THE UNIVERSITY
of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

A neurocomputational model for learning, memory consolidation and schemas

Nathalie Dupuy

Doctor of Philosophy
Institute for Adaptive and Neural Computation
School of Informatics
University of Edinburgh
2018

Abstract

This thesis investigates how through experience the brain acquires and stores memories, and uses these to extract and modify knowledge. This question is being studied by both computational and experimental neuroscientists as it is of relevance for neuroscience, but also for artificial systems that need to develop knowledge about the world from limited, sequential data. It is widely assumed that new memories are initially stored in the hippocampus, and later are slowly reorganised into distributed cortical networks that represent knowledge. This memory reorganisation is called systems consolidation. In recent years, experimental studies have revealed complex hippocampal-neocortical interactions that have blurred the lines between the two memory systems, challenging the traditional understanding of memory processes. In particular, the prior existence of cortical knowledge frameworks (also known as schemas) was found to speed up learning and consolidation, which seemingly is at odds with previous models of systems consolidation. However, the underlying mechanisms of this effect are not known.

In this work, we present a computational framework to explore potential interactions between the hippocampus, the prefrontal cortex, and associative cortical areas during learning as well as during sleep. To model the associative cortical areas, where the memories are gradually consolidated, we have implemented an artificial neural network (Restricted Boltzmann Machine) so as to get insight into potential neural mechanisms of memory acquisition, recall, and consolidation.

We analyse the network's properties using two tasks inspired by neuroscience experiments. The network gradually built a semantic schema in the associative cortical areas through the consolidation of multiple related memories, a process promoted by hippocampal-driven replay during sleep. To explain the experimental data we suggest that, as the neocortical schema develops, the prefrontal cortex extracts characteristics shared across multiple memories. We call this information meta-schema. In our model, the semantic schema and meta-schema in the neocortex are used to compute consistency, conflict and novelty signals. We propose that the prefrontal cortex uses these signals to modulate memory formation in the hippocampus during learning, which in turn influences consolidation during sleep replay.

Together, these results provide theoretical framework to explain experimental findings and produce predictions for hippocampal-neocortical interactions during learning and systems consolidation.

Acknowledgements

I was told the acknowledgements should be funny. But funny how?

I would like to start by expressing my gratitude to Mark, a.k.a. my supervisor. Thank you for your patience (I will eventually watch Goodfellas), your guidance (which pub should we go next), and understanding (what did you say about the use of brackets?).

I would also like to give thanks to my supervisors in Stockholm, Pawel and Anders. Thank you for your warm welcome and your help during my stay. My special thanks go to the researchers in the CCNS lab, Lisa, Dorothy and Richard. Your input was very valuable. Finally, I would like to thank the Eurospin Erasmus Mundus and the DTC for supporting my research, and for giving me the opportunity to attend amazing conferences.

But I would not be writing these lines if it were not for my super hero Katwoman. Thank you Katharina for being simply awesome.

I would like to thank my former partner in crime Emma. Thank you for teaching me the subtleties of the Irish language, although I wish I never knew the truth about Mickey Mouse.

I also would like to thank my fantastic office crew: Poster Boy Joe (thanks for the edit!), lovely and proud Catalan Alba, the invincible Ninja Emilia, and Genius Martino.

Thank you Akash for the shot of adrenaline that helped me get through the last months of this PhD. Thank you for reminding me that not everything is a problem.

Thank you Cole for checking in on me this last crucial month. Thumbs up. I owe you my intro by the way.

I would like to thank my companions of the epic adventure on Tiree island, where cows look like lions – Rai, Pippa (python), Janie, Jenny, Pablo, Emmy and Arlene.

My thanks to many more wonderful friends who have accompanied along the way – Simona, Paul, Wioletta, Ksenia, Duncan... and the DTC / CDT & co.

Thank you to my Forro crews in Stockholm and Edinburgh for bringing some sunshine from Brazil into my life. Thank you Laura for the edit!

I probably forgot people but it is 6AM and I must get some sleep to submit tomorrow. If I did, I will apologise with a bottle of wine (that is the French way).

Last, thank you my imaginary friend for the very fruitful discussion (although I would have appreciated a bit more of feedback). Thank you Tesco for (almost) never

failing to deliver my egg&cress sandwich, and thank you Machina Espresso for filling my cups of coffee.

Et enfin, merci Caroline de m'avoir fait découvrir que les pommes de terre existent aussi en Suède, merci Geneviève et Marc d'avoir porté mes valises jusqu'au bout du monde. Merci à vous trois d'avoir patienté que j'atteigne l'âge de la sagesse (si celle-ci s'atteint jamais) et merci de votre soutien infaillible. Promis, la prochaine destination, c'est les tropiques.

Declaration

I declare that this thesis was composed by myself and that the work contained herein is my own except where explicitly stated otherwise in the text. This work has only been submitted for a joint PhD degree between KTH, Stockholm and the University of Edinburgh within the regulations of the Erasmus Mundus EuroSPIN programme and for no other degree or professional qualification except as specified.

(*Nathalie Dupuy*)

Contents

Chapter 1

Introduction

What would you do if you saw a pink alien mammoth walking towards you?

The input the brain receives is a mix of relevant and irrelevant information, and importantly, like here, sometimes unknown information. Nevertheless, the brain has to somehow make sense of the input data. Fortunately, we do not memorise plentiful isolated facts, but instead we build mental schemas upon our experiences which allow us to confront new situations. These mental schemas are believed to help us to understand incoming information, and then help us to decide what is presumably the best course of action.

Besides guiding behaviour in situations of uncertainty, mental schemas also impact the way we learn new information. For instance, after seeing the pink alien mammoth only once, we rapidly create a new memory model of this new species, because we transfer our prior knowledge about related species (e.g. elephant). While it is clear that we assimilate new information faster when it relates to past experiences, it is unclear how the brain can incorporate new information quickly, and how the brain knows if it is necessary to do so. Indeed, the brain often learns from sparse data, but it would be risky to always learn rapidly as the brain might assimilate unreliable information, or worse, it might overwrite crucial prior information.

This thesis aims to provide a computational framework to examine these questions. In particular, we investigate potential neural mechanisms underlying the gradual acquisition of mental schemas across sparse data, and how these schemas may influence memory processing in the brain. In this chapter, we first review the relevant literature in neuroscience for memory consolidation (Section 1.1.1), sleep replay (Section 1.1.2), and memory schemas (Section 1.1.2). Next, we briefly describe computational frameworks for modelling memory acquisition and consolidation (Section 1.2). We conclude this chapter with the main objectives and outline the thesis (Section 1.3).

1.1 Declarative memory in neuroscience

Long-term memory in the brain is traditionally divided into two categories. The first category, which is the focus of this work, is called declarative or explicit memory. It includes memories of facts, so called semantic memory, such as general knowledge about elephants, as well as memories of events or episodic memory (e.g. an elephant charged you while on a safari in South Africa), a distinction introduced by Tulving (1972). By contrast, the second memory category is used without being aware, and thus is called implicit memory. Examples include procedural memory, i.e. perceptual and motor skills, conditioning, and habits.

