., Nader, K., & Schiller, D. (2017). An update on memory reconsolidation updating. *Trends in cognitive sciences*, *21*(7), 531-545.

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

Long, N. M., Lee, H., & Kuhl, B. A. (2016). Hippocampal mismatch signals are modulated by the strength of neural predictions and their similarity to outcomes. *Journal of Neuroscience*, *36*(50), 12677-12687.

McGaugh, J. L. (1966). Time-dependent processes in memory storage. Science, 153(3742), 1351–1358.

Merlo, E., Milton, A. L., Goozée, Z. Y., Theobald, D. E., & Everitt, B. J. (2014). Reconsolidation and extinction are dissociable and mutually exclusive processes: behavioral and molecular evidence. *Journal of Neuroscience*, *34*(7), 2422-2431.

Misanin, J. R., Miller, R. R., & Lewis, D. J. (1968). Retrograde amnesia produced by electroconvulsive shock after reactivation of a consolidated memory trace. Science, 160(3827), 554–555.

Monfils, M. H., Cowansage, K. K., Klann, E., & LeDoux, J. E. (2009). Extinction-reconsolidation boundaries: key to persistent attenuation of fear memories. *Science (New York, N.Y.)*, *324*(5929), 951–955. doi:10.1126/science.1167975

Monfils, M. H., & Holmes, E. A. (2018). Memory boundaries: opening a window inspired by reconsolidation to treat anxiety, trauma-related, and addiction disorders. *The Lancet Psychiatry*.

Moscovitch, M., Cabeza, R., Winocur, G., & Nadel, L. (2016). Episodic memory and beyond: the hippocampus and neocortex in transformation. *Annual review of Psychology*, 67, 105-134.

Moscovitch, M and Nadel, L. (2019) Sculpting memory: Reproduction and reconstruction of remote memory, *Trends in Cognitive Sciences.*

Müller, G., & Pilzecker, A. (1900). Experimental contributions to the theory of memory. *Z Psychol Z Angew Psychol*, *1*, 1-288.

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

Nadel, L., & O'keefe, J. (1974). The hippocampus in pieces and patches: an essay on modes of explanation in physiological psychology. In *Essays on the nervous system*(pp. 367-390). Clarendon Press Oxford.

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

O'keefe, J., & Nadel, L. (1978). The hippocampus as a cognitive map. Clarendon Press.

Orederu, T. and Schiller, D. The dynamic memory engram life cycle: Reactivation, destabilization, and reconsolidation. In Gazzaniga, M. The Cognitive Neurosciences, 2019, in press.

Phillips, R. G., & LeDoux, J. E. (1992). Differential contribution of amygdala and hippocampus to cued and contextual fear conditioning. *Behavioral neuroscience*, *106*(2), 274.

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

Postman, L., & Underwood, B.J. (1973). Critical issues in interference theory. *Memory and Cognition*, 1, 19-40.

Przybyslawski, J., & Sara, S. J. (1997). Reconsolidation of memory after its reactivation. *Behavioural Brain Research*, 84.

Riccio, D. C., Millin, P. M., & Bogart, A. R. (2006). Reconsolidation: A brief history, a retrieval view, and some recent issues. *Learning & Memory*, *13*(5), 536-544.

Rigoli, F., Michely, J., Friston, K. J., & Dolan, R. J. (2019). The role of the hippocampus in weighting expectations during inference under uncertainty. *Cortex*, *115*, 1-14.

Roediger, H. L. & Karpicke, J. D. (2006). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, 17(3), 249–255.

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

Schiller, D., Monfils, M. H., Raio, C. M., Johnson, D. C., LeDoux, J. E., & Phelps, E. A. (2010). Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature*, *463*(7277), 49.

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, 115(4), 893.

Sederberg, P. B., Gershman, S. J., Polyn, S. M., & Norman, K. A. (2011). Human memory reconsolidation can be explained using the temporal context model. *Psychonomic Bulletin & Review*, 18(3), 455-468.

Silva, M. B., & Soares, A. B. (2018). Reconsolidation of human motor memory: From boundary conditions to behavioral interventions—How far are we from clinical applications?. *Behavioural brain research*.

Simon, K.C., Gómez, R.L., and Nadel, L. (2020) Sleep's role in memory reconsolidation. *Current Opinion in Behavioral Sciences*, 33, 132-137.

Simon, K. C., Gómez, R. L., Nadel, L., & Scalf, P. E. (2017). Brain correlates of memory reconsolidation: a role for the TPJ. *Neurobiology of learning and memory*, 142, 154-161.

Sinclair, A. H., & Barense, M. D. (2018). Surprise and destabilize: prediction error influences episodic memory reconsolidation. *Learning & Memory*, *25*(8), 369-381.

Smith, T. A., Hasinski, A. E., and Sederberg, P. B. (2013). The Context Repetition Effect: Predicted Events are Remembered Better, Even When They Don't Happen. *Journal of Experimental Psychology: General, 142*(4), 1298–1308.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic bulletin & review*, *8*(2), 203-220.

Squire, L. R., Chace, P. M., & Slater, P. C. (1976). Retrograde amnesia following electroconvulsive therapy. *Nature*, *260*(5554), 775.

Squire, L. R., Cohen, N. J., & Nadel, L. (1984). The medial temporal region and memory consolidation: A new hypothesis. In H. Weingartner & E. S. Parker (Eds.), *Memory consolidation* (pp. 185–210). Hillsdale, NJ: Erlbaum..

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Tay, K. R., Flavell, C. R., Cassini, L., Wimber, M., & Lee, J. L. (2019). Postretrieval Relearning Strengthens Hippocampal Memories via Destabilization and Reconsolidation. *Journal of Neuroscience*, *39*(6), 1109-1118.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189.

Tolman, E. C. (1949). There is more than one kind of learning. *Psychological Review*, *56*(3), 144.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie canadienne*, *26*(1), 1.

Vallejo, A. G., Kroes, M. C., Rey, E., Acedo, M. V., Moratti, S., Fernández, G., & Strange, B. A. (2019). Propofol-induced deep sedation reduces emotional episodic memory reconsolidation in humans. *Science advances*, *5*(3), eaav3801.

Wagatsuma, A., Okuyama, T., Sun, C., Smith, L. M., Abe, K., & Tonegawa, S. (2018). Locus coeruleus input to hippocampal CA3 drives single-trial learning of a novel context. *Proceedings of the National Academy of Sciences*, *115*(2), E310-E316.

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

Walker, M. P., & Stickgold, R. (2016). Understanding the boundary conditions of memory reconsolidation. *Proceedings of the National Academy of Sciences*, *113*(28), E3991-E3992.

Wideman, C. E., Jardine, K. H., & Winters, B. D. (2018). Involvement of classical neurotransmitter systems in memory reconsolidation: Focus on destabilization. *Neurobiology of learning and memory*.

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.

Wymbs, Nicholas F., Amy J. Bastian, and Pablo A. Celnik. (2016) Motor skills are strengthened through reconsolidation. *Current Biology,* 26.3, 338-343

Zhang, J.J., Haubrich, J., Bernabo, M., Finnie, P.S.B. and Nader, K. (2018) Limits on lability: Boundaries of reconsolidation and the relationship to metaplasticity. Neurobiology of Learning and Memory, 154: 78-86. https://doi.org/10.1016/j.nlm.2018.02.018