Neuroscience and Biobehavioral Reviews 129 (2021) 367–388

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Replay in minds and machines

Lennart Wittkuhn [a,b,*], Samson Chien [a,b], Sam Hall-McMaster [a,b], Nicolas W. Schuck [a,b,*]

[a] Max Planck Research Group NeuroCode, Max Planck Institute for Human Development, Lentzeallee 94, D-14195 Berlin, Germany
[b] Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Lentzeallee 94, D-14195 Berlin, Germany

ARTICLE INFO

Keywords:
Replay
Reinforcement learning
Machine learning
Representation learning
Decision-making

ABSTRACT

Experience-related brain activity patterns reactivate during sleep, wakeful rest, and brief pauses from active behavior. In parallel, machine learning research has found that experience replay can lead to substantial performance improvements in artificial agents. Together, these lines of research suggest that replay has a variety of computational benefits for decision-making and learning. Here, we provide an overview of putative computational functions of replay as suggested by machine learning and neuroscientific research. We show that replay can lead to faster learning, less forgetting, reorganization or augmentation of experiences, and support planning and generalization. In addition, we highlight the benefits of reactivating abstracted internal representations rather than veridical memories, and discuss how replay could provide a mechanism to build internal representations that improve learning and decision-making.

Memory, planning and imagination are important aspects of intelligent behavior; they allow the mind to go beyond merely observing and reacting to its surroundings. But how the brain implements these functions, and how they could help to improve artificial intelligent agents, is not yet fully understood. In this review, we will provide an overview of one important candidate mechanism involved in memory, imagination and planning: Replay. The term replay is used to refer to a wide variety of mechanisms that relate to the reactivation of past memories. This reactivation has been observed in the brain and is commonly implemented in artificial agents. Replay occurs often, but not always, in sequential form, and mostly while the agent or animal is not interacting with its environment. This article asks *why* the brain and artificial agents might use replay.

1. Replay in the brain

Before we discuss the benefits of replay, we will give a brief historical overview of the subject from a neuroscientific point of view. In neuroscience, much research on memory, planning and imagination has focused on the hippocampus (e.g., Squire, 1992; Buckner, 2010). Early indications that the hippocampus gives rise to memory functioning came from studies of lesion patients (Scoville and Milner, 1957), and studies of rodent spatial navigation (O'Keefe and Nadel, 1978). Of particular importance were rodent recordings from hippocampal pyramidal neurons, known as *place cells*, that demonstrated spatial firing selectivity when animals navigated in a spatial environment (O'Keefe and Dostrovsky, 1971; O'Keefe et al., 1978). These place cells have since been regarded as a core neural substrate for a *cognitive map* (Tolman, 1948) of physical space that supports spatial navigation (O'Keefe and Nadel, 1974, 1978; Moser et al., 2008), as well as memory (Cohen and Eichenbaum, 1993; Redish and Touretzky, 1998).

It soon became clear that hippocampal place cells are also active when an animal is not engaged in a particular task, in line with theoretical proposals that a *reactivation* mechanism could support consolidation of recent memory traces into an aggregated memory store (Marr, 1971). Following early empirical support for this idea (Buzsáki, 1989; Pavlides and Winson, 1989), multi-unit recordings in rodents led to the discovery of *replay* – the finding that during periods of rest and sleep, hippocampal cells reactivate sequentially in fast bursts, as if retracing paths the animal had taken during wakefulness (Wilson and McNaughton, 1994; Skaggs and McNaughton, 1996; Kudrimoti et al., 1999; Nádasdy et al., 1999; Gerrard et al., 2001; Lee and Wilson, 2002; Louie and Wilson, 2001, for reviews of these earlier findings, see Redish, 1999; Sutherland and McNaughton, 2000). These observations were followed by a wealth of findings that established the now classic neuroscientific view of replay: replay is sequential, occurs during sleep or rest, reflects previous experience in spatial navigation and memory tasks, and happens on a temporally compressed timescale (for review, see e.g., Foster, 2017).

Over the following decades, much more became known about the

* Corresponding authors.
E-mail addresses: wittkuhn@mpib-berlin.mpg.de (L. Wittkuhn), schuck@mpib-berlin.mpg.de (N.W. Schuck).

https://doi.org/10.1016/j.neubiorev.2021.08.002
Received 22 March 2021; Received in revised form 19 July 2021; Accepted 1 August 2021
Available online 8 August 2021

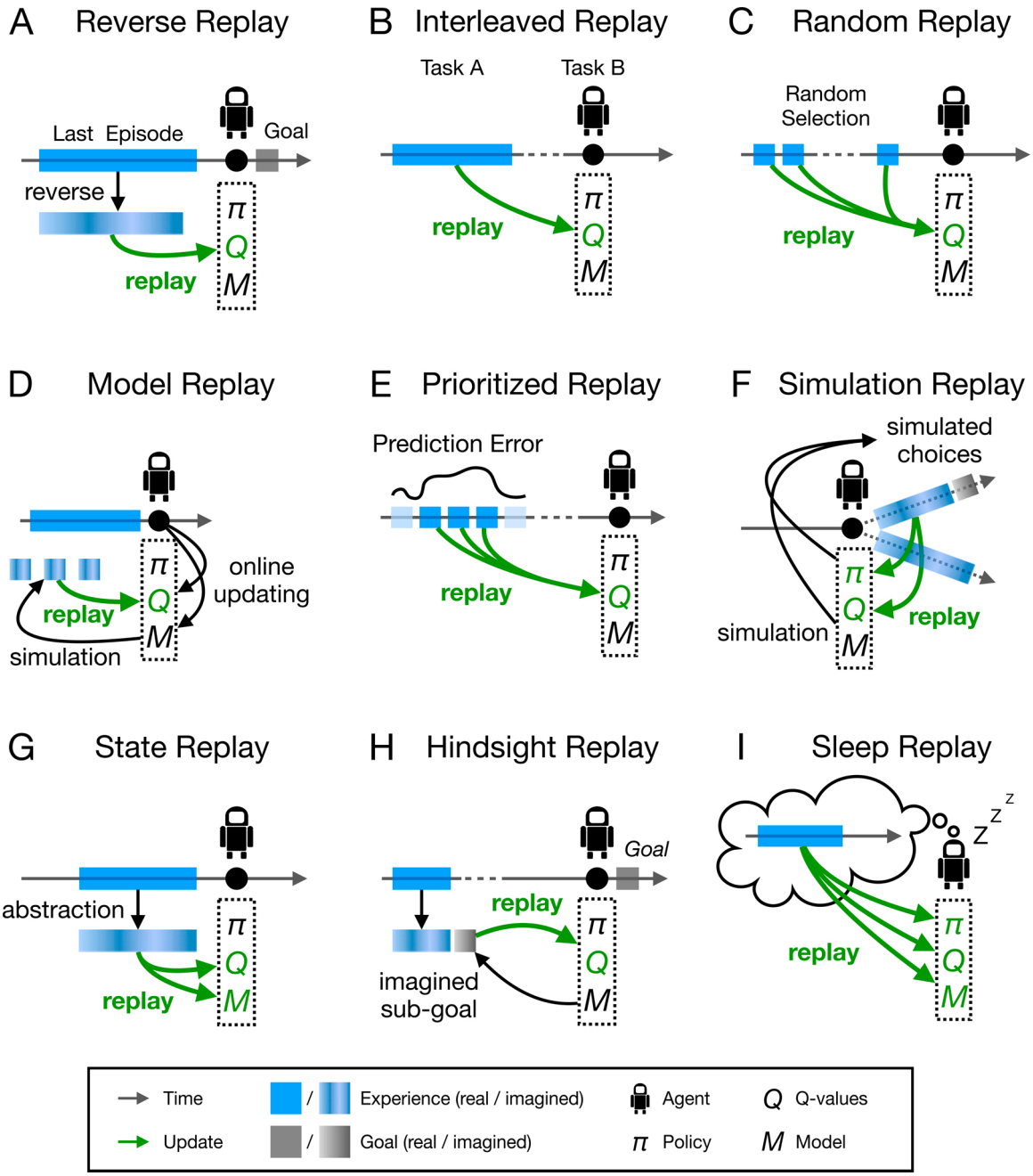

Fig. 1. Diversity of replay. Each panel outlines one example of one possible form of replay. In each case, we show an agent (black dot / robot) that stores a policy π, a value function Q and a model M see Box 1. Depicted is also the closest goal / reward (grey square), the relevant episode (blue bar), whether the episode is internally transformed (blue striped bar) and which aspect of the agent is updated through replay (green arrow). (A) A case in which encountering a goal triggers reverse replay. Reverse replay is then used to update the agent's value function, similar to Lin (1992). (B) Interleaved replay in which episodes from a previous task are replayed to prevent catastrophic forgetting, see McClelland et al. (1995). (C) Replay of uniformly / randomly selected individual transitions (Mnih et al., 2015). (D) An agent can also learn a model through online updating, and replay from the model during offline periods to update its value function (Sutton, 1991). (E) Episodes can be selected for replay based on the magnitude of prediction errors or other reward-related signals experienced during task performance (Schaul et al., 2015). (F) Instead of replaying previously experienced episodes, an agent can simulate possible episodes based on its model and policy in order to update the agent's value function ("offline policy evaluation", Sutton and Barto, 2018), or to plan the next actions (the policy π) at a choice point, without updating values. (G) Previously experienced episodes can also be abstracted before they are replayed, as done for instance when internal representations instead of observations are reactivated (Kapturowski et al., 2019). This can be used to e.g., update the agent's model and / or value function. (H) Agents can also insert imagined sub-goals into replayed episodes, in particular in order to leverage information from episodes in which the agent never reached the final goal. This is done in hindsight replay (Andrychowicz et al., 2017). (I) Replay occurring during sleep, i.e., while the agent is not engaging in any task at all. This is commonly observed in animals (Klinzing et al., 2019), but analogies from the ML literature are lacking because artificial agents do not sleep. Note, that this figure is not meant to be complete and merely illustrates some but not all aspects of the referenced algorithms.

biological aspects of replay, and many findings supported the idea that replay is important for memory. First, replay is commonly detected during brief, high-frequency oscillations called sharp wave-ripples (SWRs) (for review, see e.g., Buzsáki, 2015; Joo and Frank, 2018), which have also been found in human medial temporal lobe (MTL) (Bragin et al., 1999; Staba et al., 2002) and can be linked to memory consolidation during sleep, rest, and awake episodic memory retrieval (Axmacher et al., 2008; Staresina et al., 2015; Zhang et al., 2018; Helfrich et al., 2019; Norman et al., 2019; Vaz et al., 2019, 2020). A link to memory is also supported by findings showing that the selective disruption of SWRs during post-task rest slows learning in hippocampus-dependent spatial memory tasks (Girardeau et al., 2009; Ego-Stengel and Wilson, 2010; Jadhav et al., 2012), and that memory can be influenced by playing sounds during SWR events while an animal sleeps (Bendor and Wilson, 2012; Rothschild et al., 2016). Second, replay is much faster than wakeful experience, and this temporal compression is believed to induce the conditions that drive learning and the strengthening of memory traces through synaptic plasticity (Bliss and Collingridge, 1993; Magee and Johnston, 1997; King et al., 1999). Third, interactions between the hippocampus and prefrontal cortex (PFC) during replay events support the idea of consolidating reactivated memories in the brain (for reviews, see e.g., Tang and Jadhav, 2019; Zielinski et al., 2020).

Due to the fast and anatomically localized nature of the replay phenomenon, these insights were almost exclusively gained from invasive recordings in rodents and human patient populations. But existing studies focusing on non-invasive detection of replay in humans point to similar conclusions. Memory benefits of non-sequential reactivation during rest or sleep are well documented in humans (Staresina et al., 2013; Deuker et al., 2013; Tambini and Davachi, 2013; Tambini et al., 2010; Gruber et al., 2016). Memory consolidation in humans can also be biased by presenting learning-associated sensory cues, a technique known as targeted memory reactivation (TMR), during replay-associated sleep phases in humans (Oudiette and Paller, 2013; Lewis and Bendor, 2019). Recent progress in neuroimaging analyses have also been able to capture the sequentiality of fast replay events using magnetoencephalography (MEG) (Kurth-Nelson et al., 2016; Liu et al., 2021a) and functional magnetic resonance imaging (fMRI) (Schuck and Niv, 2019; Wittkuhn and Schuck, 2021). In combination, these findings have demonstrated that replay exists in a variety of species and support the idea that it reflects a consolidation process that strengthens memory associations (for reviews, see Sutherland and McNaughton, 2000; Rasch and Born, 2007; O'Neill et al., 2010; Diekelmann and Born, 2010; Carr et al., 2011; Zhang et al., 2017; Tambini and Davachi, 2019).

While the above findings have established a foundational knowledge of replay, our understanding of this phenomenon has undergone significant and continued change that sometimes challenged the classic picture of replay (for review, see e.g., Foster, 2017). For instance, replay seems to be significantly more frequent than initially thought, happening not only during sleep or rest but also during brief wakeful pauses from active behavior (Csicsvari et al., 2007; Davidson et al., 2009; Diba and Buzsáki, 2007; Foster and Wilson, 2006; Karlsson and Frank, 2009, for reviews of awake replay, see e.g., Carr et al., 2011; Tambini and Davachi, 2019). Replay-like sequential reactivation patterns occur at various speeds, from highly accelerated to much slower behavioral timescales (Deng et al., 2020; Denovellis et al., 2020; Tang et al., 2021). Recordings outside of the hippocampus have identified replay-like phenomena in a large number of other brain areas, including entorhinal (Ólafsdóttir et al., 2016; Ólafsdóttir et al., 2017; O'Neill et al., 2017; Trettel et al., 2019), prefrontal (Euston et al., 2007; Peyrache et al., 2009; Jadhav et al., 2016; Yu et al., 2018; Shin et al., 2019; Kaefer et al., 2020; Tang et al., 2021), visual and auditory sensory cortices (Ji and Wilson, 2006; Rothschild et al., 2016; Wittkuhn and Schuck, 2021), parietal cortex (Qin et al., 1997; Hoffman and McNaughton, 2002; Harvey et al., 2012), motor cortex (Ramanathan et al., 2015; Gulati et al., 2017), and ventral striatum (Lansink et al., 2009, 2008; Pennartz, 2004; Gomperts et al., 2015). Moreover, replay is not necessarily a faithful replication of previous behavioral sequences but can also reverse the order of experiences (Csicsvari et al., 2007; Davidson et al., 2009; Diba and Buzsáki, 2007; Foster and Wilson, 2006; Karlsson and Frank, 2009) or change the order of actual experiences according to a learned task rule (Liu et al., 2019a). It can also represent remote, non-local and never-experienced locations (Karlsson and Frank, 2009; Gupta et al., 2010; Ólafsdóttir et al., 2015), reflect non-spatial and partially observable task features (Schuck and Niv, 2019), and occur even after tasks without explicit memory requirements (Wittkuhn and Schuck, 2021). Collectively, these results suggest that replay (1) occurs during a variety of behavioral states, including rest, sleep and pausing, (2) occurs on a variety of time scales, (3) occurs in a variety of brain areas, and (4) does not only reflect previous experience, but is involved in a much broader range of cognitive functions than memory consolidation and spatial navigation alone.

Indeed, our understanding of hippocampal place cells, and the neural architecture underlying memory and spatial navigation more generally, has also evolved considerably. The "places" represented by hippocampal neurons are not exclusively determined by location in physical space, but can also incorporate other task-relevant aspects, such as sounds (Aronov et al., 2017) and time (MacDonald et al., 2011), but see also O'Keefe and Krupic (2021). Other studies have pointed out that the hippocampus may learn and predict transitions between states in the environment (Gaussier et al., 2002) or encode representations that are predictive of future locations, so called successor representations (SRs), that can be used for reinforcement learning (RL) (Stachenfeld et al., 2017), and that grid-like patterns in the entorhinal cortex and ventromedial PFC may represent coordinates of a non-spatial space (Constantinescu et al., 2016). Thus, today the cognitive map in the hippocampal-entorhinal system is often thought to represent relationships of locations and events beyond physical space, from conceptual knowledge to social cognition (for reviews and perspectives, see Khamassi and Humphries, 2012; Kaplan et al., 2017; Epstein et al., 2017; Schafer and Schiller, 2018; Behrens et al., 2018; Bellmund et al., 2018; Peer et al., 2020; Bottini and Doeller, 2020; Spiers, 2020). Map-like representations also exist beyond the hippocampus, most notably in the medial entorhinal cortex (Hafting et al., 2005; Fyhn et al., 2007; Høydal et al., 2019), and in prefrontal and orbitofrontal cortex (OFC) (Wilson et al., 2014; Schuck et al., 2016; Constantinescu et al., 2016, for a review, see e.g., Schuck et al., 2018). These findings may have important implications for our understanding of the nature of replayed representations, and suggest a mechanism that is much broader than a mere recapitulation of past observations in the hippocampus.

How can such a diverse set of findings about replay in the hippocampus and the rest of the brain be integrated? We argue that insights into this question can be gained by considering the machine learning (ML) literature, where "experience replay" was introduced in the early 90s (Lin, 1991). More recently, experience replay has become particularly popular after its importance for training deep neural networks (DNNs) to play Atari video games became clear (e.g., Mnih et al., 2013, 2015; Hessel et al., 2018). This led experience replay to rise to prominence as a crucial ingredient in building human-level intelligence in artificial agents (Kumaran et al., 2016). Despite the conceptual similarity of biological and artificial replay, research on this subject in neuroscience and ML has progressed largely in parallel. Here, we aim to connect insights from both research fields and review computational perspectives taken in ML on the replay phenomenon. Our goal is to highlight the diversity of possible computational and cognitive functions that might be served by a replay mechanism and attempt to answer the question of *why* agents would replay in the first place. Fig. 1 provides a non-exhaustive overview of different forms of replay, which differ in which experiences are selected for replay and in how replayed information affects the subsequent behavior of an agent. We will discuss these different instantiations of replay below.

Box 1
What is reinforcement learning?

Reinforcement learning (RL) theory provides a formal framework to describe how agents learn to optimize their behavior through interactions with an environment that yields rewards or punishments (Sutton and Barto, 2018). The agent-environment interaction is modelled as a Markov decision process (MDP), which consists of (1) an **environment**, described by of a set of states S, (2) a set of **actions** A available to the agent, (3) a **state transition function** $M(s_t, a_t, s_{t+1})$, also called a model, reflecting the probabilities of moving from state s_t to the next state s_{t+1} after taking action a_t, and (4) a **reward function** $R(s_t, a_t, s_{t+1})$ that maps each [state, action, next-state]-triplet to a scalar reinforcement signal r. MDPs can have continuous state or action spaces, although most applications consider finite and discrete cases.

In MDPs, the agent-environment interaction is assumed to be Markovian with respect to reward and state, which means that the state and reward at the next time point $t+1$ depend only on the state and action of the current time point t, but not on any states or actions before. The current state s_t therefore contains all relevant information from the previous history to determine the next state after an action has been performed. In brief, this means that we can think of learning from trial-and-error as the following process: the agent represents the current state of the environment, s_t and then performs an action a_t. The action will affect the environment, changing the agent's state from s_t to s_{t+1} as described by the state transition function M, and potentially yield a reward, as described by the reward function R. The agent's goal is to always perform the actions that maximize return – the expected (discounted) sum of total rewards over the course of its interaction with the environment. In value-based approaches, the agent learns values that estimate the return, and then implements a **policy** π that maximizes the values. One popular approach is to estimate values with so-called temporal difference (TD) learning, using a Q-learning algorithm (Watkins and Dayan, 1992):

$$Q_{k+1}\left(s_t, a_t\right) = Q_k\left(s_t, a_t\right) + \alpha\left[r_t + \gamma \max_{a_{t+1}} Q_k\left(s_{t+1}, a_{t+1}\right) - Q_k\left(s_t, a_t\right)\right] \quad (1)$$

where $\gamma \in [0, 1]$, the discount factor, attenuates the influence of distal rewards, and $\alpha \in [0, 1]$ is a learning rate. Based on the Bellman equation, Q-learning estimates the discounted sum of future rewards in an iterative bootstrapping process that involves the current and future Q-value. Notably, because the algorithm uses the value of the best action on the next step, rather than the value of the action that was actually performed, it is a so-called off-policy approach. Q-learning does not require a transition model of the environment, and hence is a model-free method. One important question that is not addressed by the framework of RL is what information about the environment is encoded in an agent's internal states. It is important to realize that agents mostly do not have a way to simply know the objective, "true" state of the environment, but rather must infer that state from their observations. An agent's internal state representation therefore may not be equivalent to the true state, which, as we will discuss in Section 2.6, has many repercussions. The agent could simply take its sensory data to be the states, but this is not sufficient for many tasks; rather the agent needs to supplement its internal state representations with non-observable information (e.g., Wilson et al., 2014; Schuck et al., 2018).

2. Computational benefits of replay

Replay has become a highly studied aspect of artificial agents. But why do machines need replay, and do animals and machines have the same reasons to employ this process? In the following sections, we will compare the roles of replay in both biological and artificial agents, and distill the most significant benefits of replay.

Before we begin, we would like to point out some significant aspects in which the concept of replay differs between ML and neuroscience. First, neuroscience emphasizes the sequential and often accelerated nature of replay (Genzel et al., 2020). In ML, in contrast, some methods focus on replaying sets of individual transitions (e.g., Mnih et al., 2015), rather than sequences (but see e.g., Hausknecht and Stone, 2015). The issue of replay speed has not been a major consideration in ML, as artificial agents are not bound to physical interaction with the environment and the timescales of biology. A second difference is that understanding the distinction between sleep replay and replay during wakeful pauses from active behavior has a prominence in neuroscience (and is covered extensively in previous reviews; see e.g., Findlay et al., 2021; Klinzing et al., 2019) that is not equivalently mirrored in ML research. While the contrast between sleep and wakefulness is a theme that has inspired ML research conceptually (see e.g., Hinton et al., 1995), the mere fact that artificial agents do not "sleep" in the way that biological agents do, makes it practically impossible to investigate those differences in artificial agents. While biological agents have several "modes" in which they are "off-policy" (sleeping, resting, pausing, mind-wandering, etc.), to our knowledge no comparable distinctions have been made for artificial agents. Third, ML researchers often distinguish between experience replay, which corresponds to sampling experiences from a memory buffer, and model-based methods, in which the agent internally generates new experiences from a learned model of the environment. While these model-based methods involve an offline reactivation process, they are not always called replay in the ML literature, but are often referred to as planning instead. In neuroscience, in contrast, many sequential reactivation phenomena are universally referred to as replay, whereas planning is considered to be one of the cognitive processes that might be supported by a replay mechanism.

Similar to previous work (e.g., Foster and Knierim, 2012; Cazé et al., 2018; Momennejad, 2020), our review focuses on the frameworks of RL (Sutton and Barto, 2018) and neural networks. The formalism of RL allows parallels to be drawn between reactivation of neural patterns in biological agents and replay of task states in artificial agents. The RL framework considers agents that learn from interactions with their environment and thereby gather experiences one at a time. RL techniques are designed to learn from experience gradually, through trial-and-error, using every new experience immediately to adjust the agent's knowledge about the task. This has the benefit of accruing knowledge without delay, while integrating information over all experiences gained so far, rather than using just the most recent experience to make decisions. Typically, small adjustments are made to the agent's knowledge with each new experience because large updates risk overwriting the effects of earlier learning and can limit generalization. Box 1 describes the fundamental aspects of RL. Next to RL, we will also draw on insights from (supervised) deep learning (for overviews, see e.g., LeCun et al., 2015; McClelland and Botvinick, 2020) and the successful combination of the two approaches, deep RL (Mnih et al., 2015, see Tesauro, 1995 for an earlier integration).