Early studies in humans showed that declarative memories are initially vulnerable to disruption, but gradually stabilise over time; researchers then concluded that memories must be reorganised over time, and they named this process memory consolidation (Müller and Pilzecker 1900; Lechner et al. 1999). The idea that memories need time to be stabilised had already been reported in 1881 by the psychologist Ribot. He found that patients with brain trauma forgot most memories acquired not long before the injury, but remembered most memories acquired long before. This temporal gradient of memory loss, which became known as the Ribot's law, was later linked to a specific region in the brain, namely the hippocampal region. In the mid-twentieth century, the pioneering work of Scoville and Milner on patient H.M., and the clinical studies on retrograde amnesia that followed, revealed that damage to the hippocampal region impaired the recall of recent memories, while remote memories were usually intact.

Animal studies confirmed the temporal gradient observed in amnesic patients, and provided more evidence that the hippocampal region plays a central, but temporary role in memory recollection (early studies include: object discrimination in monkeys, Zola-Morgan & Squire 1990; in rodents, social transmission of food preference, Winocur 1990, and contextual fear conditioning, Kim & Fanselow 1992).

These observations led to the standard cortico-hippocampal interaction model for systems consolidation: the acquisition, initial storage and recollection of memory require the hippocampal region, but over time memory is reorganised and permanently stored in the neocortex, so that the hippocampal region is no longer needed to support recall.

However, the numerous studies on retrograde amnesia and damage to the hippocampal region reported a wide range of temporal gradients and conflicting results: the range of temporal gradients found in humans spreads from a few years to decades,

in animals it spreads from a few days to weeks, and sometimes researchers did not find any temporal gradient at all. Although this question is presently still debated, many factors have been put forward to explain the discrepancies in results, besides the species tested and the extent of lesions. For instance, because the hippocampal formation is essential for navigation and contextual information, damage or inhibition of this region was found to impair both recent and remote memories in studies that involve spatial or context discrimination tasks (e.g. context fear conditioning in rodents, Wiltgen et al. 2010). One theory, the transformation hypothesis (previously multiple-trace theory, Nadel & Moscovitch 1997, Winocur & Moscovitch 2011), argues that the hippocampus is always required to recollect past memories with spatial and contextual details, and that only for semantic memories the hippocampus has a temporary role. This theory is nonetheless contested by evidence showing that rodents successfully performed a spatial memory task when the hippocampus was lesioned after weeks of training (Tse et al. 2007), and that some amnesic patients with extensive lesions in the hippocampal region had intact, but less detailed, remote spatial memory (e.g. Rosenbaum et al. 2000), or even intact remote autobiographical memories (Bayley et al. 2003).

The diversity of temporal gradients observed in retrograde amnesia can also be the result of memories being more or less rapidly consolidated. Indeed, recent studies have revealed complex interactions between the hippocampus and the neocortex which start at the time of acquisition, and have identified factors that influence the speed of memory consolidation (e.g. reward, sleep, schemas). Even though the two explanations are not mutually exclusive, we only investigate the latter one in this thesis. In the following subsections we briefly summarise experimental work that supports this view and justify the choices we have made for our computational model.

Important note Throughout this work, “memory consolidation” will refer to systems consolidation, the gradual reorganisation of memory across brain regions over days to years. This kind of consolidation must not be confused with synaptic consolidation, which involves cellular mechanisms, protein synthesis in particular, to fix memory within synapses and occurs over minutes to hours. Likewise, a “consolidated memory” will refer to a memory that can be recalled independently of the hippocampus.

1.1.1 Interaction between the hippocampus and neocortex during memory consolidation

This section briefly presents evidence of the reorganisation of memories across hippocampal and neocortical networks during memory consolidation. For reviews of the topic, see Frankland & Bontempi (2005), Genzel & Wixted (2017).

Formation of memories in the hippocampus and neocortex

In the contemporary understanding of declarative memory processes, the medial temporal lobe (MTL) is necessary for the acquisition of memories, and hence plays a fundamental role to later establish long-term memories. Damage to the MTL, as in H.M. was indeed found to severely impair the capacity to form new declarative memories (Milner et al. 1968). The MTL, which comprises the hippocampus, entorhinal, perirhinal, and parahippocampal cortices, has a hierarchical organisation which is believed to progressively integrate information scattered across higher-order associational cortices (Lavenex & Amaral 2000). Thus, it is widely assumed that the hippocampus binds the neural representations of sensory stimuli that can compose a memory (e.g. visual input, scents, sounds...). These representations are encoded in distributed areas in the neocortex and are usually not connected to each other initially; therefore, the hippocampus is believed to mediate memory recall while the memories are not yet consolidated.

To support this view, Tanaka et al. (2014) have monitored neurons that were recruited in the cortex and CA1 region of the hippocampus in rodents during learning of a fear conditioning paradigm. The authors have shown that silencing the hippocampal cells in CA1 during retrieval selectively disrupted the reactivation of patterns in cortical regions (entorhinal cortex and amygdala), and prevented memory retrieval.

However, Genzel & Wixted (2017) stress that memories are not literally “transferred” from the hippocampus to the cortex. Instead, studies suggest that upon learning there is already a memory trace in the neocortex, albeit it is too weak to support recall on its own. For instance, Cowansage et al. (2014) trained mice in a context fear conditioning task and found that, shortly after training, optogenetic stimulation of the retrosplenial cortex could successfully elicit freezing behaviour despite hippocampal inactivation. However, mice did not freeze in response to natural sensory stimulation (context exposure) when the hippocampus was inactivated. Their findings imply that the recall pathway was formed early in the neocortex, but was not yet able to drive behaviour and hence the hippocampus was required to support memory retrieval.

Consolidation of long-term memories in the neocortex

The general consensus posits that the involvement of the hippocampus is only temporary (but as mentioned earlier, "temporary" can be days up to years!), and long-term memories are thought to be sustained solely by the distributed cortical areas. During (systems) consolidation, the cortical-cortical connections are presumably strengthened, and thus, representations distributed in neocortical areas eventually become interconnected so that memories no longer require the hippocampus to be recalled. Experimental studies have investigated the relative engagement of the hippocampus and cortical areas during memory retrieval in humans and animals (for reviews, see Frankland & Bontempi 2005, McKenzie & Eichenbaum 2011), and have found evidence that the hippocampus is most active during the recall of recent memories, but seemingly becomes less active over time; at the same time activity increased in various cortical areas during recall of remote memories. Studies have also shown that functional connectivity was changed: while it decreased between the hippocampus and cortical areas, it increased between cortical networks.

Nonetheless, the consolidation of memories in the neocortex does not mean they are stable and immutable. Indeed, we are usually exposed to new information that is related to past experience, and hence established cortical circuits might change to incorporate newly learned material. A common idea posits that when new experience overlaps with prior knowledge, consolidated memories can be reactivated and return to a labile state, and can either be erased or "reconsolidated" (Nader et al. 2000); the latter mechanism has been proposed to mediate the reorganisation of cortical networks to assimilate new memories (for a review see McKenzie & Eichenbaum 2011). However, it remains unclear how the brain reconsolidates remote, hippocampal-independent memories in the neocortex, and whether the hippocampus is involved in this process.

1.1.2 Neural replay during wake and sleep

Neurons change their connectivity with other neurons via synaptic plasticity (Martin & Morris 2002). As a simplified summary, when neurons tend to co-activate in response to a stimulus, their connection strength increases (long-term potentiation or Hebbian plasticity), but if their activity is not synchronised their connection strength decreases (long-term depression).

However, exposure to specific sensory input from the external world is limited (say, you are attending a math class, or observing a pink elephant in the savanna). You may

still repeat the experience (study the math lesson again), but when learning something related (say, you attend a second class of math), you are not exposed to the exact, original stimulus. Therefore, to nevertheless have a strong memory, neuroscientists have hypothesised that the brain 'replays' neural activity patterns that mimic sensory experience, so that synapses can be progressively strengthened or weakened. They propose this neural replay as a mechanism underlying memory consolidation and reorganisation across hippocampal-neocortical networks.

Hippocampal replay during sharp wave-ripples (SPW-Rs)

In line with this theory, experimental studies in rats, birds, monkeys, and humans have found that patterns of brain activity related to a specific task could re-occur later on even in the absence of the stimulus. In particular, spontaneous reactivations of neuronal ensembles have been identified during sharp wave-ripples (SPW-Rs) in the hippocampus (for a review, see Buzsáki 2015). SPW-Rs are events of highly synchronised network activity that can be recorded in the local field potential in the hippocampus. They occur while the animal is awake and immobile, but also during sleep (non-REM). SPW-R complexes have been found in mice, rats, rabbits, monkeys and humans; however most studies focus on hippocampal place cells activity in animals, as they have particular spatiotemporal activity patterns that are well suited to understand replay.

Hippocampal place cells fire when the animal is in specific locations, and hence their successive activations form sequences that reflect the trajectory of the animal. The sequential structure of these place cells activity is preserved during SPW-R reactivation, albeit on a compressed time scale; thus, these hippocampal bursts of activity are hypothesised to be rapid 'replay' events, compressed versions of the neuronal activity observed during experience. Replay during SPW-R was indeed found to echo the sequences observed while the animal explored the environment (forward or backward), but replay could also contain sequences that did not correspond to experience, suggesting that these might reflect past or potential future experiences (Gupta et al. 2010, but see Chadwick et al. 2015). Therefore, replay in the hippocampus has been predicted to play a fundamental role for imagining, learning and modifying cognitive maps (see Buzsáki 2015).

When the animal temporarily stops exploration and is immobile but awake, neural replay during SPW-Rs is suggested to help to recall, predict, and plan trajectories to guide navigational behaviour (Pfeiffer & Foster 2013), as well as to support learning. In line with this, Jadhav et al. (2012) interrupted SPW-Rs with electrical stimulation

while rodents learned a spatial task, and found that it impaired performance. More recently, Roux et al. (2017) found that such disruption (optogenetic silencing) prevented memory stabilisation of spatial representations and maintenance of cognitive maps. We discuss the role of SPW-Rs during sleep next.

Hippocampal-cortical replay during sleep: dialogue for memory consolidation

During 'off-line' periods, hippocampal SPW-R events were found to occur during non-REM sleep. This period of sleep is characterised by the presence of cortical slow oscillations, synchronous events that spread across cortical areas, in contrast to REM sleep which is characterised by desynchronised dynamics (for review on sleep, see Genzel et al. 2014). Importantly, hippocampal SPW-Rs are coupled to these cortical slow oscillations. In addition, studies have suggested that replay of experiences also occurred in cortical areas, since activity patterns during sleep were similar to the task-related activity patterns observed during training, for example, in the visual cortex (Ji & Wilson 2007), the auditory cortex (Rothschild et al. 2017), the prefrontal cortex (Peyrache et al. 2009), and rapid replay of sequences was also recorded in the rat medial prefrontal cortex (Euston et al. 2007).

In particular, hippocampal SPW-R-related replay has been hypothesised to establish a dialogue with the neocortex during non-REM, slow-wave sleep, to support consolidation and formation of long-term memories in distributed neocortical areas. Evidence that SPW-Rs impact consolidation was shown for example by Girardeau et al. (2009) and Ego-Stengel & Wilson (2010). In these two studies, they suppressed SPW-Rs during sleep after rats learned a hippocampal-dependent spatial memory task, thereby blocking potential replay of place-cell sequences, and observed that performance was impaired.

Furthermore, experiments in animals have revealed that there is a correlation between hippocampal SPW-Rs and activity in multiple cortical areas (Logothetis et al. 2012), supporting the idea that synchronised replay in the hippocampus and cortex mediates memory consolidation. Ji & Wilson (2007) showed that such coordinated replay occured between the visual cortex and hippocampus in rats. In this study they recorded firing patterns during slow wave sleep in both structures, and they found that patterns reactivated during sleep were coordinated and coherent with patterns recorded while the animals actively explored a maze. Similarly, Peyrache et al. (2009) found coordinated replay during slow wave sleep between the medial prefrontal cortex and the hippocampus. They trained rats on a task in a maze with a specific rule, and then

changed the rule in the maze. They showed that during subsequent sleep, the neural activity patterns that occurred when animals learned the new rule were predominantly replayed in the prefrontal cortex and coincided with SPW-Rs in the hippocampus.

Direct evidence about the role of the dialogue between hippocampus and neocortex for memory consolidation during sleep was shown in a recent study by Maingret et al. (2016). The authors reported that memory consolidation in rats was associated with increased temporal coupling between the hippocampus and medial prefrontal cortex during slow-wave sleep. Following this observation, they applied electrical stimulation to the neocortex that was synchronised to SPW-Rs so as to boost this coupling. They found that the prefrontal cortex was subsequently more responsive to the task, and this result was accompanied by increased recall performance.

Organisation of replay and influential factors

Despite the evidence of a hippocampo-cortical coupling during sleep it is not clear where the reactivations start and what orchestrates them, and studies suggest a bidirectional dialogue between the two structures. It is commonly assumed that the hippocampus coordinates replay in the cortex because of its role in memory formation and recall (Section 1.1.1), but also because blocking hippocampal replay during during SPW-R impairs memory consolidation (cf. previous paragraph). However, recent studies suggest otherwise. For instance, Rothschild et al. (2017) has recently shown that replay during sleep seemingly originates from neocortical areas. The authors monitored patterns of activation in the hippocampus (CA1) and the auditory cortex as rodents learned a sound-guided task, and recorded the timing of reactivations in the two brain stuctures during subsequent non-REM sleep. They observed a cortical-hippocampal-cortical loop of information flow during a time window centered around hippocampal SPW-Rs: the loop began with cortical reactivation, which predicted hippocampal patterns of activity during the following SPW-R, which in turn predicted subsequent reactivation in the auditory cortex. The idea that during sleep hippocampal SPW-Rs are influenced by cortical reactivation preceding the SPW-R complex is further supported by Bendor & Wilson (2012), who successfully biased the content of hippocampal replay in rats during sleep by presenting the same auditory stimuli as during training.