Which benefits can an agent obtain from using replay? In the next sections, we will discuss five potential computational functions of replay: increasing speed and data efficiency of learning, reducing forgetting, reorganizing experiences, planning, and generalization. We do not consider these functions to be entirely separable. We distinguish them because they each offer a unique perspective on what an intelligent agent, biological or artificial, stands to gain from replaying past

experiences. This perspective also sheds light on why replay can have different properties in different study contexts, which have found replay to be sometimes backward and sometimes forward, or in some cases to occur immediately and in other cases long after the experience was acquired.

In addition to the topics above, we will consider one underexplored aspect of replay: whether replay reflects sensory memories, or past internal representations, and whether replay may also be involved in *shaping* internal representations as well. We hypothesize that the content and function of replay is determined by its interplay with the agent's current representation of the task and the representational demands of the task at hand, a notion which has recently received some computational (Russek et al., 2017; Caselles-Dupré et al., 2019; Momennejad, 2020) as well as empirical support (Schuck and Niv, 2019). In this view, replay can be understood not only as a phenomenon that retrieves relational information stored in a cognitive map, but also as a process that changes relational information and internal state representations of an agent (see Box 1 for a definition of state representations).

Each of the sections will be organized as follows. First, we will state a computational problem that any learning agent will be faced with. Then we discuss how this problem has been approached in ML using replay, highlighting both theoretical and empirical results. Finally, we will discuss empirical findings from the neuroscience literature that support a particular ML proposal, or suggest alternative mechanisms.

2.1. Faster learning and data efficiency

The gradual approach to learning in RL has many benefits, but it results in very slow learning that may need thousands of iterations to achieve the optimal policy. Even worse, the slowness of learning grows exponentially with the number of states in the task environment, a phenomenon known as the "curse of dimensionality" (Bellman, 1957). To be a feasible approach to learning in complex and changing environments, gradual methods must therefore be complemented by mechanisms that will speed up learning without sacrificing the benefits of immediate knowledge acquisition and stable long-term memory. In this light, the idea of recapitulating previous experiences seems particularly appealing for machines, because it is easy and cheap for artificial agents to relearn from past experience that is retrieved from a memory buffer.

The brain arguably faces a similar computational challenge. Humans, and other animals, often have to learn directly from the outcomes of their decisions. Yet, repeating errors can pose actual risks, which limits the usefulness of exclusively relying on a slow, trial-and-error-based learning mechanism. More generally, the number of experiences that is acquired with a particular situation in a lifetime is quite limited in relation to the complexity of the environments and the brain, which contains approximately 10^{14} synapses (Tang et al., 2001). In order to make thousands of gradual adjustments to each of these synapses, the ability to reuse experience efficiently is paramount. Replay might be one mechanism to do just that.

2.1.1. Replay can speed-up gradual learning from experience and support temporal credit assignment

In the RL literature, "experience replay" was initially introduced to address the issues of slow learning and data inefficiency (Lin, 1991, 1992, 1993). In his seminal paper, Lin (1992) wrote that "[…*Q*-learning algorithms […] are inefficient in that experiences obtained by trial-and-error are utilized to adjust the networks only once and then thrown away. […] Experiences should be reused in an effective way." (p. 299). Lin (1992) proposed that experiences can be used to update knowledge in a dual fashion; (1) immediately when experiences are acquired, and (2) at later time points, after experience itself may have long passed. Specifically, Lin (1992) proposed replaying full sequences of experiences, starting from an initial state to a final state, in backward order, and learning from these experiences, as if they were real. Lin (1992) then showed that this is a more efficient use of data that accelerates learning of an RL agent. In line with these ideas, many others have since emphasized the computational benefit of replay for maximizing data efficiency and the speed of learning (for reviews, see e.g., Hassabis et al., 2017; Kumaran et al., 2016).

There are several reasons why replay can help learning. In the real world, outcomes are often only obtained after a long sequence of events and actions but agents still need to know how to behave at the start of the sequence, as for instance, in a chess game. This problem is known in RL as the *temporal credit assignment problem* (Minsky, 1961) and replay may help to solve it. The early work by Lin (1991, 1992, 1993) pointed out that replay could help an agent to remember the sequence of previous states and actions that led to a given outcome, and assign credit for the reward to the sequence of states and actions that preceded it. This also explains why sequential replay may proceed in backward order (Lin, 1992). Another aspect is that as the agent's knowledge of the rewards becomes better with time, outcomes in the past should be re-evaluated in light of this updated knowledge (van Seijen and Sutton, 2015). Replay could serve this function by retrieving past rewards which can then be compared to current value estimates.

Several neuroscientific studies suggest an important role of replay in speeding up learning in biological agents too. First, studies in rodents reported increases in SWR-associated reactivation following initial learning in novel environments (Cheng and Frank, 2008; Eschenko et al., 2008; O'Neill et al., 2008; van de Ven et al., 2016; Tang et al., 2017), when an acceleration of learning from replay might be most beneficial. Second, several studies reported that it requires only a few experiences in a novel environment for replay to occur, and that it can be detected already during the awake state immediately after behavior (Foster and Wilson, 2006), but see Jackson et al. (2006). Third, disrupting replay-related SWRs during awake rest in rodents slows learning in a spatial navigation task (Jadhav et al., 2012).

Previous research has also suggested that backward replay reflects learning through temporal credit assignment in the brain. First, awake backward replay has indeed been frequently observed, where rewarded spatial trajectories of an animal are replayed in reverse order (Diba and Buzsáki, 2007; Foster and Wilson, 2006; Singer and Frank, 2009), and the frequency of awake backward (but not forward) replay is modulated by the change in reward magnitude (Ambrose et al., 2016; Liu et al., 2019a). Second, the rate of backward replay was observed to be more frequent in novel compared to familiar environments (Foster and Wilson, 2006; Singer and Frank, 2009) and to decrease its bias to reflect previous paths to the goal location as a function of learning (Shin et al., 2019). This could suggest that the relevant trajectory has been learned and does not need to be reinforced through replay anymore (Foster and Knierim, 2012). Interestingly, Cazé et al. (2018) have shown that in particular model-based replay will also decrease its tendency to replay paths to the goal with learning, while changes in forward planning (Johnson and Redish, 2007) might stem from a model-free process. In a task setting with a stable goal, the replay buffer of a model-free learner will increasingly accumulate rewarded episodes while a model-based learner draws on a learned model to sample episodes in a more balanced fashion. The learning-related changes discussed here might therefore reflect a shift from a model-free to a model-based process with learning – although further data will be needed, and model-based and model-free replay might be difficult to disentangle experimentally (Khamassi and Girard, 2020). The third line of support comes from computational work that shows how backward replay can strengthen forward synaptic pathways through spike timing dependent plasticity (STDP) (Haga and Fukai, 2018) and thus support forward replay during sleep and active behavior (Johnson and Redish, 2007; Pfeiffer and Foster, 2013; Wikenheiser and Redish, 2015b, 2013). Fourth, further evidence for the role of replay in assigning credit is provided by findings that show replay is coordinated with subcortical activation of brain areas related to processing reward (Lansink et al., 2009; Pennartz, 2004; Gomperts et al., 2015), which could convey reward signals to other brain regions like the hippocampus. Finally, in a recent MEG study in humans, backward replay following reward receipt was found to be related to non-local learning of task sequences leading to the reward (Liu et al., 2021b). In summary, existing empirical studies

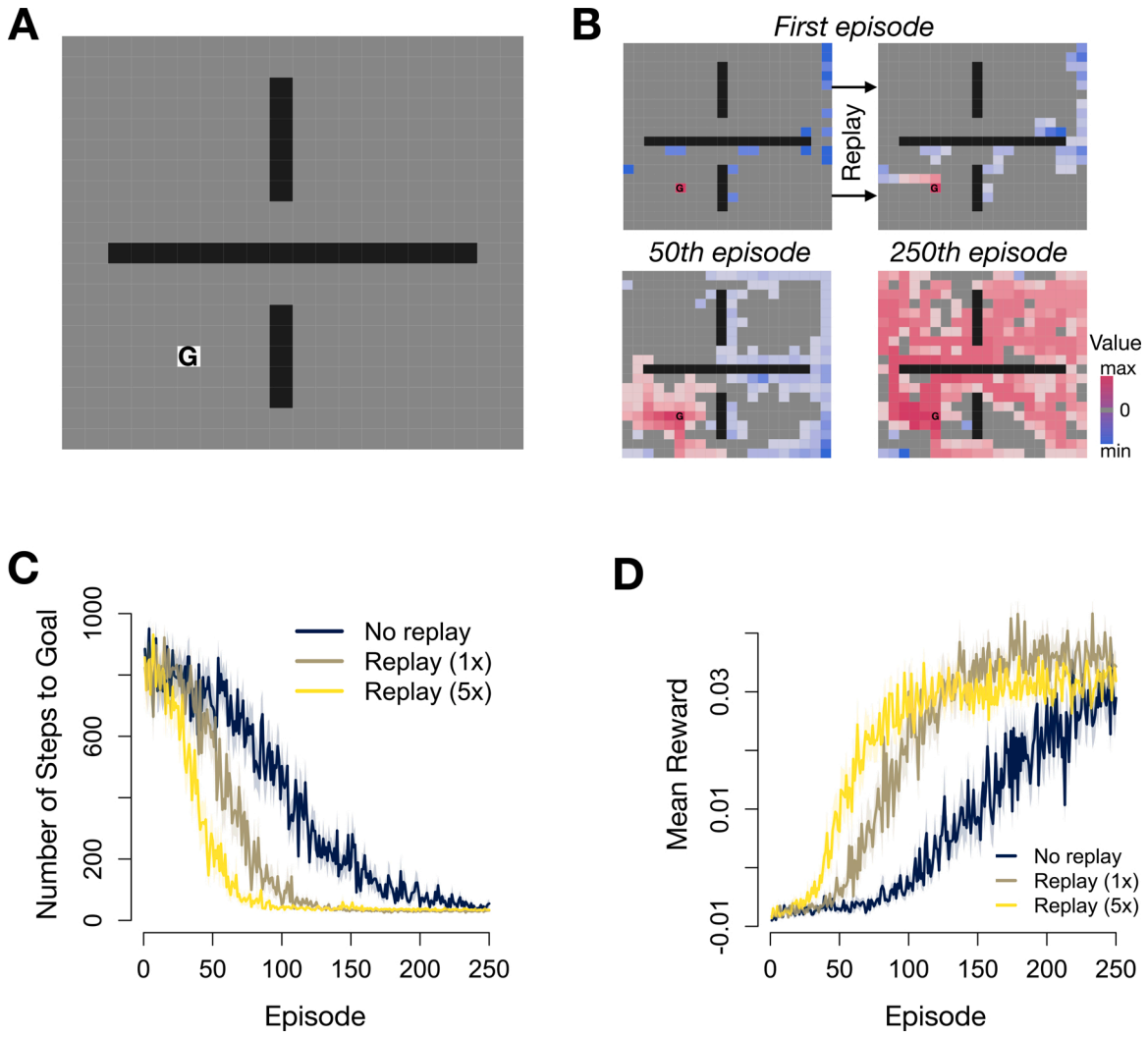

Fig. 2. Replay speeds up learning to navigate to a goal in a grid world. (A) Square environment ("grid world") with 20 × 20 tiles (shown in gray) that contains several walls (black tiles) and one goal location (white tile labelled "G") that contains a reward. At the beginning of each episode, an RL agent is placed in a random location and can move into one of four cardinal directions (up, down, left, right). The agent receives no reward for moving, a small negative reward for bumping into a wall (− 0.1), and a reward of 1 when arriving at the goal location. An episode is terminated once the agent reaches the goal location or a maximum of allowed steps per episode set to 1000. (B) Illustration of the learned value function after the first episode of experience (top left), following replay of the first episode (top right), and after the 50th and 250th episode (bottom). Colors indicate the values of locations from smallest (blue) to highest (red). Values are under the best possible policy, which is assuming that the agent would perform the value-maximizing action in each location. The increasing prevalence of red tiles after 250 episodes therefore reflects that after training the agent has learned a policy for most locations that will avoid any collisions with the wall and reach the goal within the maximum number of allowed steps. Color mapping is scaled for each plot and values smaller than 0.1 are shown as gray tiles. (C) Number of steps (y-axis) needed by the RL agent in each consecutive episode (x-axis) to reach the goal location when using no replay (blue line) between episodes, or when replaying the previous episode in backward order once (brown line) or five times (yellow line). (D) Mean reward (y-axis) achieved by the RL agent in each consecutive episode (x-axis). Colors as in (c). The computer code for the simulations is publicly available at https://github.com/nschuck/replaysim-wittkuhn-etal2021.

support the idea that awake backward replay supports temporal credit assignment by retrieving states that led to the outcome, accelerating learning for cases in which a long delay between rewards and actions must be encoded.

2.1.2. *An example simulation of backward replay*

Fig. 2 provides an illustration of how backward replay of full sequences works in the context of RL. We consider an RL agent navigating in a square environment with 20 × 20 tiles that contains several walls and one goal location with a reward (see Fig. 2A). The agent can move into one of the four cardinal directions (up, down, left, right). A small negative reward is given for bumping into a wall (−0.1), and a reward of 1 when arriving at the goal location. Otherwise no rewards are provided. The best policy in this case is to navigate to the reward with as little steps as possible, avoiding the wall. This is a well-known "grid world" problem that can be solved using RL, but might be painfully slow without replay. For illustrative purposes, we use the off-policy, model-free *Q*-learning algorithm described in Equation 1 in Box 1. The learning rate α, temperature τ and discounting factor γ were arbitrarily set to 0.3, 1 and 0.99 for the purposes of this illustration.

Algorithm 1 Q-learning with Replay

Initialize $Q(s, a) \leftarrow 0$ for all states s, actions a and parameters $\{\alpha, \gamma, \tau\}$
for $episode = 1, 2, \ldots,$ number of episodes **do**
  $s \leftarrow$ random start position farther than 10 steps from goal
  for $step = 1, 2, \ldots,$ maximum walk length **do**
    $s \leftarrow$ state at current step
    $a \leftarrow$ softmax(s, Q) with temperature τ
    Execute a, obtain next state s' and reward r
    $Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \max_{a'} Q(s', a') - Q(s, a)]$
    if goal reached **then**
      for t = step-1, step-2, ... beginning of episode **do**
        $s, a, r \leftarrow$ state, action, reward at t
        $s', a' \leftarrow$ state, action at $t + 1$
        $Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma Q(s', a') - Q(s, a)]$
      end for
      End episode and continue to next
    end if
  end for
end for

The algorithm is described in Algorithm 1. Briefly, in each episode, the agent starts in a random position and navigates until it has found the reward or the maximum search time has elapsed. The starting locations varied randomly, although start location distances to the reward location were constrained to lie at least 10 tiles away from the goal location (in order to avoid episodes which were too simple). If the agent found the reward, it internally traversed backwards through the sequence of states, actions and rewards until the beginning of the episode, updating its Q-value at each step.

The blue lines in Fig. 2C–D show the number of steps the agent needs to navigate to the goal location; about 250 episodes are needed before the the agent quickly finds the goal location from a new start position. But the speed of learning increases when we supply the agent with a simple replay mechanism described in Table 1, as can be seen in Fig. 2C–D (brown and yellow lines). Adding replay reduced the number of interactions needed to achieve ceiling performance to less than half of what was observed without replay. Note that the choice to replay the full sequence of states, actions and rewards between the start location and the goal location is not without consequence, and a variety of different definitions of what constitutes an episode are common in RL and neuroscience (see Box 2). There are multiple ways to instantiate replay in an RL agent, and the illustration in Fig. 2 only serves as a basic introduction to computational replay (see Fig. 1).

2.2. Less forgetting

Increasing the speed of learning is an important computational benefit of replay, but not the only one. Replay may also help to reduce forgetting. The problem of forgetting arises because many statistical learning mechanisms were built under the assumption that the agent encounters its environment entirely at random, and therefore can learn from examples that are independent and identically distributed (i.i.d.). Yet, experiences in real life are often not "i.i.d.". First, we typically experience the world as a sequence of related events, chunked in time. Second, some events are much rarer than others, partly because of the way we interact with the environment. These temporal auto-correlations and uneven distributions of events can be an important obstacle for learning. Why does this pose a computational challenge for gradual learning algorithms? Gradual learning mechanisms are designed to integrate experiences over longer periods, but they emphasize the most recent experience. This can cause the agent to forget about important past experiences that were not re-experienced for a long time. This problem is particularly apparent in a supervised learning setting in which neural networks that rely on stochastic gradient descent (SGD) for learning engage in two tasks in a *blocked* manner. RL-based networks also struggle with this phenomenon (Atkinson et al., 2021). If a DNN, for instance, is first trained to perform a task A and subsequently trained with another task B, performance on task A drops dramatically, as if the network forgot how to solve A. In other words: learning task B interfered with what was learned about task A. This problem is known as catastrophic forgetting, or catastrophic interference, and has long been recognized as a major problem in the ML field (McCloskey and Cohen, 1989; Ratcliff, 1990; French, 1999; Hassabis and Maguire, 2007; Kumaran et al., 2016; Parisi et al., 2019). Catastrophic forgetting is one of the main reasons why artificial agents can usually learn a single task quite well but subsequent training on a different task results in poor performance on the previously learned task. This prevents the agent from achieving competencies across multiple tasks, which comes relatively easily to humans. Catastrophic interference can also be understood as an issue threatening the stability of a cognitive map representation (Gupta et al., 2010; McClelland et al., 1995; O'Reilly and McClelland, 1994).

2.2.1. Replay can prevent overwriting of previous experiences

A potential solution to the computational problem of catastrophic interference is *interleaved learning* where new experiences are interleaved with existing knowledge to reconcile competing memory representations (McClelland et al., 1995). This influential idea, rooted in the complementary learning systems (CLS) theory (McClelland et al., 1995; O'Reilly et al., 2014; Schapiro et al., 2017), also suggests that replay may be the mechanism that agents can use to "mentally" interleave past with present experience. While the DNN tunes its connection weights to solve task A, the experienced episodes are stored in a memory buffer. During learning of task B, previous experience with task A is integrated during offline periods via a replay-like mechanism, preventing forgetting, and allowing the agent to perform well on both tasks. Shin et al. (2017), for instance, proposed an approach that learns a generative model based on experience with one classification task A. When switching to an independent classification task B, the system is retrained using a combination of new task data and fictitious sequences from the generative model, resulting in rapid generalization to the new task with little performance loss. Similarly, implementing replay in this way in DNNs can help to overcome performance deficits in incremental task learning scenarios and continuous task environments (van de Ven and Tolias, 2018; van de Ven et al., 2020). Of note, non-sequential replay has been shown to become necessary in an artificial neural network (ANN)

Box 2
What is replayed?

Replay is generally thought to represent previous experience. How is this experience stored in artificial and biological agents? In artificial agents, an experience at time t, e_t, is commonly defined as a quadruple consisting of the state s_t, the taken action a_t, the reward r_t received after taking action a_t in state s_t, and the next state s_{t+1}, together $e_t = (s_t, a_t, r_t, s_{t+1})$, effectively describing a single transition between two states as the atomic unit of an artificial replay event. Although in some cases individual transitions are replayed, such as in the Deep Q-Network (DQN) approach by Mnih et al. (2015) where the states S consisted of pre-processed versions of Atari pixel frames, other work uses sequential replay of past states (e.g., the early version of experience replay by Lin (1992), see Fig. 2, or replay in recurrent neural networks (RNNs), see e.g., Hausknecht and Stone, 2015; Kapturowski et al., 2019). Interestingly, replay techniques in ML increasingly reactivate internal state representations, rather than observations like pixel values (Hayes et al., 2021). We will discuss this aspect in more detail in Section 2.6 on representation learning.

What constitutes a replayed experience is more difficult to answer for biological agents. Unlike in artificial agents, replay in biological agents is thought to be sequential (Genzel et al., 2020), and typically involves hippocampal place cells that represent locations in a spatial environment, akin to previously experienced trajectories of locations. However, hippocampal cells appear to be quite flexible in encoding task-relevant information other than physical space, for instance sounds (Aronov et al., 2017), trial history (Wood et al., 2000; Sun et al., 2020) or abstract task states (Schuck and Niv, 2019). Moreover, a prominent theme in neuroscience emphasizes that the brain segments continuous experience into representations of distinct neural states that transition at event boundaries or shifts in context (for reviews, see e.g., Bird, 2020; Brunec et al., 2018; Maurer and Nadel, 2021; Richmond and Zacks, 2017; Shin and DuBrow, 2020). To complicate matters, this process might also happen retroactively, i.e., *after* experiences have been obtained (Clewett et al., 2019). This formation of segmented memory traces is thought to be driven by various factors, including inferred changes in the environment (DuBrow et al., 2017), prediction error signals elicited by reward outcomes (Rouhani et al., 2020) or discontinuities in the statistical structure of the environment (Gershman et al., 2014). We suggest that a practical approach for human research therefore seems to be to define events as "meaningful" units of experience (Bird, 2020) within the current experimental paradigm, and to potentially formalize them as states in an MDP, as for instance in Schuck et al. (2016). Finally, in understanding memory as a constructive process, it is important to note that neural task representations may change from perception to reactivation (Favila et al., 2020). We argue that this aspect is particularly crucial for the study of replay in humans, because activity patterns that are expected to reactivate are commonly determined based on simple localizer tasks that do not involve mnemonic task components (see e.g., Wittkuhn and Schuck, 2021). The brain might have already transformed its input data to a representation that is different from what the researcher was hoping to see re-merge from replayed activity patterns.

How many experiences are stored and for how long? In DNNs using replay, the newest experiences are stored at each time step in a memory buffer $D = e_1, \ldots, e_t$ with finite size N (e.g., Mnih et al., 2013, 2015; Zhang and Sutton, 2017). Since the success of the DNN by Mnih et al. (2015), the memory buffer is typically set to a size of $N = 10^6$ newest experiences, which continuously replace the oldest experiences (Fedus et al., 2020; Zhang and Sutton, 2017). Recently, Fedus et al. (2020) investigated the relationship between the number and age of experiences stored in the memory buffer. First, they found that increased memory capacity improved learning performance, likely due to a larger coverage of state-action pairs (Fedus et al., 2020). Second, decreasing the age of the oldest experience in the memory buffer also improved performance, likely because of older experiences that resulted from policies that are inconsistent with the current on-policy decision strategy, which is in line with earlier findings noting that experience replay is only beneficial if it is consistent with the current decision policy (Lin, 1991, 1993). An exception to this are certain Atari games that are characterized by sparse rewards and require high levels of exploration. In such tasks, sampling from older off-policy experiences is still beneficial (Fedus et al., 2020). These considerations about the size and age of the memory buffer in artificial agents point to an intriguing trade-off between the utilities of old and new memories: On the one hand, a youthful memory buffer storing only recent experiences can effectively drive the current decision policy and quickly abandon outdated and potentially inefficient behavior. On the other hand, keeping older experiences and integrating them with recent ones may foster generalization and prevent an agent from becoming stuck in a decision policy that is suboptimal.

While the size and content of a memory buffer in artificial agents can be crafted by ML researchers, determining number and nature of memories in brains is topic of ongoing debate for neuroscientists. The human brain is famously known to have a very large storage capacity, owing to the large number of modifiable synapses (Bartol Jr. et al., 2015). But forgetting is a common phenomenon. Although decay plays some role in forgetting (Hardt et al., 2013), other factors, such as interference and usage seem to be important as well (Feld and Born, 2017). Indeed forgetting might also be an important aspect of sleep, even while replay processes lead to consolidation (Feld and Born, 2017). Moreover, even if biological agents had an unlimited memory storage, selecting memories for replay from that storage would become challenging with a large amount of experiences, and, from a decision-making perspective, memory representations are only useful in so far they have utility for behavior.

when an internal model of a continuous task environment has to be learned (Aubin et al., 2018).

Despite its benefits, interleaved replay can result in problems if the agent's current policy is very different from the behavioral policy when the experiences were collected. To account for this, several authors have argued that replay needs to be corrected for such "off-policyness" using importance sampling (Meuleau et al., 2010) or other off-policy correction methods such as Retrace (Munos et al., 2016) or V-trace (Espeholt et al., 2018). These approaches essentially weight updates that result from replay in proportion to the mismatch between the policy used to generate the replay and the agent's current policy. This issue is particularly pressing in distributed replay approaches (e.g., Horgan et al., 2018), where virtual experiences are simulated in parallel and are then used for learning only with some time gap.

Given that humans and other animals do not necessarily seem to suffer from the computational problem of catastrophic interference, the question arises how the brain has apparently solved this issue and, for the purpose of this review, whether replay plays a role in the solution. Humans and animals can solve a wide set of tasks throughout their lifetime, despite temporal autocorrelation of experience and even learn well from blocked experience which troubles DNNs (Flesch et al., 2018). The idea that this ability might be related to replay (Antony and Schapiro, 2019) is supported by several studies. Karlsson and Frank (2009) for instance have observed replay of episodes from a remote spatial context. In humans, reactivation of previously learned events in the hippocampus that overlap with newly encoded memories leads to better retention (Kuhl et al., 2010).

2.2.2. Replay can amplify the influence of rare events on learning

Another challenge arises when learning must occur in environments where some events happen rarely, but are nevertheless of great significance for the agent's success or well-being. Naive DNNs will, for instance, often forget about dangerous states and revisit them (García and Fernández, 2015). This can be mitigated by replay when separate replay buffers for safe and dangerous states are maintained, such that the model cannot forget, and will frequently be reminded about dangers (Meuleau et al., 2010; Lipton et al., 2016). More generally, importance sampling techniques have been used to ensure that those experiences are sampled which are most important for the current policy of the agent, rather than those that occurred most frequently (Wang et al., 2016).

Evidence that replay might be used to mitigate this problem in animals comes from studies showing that actions which should be avoided will be reactivated, like paths to a shock zone (Wu et al., 2017) or paths to devalued outcomes (Carey et al., 2019). In addition to learning about events that should be avoided, replaying rare events that are only weakly encoded could allow the agent to form a stable representation of the entire environment even if only a smaller subset is experienced frequently. In Gupta et al. (2010) non-local replay was stronger for remote sequences if they were experienced less frequently. Using MEG in humans, Jafarpour et al. (2017) showed that stronger reactivation of one of three previously encoded stimuli was determined by how weakly the stimulus was attended to during encoding. These findings are supported by an fMRI study by Schapiro et al. (2018), who demonstrated that older, less well remembered task stimuli were selectively reactivated during a subsequent rest period resulting in memory improvement, an effect that was particularly strong in participants who slept in the 12-hr interval between test sessions, which likely offered opportunity for additional consolidation through replay. In another study, the benefits of targeted memory reactivation (TMR) were stronger for weakly learned information (Tambini et al., 2017). Further, replay-associated electroencephalography (EEG) sleep spindles during a nap following difficult (potentially weaker) but not easy (potentially stronger) memory encoding were related to improved subsequent memory performance (Schmidt et al., 2006). Together, we suggest that replay liberates an agent from needing to consider transitions only in proportion to how many times they were experienced. Instead, replay can flexibly increase or decrease the number of opportunities for learning from single episodes.

2.3. Re-inventing the past

In our introductory example (see Fig. 2), replayed content was a close reflection of past experience. Replay occurred immediately after an episode was experienced and reflected past trajectories from start to finish, albeit in reverse order. This setup stands in contrast to the ideas discussed in Section 2.2 on forgetting, which imply that replay must not necessarily respect the structure of experiences, but could, for instance, change the order and frequency of events. Beyond dealing with unevenly distributed events, replay could in fact be used to arbitrarily alter the distribution of events upon which memory is built.

Such a reorganization of experience also requires a different understanding of what constitutes an episode. In our simulation, we had assumed that the minimal unit of replayed content is one entire sequence of states, actions and rewards that occurred between a random start position and the encounter of a goal. This meant that episodes were often quite long, involving several hundreds of steps particularly early in learning (see Fig. 2C), and that the transitions between locations had to be replayed in the order in which they were experienced. But for replay to be able to reorganize experience, an episode could be divided into a much smaller unit of experience, a simple sequence of just one state, one action, one reward and the next state, known as a (s_t, a_t, r_t, s_{t+1})-tuple. Arguably, replaying such minimal experiences risks losing the benefits of temporal credit assignment, because values will not necessarily propagate along the trajectory to starting positions. But it does offer important advantages, discussed in Subsections 2.3.1 to 2.3.4. In consequence, the question of what constitutes an atomic unit of experience from the perspective of replay has important implications and is therefore actively debated (see Box 2).

2.3.1. Replay can reactivate experiences randomly

Using minimal transitions, there is a large variety of ways in which replay may alter the structure of experiences that have been discussed in the ML and neuroscience literature. One possibility is to reactivate (s_t, a_t, r_t, s_{t+1})-tuples in a random order, which artificially crafts similar conditions as during supervised learning that allows ANNs trained with stochastic gradient descent (SGD) to excel (Botvinick et al., 2020). Such uniformly sampled (s_t, a_t, r_t, s_{t+1})-tuples have therefore played an important role in adapting DNNs to RL problems, such as the famous DQN (Mnih et al., 2015). Random replay has also been found to be useful when updates are done incrementally (learning from each example as it arrives), rather than in a batch-wise manner (learning from groups of examples gathered over time), as is common in ML (Chaudhry et al., 2018). Interestingly, some animal studies have also found replay of seemingly random trajectories following exploration of a familiar open-field arena (Stella et al., 2019). Note however, that Stella et al. (2019) still observed replay of sequentially organized transitions that reflected the spatial constraints of the environment, whereas random replay used in ML can involve sets of single transitions that do not form sequential trajectories. This highlights the different understanding of replay content in ML and neuroscience. Additionally, most animal studies impose the assumption of sequentiality during data analysis, and would discard fully random activation of transitions as noise. In both ML and neuroscience, however, random replay refers to sequential reactivation that is unrelated to previously experienced action sequences, and can be seen at one extreme of a continuum describing how closely replay matches actual behavioral sequences (see Swanson et al., 2020, their Figure 2).

2.3.2. Replay can prioritize rewarding experiences

Another particularly important idea from the ML literature is to prioritize replay of transitions that led to large surprises, which often prove to be more informative than others and result in more efficient learning (Schaul et al., 2015; Horgan et al., 2018). Such prioritized replay records a prediction error (PE), the difference between the expected and actual reward, for every encountered transition and uses this signal to select experiences for replay later. This method is very similar to, and inspired by, an earlier algorithm in model-based planning known as prioritized sweeping, which selects the state to be updated according to the magnitude of the change in value upon the execution of the update (Andre et al., 1998; Moore and Atkeson, 1993; Peng and Williams, 1993). Based on the success of the prioritized replay approach, more frequent sampling of transitions with a high absolute TD error is now a common approach to train DNNs (Fedus et al., 2020). Using RL models, Mattar and Daw (2018) extended previous approaches by focusing prioritization on behaviorally relevant states that are likely to be encountered again in the future and those transitions where a policy change would yield the largest net increase in discounted future reward. Note that although prioritization algorithms assume selection of replay content on the level of individual transitions, they can, under some circumstances, still lead to sequential replay. This is true, for example, in the model from Mattar and Daw (2018), because expectations about increases in future reward are themselves often auto-correlated.

The idea that replay should be influenced by reward and surprise is in line with several animal studies. Place cell sequences associated with reward are replayed more often (Ólafsdóttir et al., 2015; Foster and Wilson, 2006; Bhattarai et al., 2019), in particular those with a high PE (Singer and Frank, 2009; Michon et al., 2019; Roscow et al., 2019), and the rate of SWRs is also influenced by reward (Ambrose et al., 2016; Singer and Frank, 2009). These results highlight replay's role in credit assignment, as discussed in Section 2.1. In human neuroimaging studies,

hippocampal activity is modulated by reward magnitude (Wolosin et al., 2012; Igloi et al., 2015). This is in line with the link between backward replay and selection of transitions based on changes in value, proposed by Mattar and Daw (2018) and Cazé et al. (2018). Replay is also more likely to contain behaviorally significant locations, such as the current goal (Gupta et al., 2010; Pfeiffer and Foster, 2013; Ólafsdóttir et al., 2015) and is biased by novelty (Cheng and Frank, 2008; Foster and Wilson, 2006). It has also been observed that optogenetic manipulation of dopaminergic input neurons, thought to signal PEs, increase replay during subsequent sleep (McNamara et al., 2014). Note that reward prediction errors might be accompanied by state prediction errors, which in the brain might both be conveyed by dopaminergic signals (see e.g., Sharpe et al., 2017; Gardner et al., 2018 Gardner et al., 2018).

2.3.3. Replay can connect experiences in novel ways

Reactivating reordered sequences can also be used to connect experiences in novel ways or strengthen weakly learned relationships. Replay can for instance correspond more closely to sampling from an internal model of the environment, rather than a veridical recapitulation of past experiences (Sutton, 1991). Among the most early ideas about replay, the Dyna architecture (Sutton, 1991) used an internal model to generate experiences that were then used to train a model-free agent. Indeed, replay can be seen as a way to blur the lines between model-free RL, such as the *Q*-learning method introduced in Box 1, and model-based RL, during which the agent stores an explicit model of the environment and can use it for planning (van Seijen and Sutton, 2015; Russek et al., 2017; Momennejad et al., 2017).

Neuroscientific evidence for reorganized experience has also been reported. During behavior, replay events can switch between reflecting immediately preceding, upcoming or more remote episodes, depending on the behavioral state of the animal at the time of replay (Pfeiffer and Foster, 2013; Ólafsdóttir et al., 2017). Even single replay events can depict more than one trajectory, such as the next one and the path the animal will take after reaching the goal location (Pfeiffer and Foster, 2013), as if representing a multi-step planning process (Foster, 2017; Miller and Venditto, 2021). Note, however, that another reason why experiences from a more distant past might be replayed could simply be that the agent is using a period during which it does not have to engage with the environment to optimize memory. This is particularly apparent for replay during sleep, when the brain has idle time to process experiences while not being actively engaged with any task. Sleep replay has frequently been observed in animals and humans, and been linked in particular to memory consolidation. Following sleep, memory interference is reduced (Baran et al., 2010; McDevitt et al., 2015 McDevitt et al., 2015) and memory integration or differentiation has been found in fMRI patterns after a delay period with sleep (Favila et al., 2016; Tompary and Davachi, 2017).

2.3.4. Replay can reuse past experiences to learn about new goals

A final aspect of reorganization relates to re-considering the usefulness of past experiences in light of one's knowledge about a goal. The RL framework presented thus far is aimed at the pursuit of a single goal (e.g., the single reward location in our grid world, see Fig. 2). However, in many real-world applications, such as the movements of a robotic arm that needs to pick and place objects, an RL strategy incorporating multiple goals would be far more beneficial. Consider again the grid world example in Fig. 2, but this time the agent can only move a finite number of steps. Since there is only one goal state that returns a reward, most of the transitions do not land in the goal state and therefore receive no reward. In such a sparse binary reward situation, where success only results from those sequences of transitions ending in the goal state, most sequences of transitions end in uninformative failures, often related to early termination without reward ("giving up"). For instance, when an agent gives up because a goal was not found after a particular amount of time, it can not know how close it was to the goal. Humans, however, can learn from failure as well as success. Inspired by this idea, an ML technique known as hindsight experience replay (Andrychowicz et al., 2017) is used to relabel the unsuccessful transitions by simply changing the goal state, such that the transitions would now be considered as successful under the new goal, thus contributing to the agent's learning. To the best of our knowledge, no directly equivalent observation has been made in the brain so far.

2.4. Planning for a better future

So far, we have mainly focused on the various ways in which replay serves learning and memory. Yet, psychological, neuroscientific and ML research has pointed out the importance of another mechanism that is crucial for goal-directed behavior: planning. A core aspect of this process is the prospective evaluation during which an agent deliberates which of the available sequences of actions and states leads to the best among several potential outcomes. In most cases, planning requires a mental model, or cognitive map (Tolman, 1938, 1948), of the environment, that describes the agent's knowledge about the transition structure of events, including the outcomes at each potential location (e.g., Moerland et al., 2020). Knowledge about the causal structure of the environment allows an agent to predict and compare the outcomes of sequences of states and actions and to choose the one that yields most reward. Yet, as we will see below, a replay buffer can be used instead of a model in order to perform planning functions too. Fig. 3 provides an illustration of how planning differs from the other two aspects of cognition we have considered so far, acting and learning.

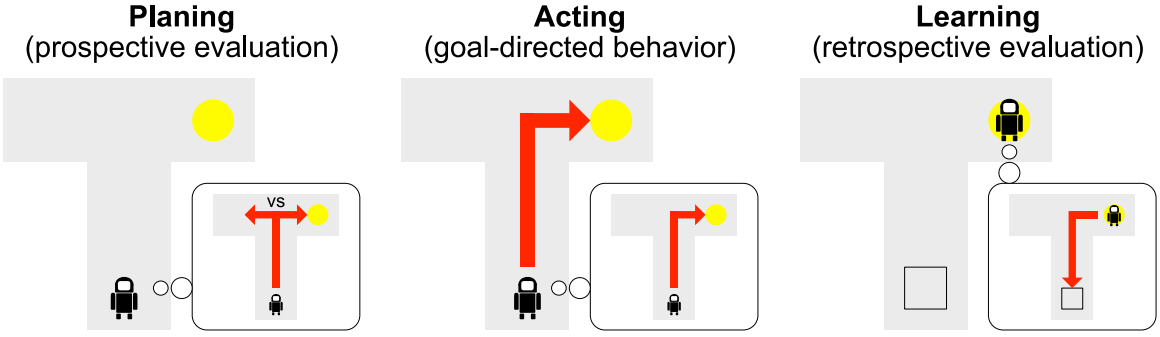

Fig. 3. Illustration of replay content. The agent is represented by the robot. The goal location is indicated by the yellow circle. Red arrows indicate behavioral trajectories of the agent in the T-maze environment or internally generated trajectories during replay. During planning (left panel) the agent engages in prospective evaluation of potential behavioral sequences in order to select the one that leads to the goal location, using forward replay. During goal-directed behavior (center panel) the agent instantiates the behavioral trajectory that is immediately relevant to prepare action using forward replay. During learning (right panel), the agent retrospectively evaluates its previous behavior, usually upon reaching a goal location, using backward replay. © Wittkuhn et al., https://doi.org/10.6084/m9.figshare.14261636.v4, CC-BY 4.0 license (https://creativecommons.org/licenses/by/4.0/).

Within the RL framework, the difference between acting based on learned (cached) values versus acting based on an internal planning process is embodied by the distinction between model-free and model-based systems (Sutton and Barto, 2018; Daw et al., 2005). In model-based RL, the agent uses experience to learn a model of the environment that is described by a function that relates the current state s_t and action a_t to the next state s_{t+1} and the reward r_t (see Box 1). The agent can use this model at decision time or during offline periods to simulate experience (Sutton, 1990). Simulated replay can be used to update cached values or to determine which action would be best to execute next, considering the rewards obtained and how the environment would change if a particular action was taken. This deliberation process has two advantages. First, planning allows the agent to remain in the safety of mental imagination and avoid the risk of suffering from potentially harmful consequences. Second, planning can be used to decide between never-experienced, entirely hypothetical courses of action (Liu et al., 2019a), a feat which would not be possible with purely experience-based replay.

Despite these differences between planning and learning, much work in RL has emphasized their similarities (Sutton, 1990, 1991; Sutton et al., 2012; van Seijen and Sutton, 2015). This research points to a function of planning that goes beyond deliberation and has shown that planning functions can be achieved without an explicit model (van Seijen and Sutton, 2015; van Hasselt et al., 2019).

As we have seen in our discussion of Lin (1992), the same learning mechanisms can be applied to real or simulated experience. Planning can thus not only be used to determine immediate behavior, but also to shape value functions, a process referred to as background planning (as opposed to decision-time planning and deliberation, see Pezzulo et al., 2019). This can be illustrated by the Dyna architecture (Sutton, 1990, 1991). Just as any model-free RL agent, a Dyna agent selects actions according to learned *Q*-values, and uses experiences to update these *Q*-values. But it also uses experiences to observe which states and rewards follow the current action, using this information to update its internal model of the world. Importantly, in Dyna the model is then used to train the model-free agent by replaying simulated episodes, and updating the agent's *Q*-values based on prediction errors, just like real experiences.

A second aspect is that replaying experiences stored in a memory buffer in some sense replaces functions that would otherwise be subserved by a model (van Seijen and Sutton, 2015; Hessel et al., 2018; van Hasselt et al., 2019), or at least enhance model-based planning functions (Eysenbach et al., 2019). van Seijen and Sutton (2015), for instance, have shown that learning value functions by a model-free method with replay can be equivalent to learning value functions with a model-based method. Empirically, van Hasselt et al. (2019) have shown that state-of-the-art replay methods, involving prioritization based on a Kullback-Leibler (KL) loss, can outperform model-based methods on Atari games (Kaiser et al., 2019), in part because an inaccurate model can lead to unstable learning. Moreover, Eysenbach et al. (2019) have shown that replay can be used to infer a graph representation of the current task that provides insights into subgoals, which in turn can be used for planning (cf. Pong et al., 2018). This is reminiscent of hindsight replay, which retroactively inserts rewards into stored replay sequences in order to facilitate learning about hierarchical subgoals (Andrychowicz et al., 2017). We note, however, that model-based planning methods remain popular in ML (Pan et al., 2018; Kaiser et al., 2019; Moerland et al., 2020). Planning methods provide the flexibility needed to generate unseen but possible transitions, and planning over long horizons can be achieved using algorithms such as tree search (Guo et al., 2014; Silver et al., 2016; Anthony et al., 2017).

While the potential benefits of replay for planning have been recognized early on in RL (Sutton, 1990), consideration of this aspect in neuroscience only appeared later, when studies demonstrated replay events in the awake state, often during short pauses from active behavior (e.g., Csicsvari et al., 2007; Diba and Buzsáki, 2007; Eldar et al., 2020; Foster and Wilson, 2006; Kudrimoti et al., 1999; Kurth-Nelson et al., 2016). This allowed researchers to draw closer correspondence between the replayed and the behavioral trajectories, and has resulted in a wealth of findings supporting the idea that replay supports model-based planning in animals as well as humans (for reviews, see e.g., Yu and Frank, 2015; Pezzulo et al., 2019; Wang et al., 2020; Tambini and Davachi, 2019; Carr et al., 2011; Ólafsdóttir et al., 2018). Disruption of awake hippocampal SWRs during a spatial alternation task specifically impaired the ability to decide between two trajectories to alternating goal locations, whereas place field representations, reactivation during rest, and other navigation behavior remained intact (Jadhav et al., 2012). Replay events in the awake state predominantly co-occur with SWRs during short pauses from ongoing exploratory behavior. Forward replay trajectories during awake SWRs often start at the current location of the animal (a well-known "initiation bias", Ambrose et al., 2016; Davidson et al., 2009; Diba and Buzsáki, 2007; Karlsson and Frank, 2009; Pfeiffer and Foster, 2013; Singer et al., 2013), and end at the goal location (Dupret et al., 2010; Pfeiffer and Foster, 2013), but not always (see e.g., Johnson and Redish, 2007).

A behavioral correlate of deliberation was already described in the 1930s in rodents (Tolman, 1926; Muenzinger and Fletcher, 1936), who tend to pause at a decision point to look back and forth between possible paths, a behavior called vicarious trial and error (VTE) (for review, see Redish, 2016). Later studies found that during VTE events, hippocampal place cells associated with theta sequences sweep ahead from the animal's current location (Johnson and Redish, 2007; Wikenheiser and Redish, 2015b; Amemiya and Redish, 2016; Papale et al., 2016). It was also found that during VTE-like behavior, place cell activity influenced the formation of place fields thought to stabilize the cognitive map (Monaco et al., 2014). Note that VTE-associated replay is often accompanied by theta sequences, which differ from SWRs in their neurophysiology. Nonetheless, both can be described as sequential activation of hippocampal cell populations, a simplifying assumption that is helpful from a computational perspective (Foster, 2017; Pezzulo et al., 2019). Recently, theta sequences have been shown to quickly cycle between possible future trajectories (Kay et al., 2020), and increases in theta power in the MTL have been observed in humans in a spatial planning task (Kaplan et al., 2020). In human fMRI, blood-oxygen-level dependent (BOLD) activity in the hippocampus has been shown to increase with deliberation time when deciding between two food items with similar value (Bakkour et al., 2019) and hippocampal activity patterns reflect routes to navigational goal locations (Brown et al., 2016). Another study has found that when humans re-learn outcomes associated with choices at lower levels of a decision tree, the extent to which higher levels of the decision tree are reactivated during rest correlates with how much their decisions change, to reach the new downstream reward states (Momennejad et al., 2018).

2.4.1. Replay can influence behavior directly or indirectly

It should be noted that although awake replay during deliberation of future choices is often related to improved task performance, the replayed trajectories do not necessarily correspond to the behavioral trajectory the animal will subsequently take, and sometimes do not end in the goal location (Johnson and Redish, 2007; Singer et al., 2013). In the study by Singer et al. (2013), hippocampal replay during SWRs that preceded correct choices reflected trajectories for the correct *and* incorrect option in a two-alternative W-maze. Once correct performance became stable (at 85% correct), replayed trajectories shifted to represent the correct future choice more frequently than the incorrect one (Singer et al., 2013). One interpretation of these findings is that the hippocampus uses replay to evaluate all potential trajectories and the behaviorally relevant trajectory is instantiated in a different brain region. Furthermore, backward replay, which backpropagates value information from the goal location, and forward replay, which samples possible trajectories ahead of the animal, might connect their trajectories as proposed by models of bidirectional planning (Khamassi and

Girard, 2020). Forward replay events have been shown to end at or close to the goal location (Pfeiffer and Foster, 2013) and might efficiently stop in states where value estimates have already been updated by a backward replay mechanism, as could be instantiated by prioritized sweeping (see Khamassi and Girard, 2020). In sum, these findings support the idea that deliberation and learning may interact. Changes in a familiar environment might increase deliberation, while the need for model updating and deliberation could diminish with learning, e.g., because decisions become more habitual and less deliberate (Dolan and Dayan, 2013). Thus, replayed trajectories in the hippocampus that evaluate all potential trajectories might be only predictive of behavior during earlier phases of learning (Singer et al., 2013) or vanish from the hippocampus (e.g., Wimmer and Büchel, 2019) when behavior becomes stereotyped. Findings by Papale et al. (2016) also demonstrate an inverse relationship between SWRs at reward sites and deliberation at choice points.

One additional complicating factor regarding the relationship between replay and subsequent behavior concerns the task setting and motivational state of the animal (e.g., Carey et al., 2019; Wu et al., 2017). Take the example of replayed place cell sequences representing the trajectory into a shock zone that is subsequently avoided (Wu et al., 2017). This might serve the purpose of learning strongly from and not forget about significant outcomes, and thus in this circumstance replay is related to avoiding rather than initiating trajectories. In line with this idea, a growing literature on computational psychiatry posits that replay could underlie symptoms like avoidance and rumination that characterize psychiatric disorders like anxiety (Gagne et al., 2018; Heller and Bagot, 2020; Mobbs et al., 2020).