Recent theories have also argued that sleep facilitates 'memory triage' (Stickgold & Walker 2013), as it appears that after sleep relevant memories are better remembered

than irrelevant ones. Such proposal is in line with recent evidence that neuromodulation, e.g. in response to novelty or reward, can influence which memory traces are subsequently replayed (Atherton et al. 2015). Stickgold & Walker (2013) suggest that memories are attached to "salience tags" at the time of learning, and thus memories can later be selectively consolidated, in particular during sleep replay. Consistent with the theories, McNamara et al. (2014) have used optogenetics to stimulate midbrain dopaminergic neurons that projected to the hippocampus while mice learned specific rewarded locations. During subsequent sleep/rest, they observed that hippocampal memory patterns related to experience were more often reactivated, and performance was improved during later recall. Similarly, Igloi et al. (2015) have found that three months after learning an associative memory task, participants who had a nap after the task better remembered pictures that were associated with high reward outcome than low reward outcome, but participants who did not sleep after the task did not show such distinction (and overall their performance was lower). Furthermore, the authors found a correlation between performance of highly rewarded memories and sleep replay mechanisms (number of sleep spindles). These results suggest that the brain prioritises processing of memories with high relevance during sleep.

Summary: neural replay

Replay of experiences, wherein specific, coherent patterns of neural activity are reactivated, has been hypothesised to mediate memory retrieval and consolidation. Such events occur when the brain is at rest during wake, or during sleep, and reactivations spread brain-wide via hippocampo-cortical coupling, suggesting a transfer of information between memory systems. Spontaneous reactivations of neural patterns sometimes reflect recently acquired information, but when they do not, we may suppose that they represent various past memories or reflect imagining new scenarios. Importantly, recent theories suggest that memory replay is organised in such a way that it promotes the integration of relevant experiences. Because of these properties, replay and in particular replay during sleep has been hypothesised to underlie not only consolidation of specific, task-related memory traces, but also to underlie the extraction of structured representations, the topic of the next section.

1.1.3 Theory of cognitive schemas

In cognitive psychology Piaget (1926) and Bartlett (1932) introduced the idea of mental frameworks, called "schemas", that guide the way we interpret the world and learn about it. They suggested that schemas can be activated to support inference, and that a schema continuously develops by either assimilating new consistent information, or by adjusting to be more conform to the new observation. Hence, studies have proposed that schemas are the basis of cognitive development, in particular since young children already demonstrate the ability of inductive reasoning. For instance, Smith et al. (2002) showed that when two-year-old children learned names of various objects, they were also able to learn what characterised each object category, without explicit indication. In their experiment, the children were presented with a training set comprising four groups of unknown objects with two exemplars in each group. Critically, exemplars of the same category had the same shape, but different texture and color. Next, they presented an exemplar object from one group, along with three other tests objects which matched one of the properties of the exemplar object, i.e. either the shape, the texture, or the color. The children were able to recognise the correct object with the same shape. More importantly, children generalised the structure common to the four categories to new objects categories, which allowed them to rapidly learn new object names.

Building schemas

Leading theories suggest that the brain builds schemas by extracting commonalities and salient information across related episodes, probably through the reorganisation of memories during consolidation. For instance, Richards et al. (2014) showed that, over time, rodents learned the statistical structure underlying multiple experiences in a water-maze memory task, where the escape platform was more likely in one area of the pool than elsewhere (the details of the experiments are discussed in Section 5.1.1). Throughout training, the animals tended to return to where they found the platform last, but weeks after the training their search behaviour matched the latent distribution of the platform locations.

Schema extraction relies on memory integration, a process by which related memories are encoded with overlapping patterns of activity (Schlichting & Preston 2015). In particular, the hippocampus presumably uses its remarkable spatio-temporal representations to organise memories into relational schemas. Such organisation has

been found for instance in rodents, which develop hierarchical memory representations in the hippocampus that capture the relative importance of the stimuli for the task (McKenzie et al. 2014).

Relational schemas in the hippocampus support rapid encoding of new related items (McKenzie et al. 2014), and they might coordinate replay and thereby could promote memory integration in neocortical areas. Specifically, Lewis & Durrant (2011) suggest that related memories are replayed together in the hippocampus, and subsequently reactivate their corresponding memories in the neocortex (Section 1.1.2). Thus, shared neural ensembles of cortical representations are strengthened, and the brain would gradually abstract schemas via this selective strengthening. Although this process could already occur during wake, Lewis & Durrant (2011) propose that it is primarily facilitated by replay during slow wave sleep. Similarly, this process would allow to incorporate new related memories as they overlap with existing representations, perhaps accompanied by reconsolidation (McKenzie & Eichenbaum 2011).

Learning with schemas

Once a schema is formed, it can influence subsequent learning. Tse et al. (2007) showed that it could also significantly speed up memory consolidation. Indeed, when rats had be trained to learn six odor-place associations in an arena, they learned two new associations in a single trial; critically, these new associations were consolidated in a time window between 24 and 48 hours, whereas for a naive animal it would probably take weeks to consolidate (the details of these experiments are in Section 3.1.1).

Why is learning fast with a schema? Similar to memory integration during the formation of schema, if a new memory relates to prior knowledge it will overlap with existing representations, and hence the new memory can be rapidly assimilated. It was suggested that new memories that are congruent with a schemas might rely less on MTL regions (van Kesteren et al. 2012), albeit in Tse et al. (2007) if the hippocampus was lesioned too early it prevented the formation of long term memories of the new associations.

Furthermore, sleep is hypothesised to facilitate the incorporation of new memories within established cortical networks (Lewis & Durrant 2011). Recent studies investigating the interaction between sleep and schemas found in particular that sleep spindles (markers of replay in the neocortex) can be associated with memory integration into prior knowledge (Tamminen et al. 2013, Hennies et al. 2016, Groch et al. 2017). Hence, Lewis & Durrant (2011) suggested that the speed of acquisition of

new schema-related memories might be influenced by the number of reactivations, the degree of overlap, and strength at encoding.

The role of the prefrontal cortex

In light of recent studies such as Tse et al. (2007), it has been suggested that the formation, and also recall, of long-term memories can not solely be explained by the interaction between the hippocampus and representational areas in the cortex. More and more experimental findings have pointed out that the prefrontal cortex is essential to both form and retrieve remote memories, albeit its exact role is still not well defined (Fernández 2017). For example, Tse et al. (2011) found that pharmacological blockade in the prefrontal cortex impaired the recall of the remote, consolidated odor-place associations.