If replay is related to planning, but the ultimate determination of behavior also depends on other brain areas, then replayed trajectories might be influenced by concurrent reactivation outside the hippocampus, such as the amygdala in the case of aversive outcomes (Girardeau et al., 2017). A number of studies sheds light on how replay in the hippocampus is coordinated with other brain regions to instantiate behavior. Replay is known to be coordinated with PFC (Jadhav et al., 2016; Pezzulo et al., 2014; Peyrache et al., 2009; Tang et al., 2017), and some work has placed particular focus on the interaction of hippocampal replay and the OFC (Schuck and Niv, 2019; Steiner and Redish, 2012). Indeed, disruption of nearby medial PFC attenuated components of hippocampal theta sequences representing the current location of the animal (Schmidt et al., 2019) and suppression of hippocampal input impaired the integration of task state structure in the OFC (Wikenheiser et al., 2017). Similarly, a recent study in humans has found that hippocampal replay at rest was not directly linked to behavior during a task (Schuck and Niv, 2019). Rather, replay at rest was linked to how well the different task-states were represented in the OFC, which in turn were linked to behavior (Schuck and Niv, 2019). Outside of the PFC, entorhinal grid cells that are thought to enable vector-based spatial navigation likely contribute to planning, as implicated in computational work (see e.g., Erdem and Hasselmo, 2012; Bush et al., 2015).

2.4.2. *Preplay can help planning in unknown environments*

While replay is mostly thought to occur *after* experiences have been made, some ideas have been put forth that assume a "preplay" mechanism, in which experiences are mapped out before they are encountered. Few models related to this idea have been proposed in ML, but perhaps the closest concepts are related to attractor dynamics or reservoir computing (for a review, see e.g., Lukoševičius and Jaeger, 2009). Indeed, recent computational work has suggested a link between preplay and efficient learning, arguing that attractor dynamics can account for replay (Corneil and Gerstner, 2015) or preexisting internal sequences could be used as a dynamical reservoir (Leibold, 2020). In work by Cazin et al. (2019), the framework of reservoir computing is used to model the PFC that is shown to integrate replayed sequences into larger sequence assemblies that can be recalled.

Preplay has also been observed in neuroscience, with some studies reporting apparent "preplay" of place cell sequences before the environment was ever experienced (Dragoi and Tonegawa, 2011, 2013). While preplay seems reminiscent of a planning process, most findings highlight that apparent sequentiality can also reflect hippocampal cell assemblies that are connected in a way that constrains sequential firing, even prior to experience of a new maze. Nevertheless, other findings indicate that previous experience is required for such spontaneous sequential activation to occur (Silva et al., 2015). The extent to which the hippocampus is able to seemingly preplay novel experiences could depend on the similarity between pre-existing hippocampal representations and new memories about to be formed (Eichenbaum, 2015). Methodologically, this nevertheless highlights the necessity of comparing pre- versus post-task replay (e.g., Buhry et al., 2011), as shown by recent research that observed pre-vs.-post changes in replay can indeed be explained by cell activation and firing rate correlations during experience (Farooq et al., 2019).

2.5. *Inference and generalization*

Although past information provides a glimpse into what we might expect in the future, every new experience is different from the past in some form or another. In order to use experiences effectively, agents must therefore know how to abstract from their details and store, and replay, information which could generalize best to future challenges. Past experiences should also be used to perform inferences that give novel insights that go beyond what has been observed.

2.5.1. *Replay can reflect generalizable information and transition structure*

Apart from its role in learning and planning, recent developments in ML and neuroscience research suggest that replay also contributes to inference and generalization (for previous reviews and perspectives, see Kumaran, 2012; Kumaran and McClelland, 2012; Cazé et al., 2018; Herszage and Censor, 2018; Lewis et al., 2018; Momennejad, 2020). One theme in this domain has been to build artificial agents that learn generative models from experience, which can then be used to infer new connections based on latent structural rules (Evans and Burgess, 2019), infer the correct context when given new data (Stoianov et al., 2020) and generalize information to new tasks to mitigate performance losses (Shin et al., 2017). In the model proposed by Stoianov et al. (2020), for instance, trajectories through a maze are used to learn a generative model, which can produce new trajectories consistent with the current maze structure during offline periods. As new mazes are learned, novel trajectories continue to be generated offline, but from all the mazes that have been experienced, preventing information about any one maze from being lost (similar to our considerations about forgetting in Section 2.2). The hierarchical structure of the model results in trajectories being clustered into distinct maze contexts, which allows maze categories to be inferred when presented with new data. Unlike replay used in other contexts, the model by Stoianov et al. (2020) does not suggest that prioritized replay helps to improve behavioral outcomes. Generative replay that was prioritized based on how surprising observations were under the generative model increased the number of reactivation events that contained important goal locations but did not further improve inference performance (Stoianov et al., 2020).

The clustering of trajectories seen in the model above is related to a broader theoretical view, which has emphasized that separate encoding of transition information and sensory information during learning will allow knowledge about transitions to be reused across situations with structural similarities but new sensory specifics (Behrens et al., 2018; Baram et al., 2020; Whittington et al., 2020). Because replay provides a strong candidate mechanism for learning about transition structure (Stoianov et al., 2020), replay of abstract (sensory-independent) transition information could help to build representations of task structure that can be generalized and used to guide behaviour in new sensory environments (Liu et al., 2019b) or combined with sensory observations to make inferences about the current environment (Evans and Burgess, 2019; Stoianov et al., 2020).

Another major computational approach has focused on replay as a mechanism to learn successor representations, a predictive representation that reflects the expected future visitation of states, given the current state (Dayan, 1993). Unlike the one-step transition matrices that are known as models in model-based RL, the successor matrix can reflect non-adjacent dependencies. This allows the agent to understand relationships between a state and multiple successor states, knowledge which can be used to solve inference problems, such as finding the shortest path to a new reward location (Russek et al., 2017; Momennejad et al., 2017). The eigenvectors of a successor matrix can also partition the environment into clusters that help planning (Stachenfeld et al., 2017). Critically, replay of past experiences could be used to update the successor matrix during offline periods (e.g., Russek et al., 2017), resonating with the general theme of using replay for model updating (Aubin et al., 2018). In recent work, Russek et al. (2017) proposed replay of (s_t, a_t, s_{t+1})-tuples that are prioritized by recency, in which rewards are not needed to update the successor matrix. Using this approach, it was shown that learning SRs with offline replay gives an agent unique benefits compared to agents without replay. In particular, the fast updating of successor states through replay allowed the agent to quickly infer policy updates needed to adapt to changes in the task environment, like a new barrier, that affect the state transition structure.

A potential role of replay in generalization and inference has also been suggested by neuroscientific studies. In a recent study from Barron et al. (2020), hippocampal cells selective to cues and rewarding outcomes that have not been directly experienced together, but whose relationship can be inferred based on sensory pre-conditioning (Brogden, 1939), were found to be co-active during SWR events (Barron et al., 2020). These cells also tended to be reactivated during SWR events in a specific order, with reward selective cells reactivated prior to cue selective cells akin to backward replay. Using MEG, researchers have discovered that visual stimuli are reactivated in a non-experienced order that was based on prior learning of a rule about how items should be reported (Liu et al., 2019a). This indicates that sequential reactivation is able to combine prior learning with new sensory inputs to produce behavior relevant to new environments. Recordings in the hippocampus and PFC have also shown that hippocampal place cells are reactivated with subsets of prefrontal cells that encode generalizable task elements (Yu et al., 2018) and that at least some medial PFC neurons involved in replay have generalized firing fields that cover multiple starting locations or multiple goal locations within a maze (Kaefer et al., 2020), suggesting that replay could also contribute to generalization through coordinating the appropriate reactivation of PFC neurons. Finally, neuroscientific studies have found supporting evidence for SRs, which have similar properties to place fields, skewing in the opposite direction of travel and over-representing goal locations, while the eigenvectors of SRs can account for entorhinal grid cells in some spatial contexts (Stachenfeld et al., 2017). Hippocampal-entorhinal fMRI signals have also been shown to reflect relationships between successive non-spatial objects organized in a graph (Schapiro et al., 2013; Garvert et al., 2017), consistent with the SR (Stachenfeld et al., 2017).

2.6. Representation learning

A final important aspect for understanding replay in minds and machines concerns states, the internal representations agents use to describe their environment. In this last section, we highlight findings indicating that replay is not specific to spatial locations or sensory observations, but might instead involve task-dependent state representations. We argue that replaying states has unique benefits as opposed to replaying only observations. Moreover, we speculate that replay might also have a role in *learning* the representations that guide behavior. As such, replay could offer a window into the operations the brain performs to craft useful representations of the possible task states.

2.6.1. Replay can reflect state representations

The internal states of an agent are a major determinant of its success (see Box 1). In most environments, sensory input alone will be neither fully necessary nor fully sufficient for predicting outcomes. It contains too much task-irrelevant information, and what is needed to determine the best action can often not be observed (a property called "partial observability"). Even when doing a mundane task such as crossing the street, there will be many perceived aspects you can safely ignore (the color of the cars, the behavior of passers-by, etc.), but also factors that are very important for your decision that might *not* be in your current sensory input, such as your expectation that cars can appear quickly from behind a sharp bend. Hence, representing sensory input alone is often insufficient as a state representation. As Dayan (1993) has put it: "difficult problems can be rendered trivial if looked at in the correct way" (p. 613).

Moreover, since the agent does not know what the true states of the environment are, *learning* useful state representations constitutes a major challenge (Bengio et al., 2013; Niv, 2019). This learning involves focusing attention on task-relevant dimensions (e.g., Niv et al., 2015; Leong et al., 2017), representing non-observable context, such as past events, in combination with current observations (e.g., Wilson et al., 2014; Schuck et al., 2016, 2018), and leveraging similarity among the states to determine which experiences might reflect the same hidden causes and which information can be generalized (e.g., Gershman and Niv, 2010).

Many replay algorithms for DNNs store past observations, and during replay internally convert observations into a suitable feature space using a previously learned transformation, such as a convolutional network (e.g., Mnih et al., 2015). But the benefits of directly storing internal representations for replay are increasingly acknowledged (Kapturowski et al., 2019; Iscen et al., 2020; Caccia et al., 2019; Hayes et al., 2019, 2021; van de Ven et al., 2020; Pellegrini et al., 2019). Amongst others, storing internal representations is often more memory efficient (Iscen et al., 2020; Hayes et al., 2019), while observations can still be recreated from compressed internal representations if they are needed (van de Ven et al., 2020). Moreover, representational replay can capture unobservable context that was necessary to process a given observation when it was made (Kapturowski et al., 2019).

However, representational replay, also called state replay, comes with its own set of challenges, in particular in the context of recurrent networks. Specifically, the problem of partial observability is often addressed by combining long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997) with DNNs, an architecture known as Deep Recurrent Q-Networks (Hausknecht and Stone, 2015). Because the agent's internal state representations in these networks depend on the history of previous observations, replay of past observations risks being out of context, and replay of internal states becomes necessary. Yet, as agents continue to learn from their experience, the way external inputs are mapped onto internal states changes too; in consequence, the way observations were represented internally in the past might be outdated, a phenomenon known as representational drift. It has therefore been suggested that while replaying internal representations, offline learning should be regularized in a way that captures the amount of representational drift since the replay episode (Pomponi et al., 2020; Balaji et al., 2020). The problem of representational drift will be most severe for RNNs, where observational replay will not lead to useful updates if recreated internal states do not match the internal states when the observation was made originally (e.g., Kapturowski et al., 2019). Although initial research has suggested that simply "zeroing" the agents internal state at the start of a replay event is useful in some circumstances (Hausknecht and Stone, 2015), this makes learning longer temporal dependencies more difficult. Accordingly, Kapturowski et al. (2019) have shown that it is beneficial for an agent to store its own past internal states and re-initialize the appropriate state at the start of a replay event. To account for representational drift, a part of the stored state sequences can first be replayed without updating, to reach a more

appropriate internal state, and only the remainder of the sequence is then used for offline learning. In sum, replay can be beneficial for learning if it involves not only past observations but also past states.

In animals, several findings indicate that a large variety of representations, including non-spatial sensory as well as state-like representations, might be replayed. First, the firing of hippocampal "place" cells can reflect a number of non-spatial aspects of the environment, if they are task-relevant, such as sounds (Aronov et al., 2017), time (MacDonald et al., 2011), accumulated evidence for a choice (Nieh et al., 2021), or successor representations (Stachenfeld et al., 2017), but see O'Keefe and Krupic (2021). In fact, findings by Cabral et al. (2014) show that hippocampal neurons in mice flexibly switch between representations of spatial or temporal aspects of a task, depending on which strategy was needed to solve it. More directly, one fMRI study by Schuck and Niv, 2019 has found that sequential hippocampal replay during post-task rest reflected the non-spatial states of a sequential decision-making task. Importantly, observed transitions between decoded replay events were best explained by replay of states that include non-observable task aspects, such as information from the previous trial, rather than by replay of sensory features of the task stimuli alone. This study therefore provides direct evidence for the idea that replay involves state representations that are optimized for the operation of RL algorithms. In an MEG-study by Liu et al. (2019b), human participants first learned an abstract rule governing how objects should be ordered in a sequence and later replayed a novel set of objects according to the learned rule rather than in order of experience. Replayed sequences consisted of factorized representations of sensory objects, the identity of the sequence they belonged to, as well as the position within that sequence, supporting the notion that replay is not limited to one kind of information. Moreover, Jadhav et al. (2012) showed that disruption of SWRs in a spatial alternation task impaired navigation when it required unobservable knowledge of the previous trial, thus hinting at the activation of state representations rather than observations during replay.

Interestingly, much evidence in neuroscience indicates that replay involves multiple representations which are reactivated in parallel, possibly suggesting that observations might be recreated at the time of replay (van de Ven et al., 2020). These representations reflect visual (Ji and Wilson, 2006; Wittkuhn and Schuck, 2021), auditory (Rothschild et al., 2016) or grid-like (Ólafsdóttir et al., 2016; Ólafsdóttir et al., 2017; O'Neill et al., 2017) information. These reactivated offline and online representations might interact, as it has been observed for the case of hippocampus and OFC (Schuck and Niv, 2019). This interaction between the OFC and the hippocampus (for reviews, see Wikenheiser and Redish, 2015a; Wikenheiser and Schoenbaum, 2016Wikenheiser and Redish, 2015a; Wikenheiser and Schoenbaum, 2016) is particularly interesting given that the OFC might store an agent's task state representations (Schuck et al., 2016, see also Kaplan et al., 2017). Disruption of the medial PFC particularly attenuated components of hippocampal theta sequences representing the current location of the animal (Schmidt et al., 2019) and suppression of hippocampal input to the OFC impaired the integration of task state structure (Wikenheiser et al., 2017). Conversely, disruption of SWRs during sleep impaired the integrity of hippocampal maps but they re-emerged following re-learning (Gridchyn et al., 2020) suggesting that relevant maps are stored in brain areas other than the hippocampus (Niethard and Born, 2020). Despite these interactions, it should also be noted that a number of investigations have shown replay events outside the hippocampus need not be coordinated with hippocampal activity (O'Neill et al., 2017; Kaefer et al., 2020; Wittkuhn and Schuck, 2021). In sum, hippocampal "place cell" firing can reflect a variety of non-spatial but task-relevant aspects (e.g., Aronov et al., 2017), replay occurs in a wide variety of interacting brain areas that reflect an animal's understanding of what is task-relevant, and replay has also be found to directly reflect partially observable task states (Schuck and Niv, 2019).

2.6.2. Can replay support learning useful representations?

Simultaneous replay on different levels of representation, including states and sensory observations, might convey benefits beyond those discussed so far; it might help to build better state representations. One interesting instance concerns successor representation (SR), a form of state representation that provides an efficient way to incorporate knowledge about the transitions between states into the state definition. Computational work by Russek et al. (2017) has shown that SRs can also be learned and updated through replay. More generally, information about state transitions can give rise to further graph analytical insights that are known to provide a good basis for state representations (Mahadevan and Maggioni, 2007). Sequential replay is a natural match as a mechanism to learn states that encode transitional information. Possibly, it could also be used to extract graph properties from experienced transition structures, such as bottleneck states, which then become integrated into state representations. A similar approach has been proposed by Eysenbach et al. (2019), who used replay to infer graph representations that can be used for planning. Other work has highlighted that representations which predict latent embeddings of future observations are particularly useful (Guo et al., 2020). An evaluation of predictiveness could therefore be an important contribution of replay to state representation learning.

More speculatively, coordinated replay across several levels might serve as a mechanism to identify which aspects of sensory observations exhibit transitions that are uncorrelated with state transitions of stored outcomes. Note that RL models benefit from transition information between states, but they could be affected adversely if transitions of task-irrelevant aspects influenced the agent's internal model. For example, representing states as specific locations in physical space will result in a transition matrix that is different from a transition matrix of more abstract task states, but if spatial position is irrelevant to the task at hand, then transitions between locations could be harmful for learning and planning. If replay can be used to find unattended aspects of sensory observations that correlate with reward or relevant transitions, in turn it might be used to determine which dimensions of observed input are task-irrelevant (see e.g., Schuck et al., 2015, for an example of how recognizing correlations in the environment could lead to changes in state representations). To the best of our knowledge, this idea has not been evaluated yet.

Other evidence suggests that the role of replay for state learning could go beyond information about observed transitions. SRs, for instance, can be extended to deal with partially observable task environments (Vértes et al., 2019). Caselles-Dupré and colleagues proposed another interesting account that involves variational autoencoders (VAEs) (Caselles-Dupré et al., 2018; Caselles-Dupré et al., 2019). Building on earlier work that used generative models to circumvent the memory requirements of observational replay (Shin et al., 2017), Caselles-Dupré et al. (2019) proposed storing latent representations rather than observations, and using past experiences in this form to continually train a VAE that acts as a state model. Importantly, only by replaying past episodes can the VAE learn to form a state representation that allows the agent to act efficiently across more than one environment.

In the brain, only some evidence so far suggests that replay can change state representations. Schuck and Niv, 2019 observed that replay in the hippocampus during rest was related to better decodability of partially observable state representations from the OFC during the task. Moreover, decoding of state representations in the OFC increased over time, suggesting that representation learning continued during the task, and perhaps was related to replay. Although this evidence is correlational, it hints at a relationship between replay and state representation learning. Yet, much work is left to do, and uncovering representational changes during or following replay will require new analytical approaches which for instance do not use localizer tasks. It is therefore still unclear, whether state learning mechanisms provide a realistic account for biological replay.

3. Goal-directed behavior without replay?

In this review, we have outlined the myriad ways in which replay-like mechanisms can support intelligent behavior. But can the function of replay really be so broad, or has replay simply become a scientific bandwagon? One part of the problem is that replay has no definition that is universally agreed on by the whole scientific community, and given the popularity of the topic, this has led to subsuming a vast variety of phenomena under the same term (but see Genzel et al., 2020, for a consensus statement). Moreover, the search for an inclusive understanding across the ML and neuroscience communities has probably led to further broadening of the concept. Against this backdrop, the concept of replay has occupied a large share of the scientific study of memory, and memory undoubtedly is a very fundamental building block of intelligent behavior in minds and machines alike.

Although our review covers the broad range of functions associated with replay, it is meant as an attempt to differentiate the debate about the topic. We belief that the field needs to be careful with not overburdening a single concept. In this spirit, we hope to have elucidated how, for instance, interleaved replay of past task experiences differs from replay observed during planning, or coordinated replay during offline periods. These differences can be both computational and implementational in nature: replay in these scenarios presumably serves a different function, and it is implemented in the brain, and in computers, in different ways.

In addition, we would like to underline that despite replay's continued popularity in ML, many state-of-the-art techniques exist which do not use replay. Efficient learning can be achieved without replay, for instance using Asynchronous Advantage Actor-Critic (A3C; Mnih et al., 2016) or on-policy policy gradient optimization (V-MPO; Song et al., 2019). Moreover, novel transformer models (Vaswani et al., 2017), which dispense of the need for convolutional and recurrent computations, have emerged as a powerful framework for solving complex tasks such as language processing (Dai et al., 2019) or RL (Parisotto et al., 2019). Some transformer models incorporate replay (Wu et al., 2020), but many powerful transformers have been proposed which do not require replay, including for RL problems (Parisotto et al., 2019).

Where replay ends and other forms of memory access start is often unclear as well. Consider, for instance, approaches in which agents rely directly on specific single episodes for behavioral control, such as in the context of *episodic RL* (for recent reviews, see e.g., Gershman and Daw, 2017; Botvinick et al., 2019). During episodic RL, specific single episodes are stored in memory and retrieved to directly determine behavior when the same or a similar situation is encountered again (Lengyel and Dayan, 2007; Gershman and Daw, 2017; Botvinick et al., 2019). In humans, the retrieval of single experiences in decision-making is associated with the hippocampus (Bornstein and Norman, 2017; Lee et al., 2015; Wimmer and Büchel, 2020) and reinstatement of information from past choice trials at decision-time biases present choices towards decisions made previously in the reinstated context (Duncan and Shohamy, 2016; Bornstein et al., 2017; Bornstein and Norman, 2017). To what extent these retrievals of single episodes are supported by sequential replay remains an open question.

Moreover, even if complex memory computations are needed to solve a task, external memory architectures, such as MERLIN (Wayne et al., 2018), can store past experiences in a memory buffer and learn how to read out only relevant experiences when needed. Memory storage in this model can still be efficient as the model can learn to store lower dimensional state representations instead of raw observations, and memory access is targeted to only the currently needed past information. MERLIN has been shown to outperform the LSTM architectures discussed above. But is MERLIN a replay mechanism? In some ways yes, and in others not. But a more important question is which predictions the algorithm makes, and whether they might fit neuroscientific observations. This can be nicely illustrated in the case of the MERLIN algorithm: although this model is computationally distinct from the "traditional" replay-based architectures, MERLIN predicts sophisticated reactivation phenomena. In a task in which the agent had to navigate to a goal location, for instance, the agent's memory read-out alternated between the subgoals along the way to the goal. In our opinion, asking whether this prediction is true in the brain, and in which environments such a mechanism could be helpful, rather than labeling it as replay or not, would be the most fruitful way forward. This also illustrates that replay is not a single testable theory, but rather a framework within which memory, planning and imagination-related functions, as well as their relationships, can be understood.

4. Conclusion and outlook

In this review, we have summarized the literature on replay in neuroscience and ML to showcase which computational benefits biological and artificial agents can gain from replaying previous experience. We have discussed five main computational benefits that, although overlapping, provide useful categories for thinking about what might motivate an agent to employ replay: faster learning and increased data efficiency, less forgetting, the reorganization of experience, planning and generalization. In addition we have argued that replayed content is much richer than a sequence of locations, and could reflect the agent's current state representation. State representations are often task- and context-dependent, being influenced by a range of factors, including the goal-relevant aspects of the agents observations, the transition structure of states, the location, number and value of goal locations and the motivational and metabolic state of the animal. We have argued that RL theory provides useful guidance to understand which form state representations might take in a given task, and which implications a particular state representation would have for an agent's behavior. Finally, we have discussed how replay might not only reflect but could help the agent to learn those states to begin with. While many questions in particular regarding the latter idea still remain, considering these factors will greatly help to determine what replayed representations represent and how replay updates decision-making policies that are used to control behavior.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgements

This work was supported by an Independent Max Planck Research Group grant awarded to N.W.S by the Max Planck Society (M.TN.A. BILD0004), a Starting Grant awarded to N.W.S by the European Union (ERC-2019-StG REPLAY-852669), and a Humboldt Research Fellowship awarded to S.H.M. by the Alexander von Humboldt Foundation. We thank members of the Max Planck Research Group NeuroCode for helpful discussions about the contents of this manuscript. We thank Anika Löwe (https://orcid.org/0000-0003-3132-5767) for feedback on a previous version of this manuscript. L.W. is a pre-doctoral fellow of the International Max Planck Research School on Computational Methods in Psychiatry and Ageing Research (IMPRS COMP2PSYCH). The participating institutions are the Max Planck Institute for Human Development, Berlin, Germany, and University College London, London, UK. For more information, see https://www.mps-ucl-centre.mpg.de/en/comp2psych.

References

Ambrose, R. Ellen, Pfeiffer, Brad E., Foster, David J., 2016. Reverse replay of hippocampal place cells is uniquely modulated by changing reward. Neuron 91 (5), 1124–1136. https://doi.org/10.1016/j.neuron.2016.07.047. ISSN 0896-6273.

Amemiya, Seiichiro, Redish, Aaron David, 2016. Manipulating decisiveness in decision making: effects of clonidine on hippocampal search strategies. J. Neurosci. 36 (3), 814–827. https://doi.org/10.1523/JNEUROSCI.2595-15.2016. ISSN 0270-6474.
Andre, David, Friedman, Nir, Parr, Ronald, 1998. Generalized prioritized sweeping. Advances in Neural Information Processing Systems, 10. MIT Press, pp. 1001–1007. In: https://proceedings.neurips.cc/paper/1997/file/7b5b23f4aadf9513306bcd59afb6e4c9-Paper.pdf.
Andrychowicz, Marcin, Wolski, Filip, Ray, Alex, Schneider, Jonas, Fong, Rachel, Welinder, Peter, McGrew, Bob, Tobin, Josh, Abbeel, Pieter, Zaremba, Wojciech, 2017. Hindsight experience replay. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, 30. Curran Associates, Inc, pp. 5048–5058. http://papers.nips.cc/paper/7090-hindsight-experience-replay.pdf.
Anthony, Thomas, Tian, Zheng, Barber, David, 2017. Thinking Fast and Slow with Deep Learning and Tree Search (May) arXiv e-prints, art. arXiv:1705.08439.
Antony, James W., Schapiro, Anna C., 2019. Active and effective replay: systems consolidation reconsidered again. Nat. Rev. Neurosci. 20 (8), 506–507. https://doi.org/10.1038/s41583-019-0191-8. ISSN 1471-0048.
Aronov, Dmitriy, Nevers, Rhino, Tank, David W., 2017. Mapping of a non-spatial dimension by the hippocampal-entorhinal circuit. Nature 543 (7647), 719–722. https://doi.org/10.1038/nature21692. ISSN 1476-4687.
Atkinson, Craig, McCane, Brendan, Szymanski, Lech, Robins, Anthony, 2021. Pseudo-rehearsal: achieving deep reinforcement learning without catastrophic forgetting. Neurocomputing 428 (7), 291–307. https://doi.org/10.1016/j.neucom.2020.11.050. ISSN 0925-2312.
Aubin, Lise, Khamassi, Mehdi, Girard, Benoît, 2018. Prioritized sweeping neural DynaQ with multiple predecessors, and hippocampal replays. In: Vouloutsi, Vasiliki, Halloy, José, Mura, Anna, Mangan, Michael, Lepora, Nathan, Prescott, Tony J., Verschure, Paul F.M.J. (Eds.), Biomimetic and Biohybrid Systems. Springer International Publishing, Cham, pp. 16–27. ISBN 978-3-319-95972-6.
Axmacher, Nikolai, Elger, Christian E., Fell, Juergen, 2008. Ripples in the medial temporal lobe are relevant for human memory consolidation. Brain 131 (7), 1806–1817. https://doi.org/10.1093/brain/awn103. ISSN 1460-2156.
Bakkour, Akram, Palombo, Daniela J., Zylberberg, Ariel, Kang, Yul H.R., Reid, Allison, Verfaellie, Mieke, Shadlen, Michael N., Shohamy, Daphna, 2019. The hippocampus supports deliberation during value-based decisions. eLife 8, e46080. https://doi.org/10.7554/eLife.46080. ISSN 2050-084X.
Balaji, Yogesh, Farajtabar, Mehrdad, Yin, Dong, Mott, Alex, Li, Ang, 2020. The Effectiveness of Memory Replay in Large Scale Continual Learning (October) arXiv e-prints, art. arXiv:2010.02418.
Baram, Alon Boaz, Muller, Timothy Howard, Nili, Hamed, Garvert, Mona Maria, Behrens, Timothy Edward John, 2020. Entorhinal and ventromedial prefrontal cortices abstract and generalize the structure of reinforcement learning problems. Neuron 109 (4), 713–723.e7. https://doi.org/10.1016/j.neuron.2020.11.024. ISSN 0896-6273.
Baran, Bengi, Wilson, Jessica, Spencer, Rebecca M.C., 2010. REM-dependent repair of competitive memory suppression. Exp. Brain Res. 203 (2), 471–477. https://doi.org/10.1007/s00221-010-2242-2.
Barron, Helen C., Reeve, Hayley M., Koolschijn, Renée S., Perestenko, Pavel V., Shpektor, Anna, Nili, Hamed, Rothaermel, Roman, Campo-Urriza, Natalia, O'Reilly, Jill X., Bannerman, David M., Behrens, Timothy E.J., Dupret, David, 2020. Neuronal computation underlying inferential reasoning in humans and mice. Cell 183 (1), 228–243e21. https://doi.org/10.1016/j.cell.2020.08.035. ISSN 0092-8674.
Bartol Jr, Thomas M., Bromer, Cailey, Kinney, Justin, Chirillo, Michael A., Bourne, Jennifer N., Harris, Kristen M., Sejnowski, Terrence J., 2015 nov. Nanoconnectomic upper bound on the variability of synaptic plasticity. eLife 4, e10778. https://doi.org/10.7554/eLife.10778. ISSN 2050-084X.
Behrens, Timothy E.J., Muller, Timothy H., Whittington, James C.R., Mark, Shirley, Baram, Alon B., Stachenfeld, Kimberly L., Kurth-Nelson, Zeb, 2018. What is a cognitive map? Organizing knowledge for flexible behavior. Neuron 100 (2), 490–509. https://doi.org/10.1016/j.neuron.2018.10.002. ISSN 0896-6273.
Bellman, Richard, 1957. Dynamic Programming. Princeton University Press.
Bellmund, Jacob L.S., Gärdenfors, Peter, Moser, Edvard I., Doeller, Christian F., 2018. Navigating cognition: Spatial codes for human thinking. Science 362 (6415), eaat6766. https://doi.org/10.1126/science.aat6766. ISSN 0036-8075.
Bendor, Daniel, Wilson, Matthew A., 2012. Biasing the content of hippocampal replay during sleep. Nat. Neurosci. 15 (10), 1439–1444. https://doi.org/10.1038/nn.3203. ISSN 1546-1726.
Bengio, Yoshua, Courville, Aaron, Vincent, Pascal, 2013. Representation learning: A review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. 35 (8), 1798–1828. https://doi.org/10.1109/tpami.2013.50. ISSN 2160-9292.
Bhattarai, Baburam, Lee, Jong Won, Jung, Min Whan, 2019. Distinct effects of reward and navigation history on hippocampal forward and reverse replays. Proc. Natl. Acad. Sci. 117, p. 201912533 ISSN 1091-6490.
Bird, Chris M., 2020. How do we remember events? Curr. Opin. Behav. Sci. 32, 120–125. https://doi.org/10.1016/j.cobeha.2020.01.020. ISSN 2352-1546.
Bliss, Timothy Vivian Pelham, Collingridge, Graham Leon, 1993. A synaptic model of memory: long-term potentiation in the hippocampus. Nature 361 (6407), 31–39. https://doi.org/10.1038/361031a0.
Bornstein, Aaron M., Khaw, Mel W., Shohamy, Daphna, Daw, Nathaniel D., 2017. Reminders of past choices bias decisions for reward in humans. Nat. Commun. 8 (1), 15958. https://doi.org/10.1038/ncomms15958.
Bornstein, Aaron M., Norman, Kenneth A., 2017. Reinstated episodic context guides sampling-based decisions for reward. Nat. Neurosci. 20 (7), 997–1003. https://doi.org/10.1038/nn.4573.
Bottini, Roberto, Doeller, Christian F., 2020. Knowledge across reference frames: Cognitive maps and image spaces. Trends Cogn. Sci. 24 (8), 606–619. https://doi.org/10.1016/j.tics.2020.05.008. ISSN 1364-6613.
Botvinick, Matthew, Ritter, Sam, Wang, Jane X., Kurth-Nelson, Zeb, Blundell, Charles, Hassabis, Demis, 2019. Reinforcement learning, fast and slow. Trends Cogn. Sci. 23 (5), 408–422. https://doi.org/10.1016/j.tics.2019.02.006. ISSN 1364-6613.
Botvinick, Matthew, Wang, Jane X., Dabney, Will, Miller, Kevin J., Kurth-Nelson, Zeb, 2020. Deep reinforcement learning and its neuroscientific implications. Neuron 107 (4), 603–616. https://doi.org/10.1016/j.neuron.2020.06.014. ISSN 0896-6273.
Bragin, Anatol, Engel, Jerome, Wilson, Charles L., Fried, Itzhak, Buzsáki, György, 1999. High-frequency oscillations in human brain. Hippocampus 9 (2), 137–142. https://doi.org/10.1002/(sici)1098-1063(1999)9:2<137::aid-hipo5>3.0.co;2-0. ISSN 1098-1063.
Brogden, Wilfred J., 1939. Sensory pre-conditioning. J. Exp. Psychol. 25 (4), 323–332. https://doi.org/10.1037/h0058944.
Brown, Thackery I., Carr, Valerie A., LaRocque, Karen F., Favila, Serra E., Gordon, Alan M., Bowles, Ben, Bailenson, Jeremy N., Wagner, Anthony D., 2016. Prospective representation of navigational goals in the human hippocampus. Science 352 (6291), 1323–1326. https://doi.org/10.1126/science.aaf0784. ISSN 1095-9203.
Brunec, Iva K., Moscovitch, Morris, Barense, Morgan D., 2018. Boundaries shape cognitive representations of spaces and events. Trends Cogn. Sci. 22 (7), 637–650. https://doi.org/10.1016/j.tics.2018.03.013. ISSN 1364-6613.
Buckner, Randy Lee, 2010. The role of the hippocampus in prediction and imagination. Annu. Rev. Psychol. 61 (1), 27–48. https://doi.org/10.1146/annurev.psych.60.110707.163508.
Buhry, Laure, Azizi, Amir H., Cheng, Sen, 2011. Reactivation, replay, and preplay: How it might all fit together. Neural Plas. 2011, 1–11. https://doi.org/10.1155/2011/203462. ISSN 1687-5443.
Bush, Daniel, Barry, Caswell, Manson, Daniel, Burgess, Neil, 2015. Using grid cells for navigation. Neuron 87 (3), 507–520. https://doi.org/10.1016/j.neuron.2015.07.006. ISSN 0896-6273.
Buzsáki, György, 1989. Two-stage model of memory trace formation: a role for "noisy" brain states. Neuroscience 31 (3), 551–570. https://doi.org/10.1016/0306-4522(89)90423-5. ISSN 0306-4522.
Cabral, Henrique O., Vinck, Martin, Fouquet, Celine, Pennartz, Cyriel M.A., Rondi-Reig, Laure, Battaglia, Francesco P., 2014. Oscillatory dynamics and place field maps reflect hippocampal ensemble processing of sequence and place memory under nmda receptor control. Neuron 81 (2), 402–415. https://doi.org/10.1016/j.neuron.2013.11.010. ISSN 0896-6273.
Caccia, Lucas, Belilovsky, Eugene, Caccia, Massimo, Pineau, Joelle, 2019. Online Learned Continual Compression with Adaptive Quantization Modules (November) arXiv e-prints, page arXiv:1911.08019.
Carey, Alyssa A., Tanaka, Youki, van der Meer, Matthijs A.A., 2019. Reward revaluation biases hippocampal replay content away from the preferred outcome. Nat. Neurosci. 22 (9), 1450–1459. https://doi.org/10.1038/s41593-019-0464-6. ISSN 1546-1726.
Carr, Margaret F., Jadhav, Shantanu P., Frank, Loren M., 2011. Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. Nat. Neurosci. 14 (2), 147–153. https://doi.org/10.1038/nn.2732. ISSN 1546-1726.
Caselles-Dupré, Hugo, Garcia-Ortiz, Michael, Filliat, David, 2018. Continual State Representation Learning for Reinforcement Learning Using Generative Replay (October) arXiv e-prints, page arXiv:1810.03880, URL https://ui.adsabs.harvard.edu/abs/2018arXiv181003880C.
Caselles-Dupré, Hugo, Garcia-Ortiz, Michael, Filliat, David, 2019. S-TRIGGER: Continual State Representation Learning via Self-Triggered Generative Replay (February) arXiv e-prints, art. arXiv:1902.09434.
Cazé, Romain, Khamassi, Mehdi, Aubin, Lise, Girard, Benoît, 2018. Hippocampal replays under the scrutiny of reinforcement learning models. J. Neurophysiol. 120 (6), 2877–2896. https://doi.org/10.1152/jn.00145.2018. ISSN 1522-1598.
Cazin, Nicolas, Alonso, Martin Llofriu, Chiodi, Pablo Scleidorovich, Pelc, Tatiana, Harland, Bruce, Weitzenfeld, Alfredo, Fellous, Jean-Marc, Dominey, Peter Ford, 2019. Reservoir computing model of prefrontal cortex creates novel combinations of previous navigation sequences from hippocampal place-cell replay with spatial reward propagation. PLOS Comput. Biol. 15 (7), e1006624. https://doi.org/10.1371/journal.pcbi.1006624. ISSN 1553-7358.
Chaudhry, Arslan, Dokania, Puneet K., Ajanthan, Thalaiyasingam, Torr, Philip H.S., 2018. Riemannian Walk for Incremental Learning: Understanding Forgetting and Intransigence (January) arXiv e-prints, page arXiv:1801.10112.
Cheng, Sen, Frank, Loren M., 2008. New experiences enhance coordinated neural activity in the hippocampus. Neuron 57 (2), 303–313. https://doi.org/10.1016/j.neuron.2007.11.035. ISSN 0896-6273.
Clewett, David, DuBrow, Sarah, Davachi, Lila, 2019. Transcending time in the brain: how event memories are constructed from experience. Hippocampus 29 (3), 162–183. https://doi.org/10.1002/hipo.23074. ISSN 1050-9631.
Cohen, Neal J., Eichenbaum, Howard, 1993. Memory, Amnesia, and the Hippocampal System. MIT Press. ISBN 9780262531320.
Constantinescu, Alexandra O., OReilly, Jill X., Behrens, Timothy E.J., 2016. Organizing conceptual knowledge in humans with a gridlike code. Science 352 (6292), 1464–1468. https://doi.org/10.1126/science.aaf0941. ISSN 1095-9203.
Corneil, Dane S., Gerstner, Wulfram, 2015. Attractor network dynamics enable preplay and rapid path planning in maze–like environments. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, 28. Curran Associates, Inc. In: URL https://proceedings.neurips.cc/paper/2015/file/e515df0d202ae52fcebb14295743063b-Paper.pdf.
Csicsvari, Jozsef, O'Neill, Joseph, Allen, Kevin, Senior, Timothy, 2007. Place-selective firing contributes to the reverse-order reactivation of CA1 pyramidal cells during

sharp waves in open-field exploration. Eur. J. Neurosci. 26 (3), 704–716. https://doi.org/10.1111/j.1460-9568.2007.05684.x. ISSN 0953-816X.
Dai, Zihang, Yang, Zhilin, Yang, Yiming, Carbonell, Jaime, Le, Quoc V., Salakhutdinov, Ruslan, 2019. Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context (January) arXiv e-prints, art. arXiv:1901.02860.
Davidson, Thomas J., Kloosterman, Fabian, Wilson, Matthew A., 2009. Hippocampal replay of extended experience. Neuron 63 (4), 497–507. https://doi.org/10.1016/j.neuron.2009.07.027. ISSN 0896-6273.
Daw, Nathaniel D., Niv, Yael, Dayan, Peter, 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci. 8 (12), 1704–1711. https://doi.org/10.1038/nn1560.
Dayan, Peter, 1993. Improving generalization for temporal difference learning: The successor representation. Neural Comput. 5 (4), 613–624. https://doi.org/10.1162/neco.1993.5.4.613. ISSN 1530-888X.
Deng, Xinyi, Chen, Shizhe, Sosa, Marielena, Karlsson, Mattias P., Wei, Xue-Xin, Frank, Loren M., 2020. A variable clock underlies internally generated hippocampal sequences. bioRxiv. https://doi.org/10.1101/2020.04.10.035980.
Denovellis, Eric L., Gillespie, Anna K., Coulter, Michael E., Sosa, Marielena, Chung, Jason E., Eden, Uri T., Frank, Loren M., 2020. Hippocampal replay of experience at real-world speeds. bioRxiv. https://doi.org/10.1101/2020.10.20.347708.
Deuker, Lorena, Olligs, J., Fell, J., Kranz, T.A., Mormann, F., Montag, C., Reuter, M., Elger, C.E., Axmacher, Nikolai, 2013. Memory consolidation by replay of stimulus-specific neural activity. J. Neurosci. 33 (49), 19373–19383. https://doi.org/10.1523/jneurosci.0414-13.2013. ISSN 1529-2401.
Diba, Kamran, Buzsáki, György, 2007. Forward and reverse hippocampal place-cell sequences during ripples. Nat. Neurosci. 10 (10), 1241–1242. https://doi.org/10.1038/nn1961. ISSN 1546-1726.
Diekelmann, Susanne, Born, Jan, 2010. The memory function of sleep. Nat. Rev. Neurosci. 11 (2), 114–126. https://doi.org/10.1038/nrn2762. ISSN 1471-0048.
Dolan, Ray J., Dayan, Peter, 2013. Goals and habits in the brain. Neuron 80 (2), 312–325. https://doi.org/10.1016/j.neuron.2013.09.007. ISSN 0896-6273.
Dragoi, George, Tonegawa, Susumu, 2011. Preplay of future place cell sequences by hippocampal cellular assemblies. Nature 469 (7330), 397–401. https://doi.org/10.1038/nature09633. ISSN 1476-4687.
Dragoi, George, Tonegawa, Susumu, 2013. Distinct preplay of multiple novel spatial experiences in the rat. Proc. Natl. Acad. Sci. 110 (22), 9100–9105. https://doi.org/10.1073/pnas.1306031110. ISSN 1091-6490.
DuBrow, Sarah, Rouhani, Nina, Niv, Yael, Norman, Kenneth A., 2017. Does mental context drift or shift? Curr. Opin. Behav. Sci. 17, 141–146. https://doi.org/10.1016/j.cobeha.2017.08.003. ISSN 2352-1546.
Duncan, Katherine D., Shohamy, Daphna, 2016. Memory states influence value-based decisions. J. Exp. Psychol. Gen. 145 (11), 1420–1426. https://doi.org/10.1037/xge0000231.
Dupret, David, O'Neill, Joseph, Pleydell-Bouverie, Barty, Csicsvari, Jozsef, 2010. The reorganization and reactivation of hippocampal maps predict spatial memory performance. Nat. Neurosci. 13 (8), 995–1002. https://doi.org/10.1038/nn.2599. ISSN 1546-1726.
Ego-Stengel, Valérie, Wilson, Matthew A., 2010. Disruption of ripple-associated hippocampal activity during rest impairs spatial learning in the rat. Hippocampus 20 (1), 1–10. https://doi.org/10.1002/hipo.20707. ISSN 1098-1063.
Eichenbaum, Howard, 2015. Does the hippocampus preplay memories? Nat. Neurosci. 18 (12), 1701–1702. https://doi.org/10.1038/nn.4180. ISSN 1546-1726.
Eldar, Eran, Lièvre, Gaëlle, Dayan, Peter, Dolan, Raymond J., 2020. The roles of online and offline replay in planning. eLife 9. https://doi.org/10.7554/elife.56911. ISSN 2050-084X.
Epstein, Russell A., Patai, Eva Zita, Julian, Joshua B., Spiers, Hugo J., 2017. The cognitive map in humans: Spatial navigation and beyond. Nat. Neurosci. 20 (11), 1504–1513. https://doi.org/10.1038/nn.4656. ISSN 1546-1726.
Erdem, Uğur M., Hasselmo, Michael, 2012. A goal-directed spatial navigation model using forward trajectory planning based on grid cells. Eur. J. Neurosci. 35 (6), 916–931. https://doi.org/10.1111/j.1460-9568.2012.08015.x.
Eschenko, Oxana, Ramadan, Wiâm, Mölle, Matthias, Born, Jan, Sara, Susan J., 2008. Sustained increase in hippocampal sharp-wave ripple activity during slow-wave sleep after learning. Learn. Mem. 15 (4), 222–228. https://doi.org/10.1101/lm.726008.
Espeholt, Lasse, Soyer, Hubert, Munos, Remi, Simonyan, Karen, Mnih, Volodymir, Ward, Tom, Doron, Yotam, Firoiu, Vlad, Harley, Tim, Dunning, Iain, Legg, Shane, Kavukcuoglu, Koray, 2018. IMPALA: Scalable Distributed Deep-RL with Importance Weighted Actor-Learner Architectures (February) arXiv e-prints, page arXiv:1802.01561. URL https://ui.adsabs.harvard.edu/abs/2018arXiv180201561E.
Euston, David R., Tatsuno, Masami, McNaughton, Bruce L., 2007. Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. Science 318 (5853), 1147–1150. https://doi.org/10.1126/science.1148979. ISSN 1095-9203.
Evans, Talfan, Burgess, Neil, 2019. Coordinated hippocampal-entorhinal replay as structural inference. Adv. Neural Inf. Process. Syst. 1729–1741. URL http://papers.nips.cc/paper/8450-coordinated-hippocampal-entorhinal-replay-as-structural-inference.pdf.
Eysenbach, Benjamin, Salakhutdinov, Ruslan, Levine, Sergey, 2019. Search on the Replay Buffer: Bridging Planning and Reinforcement Learning (June) arXiv e-prints, page arXiv:1906.05253. URL https://arxiv.org/abs/1906.05253.
Farooq, Usman, Sibille, Jeremie, Liu, Kefei, Dragoi, George, 2019. Strengthened temporal coordination within pre-existing sequential cell assemblies supports trajectory replay. Neuron 103 (4), 719–733.e7. https://doi.org/10.1016/j.neuron.2019.05.040. ISSN 0896-6273.
Favila, Serra E., Chanales, Avi J.H., Kuhl, Brice A., 2016. Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. Nat. Commun. 7 (1), 11066. https://doi.org/10.1038/ncomms11066.
Favila, Serra E., Lee, Hongmi, Kuhl, Brice A., 2020. Transforming the concept of memory reactivation. Trends Neurosci. 43 (12), 939–950. https://doi.org/10.1016/j.tins.2020.09.006.
Fedus, William, Ramachandran, Prajit, Agarwal, Rishabh, Bengio, Yoshua, Larochelle, Hugo, Rowland, Mark, Dabney, Will, 2020. Revisiting Fundamentals of Experience Replay (July) arXiv e-prints, art. arXiv:2007.06700, URL https://ui.adsabs.harvard.edu/abs/2020arXiv200706700F.
Feld, Gordon B., Born, Jan, 2017. Sculpting memory during sleep: Concurrent consolidation and forgetting. Curr. Opin. Neurobiol. 44, 20–27. https://doi.org/10.1016/j.conb.2017.02.012. ISSN 0959-4388.
Findlay, Graham, Tononi, Giulio, Cirelli, Chiara, 2021. The evolving view of replay and its functions in wake and sleep. SLEEP Advances 1 (1). https://doi.org/10.1093/sleepadvances/zpab002. ISSN 2632-5012.
Flesch, Timo, Balaguer, Jan, Dekker, Ronald, Nili, Hamed, Summerfield, Christopher, 2018. Comparing continual task learning in minds and machines. Proc. Natl. Acad. Sci. 115 (44), E10313–E10322. https://doi.org/10.1073/pnas.1800755115. ISSN 0027-8424.
Foster, David J., 2017. Replay comes of age. Annu. Rev. Neurosci. 40 (1), 581–602. https://doi.org/10.1146/annurev-neuro-072116-031538.
Foster, David J., Knierim, James J., 2012. Sequence learning and the role of the hippocampus in rodent navigation. Curr. Opin. Neurobiol. 22 (2), 294–300. https://doi.org/10.1016/j.conb.2011.12.005. ISSN 0959-4388.
Foster, David J., Wilson, Matthew A., 2006. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. Nature 440 (7084), 680–683. https://doi.org/10.1038/nature04587. ISSN 1476-4687.
French, Robert M., 1999. Catastrophic forgetting in connectionist networks. Trends Cogn. Sci. 3 (4), 128–135. https://doi.org/10.1016/s1364-6613(99)01294-2.
Fyhn, Marianne, Hafting, Torkel, Treves, Alessandro, Moser, May-Britt, Moser, Edvard I., 2007. Hippocampal remapping and grid realignment in entorhinal cortex. Nature 446 (7132), 190–194. https://doi.org/10.1038/nature05601. ISSN 1476-4687.
Gagne, Christopher, Dayan, Peter, Bishop, Sonia J., 2018. When planning to survive goes wrong: predicting the future and replaying the past in anxiety and PTSD. Curr. Opin. Behav. Sci. 24, 89–95. https://doi.org/10.1016/j.cobeha.2018.03.013. ISSN 2352-1546.
García, Javier, Fernández, Fernando, 2015. A comprehensive survey on safe reinforcement learning. J. Mach. Learn. Res. 16 (42), 1437–1480. http://jmlr.org/papers/v16/garcia15a.html.
Gardner, Matthew P.H., Schoenbaum, Geoffrey, Gershman, Samuel J., 2018. Rethinking dopamine as generalized prediction error. Proc. R. Soc. B: Biol. Sci. 285 (1891), 20181645. https://doi.org/10.1098/rspb.2018.1645.
Garvert, Mona M., Dolan, Raymond J., Behrens, Timothy E.J., 2017. A map of abstract relational knowledge in the human hippocampal-entorhinal cortex. eLife 6. https://doi.org/10.7554/elife.17086. ISSN 2050-084X.
Gaussier, Philippe, Revel, A., Banquet, J.P., Babeau, Vincent, 2002. From view cells and place cells to cognitive map learning: processing stages of the hippocampal system. Biol. Cybern. 86 (1), 15–28. https://doi.org/10.1007/s004220100269. ISSN 1432-0770.
Genzel, Lisa, Dragoi, George, Frank, Loren, Ganguly, Karunesh, Prida, Liset de la, Pfeiffer, Brad, Robertson, Edwin, 2020. A consensus statement: Defining terms for reactivation analysis. Philos. Trans. R. Soc. B Biol. Sci. 375 (1799), 20200001. https://doi.org/10.1098/rstb.2020.0001.
Gerrard, Jason L., Kudrimoti, Hemant, McNaughton, Bruce L., Barnes, Carol A., 2001. Reactivation of hippocampal ensemble activity patterns in the aging rat. Behav. Neurosci. 115 (6), 1180–1192. https://doi.org/10.1037/0735-7044.115.6.1180. ISSN 0735-7044.
Gershman, Samuel J., Daw, Nathaniel D., 2017. Reinforcement learning and episodic memory in humans and animals: An integrative framework. Annu. Rev. Psychol. 68 (1), 101–128. https://doi.org/10.1146/annurev-psych-122414-033625. ISSN 1545-2085.
Gershman, Samuel J., Niv, Yael, 2010. Learning latent structure: Carving nature at its joints. Curr. Opin. Neurobiol. 20 (2), 251–256. https://doi.org/10.1016/j.conb.2010.02.008.
Gershman, Samuel J., Radulescu, Angela, Norman, Kenneth A., Niv, Yael, 2014. Statistical computations underlying the dynamics of memory updating. PLOS Comput. Biol. 10 (11), 1–13. https://doi.org/10.1371/journal.pcbi.1003939.
Girardeau, Gabrielle, Benchenane, Karim, Wiener, Sidney I., Buzsáki, György, Zugaro, Michaël B., 2009. Selective suppression of hippocampal ripples impairs spatial memory. Nat. Neurosci. 12 (10), 1222–1223. https://doi.org/10.1038/nn.2384. ISSN 1546-1726.
Girardeau, Gabrielle, Inema, Ingrid, Buzsáki, György, 2017. Reactivations of emotional memory in the hippocampus-amygdala system during sleep. Nat. Neurosci. 20 (11), 1634–1642. https://doi.org/10.1038/nn.4637.
Gomperts, Stephen N., Kloosterman, Fabian, Wilson, Matthew A., 2015. VTA neurons coordinate with the hippocampal reactivation of spatial experience. eLife 4. https://doi.org/10.7554/elife.05360. ISSN 2050-084X.
Gridchyn, Igor, Schoenenberger, Philipp, O'Neill, Joseph, Csicsvari, Jozsef, 2020. Assembly-specific disruption of hippocampal replay leads to selective memory deficit. Neuron 106 (2), 291–300.e6. https://doi.org/10.1016/j.neuron.2020.01.021. ISSN 0896-6273.
Gruber, Matthias J., Ritchey, Maureen, Wang, Shao-Fang, Doss, Manoj K., Ranganath, Charan, 2016. Post-learning hippocampal dynamics promote preferential retention of rewarding events. Neuron 89 (5), 1110–1120. https://doi.org/10.1016/j.neuron.2016.01.017. ISSN 0896-6273.