The medial prefrontal cortex (mPFC) has been hypothesised to coordinate the activation of schemas (Ghosh & Gilboa 2014). For instance, a memory disorder following damage to the mPFC is confabulation, which seemingly results from an inappropriate use of schemas during recall (Ghosh & Gilboa 2014). In particular, the mPFC is thought to bias memory retrieval in the hippocampus so as to reactivate representations that are relevant in a given context (Preston & Eichenbaum 2013). While such interaction might guide behaviour, it was also suggested that the mPFC might influence integration of related memories in the hippocampus (Schlichting & Preston 2015). Evidence of coordinated replay between the two structures during wake and sleep further supports this hypothesis (Peyrache et al. 2009, Tang & Jadhav 2018).

Researchers have thus predicted that the mPFC plays a role during the acquisition of new memories related to a schema. To test this, van Kesteren et al. (2013) investigated the activity of the mPFC and MTL of participants while they learned associations between pictures of a scene, which supposedly had an associated schema (e.g. a tennis court), and a picture of an object that fitted more or less in the setting (e.g. an umbrella). Activity in the mPFC increased when the item was congruent with prior knowledge, and decreased when the item was less congruent, while the MTL showed the opposite activation trend. These results suggest that the mPFC may monitor the congruency of a new experience with the schema it relates to, according to which memory is processed differently.

In addition, Tse et al. (2011) inactivated the mPFC while rats learned new associations related to a pre-existing schema, and they found that recall was impaired 24h

later. This result suggest that the mPFC was critical at the time of learning for rapid memory acquisition. However, such result is not limited to new information that fits into a knowledge structure, but rather if it is relevant. For instance, in the experiment of Richards et al. (2014), mice learned very rapidly that a platform was located in one side of the water-maze while they had been previously trained for several days to swim to the other side of the pool (Section 5.1.1); yet, inhibition of the mPFC prevented this rapid learning, and mice revisited also the old locations. Bero et al. (2014) showed that optogenetic silencing of excitatory mPFC neurons during contextual fear conditioning impaired the activation in the entorhinal cortex and hippocampus, and also impaired recall both after 1 day and 28 days.

In summary, these results suggest that the prefrontal cortex is crucial to establish and retrieve long-term memories and that it might control memory encoding in the hippocampus.

1.2 Computational models of declarative memory

In this section we review computational frameworks that were developed to study declarative memory at the system-level. The first models presented are based on artificial neural networks and as such are usually referred to as 'connectionist' models. We particularly emphasise the model of Káli & Dayan (2004) that we used to build our framework, and the recent update of the Complementary Learning Systems Theory by McClelland (2013) to take the recent findings about schemas into account.

1.2.1 Neural networks models for memory consolidation

Connectionist models are artificial neural networks that are inspired by the brain architecture and explore how information could be processed by the nervous system. Very briefly, these networks are composed of multiple simplified neuron-like units interlinked by weighted connections. The neural activity is typically represented by a firing rate or a (probabilistic) binary variable. The connectivity is typically feedforward, or symmetric, so that the activity is guaranteed to settle in an attractor state. When judiciously wired, a network can thus recognise patterns of activity, and some networks can also reconstruct partial patterns (e.g. Hopfield 1982).

Although the units are not detailed models of real neurons, and although biological processes are greatly simplified, these artificial networks are useful to simulate and

gain insight into neural mechanisms that may underlie memory consolidation. We first review early computational models that investigated the interaction between the medial temporal lobe (MTL) and neocortex during memory consolidation, while Sections 1.2.2 and 1.2.3 describe more recent models.

Model for long-term memory and amnesia by Alvarez and Squire

Alvarez & Squire (1994) developed a simple network for studying long-term memory and amnesia. Their model consisted of three groups of units, with two groups representing two distinct cortical areas (8 units each) and one group representing the MTL region (4 units). Each unit in one area was connected to all units in other areas, but within each group the activations were regulated by winner-take-all inhibition. A memory was composed of two patterns of activity, one in each cortical area. However at the start of training the cortico-cortical connections were weak and, since the intracortical weights changed slowly, the network could not initially retrieve a memory given a partial input via this cortical recall route. On the other hand, the connections to MTL were rapidly changing, and hence the MTL quickly supported memory recall by binding the two cortical patterns: if one pattern was presented in one cortical region, the activity in the MTL helped restore the missing pattern in the other cortical area. Over time, the cortico-cortical connections strengthened via multiple presentations of the patterns in the cortical areas, and eventually recall was independent of MTL. Interestingly, Alvarez & Squire (1994) already incorporated the notion of replay to support consolidation: the patterns were presented only twice given external inputs, and subsequently repetitive presentations relied on random reactivations of activity in the MTL area. Therefore, the model of Alvarez & Squire (1994) provided a computational framework of the hippocampal-neocortical memory system which could account for retrograde amnesia.

The Complementary Learning Systems Theory

Contemporary and in parallel to Alvarez & Squire (1994), McClelland et al. (1995) developed a computational framework called the Complementary Learning Systems Theory (CLST). The prominent prediction of CLST model was that the brain needs distinct fast and slow learning systems to prevent catastrophic forgetting.

In McClelland et al. (1995) the hippocampus supports rapid learning, while the neocortex slowly incorporates memories. The neocortex was modelled with a multi-

layer, feed-forward neural network (Rumelhart network). This network architecture had more representational power compared to the simple model of Alvarez & Squire (1994), which allowed McClelland et al. (1995) to study the acquisition of knowledge and how the brain learns new items. Their network had to interleave the learning of multiple examples to discover knowledge, and hence learning in the neocortex was slow. In particular, McClelland et al. (1995) showed that slow, interleaved learning was crucial to incorporate new memories in the neocortex without overwriting existing knowledge.

However, in contrast with Alvarez & Squire (1994), the model of McClelland et al. (1995) was not explicit on the mechanisms by which the hippocampus interacts with the neocortex, and hence it is difficult to explore hippocampal contribution to memory consolidation.

The CLST framework has since been used as theoretical background for other computational models studying the roles of hippocampus and neocortex for recognition memory (recall versus familiarity, Norman & O'Reilly 2003), and autonomous and independent reactivations in the hippocampus and neocortex during REM sleep to reduce interference (Norman et al. 2005). Fiebig & Lansner (2014) extended the framework to a three-stage model for memory consolidation. The authors included the prefrontal cortex to mimic working memory and bridge the temporal gap (seconds to decades) between early memory acquisition to consolidation, but it had no role in mental schemas.

1.2.2 Model of hippocampal-neocortical interactions of Káli & Dayan

Káli & Dayan (2004) developed a model that shares the same principles as the model of Alvarez & Squire (1994) (Section 1.2.1), namely (i) the hippocampus for rapid storage, (ii) memories consolidated across distributed cortical areas, and (iii) off-line replay driven by the hippocampus. However, the model of Alvarez & Squire (1994) was extremely simple and hence its predictive power was limited. Thus Káli & Dayan (2004) implemented a more sophisticated architecture and data. In particular, the neural network has a hierarchical structure (as an illustration see Fig. 2.6) where three cortical sensory areas are connected to a cortical area in the MTL, which in turn is connected to the hippocampus. Hence, Káli & Dayan (2004) could investigate memory consolidation in a framework where information flow was similar to the cortex: a bottom-up stream from sensory cortices, through associative cortices and converging to the hippocampus, and then a top-down stream from the hippocampus, again

through associative cortices, and reconstructing patterns of activity in the sensory cortices. Such structure contrasted with the feed-forward approach of McClelland et al. (1995).