Gulati, Tanuj, Guo, Ling, Ramanathan, Dhakshin S., Bodepudi, Anitha, Ganguly, Karunesh, 2017. Neural reactivations during sleep determine network credit assignment. Nat. Neurosci. 20 (9), 1277–1284. https://doi.org/10.1038/nn.4601. ISSN 1546-1726.

Guo, Xiaoxiao, Singh, Satinder, Lee, Honglak, Lewis, Richard L., Wang, Xiaoshi, 2014. Deep learning for real-time atari game play using offline monte-carlo tree search planning. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K. Q. (Eds.), Advances in Neural Information Processing Systems, 27. Curran Associates, Inc. In: https://proceedings.neurips.cc/paper/2014/file/8bb88f80d334b1869781beb89f7b73be-Paper.pdf.

Guo, Zhaohan Daniel, Pires, Bernardo Avila, Piot, Bilal, Grill, Jean-Bastien, Altché, Florent, Munos, Remi, Azar, Mohammad Gheshlaghi, 2020. Bootstrap latent-predictive representations for multitask reinforcement learning. In: Daumé, HalJIII, Singh, Aarti (Eds.), Proceedings of the 37th International Conference on Machine Learning, volume 119 of Proceedings of Machine Learning Research. PMLR, pp. 3875–3886 (13-18 July) URL http://proceedings.mlr.press/v119/guo20g.html.

Gupta, Anoopum S., van der Meer, Matthijs A.A., Touretzky, David S., Redish, Aaron David, 2010. Hippocampal replay is not a simple function of experience. Neuron 65 (5), 695–705. https://doi.org/10.1016/j.neuron.2010.01.034. ISSN 0896-6273.

Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. Hippocampus 25 (10), 2015, 1073–1188. https://doi.org/10.1002/hipo.22488. ISSN 1050-9631.

Hafting, Torkel, Fyhn, Marianne, Molden, Sturla, Moser, May-Britt, Moser, Edvard I., 2005. Microstructure of a spatial map in the entorhinal cortex. Nature 436 (7052), 801–806. https://doi.org/10.1038/nature03721. ISSN 1476-4687.

Haga, Tatsuya, Fukai, Tomoki, 2018. Recurrent network model for learning goal-directed sequences through reverse replay. eLife 7. https://doi.org/10.7554/elife.34171. ISSN 2050-084X.

Hardt, Oliver, Nader, Karim, Nadel, Lynn, 2013. Decay happens: The role of active forgetting in memory. Trends Cogn. Sci. 17 (3), 111–120. https://doi.org/10.1016/j.tics.2013.01.001. ISSN 1364-6613.

Harvey, Christopher D., Coen, Philip, Tank, David W., 2012. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. Nature 484 (7392), 62–68. https://doi.org/10.1038/nature10918. ISSN 1476-4687.

Hassabis, Demis, Kumaran, Dharshan, Summerfield, Christopher, Botvinick, Matthew, 2017. Neuroscience-inspired artificial intelligence. Neuron 95 (2), 245–258. https://doi.org/10.1016/j.neuron.2017.06.011. ISSN 0896-6273.

Hassabis, Demis, Maguire, Eleanor A., 2007. Deconstructing episodic memory with construction. Trends Cogn. Sci. 11 (7), 299–306. https://doi.org/10.1016/j.tics.2007.05.001. ISSN 1364-6613.

Hausknecht, Matthew, Stone, Peter, 2015. Deep Recurrent Q-Learning for Partially Observable MDPs (July) arXiv e-prints, page arXiv:1507.06527, URL https://ui.adsabs.harvard.edu/abs/2015arXiv150706527H.

Hayes, Tyler L., Kafle, Kushal, Shrestha, Robik, Acharya, Manoj, Kanan, Christopher, 2019. REMIND your Neural Network to Prevent Catastrophic Forgetting (October) arXiv e-prints, page arXiv:1910.02509.

Hayes, Tyler L., Krishnan, Giri P., Bazhenov, Maxim, Siegelmann, Hava T., Sejnowski, Terrence J., Kanan, Christopher, 2021. Replay in Deep Learning: Current Approaches and Missing Biological Elements (April) arXiv e-prints, page arXiv:2104.04132.

Helfrich, Randolph F., Lendner, Janna D., Mander, Bryce A., Guillen, Heriberto, Paff, Michelle, Mnatsakanyan, Lilit, Vadera, Sumeet, Walker, Matthew P., Lin, Jack J., Knight, Robert T., 2019. Bidirectional prefrontal-hippocampal dynamics organize information transfer during sleep in humans. Nat. Commun. 10 (1) https://doi.org/10.1038/s41467-019-11444-x. ISSN 2041-1723.

Heller, Aaron S., Bagot, Rosemary C., 2020. Is hippocampal replay a mechanism for anxiety and depression? JAMA Psychiatry 77 (4), 431–432. https://doi.org/10.1001/jamapsychiatry.2019.4788. ISSN 2168-622X.

Herszage, Jasmine, Censor, Nitzan, 2018. Modulation of learning and memory: a shared framework for interference and generalization. Neuroscience 392, 270–280. https://doi.org/10.1016/j.neuroscience.2018.08.006. ISSN 0306-4522.

Hessel, Matteo, Modayil, Joseph, van Hasselt, Hado, Schaul, Tom, Ostrovski, Georg, Dabney, Will, Horgan, Dan, Piot, Bilal, Azar, Mohammad, Silver, David, 2018. Rainbow: combining improvements in deep reinforcement learning. Proceedings of the AAAI Conference on Artificial Intelligence 32 (1). https://ojs.aaai.org/index.php/AAAI/article/view/11796.

Hinton, G.E., Dayan, P., Frey, B.J., Neal, R.M., 1995. The “wake-sleep” algorithm for unsupervised neural networks. Science 268 (5214), 1158–1161. https://doi.org/10.1126/science.7761831. ISSN 0036-8075.

Hochreiter, Sepp, Schmidhuber, Jürgen, 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735. ISSN 0899-7667.

Hoffman, Kari L., McNaughton, Bruce L., 2002. Coordinated reactivation of distributed memory traces in primate neocortex. Science 297 (5589), 2070–2073. https://doi.org/10.1126/science.1073538. ISSN 1095-9203.

Horgan, Dan, Quan, John, Budden, David, Barth-Maron, Gabriel, Hessel, Matteo, van Hasselt, Hado, Silver, David, 2018. Distributed Prioritized Experience Replay (March) arXiv e-prints, page arXiv:1803.00933, URL https://ui.adsabs.harvard.edu/abs/2018arXiv180300933H.

Høydal, Øyvind Arne, Skytøen, Emilie Ranheim, Andersson, Sebastian Ola, Moser, May-Britt, Moser, Edvard I., 2019. Object-vector coding in the medial entorhinal cortex. Nature 568 (7752), 400–404. https://doi.org/10.1038/s41586-019-1077-7. ISSN 1476-4687.

Igloi, Kinga, Gaggioni, Giulia, Sterpenich, Virginie, Schwartz, Sophie, 2015. A nap to recap or how reward regulates hippocampal-prefrontal memory networks during daytime sleep in humans. eLife 4, e07903. https://doi.org/10.7554/eLife.07903. ISSN 2050-084X.

Iscen, Ahmet, Zhang, Jeffrey, Lazebnik, Svetlana, Schmid, Cordelia, 2020. Memory-Efficient Incremental Learning Through Feature Adaptation (April) arXiv e-prints, page arXiv:2004.00713.

Jackson, Jadin C., Johnson, Adam, Redish, Aaron David, 2006. Hippocampal sharp waves and reactivation during awake states depend on repeated sequential experience. J. Neurosci. 26 (48), 12415–12426. https://doi.org/10.1523/jneurosci.4118-06.2006. ISSN 1529-2401.

Jadhav, Shantanu P., Kemere, Celeb, German, P. Walter, Frank, Loren M., 2012. Awake hippocampal sharp-wave ripples support spatial memory. Science 336 (6087), 1454–1458. https://doi.org/10.1126/science.1217230. ISSN 1095-9203.

Jadhav, Shantanu P., Rothschild, Gideon, Roumis, Demetris K., Frank, Loren M., 2016. Coordinated excitation and inhibition of prefrontal ensembles during awake hippocampal sharp-wave ripple events. Neuron 90 (1), 113–127. https://doi.org/10.1016/j.neuron.2016.02.010. ISSN 0896-6273.

Jafarpour, Anna, Penny, Will, Barnes, Gareth, Knight, Robert T., Duzel, Emrah, 2017. Working memory replay prioritizes weakly attended events. eNeuro 4 (4), 1–11. https://doi.org/10.1523/eneuro.0171-17.2017. ISSN 2373-2822.

Ji, Daoyun, Wilson, Matthew A., 2006. Coordinated memory replay in the visual cortex and hippocampus during sleep. Nat. Neurosci. 10 (1), 100–107. https://doi.org/10.1038/nn1825. ISSN 1546-1726.

Johnson, Adam, Redish, Aaron David, 2007. Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. J. Neurosci. 27 (45), 12176–12189. https://doi.org/10.1523/jneurosci.3761-07.2007. ISSN 1529-2401.

Joo, Hannah R., Frank, Loren M., 2018. The hippocampal sharp wave-ripple in memory retrieval for immediate use and consolidation. Nat. Rev. Neurosci. 19 (12), 744–757. https://doi.org/10.1038/s41583-018-0077-1. ISSN 1471-0048.

Kaefer, Karola, Nardin, Michele, Blahna, Karel, Csicsvari, Jozsef, 2020. Replay of behavioral sequences in the medial prefrontal cortex during rule switching. Neuron 106 (1), 154–165.e6. https://doi.org/10.1016/j.neuron.2020.01.015. ISSN 0896-6273.

Kaiser, Lukasz, Babaeizadeh, Mohammad, Milos, Piotr, Osinski, Blazej, Campbell, Roy H., Czechowski, Konrad, Erhan, Dumitru, Finn, Chelsea, Kozakowski, Piotr, Levine, Sergey, Mohiuddin, Afroz, Sepassi, Ryan, Tucker, George, Michalewski, Henryk, 2019. Model-Based Reinforcement Learning for Atari (March) arXiv e-prints, page arXiv:1903.00374.

Kaplan, Raphael, Schuck, Nicolas W., Doeller, Christian F., 2017. The role of mental maps in decision-making. Trends Neurosci. 40 (5), 256–259. https://doi.org/10.1016/j.tins.2017.03.002. ISSN 0166-2236.

Kaplan, Raphael, Campo, Adrià Tauste, Bush, Daniel, King, John, Principe, Alessandro, Koster, Raphael, Nacher, Miguel Ley, Rocamora, Rodrigo, Friston, Karl J., 2020. Human hippocampal theta oscillations reflect sequential dependencies during spatial planning. Cogn. Neurosci. 11 (3), 122–131. https://doi.org/10.1080/17588928.2019.1676711.

Kapturowski, Steven, Ostrovski, Georg, Dabney, Will, Quan, John, Munos, Remi, 2019. Recurrent experience replay in distributed reinforcement learning. In: International Conference on Learning Representations. OpenReview. https://openreview.net/forum?id=r1lyTjAqYX.

Karlsson, Mattias P., Frank, Loren M., 2009. Awake replay of remote experiences in the hippocampus. Nat. Neurosci. 12 (7), 913–918. https://doi.org/10.1038/nn.2344. ISSN 1546-1726.

Kay, Kenneth, Chung, Jason E., Sosa, Marielena, Schor, Jonathan S., Karlsson, Mattias P., Larkin, Margaret C., Liu, Daniel F., Frank, Loren M., 2020. Constant sub-second cycling between representations of possible futures in the hippocampus. Cell 180 (3), 552–567.e25. https://doi.org/10.1016/j.cell.2020.01.014. ISSN 0092-8674.

Khamassi, Mehdi, Girard, Benoît, 2020. Modeling awake hippocampal reactivations with model-based bidirectional search. Biol. Cybern. 114, 231–248. https://doi.org/10.1007/s00422-020-00817-x. ISSN 1432-0770.

Khamassi, Mehdi, Humphries, Mark D., 2012. Integrating cortico-limbic-basal ganglia architectures for learning model-based and model-free navigation strategies. Front. Behav. Neurosci. 6 https://doi.org/10.3389/fnbeh.2012.00079. ISSN 1662-5153.

King, Charles, Henze, Darrell A., Leinekugel, Xavier, Buzsáki, György, 1999. Hebbian modification of a hippocampal population pattern in the rat. J. Physiol. 521 (1), 159–167. https://doi.org/10.1111/j.1469-7793.1999.00159.x. ISSN 0022-3751.

Klinzing, Jens G., Niethard, Niels, Born, Jan, 2019. Mechanisms of systems memory consolidation during sleep. Nat. Neurosci. 22 (10), 1598–1610. https://doi.org/10.1038/s41593-019-0467-3. ISSN 1546-1726.

Kudrimoti, Hemant S., Barnes, Carol A., McNaughton, Bruce L., 1999. Reactivation of hippocampal cell assemblies: Effects of behavioral state, experience, and EEG dynamics. J. Neurosci. 19 (10), 4090–4101. https://doi.org/10.1523/jneurosci.19-10-04090.1999. ISSN 1529-2401.

Kuhl, Brice A., Shah, Arpeet T., DuBrow, Sarah, Wagner, Anthony D., 2010. Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. Nat. Neurosci. 13 (4), 501–506. https://doi.org/10.1038/nn.2498. ISSN 1546-1726.

Kumaran, Dharshan, 2012. What representations and computations underpin the contribution of the hippocampus to generalization and inference? Front. Hum. Neurosci. 6, 157. https://doi.org/10.3389/fnhum.2012.00157. ISSN 1662-5161.

Kumaran, Dharshan, McClelland, James L., 2012. Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. Psychol. Rev. 119 (3), 573–616. https://doi.org/10.1037/a0028681. ISSN 0033-295X.

Kumaran, Dharshan, Hassabis, Demis, McClelland, James L., 2016. What learning systems do intelligent agents need? Complementary learning systems theory updated. Trends Cogn. Sci. 20 (7), 512–534. https://doi.org/10.1016/j.tics.2016.05.004. ISSN 1364-6613.

Kurth-Nelson, Zeb, Economides, Marcos, Dolan, Raymond J., Dayan, Peter, 2016. Fast sequences of non-spatial state representations in humans. Neuron 91 (1), 194–204. https://doi.org/10.1016/j.neuron.2016.05.028. ISSN 10974199.

Lansink, Carien S., Goltstein, P.M., Lankelma, J.V., Joosten, R.N.J.M.A., McNaughton, Bruce L., Pennartz, Cyriel M.A., 2008. Preferential reactivation of motivationally relevant information in the ventral striatum. J. Neurosci. 28 (25), 6372–6382. https://doi.org/10.1523/jneurosci.1054-08.2008. ISSN 1529-2401.

Lansink, Carien S., Goltstein, Pieter M., Lankelma, Jan V., McNaughton, Bruce L., Pennartz, Cyriel M.A., 2009. Hippocampus leads ventral striatum in replay of place-reward information. PLoS Biol. 7 (8), e1000173. https://doi.org/10.1371/journal.pbio.1000173. ISSN 1545-7885.

LeCun, Yann, Bengio, Yoshua, Hinton, Geoffrey, 2015. Deep learning. Nature 521 (7553), 436–444. https://doi.org/10.1038/nature14539.

Lee, Albert K., Wilson, Matthew A., 2002. Memory of sequential experience in the hippocampus during slow wave sleep. Neuron 36 (6), 1183–1194. https://doi.org/10.1016/s0896-6273(02)01096-6. ISSN 0896-6273.

Lee, Sang Wan, O'Doherty, John P., Shimojo, Shinsuke, 2015. Neural computations mediating one-shot learning in the human brain. PLoS Biol. 13 (4), 1–36. https://doi.org/10.1371/journal.pbio.1002137.

Leibold, Christian, 2020. A model for navigation in unknown environments based on a reservoir of hippocampal sequences. Neural Netw. 124, 328–342. https://doi.org/10.1016/j.neunet.2020.01.014. ISSN 0893-6080.

Lengyel, Máté, Dayan, Peter, 2007. Hippocampal contributions to control: The third way. Proceedings of the 20th International Conference on Neural Information Processing Systems. Curran Associates Inc, Red Hook, NY, USA, pp. 889–896. ISBN 9781605603520.

Leong, Yuan Chang, Radulescu, Angela, Daniel, Reka, DeWoskin, Vivian, Niv, Yael, 2017. Dynamic interaction between reinforcement learning and attention in multidimensional environments. Neuron 93 (2), 451–463. https://doi.org/10.1016/j.neuron.2016.12.040. ISSN 0896-6273.

Lewis, Penelope A., Bendor, Daniel, 2019. How targeted memory reactivation promotes the selective strengthening of memories in sleep. Curr. Biol. 29 (18), R906–R912. https://doi.org/10.1016/j.cub.2019.08.019. ISSN 0960-9822.

Lewis, Penelope A., Knoblich, Günther, Poe, Gina, 2018. How memory replay in sleep boosts creative problem-solving. Trends Cogn. Sci. 22 (6), 491–503. https://doi.org/10.1016/j.tics.2018.03.009. ISSN 1364-6613.

Lin, Long Ji, 1991. Programming robots using reinforcement learning and teaching. Association for the Advancement of Artificial Intelligence, pp. 781–786. URL https://www.aaai.org/Library/AAAI/1991/aaai91-122.php.

Lin, Long-Ji, 1992. Self-improving reactive agents based on reinforcement learning, planning and teaching. Mach. Learn. 8, 293–321. https://doi.org/10.1007/bf00992699. ISSN 1573-0565.

Lin, Long-Ji, 1993. Reinforcement Learning for Robots Using Neural Networks. PhD thesis. School of Computer Science, USA. URL https://dl.acm.org/doi/book/10.5555/168871.

Lipton, Zachary C., Azizzadenesheli, Kamyar, Kumar, Abhishek, Li, Lihong, Gao, Jianfeng, Deng, Li, 2016. Combating Reinforcement Learning's Sisyphean Curse with Intrinsic Fear (November)arXiv e-prints, page arXiv:1611.01211.