Káli & Dayan (2004) created a semantic domain by defining a set of possible activity patterns for each of the three cortical sensory areas. The network learned these patterns and thus established the semantic background knowledge. Subsequently, episodes were defined as specific associations between three sensory patterns. These episodes had corresponding codes in the MTL, which were stored in the hippocampus. The hippocampus replayed these patterns during off-line replay via the top-down stream, allowing their consolidation in the neocortex. When patterns of activity that belonged to a new domain were presented to the network, Káli & Dayan (2004) showed that presumably consolidated memories were vulnerable to ongoing plasticity in the neocortex, and that off-line replay of the memories stored in the hippocampus prevented this effect. Furthermore, due to the hierarchical structure of network, their results revealed a different kind of interference: the codes in the MTL, obtained via the bottom-up stream from sensory cortices, were changed because of cortical plasticity, and hence no longer matched with the old codes that were stored in the hippocampus. The authors suggested that off-line replay has two functions: first, to reactivate memories in the sensory cortex for consolidation in the neocortex, and second, to allow the update of the codes stored in the hippocampus so that their correspondence with cortical memories was maintained.

1.2.3 The Complementary Learning Systems Theory and schemas

The CLST (McClelland et al. 1995) stipulates that consolidation should be slow so as to prevent catastrophic forgetting of established knowledge, yet Tse et al. (2007) showed that rodents very rapidly consolidated new associations if they possessed a framework of knowledge (schema) relevant to the task.

In a more recent computational work McClelland (2013) clarified the CLST in light of these findings. In particular, he stressed that in the original work on CLST the network learned a new item that was *inconsistent* with prior knowledge. Indeed, the network had initially acquired knowledge about the properties of living things, e.g. plant, animals, birds, fish, etc... and subsequently was presented with a new bird 'penguin'. This new item challenged existing knowledge, because as a bird it

input layer		output layer
item	relation	attributes
canary	is	living thing, animal, bird, yellow
canary	can	grow, move, fly, sing
daisy	is	living thing, plant, flower, pretty, yellow

Table 1.1: Example data for the simulations in the The Complementary Learning Systems framework (McClelland 2013).

was expected to fly yet it did not, and instead it could swim. Therefore McClelland predicted that catastrophic interference arised because the new item overlapped with two distinct categories. On the other hand, he argued that in the experiment of Tse et al. (2007) the rats learned rapidly new associations because these were *consistent* with prior knowledge.

To test this prediction, McClelland (2013) compared the acquisition of new *schema-consistent* versus new *schema-inconsistent* items in the same framework as McClelland et al. (1995). The model was a feed-forward, multi-layer neural network that learned to map input labels with output attributes (see Table 1.1). The input label consisted of the item (e.g. canary) with a relation (e.g. can), and the corresponding output was a set of features (e.g. {grow, move, fly, sing}). Each item, relation, and feature was represented by one unit in the network.

McClelland (2013) interleaved the presentation of the training examples so that the network gradually incorporated the associations. This process was the analogue of the initial training in the experiment of Tse et al. (2007), when rodents acquired slowly a set of flavour-place associations in the arena. Next, McClelland (2013) trained the network on a new *schema-inconsistent* item, the infamous bird penguin. In this case, training was slow and led to catastrophic interference, as was reported in McClelland et al. (1995). Again, interleaved learning of the new item along with the old patterns helped preserve existing knowledge, albeit this procedure greatly impeded the acquisition of the new item. On the other hand, when the network learned a new *schema-consistent* item (e.g. bird cardinal), the network rapidly learned the new item without catastrophic interference.

Therefore, McClelland (2013) concluded that the speed of learning was influenced by prior knowledge in the neocortex, and as such experimental findings were fully compatible with the CLST framework.

Why was learning rapid in the consistent case? During the first part of training, the network learned specific examples of categories but also extracted the hierarchical organisation of the data: for instance, when the network learned specific birds, it also learned that birds in general can fly. Thus, for the new training the knowledge was automatically passed on to the new consistent item. For example, the cardinal is a bird, and hence cardinal-can-fly was easily incorporated because it did not require existing connections to change. By contrast, in the inconsistent case the network had to re-arrange its connectivity.

However, the model of McClelland (2013) does not learn abstract knowledge about each category, e.g. it did not capture the idea that flowers can have many colours whereas trees have little colour variation. Yet, we believe this abstract information is important to account for experimental studies in neuroscience and cognitive psychology. For instance, if we used this framework to replicate the experimental study of Smith et al. (2002) described in Section 1.1.3, we predict that when the model learns a new object category it will associate the new material and the new shape equally fast, which would be wrong as the material is likely to change within the category.

1.3 Aim and outline of this thesis

Mental schemas are well defined at the cognitive level, but their implementation at the neural level remains unknown. Indeed, the neurobiology of schemas is particularly hard to study because memory acquisition, retrieval and consolidation involve complex interactions between the hippocampus, the prefrontal cortex and higher sensory cortical areas. Thus, our aim was to develop a computational model to investigate and have a better understanding of neural mechanisms that may underlie the acquisition of schemas, their adaptation and their influence on memory processing.

The neural network models of hippocampal-neocortical interactions that we have introduced so far do not learn abstract information about sensory experiences. On the other hand, cognitive computational models, such as hierarchical Bayesian models, are able to represent knowledge at various levels of abstraction. They learn abstract principles that define how knowledge should look like, and hence these models can explain inductive learning (Kemp et al. 2007, Tenenbaum et al. 2011). Without going into details about the modelling, we briefly describe a simple example to illustrate the approach.