Liu, Yunzhe, Dolan, Raymond J., Kurth-Nelson, Zeb, Behrens, Timothy E.J., 2019a. Human replay spontaneously reorganizes experience. Cell 178 (3), 640–652. https://doi.org/10.1016/j.cell.2019.06.012. ISSN 0092-8674.

Liu, B., Ye, X., Gao, Y., Dong, X., Wang, X., Liu, B., 2019b. Forward-looking imaginative planning framework combined with prioritized-replay double DQN. In 2019 5th International Conference on Control, Automation and Robotics (ICCAR) 336–341. https://doi.org/10.1109/ICCAR.2019.8813352. April.

Liu, Yunzhe, Dolan, Raymond J., Higgins, Cameron, Penagos, Hector, Woolrich, Mark W., Ólafsdóttir, H. Freyja, Barry, Caswell, Kurth-Nelson, Zeb, Behrens, Timothy E., 2021a. Temporally delayed linear modelling (TDLM) measures replay in both animals and humans. eLife 10 (June), e66917. https://doi.org/10.7554/eLife.66917. ISSN 2050-084X.

Liu, Yunzhe, Mattar, Marcelo G., Behrens, Timothy E.J., Daw, Nathaniel D., Dolan, Raymond J., 2021b. Experience replay is associated with efficient nonlocal learning. Science 372 (6544). https://doi.org/10.1126/science.abf1357. ISSN 0036-8075.

Louie, Kenway, Wilson, Matthew A., 2001. Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. Neuron 29 (1), 145–156. https://doi.org/10.1016/s0896-6273(01)00186-6. ISSN 0896-6273.

Lukoševičius, Mantas, Jaeger, Herbert, 2009. Reservoir computing approaches to recurrent neural network training. Comput. Sci. Rev. 3 (3), 127–149. https://doi.org/10.1016/j.cosrev.2009.03.005. ISSN 1574-0137.

MacDonald, Christopher J., Lepage, Kyle Q., Eden, Uri T., Eichenbaum, Howard, 2011. Hippocampal "time cells" bridge the gap in memory for discontiguous events. Neuron 71 (4), 737–749. https://doi.org/10.1016/j.neuron.2011.07.012. ISSN 0896-6273.

Magee, Jeffrey C., Johnston, Daniel, 1997. A synaptically controlled, associative signal for hebbian plasticity in hippocampal neurons. Science 275 (5297), 209–213. https://doi.org/10.1126/science.275.5297.209. ISSN 1095-9203.

Mahadevan, Sridhar, Maggioni, Mauro, 2007. Proto-value functions: a laplacian framework for learning representation and control in markov decision processes. J. Mach. Learn. Res. 8 (74), 2169–2231. URL http://jmlr.org/papers/v8/mahadevan07a.html.

Marr, David, 1971. Simple memory: a theory for archicortex. Philos. Trans. Royal Soc. B. Biol. Sci. 262 (841), 23–81. https://doi.org/10.1098/rstb.1971.0078. ISSN 1471-2970.

Mattar, Marcelo G., Daw, Nathaniel D., 2018. Prioritized memory access explains planning and hippocampal replay. Nat. Neurosci. 21 (11), 1609–1617. https://doi.org/10.1038/s41593-018-0232-z. ISSN 1546-1726.

McClelland, James L., McNaughton, Bruce L., O'Reilly, Randall 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. Psychol. Rev. 102 (July(3)), 419–457. https://doi.org/10.1037/0033-295x.102.3.419. ISSN 0033-295X.

Maurer, Andrew P., Nadel, Lynn, 2021. The continuity of context: A role for the hippocampus. Trends Cogn. Sci. 25 (3), 187–199. https://doi.org/10.1016/j.tics.2020.12.007. ISSN 1364-6613.

McClelland, James Lloyd, Botvinick, Matthew M., 2020. Deep learning: Implications for human learning and memory. PsyArXiv. https://doi.org/10.31234/osf.io/3m5sb.

McCloskey, Michael, Cohen, Neal J., 1989. Catastrophic interference in connectionist networks: The sequential learning problem. Psychol. Learn. Motiv. 24, 109–165. https://doi.org/10.1016/s0079-7421(08)60536-8. ISSN 0079-7421.

McDevitt, Elizabeth A., Duggan, Katherine A., Mednick, Sara C., 2015. REM sleep rescues learning from interference. Neurobiol. Learn. Mem. 122, 51–62. https://doi.org/10.1016/j.nlm.2014.11.015.

McNamara, Colin G., Tejero-Cantero, álvaro, Trouche, Stéphanie, Campo-Urriza, Natalia, Dupret, David, 2014. Dopaminergic neurons promote hippocampal reactivation and spatial memory persistence. Nat. Neurosci. 17 (12), 1658–1660. https://doi.org/10.1038/nn.3843. ISSN 1546-1726.

Meuleau, Nicolas, Plaunt, Christian, Smith, D., Smith, Tristan, 2010. A POMDP for optimal motion planning with uncertain dynamics. ICAPS-10: POMDP Practitioners Workshop.

Michon, Frédéric, Sun, Jyh-Jang, Kim, Chae Young, Ciliberti, Davide, Kloosterman, Fabian, 2019. Post-learning hippocampal replay selectively reinforces spatial memory for highly rewarded locations. Curr. Biol. 29 (9), 1436–1444.e5. https://doi.org/10.1016/j.cub.2019.03.048. ISSN 0960-9822.

Miller, Kevin J., Venditto, Sarah Jo C, 2021. Multi-step planning in the brain. Curr. Opin. Behav. Sci. 38, 29–39. https://doi.org/10.1016/j.cobeha.2020.07.003. ISSN 2352-1546.

Minsky, Marvin, 1961. Steps toward artificial intelligence. Proceedings of the IRE 49 (1), 8–30. https://doi.org/10.1109/JRPROC.1961.287775. ISSN 2162-6634.

Mnih, Volodymyr, Kavukcuoglu, Koray, Silver, David, Graves, Alex, Antonoglou, Ioannis, Wierstra, Daan, Riedmiller, Martin, 2013. Playing Atari with Deep Reinforcement Learning (December) arXiv e-prints, art. arXiv:1312.5602.

Mnih, Volodymyr, Kavukcuoglu, Koray, Silver, David, Rusu, Andrei A., Veness, Joel, Bellemare, Marc G., Graves, Alex, Riedmiller, Martin, Fidjeland, Andreas K., Ostrovski, Georg, et al., 2015. Human-level control through deep reinforcement learning. Nature 518 (7540), 529–533. https://doi.org/10.1038/nature14236. ISSN 1476-4687.

Mnih, Volodymyr, Badia, Adriá Puigdoménech, Mirza, Mehdi, Graves, Alex, Lillicrap, Timothy P., Harley, Tim, Silver, David, Kavukcuoglu, Koray, 2016. Asynchronous Methods for Deep Reinforcement Learning (February)arXiv e-prints, art. arXiv:1602.01783.

Mobbs, Dean, Headley, Drew B., Ding, Weilun, Dayan, Peter, 2020. Space, time, and fear: Survival computations along defensive circuits. Trends Cogn. Sci. 34 (3), 228–241. https://doi.org/10.1016/j.tics.2019.12.016. ISSN 1364-6613.

Moerland, Thomas M., Broekens, Joost, Jonker, Catholijn M., 2020. Model-Based Reinforcement Learning: A Survey (June) arXiv e-prints, page arXiv:2006.16712.

Momennejad, Ida, 2020. Learning structures: Predictive representations, replay, and generalization. Curr. Opin. Behav. Sci. 32, 155–166. https://doi.org/10.1016/j.cobeha.2020.02.017. ISSN 2352-1546.

Momennejad, Ida, Russek, Evan M., Cheong, J.H., Botvinick, Matthew M., Daw, Nathaniel D., Gershman, Samuel J., 2017. The successor representation in human reinforcement learning. Nat. Hum. Behav. 1 (9), 680–692. https://doi.org/10.1038/s41562-017-0180-8. ISSN 2397-3374.

Momennejad, Ida, Otto, A. Ross, Daw, Nathaniel D., Norman, Kenneth A., 2018. Offline replay supports planning in human reinforcement learning. eLife 7, e32548. https://doi.org/10.7554/eLife.32548.

Monaco, Joseph D., Rao, Geeta, Roth, Eric D., Knierim, James J., 2014. Attentive scanning behavior drives one-trial potentiation of hippocampal place fields. Nat. Neurosci. 17 (5), 725–731. https://doi.org/10.1038/nn.3687. ISSN 1546-1726.

Moore, Andrew W., Atkeson, Christopher G., 1993. Prioritized sweeping: Reinforcement learning with less data and less time. Mach. Learn. 13 (1), 103–130. https://doi.org/10.1007/bf00993104. ISSN 1573-0565.

Moser, Edvard I., Kropff, Emilio, Moser, May-Britt, 2008. Place cells, grid cells, and the brain's spatial representation system. Annu. Rev. Neurosci. 31 (1), 69–89. https://doi.org/10.1146/annurev.neuro.31.061307.090723. ISSN 0147-006X.

Muenzinger, Karl F., Fletcher, F. Milford, 1936. Motivation in learning. VI. Escape from electric shock compared with hunger-food tension in the visual discrimination habit. J. Comp. Psychol. 22 (1), 79–91. https://doi.org/10.1037/h0057664.

Munos, Rémi, Stepleton, Tom, Harutyunyan, Anna, Bellemare, Marc G., 2016. Safe and Efficient Off-Policy Reinforcement Learning (June) arXiv e-prints, art. arXiv:1606.02647.

Nádasdy, Zoltán, Hirase, Hajime, Czurkó, András, Csicsvari, Jozsef, Buzsáki, György, 1999. Replay and time compression of recurring spike sequences in the hippocampus. J. Neurosci. 19 (21), 9497–9507. https://doi.org/10.1523/jneurosci.19-21-09497.1999. ISSN 1529-2401.

Nieh, Edward H., Schottdorf, Manuel, Freeman, Nicolas W., Low, Ryan J., Lewallen, Sam, Koay, Sue Ann, Pinto, Lucas, Gauthier, Jeffrey L., Brody, Carlos D., Tank, David W., 2021. Geometry of abstract learned knowledge in the hippocampus. Nature 595, 80–84. https://doi.org/10.1038/s41586-021-03652-7.

Niethard, Niels, Born, Jan, 2020. A backup of hippocampal spatial code outside the hippocampus? New light on systems memory consolidation. Neuron 106 (2), 204–206. https://doi.org/10.1016/j.neuron.2020.03.034. ISSN 0896-6273.

Niv, Yael, 2019. Learning task-state representations. Nat. Neurosci. 22 (10), 1544–1553. https://doi.org/10.1038/s41593-019-0470-8. ISSN 1546-1726.

Niv, Yael, Daniel, Reka, Geana, Andra, Gershman, Samuel J., Leong, Yuan Chang, Radulescu, Angela, Wilson, Robert C., 2015. Reinforcement learning in multidimensional environments relies on attention mechanisms. J. Neurosci. 35 (21), 8145–8157. https://doi.org/10.1523/JNEUROSCI.2978-14.2015. ISSN 0270-6474.

Norman, Yitzhak, Yeagle, Erin M., Khuvis, Simon, Harel, Michal, Mehta, Ashesh D., Malach, Rafael, 2019. Hippocampal sharp-wave ripples linked to visual episodic recollection in humans. Science 365 (6454), eaax1030. https://doi.org/10.1126/science.aax1030. ISSN 1095-9203.

O'Keefe, John, Conway, Dulcie H., 1978. Hippocampal place units in the freely moving rat: why they fire where they fire. Exp. Brain Res. 31 (4), 573–590. https://doi.org/10.1007/BF00239813. ISSN 1432-1106.

O'Keefe, John, Dostrovsky, Jonathan, 1971. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res. 34 (1), 171–175. https://doi.org/10.1016/0006-8993(71)90358-1. ISSN 0006-8993.

O'Keefe, John, Krupic, Julija, 2021. Do hippocampal pyramidal cells respond to non-spatial stimuli? Physiol. Rev. 101 (3), 1427–1456. https://doi.org/10.1152/physrev.00014.2020.

O'Keefe, John, Nadel, Lynn, 1974. Maps in the brain. New Scientist 62 (903), 749–751.

O'Keefe, John, Nadel, Lynn, 1978. The Hippocampus as a Cognitive Map. Clarendon Press, Oxford.

Ólafsdóttir, H. Freyja, Barry, Caswell, Saleem, Aman B., Hassabis, Demis, Spiers, Hugo J., 2015. Hippocampal place cells construct reward related sequences through unexplored space. eLife 4. https://doi.org/10.7554/elife.06063. ISSN 2050-084X.

Ólafsdóttir, H. Freyja, Carpenter, Francis, Barry, Caswell, 2016. Coordinated grid and place cell replay during rest. Nat. Neurosci. 19 (6), 792–794. https://doi.org/10.1038/nn.4291. ISSN 1546-1726.

Ólafsdóttir, H. Freyja, Carpenter, Francis, Barry, Caswell, 2017. Task demands predict a dynamic switch in the content of awake hippocampal replay. Neuron 96 (4), 925–935.e6. https://doi.org/10.1016/j.neuron.2017.09.035. ISSN 0896-6273.

Ólafsdóttir, H. Freyja, Bush, Daniel, Barry, Caswell, 2018. The role of hippocampal replay in memory and planning. Curr. Biol. 28 (1), R37–R50. https://doi.org/10.1016/j.cub.2017.10.073. ISSN 0960-9822.

O'Neill, Joseph, Senior, Timothy J., Allen, Kevin, Huxter, John R., Csicsvari, Jozsef, 2008. Reactivation of experience-dependent cell assembly patterns in the hippocampus. Nat. Neurosci. 11 (2), 209–215. https://doi.org/10.1038/nn2037. ISSN 1546-1726.

O'Neill, Joseph, Pleydell-Bouverie, Barty, Dupret, David, Csicsvari, Jozsef, 2010. Play it again: Reactivation of waking experience and memory. Trends Neurosci. 33 (5), 220–229. https://doi.org/10.1016/j.tins.2010.01.006. ISSN 0166-2236.

O'Neill, Joseph, Boccara, Charlotte N., Stella, Federico, Schoenenberger, Philipp, Csicsvari, Jozsef, 2017. Superficial layers of the medial entorhinal cortex replay independently of the hippocampus. Science 355 (6321), 184–188. https://doi.org/10.1126/science.aag2787. ISSN 1095-9203.

O'Reilly, Randall C., McClelland, James L., 1994. Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off. Hippocampus 4 (6), 661–682. https://doi.org/10.1002/hipo.450040605.

O'Reilly, Randall C., Bhattacharyya, Rajan, Howard, Michael D., Ketz, Nicholas, 2014. Complementary learning systems. Cogn. Sci. 38 (6), 1229–1248. https://doi.org/10.1111/j.1551-6709.2011.01214.x.

Oudiette, Delphine, Paller, Ken A., 2013. Upgrading the sleeping brain with targeted memory reactivation. Trends Cogn. Sci. 17 (3), 142–149. https://doi.org/10.1016/j.tics.2013.01.006. ISSN 1364-6613.

Pan, Yangchen, Zaheer, Muhammad, White, Adam, Patterson, Andrew, White, Martha, 2018. Organizing Experience: A Deeper Look at Replay Mechanisms for Sample-Based Planning in Continuous State Domains (June), arXiv e-prints, page arXiv:1806.04624, URL https://arxiv.org/abs/1806.04624.

Papale, Andrew E., Zielinski, Mark C., Frank, Loren M., Jadhav, Shantanu P., Redish, Aaron David, 2016. Interplay between hippocampal sharp-wave-ripple events and vicarious trial and error behaviors in decision making. Neuron 92 (5), 975–982. https://doi.org/10.1016/j.neuron.2016.10.028. ISSN 0896-6273.

Parisi, German I., Kemker, Ronald, Part, Jose L., Kanan, Christopher, Wermter, Stefan, 2019. Continual lifelong learning with neural networks: A review. Neural Netw. 113, 54–71. https://doi.org/10.1016/j.neunet.2019.01.012. ISSN 0893-6080.

Parisotto, Emilio, Song, H. Francis, Rae, Jack W., Pascanu, Razvan, Gulcehre, Caglar, Jayakumar, Siddhant M., Jaderberg, Max, Kaufman, Raphael Lopez, Clark, Aidan, Noury, Seb, Botvinick, Matthew M., Heess, Nicolas, Hadsell, Raia, 2019. Stabilizing Transformers for Reinforcement Learning (October) arXiv e-prints, page arXiv:1910.06764.

Pavlides, Constantine, Winson, Jonathan, 1989. Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. J. Neurosci. 9 (8), 2907–2918. https://doi.org/10.1523/jneurosci.09-08-02907.1989. ISSN 1529-2401.

Peer, Michael, Brunec, Iva K., Newcombe, Nora S., Epstein, Russell A., 2020. Structuring knowledge with cognitive maps and cognitive graphs. Trends Cogn. Sci. 25 (1), 37–54. https://doi.org/10.1016/j.tics.2020.10.004. ISSN 1364-6613.

Pellegrini, Lorenzo, Graffieti, Gabriele, Lomonaco, Vincenzo, Maltoni, Davide, 2019. Latent Replay for Real-Time Continual Learning (December) arXiv e-prints, page arXiv:1912.01100, URL https://arxiv.org/abs/1912.01100.

Peng, Jing, Williams, Ronald J., 1993. Efficient learning and planning within the Dyna framework. Adapt. Behav. 1 (4), 437–454. https://doi.org/10.1177/105971239300100403. ISSN 1059-7123.

Pennartz, C.M.A., 2004. The ventral striatum in off-line processing: Ensemble reactivation during sleep and modulation by hippocampal ripples. J. Neurosci. 24 (29), 6446–6456. https://doi.org/10.1523/jneurosci.0575-04.2004. ISSN 1529-2401.

Peyrache, Adrien, Khamassi, Mehdi, Benchenane, Karim, Wiener, Sidney I., Battaglia, Francesco P., 2009. Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. Nat. Neurosci. 12 (7), 919–926. https://doi.org/10.1038/nn.2337. ISSN 1546-1726.

Pezzulo, Giovanni, van der Meer, Matthijs A.A., Lansink, Carien S., Pennartz, Cyriel M.A., 2014. Internally generated sequences in learning and executing goal-directed behavior. Trends Cogn. Sci. 18 (12), 647–657. https://doi.org/10.1016/j.tics.2014.06.011. ISSN 1364-6613.

Pezzulo, Giovanni, Donnarumma, Francesco, Maisto, Domenico, Stoianov, Ivilin, 2019. Planning at decision time and in the background during spatial navigation. Curr. Opin. Behav. Sci. 29, 69–76. https://doi.org/10.1016/j.cobeha.2019.04.009. ISSN 2352-1546.

Pfeiffer, Brad E., Foster, David J., 2013. Hippocampal place-cell sequences depict future paths to remembered goals. Nature 497 (7447), 74–79. https://doi.org/10.1038/nature12112. ISSN 1476-4687.

Pomponi, Jary, Scardapane, Simone, Lomonaco, Vincenzo, Uncini, Aurelio, 2020. Efficient continual learning in neural networks with embedding regularization. Neurocomputing 397, 139–148. https://doi.org/10.1016/j.neucom.2020.01.093. ISSN 0925-2312.

Pong, Vitchyr, Gu, Shixiang, Dalal, Murtaza, Levine, Sergey, 2018. Temporal Difference Models: Model-Free Deep Rl for Model-Based Control (February)arXiv e-prints, art. arXiv:1802.09081.

Qin, Yu-Lin, Mcnaughton, Bruce L., Skaggs, William E., Barnes, Carol A., 1997. Memory reprocessing in corticocortical and hippocampocortical neuronal ensembles. Philos. Trans. R. Soc. Lond. B: Biol. Sci. 352 (1360), 1525–1533. https://doi.org/10.1098/rstb.1997.0139. ISSN 1471-2970.

Ramanathan, Dhakshin S., Gulati, Tanuj, Ganguly, Karunesh, 2015. Sleep-dependent reactivation of ensembles in motor cortex promotes skill consolidation. PLoS Biol. 13 (9), 1–25. https://doi.org/10.1371/journal.pbio.1002263. ISSN 1545-7885.

Rasch, Björn, Born, Jan, 2007. Maintaining memories by reactivation. Curr. Opin. Neurobiol. 17 (6), 698–703. https://doi.org/10.1016/j.conb.2007.11.007. ISSN 0959-4388.

Ratcliff, Roger, 1990. Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions. Psychol. Rev. 97 (2), 285–308. https://doi.org/10.1037/0033-295x.97.2.285.

Redish, Aaron David, 1999. Beyond the Cognitive Map: From Place Cells to Episodic Memory. MIT Press.

Redish, Aaron David, 2016. Vicarious Trial and Error. Nat. Rev. Neurosci. 17 (3), 147–159. https://doi.org/10.1038/nrn.2015.30. ISSN 1471-0048.

Redish, Aaron David, Touretzky, David S., 1998. The role of the hippocampus in solving the morris water maze. Neural Comput. 10 (1), 73–111. https://doi.org/10.1162/089976698300017908. ISSN 1530-888X.

Richmond, Lauren L., Zacks, Jeffrey M., 2017. Constructing experience: event models from perception to action. Trends Cogn. Sci. 21 (12), 962–980. https://doi.org/10.1016/j.tics.2017.08.005. ISSN 1364-6613.

Roscow, Emma L., Jones, Matthew W., Lepora, Nathan F., 2019. Behavioural and computational evidence for memory consolidation biased by reward-prediction errors. bioRxiv. https://doi.org/10.1101/716290.

Rothschild, Gideon, Eban, Elad, Frank, Loren M., 2016. A cortical-hippocampal-cortical loop of information processing during memory consolidation. Nat. Neurosci. 20 (2), 251–259. https://doi.org/10.1038/nn.4457. ISSN 1546-1726.

Rouhani, Nina, Norman, Kenneth A., Niv, Yael, Bornstein, Aaron M., 2020. Reward prediction errors create event boundaries in memory. Cognition 203, 104269. https://doi.org/10.1016/j.cognition.2020.104269. ISSN 0010-0277.

Russek, Evan M., Momennejad, Ida, Botvinick, Matthew M., Gershman, Samuel J., Daw, Nathaniel D., 2017. Predictive representations can link model-based reinforcement learning to model-free mechanisms. PLoS Comput. Biol. 13 (9), e1005768. https://doi.org/10.1371/journal.pcbi.1005768. ISSN 1553-7358.

Schafer, Matthew, Schiller, Daniela, 2018. Navigating social space. Neuron 100 (2), 476–489. https://doi.org/10.1016/j.neuron.2018.10.006. ISSN 0896-6273.

Schapiro, Anna C., Rogers, Timothy T., Cordova, Natalia I., Turk-Browne, Nicholas B., Botvinick, Matthew M., 2013. Neural representations of events arise from temporal community structure. Nat. Neurosci. 16 (4), 486–492. https://doi.org/10.1038/nn.3331. ISSN 1546-1726.

Schapiro, Anna C., Turk-Browne, Nicholas B., Botvinick, Matthew M., Norman, Kenneth A., 2017. Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. Philos. Trans. Royal Soc. B. Biol. Sci. 372 (1711) https://doi.org/10.1098/rstb.2016.0049. ISSN 1471-2970.

Schapiro, Anna C., McDevitt, Elizabeth A., Rogers, Timothy T., Mednick, Sara C., Norman, Kenneth A., 2018. Human hippocampal replay during rest prioritizes weakly learned information and predicts memory performance. Nat. Commun. 9 (1) https://doi.org/10.1038/s41467-018-06213-1. ISSN 2041-1723.