Say, we observe some examples of the category Bird, and some examples of the

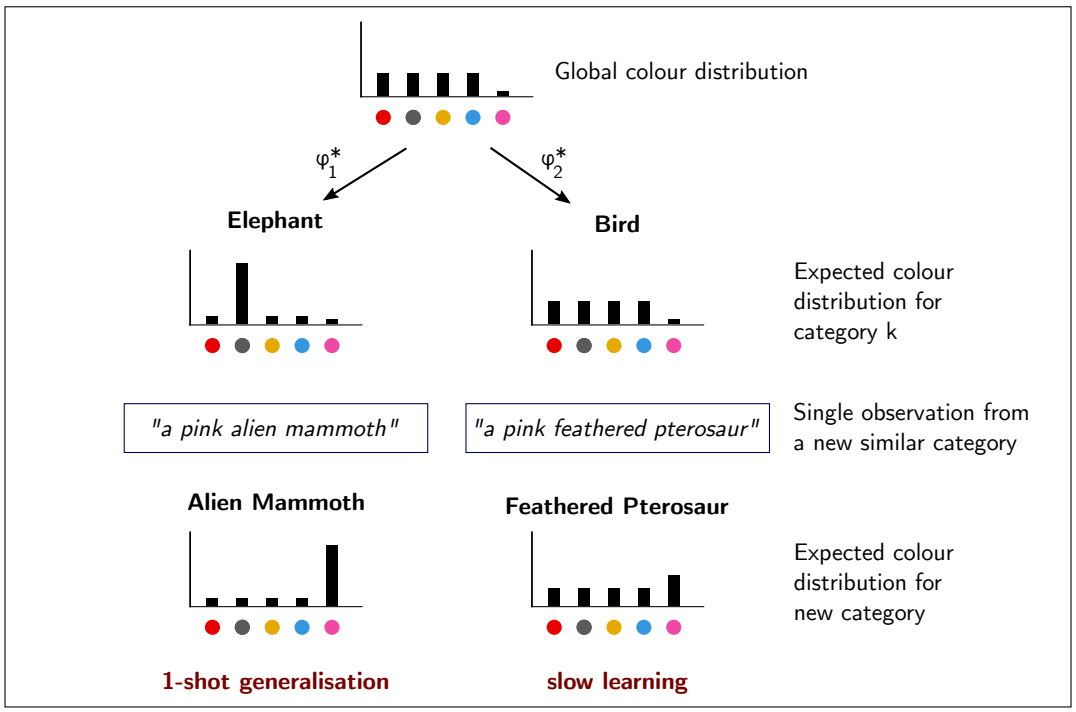


Figure 1.1: Conceptual examples to illustrate generalisation from a single example. The meta-parameter ϕ^* represents the variability in color for each category. *Adapted from examples given in* Tenenbaum et al. (2011).

category Elephant. The birds are of various colours, e.g. red, grey, yellow and blue, while all the elephants are grey. From these observations, we learn a prior distribution over the possible colours that we have encountered as well as the colour distribution within each category (Fig. 1.1). However, we also learn something more abstract: there is a high colour variability for the Bird category, while the spectrum is focused on a single colour for the Elephant category. This abstract knowledge can be represented by a meta-parameter ϕ^*, which has a different value for each category to reflect the variability of the colours within the category. Next, say we observe an example of a new category: "a pink alien mammoth". Because of its resemblance, we group it with the category Elephant, and thus transfer the abstract knowledge ϕ_1^* that we have acquired about it (Fig. 1.1). Since we expect the colour to be consistent, we can rapidly generalise from a single observation that all the examples in the new category Alien Mammoth are pink (one-shot learning). By contrast, say we observe an example of another new category: "a pink feathered pterosaur". We transfer the abstract knowledge ϕ_2^* that we have acquired about Bird category, which tells us that the colour variability is high (Fig. 1.1). Hence we do not generalise the colour pink to the Feathered Pterosaur category, but instead we wait for more observations to gradually update our

belief about possible colours of this new category.

Bayesian frameworks are well suited to model schemas in the brain at the cognitive level, but they do not allow us to investigate the underlying neural mechanisms. For instance, the symbolic representations would not capture the difficulty of incorporating the new feature colour 'pink' into cortical networks. Hence, while we did not use this modelling approach, throughout this thesis we will keep it in mind as inspiration.

Instead we built our hippocampal-neocortical model based on the framework of Káli & Dayan (2004) which we believed had the right level of complexity. Furthermore, we examined how the abstract knowledge described above could be incorporated in our framework, as we believed it was fundamental to account for experimental findings. Importantly, we assumed that regions in the prefrontal cortex coordinate the extraction and use of this abstract knowledge.

In Chapter 2 we introduce the neural network that represents the neocortex. This network, a Restricted Boltzmann Machine, is the core of our model because it extracts the knowledge which is thereafter processed by the hippocampus and prefrontal cortex. In Chapters 3&4 we present our simulations of the experiments of Tse et al. (2007). In Chapter 3 we developed two models to clarify the definition of mental schemas, their acquisition and impact on subsequent learning and consolidation. Based on our findings we developed a third model, introduced in Chapter 4, which captured the interplay between the hippocampus, prefrontal cortex and associative cortex during schema formation and new memory acquisition. In Chapter 5, we simulated the experiments of Richards et al. (2014) to gain insight into the role of plasticity and sleep replay in the context of schemas. Finally, in Chapter 6 we discuss our theoretical framework.

Chapter 2

Restricted Boltzmann Machines as models for the neocortex

In this chapter we introduce the Restricted Boltzmann Machine (RBM), an artificial neural network with properties relevant to modelling system-level memory functions in the neocortex. A RBM is a probabilistic generative model: given a set of variables, it defines a probability distribution P over all the configurations of these variables. For example, the variables could be a set of features say, NAME, SIZE, and COLOR, and the data would be specific configuration patterns such as {`elephant`, `big`, `grey`} and {`rose`, `small`, `pink`}, that follow a certain probability distribution Q. The RBM will approximate this distribution Q so as to represent the dependencies between the variables as faithfully as possible. Thus, given an incomplete configuration, for instance {`elephant`, ?, ?}, the RBM will reconstruct the associated features, SIZE and COLOR, according to the probability distribution P. For this specific example, {`big`, `grey`} will have a higher probability of being sampled than {`small`, `pink`}. Additionally, the RBM can make inference to interpret configurations that it has never seen before: when the network observes {`elephant`, ?, `pink`}, it can evaluate the probabilities of each possible answer for the feature SIZE. Importantly, unlike the popular standard feedforward neural networks, inference can go in any direction since a RBM specifies a joint distribution over the variables. Therefore, the RBM contains a probabilistic model of the world. This makes it well suited as model of semantic memory in the brain, better than a multi-layer perceptron, which does not allow for bidirectional association, and a Hopfield network, which can only memorise patterns.

Our choice for the RBM was also motivated by the fact that while it can be used as cognitive model, it is on the other hand perhaps not too far removed from biology.

Hence, it allows us to make sense of experimental behavioural data and gain insight into potential neural mechanisms. Although the units of the network and their interactions are too simple to be realistic models of neurons and synapses, they nonetheless bear similarity with biological processes that underlie learning and memory in the brain. We will highlight the similarities in this chapter.

Numerous computational models have used neural networks to model cortical memory and have demonstrated the benefits of this approach (see Section 1.2.1). The RBM model of Káli & Dayan (2004) in particular caught our attention, because the level of description of the hippocampal-neocortical interactions suited our application, and it had a simple implementation that enabled to modify and expand on it. The main difference with their framework, apart from the training protocols, is the inclusion of a new module to perform operations that we suppose are mediated by the prefrontal cortex in the brain.

Section 2.1 of this chapter describes the Boltzmann Machine (BM) of which the RBM is a sub-class, and we highlight the relationship with semantic memory. Our description of the BM was inspired by the earliest work on the subject (Hinton et al. 1984), which we believe has a more intuitive interpretation of the network as computational model of the brain. However, the reader can find alternative descriptions in the machine learning literature (i.e. probabilistic graphical models, see introduction by Fischer & Igel 2014). Section 2.2 focuses on the RBM in particular, and our implementation of it to model neocortical associative areas. We outline how this network learns, how we can probe its memory, and we conclude this section by presenting the constraints we had to consider for setting hyper-parameters. Finally in Section 2.3, we describe the main lines of our computational framework for systems consolidation, specifically the interaction between the associative neocortex (RBM) and the hippocampal memory system, and sleep replay. While we describe the different frameworks in more detail in their respective chapters, here we present the foundations of the model.