Schaul, Tom, Quan, John, Antonoglou, Ioannis, Silver, David, 2015. Prioritized Experience Replay (November) arXiv e-prints, page arXiv:1511.05952, URL https://ui.adsabs.harvard.edu/abs/2015arXiv151105952S.

Schmidt, Brandy, Duin, Anneke A., Redish, Aaron David, 2019. Disrupting the medial prefrontal cortex alters hippocampal sequences during deliberative decision making. J. Neurophysiol. 121 (6), 1981–2000. https://doi.org/10.1152/jn.00793.2018. ISSN 1522-1598.

Schmidt, Christina, Peigneux, Philippe, Muto, Vincenzo, Schenkel, Maja, Knoblauch, Vera, Münch, Mirjam, de Quervain, Dominique J.-F., Wirz-Justice, Anna, Cajochen, Christian, 2006. Encoding difficulty promotes postlearning changes in sleep spindle activity during napping. J. Neurosci. 26 (35), 8976–8982. https://doi.org/10.1523/JNEUROSCI.2464-06.2006. ISSN 0270-6474.

Schuck, Nicolas W., Gaschler, Robert, Wenke, Dorit, Heinzle, Jakob, Frensch, Peter A., Haynes, John-Dylan, Reverberi, Carlo, 2015. Medial prefrontal cortex predicts internally driven strategy shifts. Neuron 86 (1), 331–340. https://doi.org/10.1016/j.neuron.2015.03.015. ISSN 0896-6273.

Schuck, Nicolas W., Cai, Ming Bo, Wilson, Robert C., Niv, Yael, 2016. Human orbitofrontal cortex represents a cognitive map of state space. Neuron 91 (6), 1402–1412. https://doi.org/10.1016/j.neuron.2016.08.019. ISSN 0896-6273.

Schuck, Nicolas W., Niv, Yael, 2019. Sequential replay of nonspatial task states in the human hippocampus. Science 364 (6447), eaaw5181. https://doi.org/10.1126/science.aaw5181.

Schuck, Nicolas W., Wilson, Robert, Niv, Yael, 2018. A state representation for reinforcement learning and decision-making in the orbitofrontal cortex. In: Morris, Richard, Bornstein, Aaron, Shenhav, Amitai (Eds.), Goal-Directed Decision Making, chapter 12, 1 edition. Academic Press, pp. 259–278. https://doi.org/10.1016/B978-0-12-812098-9.00012-7. ISBN 978-0-12-812098-9.

Scoville, William Beecher, Milner, Brenda, 1957. Loss of recent memory after bilateral hippocampal lesions. J. Neurol. Neurosurg. Psychiatry 20 (1), 11–21. https://doi.org/10.1136/jnnp.20.1.11.

Sharpe, Melissa J., Chang, Chun Yun, Liu, Melissa A., Batchelor, Hannah M., Mueller, Lauren E., Jones, Joshua L., Niv, Yael, Schoenbaum, Geoffrey, 2017. Dopamine transients are sufficient and necessary for acquisition of model-based associations. Nat. Neurosci. 20 (5), 735–742. https://doi.org/10.1038/nn.4538. ISSN 1546-1726.

Shin, Hanul, Lee, Jung Kwon, Kim, Jaehong, Kim, Jiwon, 2017. Continual learning with deep generative replay. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, 30. Curran Associates, Inc, pp. 2990–2999. URL http://papers.nips.cc/paper/6892-continual-learning-with-deep-generative-replay.pdf.

Shin, Justin D., Tang, Wenbo, Jadhav, Shantanu P., 2019. Dynamics of awake hippocampal-prefrontal replay for spatial learning and memory-guided decision making. Neuron 104 (6), 1110–1125. https://doi.org/10.1016/j.neuron.2019.09.012. ISSN 0896-6273.

Shin, Yeon Soon, DuBrow, Sarah, 2020. Structuring memory through inference-based event segmentation. Top. Cogn. Sci. 13 (1), 106–127. https://doi.org/10.1111/tops.12505. ISSN 1756-8757.

Silva, Delia, Feng, Ting, Foster, David J., 2015. Trajectory events across hippocampal place cells require previous experience. Nat. Neurosci. 18 (12), 1772–1779. https://doi.org/10.1038/nn.4151. ISSN 1546-1726.

Silver, David, Huang, Aja, Maddison, Chris J., Guez, Arthur, Sifre, Laurent, van den Driessche, George, Schrittwieser, Julian, Antonoglou, Ioannis, Panneershelvam, Veda, Lanctot, Marc, Dieleman, Sander, Grewe, Dominik, Nham, John, Kalchbrenner, Nal, Sutskever, Ilya, Lillicrap, Timothy, Leach, Madeleine, Kavukcuoglu, Koray, Graepel, Thore, Hassabis, Demis, 2016. Mastering the game of go with deep neural networks and tree search. Nature 529 (7587), 484–489. https://doi.org/10.1038/nature16961. ISSN 1476-4687.

Singer, Annabelle C., Frank, Loren M., 2009. Rewarded outcomes enhance reactivation of experience in the hippocampus. Neuron 64 (6), 910–921. https://doi.org/10.1016/j.neuron.2009.11.016. ISSN 0896-6273.

Singer, Annabelle C., Carr, Margaret F., Karlsson, Mattias P., Frank, Loren M., 2013. Hippocampal SWR activity predicts correct decisions during the initial learning of an alternation task. Neuron 77 (6), 1163–1173. https://doi.org/10.1016/j.neuron.2013.01.027. ISSN 0896-6273.

Skaggs, William E., McNaughton, Bruce L., 1996. Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. Science 271 (5257), 1870–1873. https://doi.org/10.1126/science.271.5257.1870. ISSN 1095-9203.

Song, H. Francis, Abdolmaleki, Abbas, Springenberg, Jost Tobias, Clark, Aidan, Soyer, Hubert, Rae, Jack W., Noury, Seb, Ahuja, Arun, Liu, Siqi, Tirumala, Dhruva, Heess, Nicolas, Belov, Dan, Riedmiller, Martin, Botvinick, Matthew M., 2019. V-MPO: On-Policy Maximum A Posteriori Policy Optimization for Discrete and Continuous Control (September)arXiv e-prints, art. arXiv:1909.12238.

Spiers, Hugo J., 2020. The hippocampal cognitive map: one space or many? Trends Cogn. Sci. 24 (3), 168–170. https://doi.org/10.1016/j.tics.2019.12.013. ISSN 1364-6613.

Squire, Larry Ryan, 1992. Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. Psychol. Rev. 99 (2), 195–231. https://doi.org/10.1037/0033-295x.99.2.195.

Staba, Richard J., Wilson, Charles L., Bragin, Anatol, Fried, Itzhak, Engel, Jerome, 2002. Quantitative analysis of high-frequency oscillations (80-500 Hz) recorded in human epileptic hippocampus and entorhinal cortex. J. Neurophysiol. 88 (4), 1743–1752. https://doi.org/10.1152/jn.2002.88.4.1743. ISSN 1522-1598.

Stachenfeld, Kimberly L., Botvinick, Matthew M., Gershman, Samuel J., 2017. The hippocampus as a predictive map. Nat. Neurosci. 20 (11), 1643–1653. https://doi.org/10.1038/nn.4650. ISSN 1546-1726.

Staresina, Bernhard P., Alink, Arjen, Kriegeskorte, Nikolaus, Henson, Richard N., 2013. Awake reactivation predicts memory in humans. Proc. Natl. Acad. Sci. 110 (52), 21159–21164. https://doi.org/10.1073/pnas.1311989110. ISSN 1091-6490.

Staresina, Bernhard P., Bergmann, Til Ole, Bonnefond, Mathilde, van der Meij, Roemer, Jensen, Ole, Deuker, Lorena, Elger, Christian E., Axmacher, Nikolai, Fell, Juergen, 2015. Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. Nat. Neurosci. 18 (11), 1679–1686. https://doi.org/10.1038/nn.4119. ISSN 1546-1726.

Steiner, Adam, Redish, Aaron David, 2012. The road not taken: neural correlates of decision making in orbitofrontal cortex. Front. Neurosci. 6, 131. https://doi.org/10.3389/fnins.2012.00131. ISSN 1662-453X.

Stella, Federico, Baracskay, Peter, O'Neill, Joseph, Csicsvari, Jozsef, 2019. Hippocampal reactivation of random trajectories resembling brownian diffusion. Neuron 102 (2), 450–461. https://doi.org/10.1016/j.neuron.2019.01.052. ISSN 0896-6273.

Stoianov, Ivilin, Maisto, Domenico, Pezzulo, Giovanni, 2020. The hippocampal formation as a hierarchical generative model supporting generative replay and continual learning. bioRxiv. https://doi.org/10.1101/2020.01.16.908889.

Sun, Chen, Yang, Wannan, Martin, Jared, Tonegawa, Susumu, 2020. Hippocampal neurons represent events as transferable units of experience. Nat. Neurosci. 23, 651–663. https://doi.org/10.1038/s41593-020-0614-x. ISSN 1546-1726.

Sutherland, Gary R., McNaughton, Bruce, 2000. Memory trace reactivation in hippocampal and neocortical neuronal ensembles. Curr. Opin. Neurobiol. 10 (2), 180–186. https://doi.org/10.1016/S0959-4388(00)00079-9. ISSN 0959-4388.

Sutton, Richard S., 1990. Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In: Porter, Bruce, Mooney, Raymond (Eds.), Machine Learning Proceedings. Morgan Kaufmann, San Francisco (CA), pp. 216–224. ISBN 978-1-55860-141-3.

Sutton, Richard S., 1991. Dyna, an integrated architecture for learning, planning, and reacting. ACM SIGART Bulletin 2 (4), 160–163. https://doi.org/10.1145/122344.122377. ISSN 0163-5719.

Sutton, Richard S., Barto, Andrew G., 2018. Reinforcement Learning: An Introduction, 2 edition. MIT Press. URL http://incompleteideas.net/book/the-book-2nd.html.

Sutton, Richard S., Szepesvári, Csaba, Geramifard, Alborz, Bowling, Michael, 2012. Dyna-Style Planning with Linear Function Approximation and Prioritized Sweeping arXiv e-prints, abs/1206.3285, URL http://arxiv.org/abs/1206.3285.

Swanson, Rachel A., Levenstein, Daniel, McClain, Kathryn, Tingley, David, Buzsáki, György, 2020. Variable specificity of memory trace reactivation during hippocampal sharp wave ripples. Curr. Opin. Behav. Sci. 32, 126–135. https://doi.org/10.1016/j.cobeha.2020.02.008. ISSN 2352-1546.

Tambini, Arielle, Davachi, Lila, 2013. Persistence of hippocampal multivoxel patterns into postencoding rest is related to memory. Proc. Natl. Acad. Sci. 110 (48), 19591–19596. https://doi.org/10.1073/pnas.1308499110. ISSN 1091-6490.

Tambini, Arielle, Davachi, Lila, 2019. Awake reactivation of prior experiences consolidates memories and biases cognition. Trends Cogn. Sci. 23 (October(10)), 876–890. https://doi.org/10.1016/j.tics.2019.07.008. ISSN 1364-6613.

Tambini, Arielle, Ketz, Nicholas, Davachi, Lila, 2010. Enhanced brain correlations during rest are related to memory for recent experiences. Neuron 65 (2), 280–290. https://doi.org/10.1016/j.neuron.2010.01.001. ISSN 0896-6273.

Tambini, Arielle, Berners-Lee, Alice, Davachi, Lila, 2017. Brief targeted memory reactivation during the awake state enhances memory stability and benefits the weakest memories. Sci. Rep. 7 (15325) https://doi.org/10.1038/s41598-017-15608-x. ISSN 2045-2322.

Tang, Wenbo, Jadhav, Shantanu P., 2019. Sharp-wave ripples as a signature of hippocampal-prefrontal reactivation for memory during sleep and waking states. Neurobiol. Learn. Mem. 160, 11–20. https://doi.org/10.1016/j.nlm.2018.01.002. ISSN 1074-7427.

Tang, Wenbo, Shin, Justin D., Frank, Loren M., Jadhav, Shantanu P., 2017. Hippocampal-prefrontal reactivation during learning is stronger in awake compared with sleep states. J. Neurosci. 37 (49), 11789–11805. https://doi.org/10.1523/JNEUROSCI.2291-17.2017. ISSN 0270-6474.

Tang, Wenbo, Shin, Justin D., Jadhav, Shantanu P., 2021. Multiple time-scales of decision making in the hippocampus and prefrontal cortex. eLife 10, e66227. https://doi.org/10.7554/eLife.66227. ISSN 2050-084X.

Tang, Yong, Nyengaard, Jens R., De Groot, Didima M.G., Gundersen, Hans Jørgen G., 2001. Total regional and global number of synapses in the human brain neocortex. Synapse 41 (3), 258–273. https://doi.org/10.1002/syn.1083.

Tesauro, Gerald, 1995. Temporal difference learning and TD-gammon. CACM 38 (3), 58–68. https://doi.org/10.1145/203330.203343. ISSN 0001-0782.

Tolman, Edward C., 1926. A behavioristic theory of ideas. Psychol. Rev. 33 (5), 352. https://doi.org/10.1037/h0070532.

Tolman, Edward C., 1938. The determiners of behavior at a choice point. Psychol. Rev. 45 (1), 1–41. https://doi.org/10.1037/h0062733. ISSN 0033-295X.

Tolman, Edward C., 1948. Cognitive maps in rats and men. Psychol. Rev. 55 (4), 189–208. https://doi.org/10.1037/h0061626. ISSN 0033-295X.

Tompary, Alexa, Davachi, Lila, 2017. Consolidation promotes the emergence of representational overlap in the hippocampus and medial prefrontal cortex. Neuron 96 (1), 228–241. https://doi.org/10.1016/j.neuron.2017.09.005.

Trettel, Sean G., Trimper, John B., Hwaun, Ernie, Fiete, Ila R., Colgin, Laura Lee, 2019. Grid cell co-activity patterns during sleep reflect spatial overlap of grid fields during active behaviors. Nat. Neurosci. 22 (4), 609–617. https://doi.org/10.1038/s41593-019-0359-6. ISSN 1546-1726.

van de Ven, Gido M., Tolias, Andreas S., 2018. Generative Replay with Feedback Connections as a General Strategy for Continual Learning (September) arXiv e-prints, page arXiv:1809.10635, URL https://ui.adsabs.harvard.edu/abs/2018arXiv180910635V.

van de Ven, Gido M., Trouche, Stéphanie, McNamara, Colin G., Allen, Kevin, Dupret, David, 2016. Hippocampal offline reactivation consolidates recently formed cell assembly patterns during sharp wave-ripples. Neuron 92 (5), 968–974. https://doi.org/10.1016/j.neuron.2016.10.020. ISSN 0896-6273.

van de Ven, Gido M., Siegelmann, Hava T., Tolias, Andreas S., 2020. Brain-inspired replay for continual learning with artificial neural networks. Nat. Commun. 11 (4069) https://doi.org/10.1038/s41467-020-17866-2.

van Hasselt, Hado, Hessel, Matteo, Aslanides, John, 2019. When to Use Parametric Models in Reinforcement Learning? (June) arXiv e-prints, art. arXiv:1906.05243.

van Seijen, Harm, Sutton, Rich, 2015. A deeper look at planning as learning from replay. In: International Conference on Machine LearningInternational Conference on Machine Learning, 37, pp. 2314–2322. In: URL http://proceedings.mlr.press/v37/vanseijen15.html.

Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N., Kaiser, Lukasz, Polosukhin, Illia, 2017. Attention is All You Need (June) arXiv e-prints, page arXiv:1706.03762.

Vaz, Alex P., Inati, Sara K., Brunel, Nicolas, Zaghloul, Kareem A., 2019. Coupled ripple oscillations between the medial temporal lobe and neocortex retrieve human memory. Science 363 (6430), 975–978. https://doi.org/10.1126/science.aau8956. ISSN 1095-9203.

Vaz, Alex P., Wittig, John H., Inati, Sara K., Zaghloul, Kareem A., 2020. Replay of cortical spiking sequences during human memory retrieval. Science 367 (6482), 1131–1134. https://doi.org/10.1126/science.aba0672. ISSN 1095-9203.

Vértes, Eszter, Sahani, Maneesh, 2019. A neurally plausible model learns successor representations in partially observable environments. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 32. Curran Associates, Inc, pp. 13714–13724. URL http://papers.nips.cc/paper/9522-a-neurally-plausible-model-learns-successor-representations-in-partially-observable- environments.pdf.

Wang, Shaoming, Feng, Samuel Franklin, Bornstein, Aaron, 2020. Mixing memory and desire: how memory reactivation supports deliberative decision-making. PsyArXiv. https://doi.org/10.31234/osf.io/5vksj.

Wang, Ziyu, Bapst, Victor, Heess, Nicolas, Mnih, Volodymyr, Munos, Remi, Kavukcuoglu, Koray, de Freitas, Nando, 2016. Sample Efficient Actor-Critic with Experience Replay (November) arXiv e-prints, art. arXiv:1611.01224.

Watkins, J.C.H. Christopher, Dayan, Peter, 1992. Q-learning. Mach. Learn. 8 (3–4), 279–292. https://doi.org/10.1007/BF00992698.

Wayne, Greg, Hung, Chia-Chun, Amos, David, Mirza, Mehdi, Ahuja, Arun, Grabska-Barwinska, Agnieszka, Rae, Jack, Mirowski, Piotr, Leibo, Joel Z., Santoro, Adam, Gemici, Mevlana, Reynolds, Malcolm, Harley, Tim, Abramson, Josh, Mohamed, Shakir, Rezende, Danilo, Saxton, David, Cain, Adam, Hillier, Chloe, Silver, David, Kavukcuoglu, Koray, Botvinick, Matt, Hassabis, Demis, Lillicrap, Timothy, 2018. Unsupervised Predictive Memory in a Goal-Directed Agent (March) arXiv e-prints, page arXiv:1803.10760, URL https://ui.adsabs.harvard.edu/abs/2018arXiv180310760W.

Whittington, James C.R., Muller, Timothy H., Mark, Shirley, Chen, Guifen, Barry, Caswell, Burgess, Neil, Behrens, Timothy E.J., 2020. The Tolman-Eichenbaum machine: Unifying space and relational memory through generalization in the hippocampal formation. Cell 183 (5), 1249–1263.e23. https://doi.org/10.1016/j.cell.2020.10.024. ISSN 0092-8674.

Wikenheiser, Andrew M., Redish, Aaron David, 2013. The balance of forward and backward hippocampal sequences shifts across behavioral states. Hippocampus 23 (1), 22–29. https://doi.org/10.1002/hipo.22049.

Wikenheiser, Andrew M., Redish, Aaron David, 2015a. Decoding the cognitive map: ensemble hippocampal sequences and decision making. Curr. Opin. Neurobiol. 32, 8–15. https://doi.org/10.1016/j.conb.2014.10.002. ISSN 0959-4388.

Wikenheiser, Andrew M., Redish, Aaron David, 2015b. Hippocampal theta sequences reflect current goals. Nat. Neurosci. 18 (2), 289–294. https://doi.org/10.1038/nn.3909.

Wikenheiser, Andrew M., Schoenbaum, Geoffrey, 2016. Over the river, through the woods: Cognitive maps in the hippocampus and orbitofrontal cortex. Nat. Rev. Neurosci. 17 (8), 513–523. https://doi.org/10.1038/nrn.2016.56. ISSN 1471-0048.

Wikenheiser, Andrew M., Marrero-Garcia, Y., Schoenbaum, G., 2017. Suppression of ventral hippocampal output impairs integrated orbitofrontal encoding of task structure. Neuron 95 (5), 1197–1207.e3. https://doi.org/10.1016/j.neuron.2017.08.003.

Wilson, Matthew A., McNaughton, Bruce L., 1994. Reactivation of hippocampal ensemble memories during sleep. Science 265 (5172), 676–679. https://doi.org/10.1126/science.8036517. ISSN 1095-9203.

Wilson, Robert C., Takahashi, Yuji K., Schoenbaum, Geoffrey, Niv, Yael, 2014. Orbitofrontal cortex as a cognitive map of task space. Neuron 81 (2), 267–279. https://doi.org/10.1016/j.neuron.2013.11.005. ISSN 0896-6273.

Wimmer, Elliott G., Büchel, Christian, 2019. Learning of distant state predictions by the orbitofrontal cortex in humans. Nat. Commun. 10 (1) https://doi.org/10.1038/s41467-019-10597-z. ISSN 2041-1723.

Wimmer, G. Elliott, Büchel, Christian, 2020. Reactivation of pain-related patterns in the hippocampus from single past episodes relates to successful memory-based decision making. bioRxiv. https://doi.org/10.1101/2020.05.29.123893.

Wittkuhn, Lennart, Schuck, Nicolas W., 2021. Dynamics of fMRI patterns reflect sub-second activation sequences and reveal replay in human visual cortex. Nat. Commun. 12 (1795) https://doi.org/10.1038/s41467-021-21970-2.

Wolosin, Sasha M., Zeithamova, Dagmar, Preston, Alison R., 2012. Reward modulation of hippocampal subfield activation during successful associative encoding and retrieval. J. Cogn. Neurosci. 24 (7), 1532–1547. https://doi.org/10.1162/jocn_a_00237.

Wood, Emma R., Dudchenko, Paul A., Robitsek, R. Jonathan, Eichenbaum, Howard, 2000. Hippocampal neurons encode information about different types of memory episodes occurring in the same location. Neuron 27 (3), 623–633. https://doi.org/10.1016/s0896-6273(00)00071-4. ISSN 0896-6273.

Wu, Chun-Ting, Haggerty, Daniel, Kemere, Caleb, Ji, Daoyun, 2017. Hippocampal awake replay in fear memory retrieval. Nat. Neurosci. 20 (4), 571–580. https://doi.org/10.1038/nn.4507. ISSN 1546-1726.

Wu, Qingyang, Lan, Zhenzhong, Gu, Jing, Yu, Zhou, 2020. Memformer: The Memory-Augmented Transformer (October) arXiv e-prints, page arXiv:2010.06891, URL https://ui.adsabs.harvard.edu/abs/2020arXiv201006891W.

Yu, Jai Y., Frank, Loren M., 2015. Hippocampal-cortical interaction in decision making. Neurobiol. Learn. Mem. 117, 34–41. https://doi.org/10.1016/j.nlm.2014.02.002. ISSN 1074-7427.

Yu, Jai Y., Liu, Daniel F., Loback, Adrianna, Grossrubatscher, Irene, Frank, Loren M., 2018. Specific hippocampal representations are linked to generalized cortical representations in memory. Nat. Commun. 9 (2209) https://doi.org/10.1038/s41467-018-04498-w. ISSN 2041-1723.

Zhang, Hui, Deuker, Lorena, Axmacher, Nikolai, 2017. Replay in humans - first evidence and open questions. In: Axmacher, N., Rasch, B. (Eds.), Cognitive Neuroscience of Memory Consolidation. Springer, pp. 251–263. https://doi.org/10.1007/978-3-319-45066-7_15.

Zhang, Hui, Fell, Juergen, Axmacher, Nikolai, 2018. Electrophysiological mechanisms of human memory consolidation. Nat. Commun. 9 (4103) https://doi.org/10.1038/s41467-018-06553-y. ISSN 2041-1723.

Zhang, Shangtong, Sutton, Richard S., 2017. A Deeper Look at Experience Replay (December) arXiv e-prints, art. arXiv:1712.01275.

Zielinski, Mark C., Tang, Wenbo, Jadhav, Shantanu P., 2020. The role of replay and theta sequences in mediating hippocampal-prefrontal interactions for memory and cognition. Hippocampus 30 (1), 60–72. https://doi.org/10.1002/hipo.22821. ISSN 1050-9631.