2.1 Boltzmann Machines and semantic knowledge

The Boltzmann Machine (BM) (Hinton et al. 1984) is a stochastic neural network with one layer of visible units which are observable, and one layer of hidden units which are not connected to the external world (Fig. 2.1). Throughout this work and as is common, we only use **binary units** as a simplification of neuronal states: 1/'on' if the neuron fires, 0/'off' if the neuron remains silent. Other types of units could be used, such as softmax units to model more than two discrete states, or Gaussian units to model real values. The state of the binary units is determined stochastically according to a transition probability, the Boltzmann distribution, that we will describe later. The BM can be seen as a generalisation of the Hopfield network (1982) in that it has stochastic updates and hidden units.

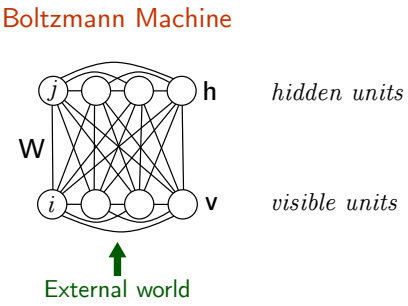


Figure 2.1: Boltzmann Machine.

In the following parts, we use the notation $v_i \in \{0,1\}$ to refer to the state of the visible unit i, $h_j \in \{0,1\}$ the state of the hidden unit j, and s_i when we refer to either a visible or a hidden unit. We denote $\mathbf{v} \in \{0,1\}^{D_v}$ the vector of dimension D_v with elements v_i, and $\mathbf{v}^\alpha$ when we refer to a particular configuration α of the visible units. Similarly, $\mathbf{h} \in \{0,1\}^{D_h}$ denotes the vector of dimension D_h with elements h_j, and $\mathbf{h}^\gamma$ when the hidden units are in a particular configuration γ. Finally, $\mathbf{s} \in \{0,1\}^D$ is the vector with the state of all the units in the network, with $D = D_v + D_h$, and $\mathbf{s}^{\alpha\gamma}$ when the visible and hidden units are in particular configurations α and γ respectively.

In line with connectionists models, the units in a BM are connected to each other by weighted links. These weights are in analogy to the synaptic strengths between pairs of neurons, which are believed to underlie long-term memory in the brain. The weights of a BM can take any real value, positive for excitatory connections and negative for inhibitory connections. Unlike multilayer perceptrons, the connections are bidirectional and hence a BM forms a recurrent network. We denote $\mathbf{W}$ the weight matrix with elements w_{ij} connecting unit i and unit j (visible or hidden). Similar to the

Hopfield network, there is no self-connection ($w_{ii} = 0$) and the connections are symmetric ($w_{ij} = w_{ji}$). (We note that both the lack of sign constraint and the symmetry are not necessarily true in biology. For instance, Dale's principle says that synaptic connections cannot change sign. However, on the network level, it might still be a reasonable assumption. It would be very interesting to know when these assumptions hold and when not.)

The total input, or field, of unit i is calculated by:

$$x_i = b_i + \sum_{j=1}^{D} w_{ij} s_j \tag{2.1}$$

which is the weighted sum of all the active units connected to unit i, plus the bias b_i of the unit.

The hidden units in a BM improve the representational power of the Hopfield network, which is fundamental to extract complex structures underlying data. Indeed, the Hopfield network has a limited representational power since the pairwise connections between observable units cannot capture interactions higher than second-order. For example, a Hopfield network cannot learn the states (patterns) with 3 observable units shown in Table 2.1, which corresponds to XOR function from the point of view of the third unit.

Instead, given sufficient hidden units, the BM can represent an arbitrary probability distribution (Le Roux & Bengio 2008). The hidden units in the BM are not observed and hence can be freely used to represent higher order interactions between groups of units. They act as "feature detectors", their state reflecting whether some specific pattern of activation is present or not, and the information they carry can then be used by the other units of the network. These "internal representations" will naturally be discovered by the network during training.

unit	1	2	3
state 1	0	0	0
state 2	1	0	1
state 3	0	1	1
state 4	1	1	0

Table 2.1: XOR function cannot be learned in a Hopfield network, but can be learned by a Boltzmann Machine with hidden units.

2.1.1 Stochastic units and energy

The BM has an *energy function* that associates a scalar value (energy) to each configuration $(\mathbf{v}^\alpha, \mathbf{h}^\gamma)$ given the network parameters. The concept of the energy function was introduced for neural networks in general by Hopfield (1982), and the existence of this function is guaranteed by the symmetric connection weights. The idea is that low-energy configurations are more consistent with the structure of domain (which is set by the weights and biases), whereas high-energy configurations are more conflicting with the constraints, that is, are less likely. The energy of a configuration, or global state, is given by:

$$\begin{aligned} E_{\alpha\gamma} &= -\mathbf{b}_v^\intercal \mathbf{v}^\alpha - \mathbf{b}_h^\intercal \mathbf{h}^\gamma - (\mathbf{v}^\alpha)^\intercal \mathbf{W}_{vh} \mathbf{h}^\gamma - \frac{1}{2} (\mathbf{v}^\alpha)^\intercal \mathbf{W}_v \mathbf{v}^\alpha - \frac{1}{2} (\mathbf{h}^\gamma)^\intercal \mathbf{W}_h \mathbf{h}^\gamma \\ &= -\mathbf{b}^\intercal \mathbf{s}^{\alpha\gamma} - \frac{1}{2} (\mathbf{s}^{\alpha\gamma})^\intercal \mathbf{W} \mathbf{s}^{\alpha\gamma} \end{aligned} \tag{2.2}$$

where $E_{\alpha\gamma} = E(\mathbf{v}^\alpha, \mathbf{h}^\gamma)$ is the energy of the configuration $(\mathbf{v}^\alpha, \mathbf{h}^\gamma)$, $\mathbf{W}_{vh}$ is the weight matrix with elements w_{ij} connecting the visible and hidden units, $\mathbf{W}_v$ and $\mathbf{W}_h$ are the matrices with connection weights visible to visible and hidden to hidden units respectively, $\mathbf{b}_v$ and $\mathbf{b}_h$ are the vectors with the bias terms of visible and hidden units respectively; the second formulation is the compact version of the parameters, where we put all the bias terms in the vector $\mathbf{b}$ and all the weights in the matrix[1] $\mathbf{W}$.

When a unit switches state from 'off' ($s_i = 0$) to 'on' ($s_i = 1$), the effect on the global energy can be computed locally as:

$$\begin{aligned} \Delta E_i &= E(s_i = 0) - E(s_i = 1) \\ &= b_i + \sum_{j=1, j\neq i}^{D} w_{ij} s_j^{\alpha\gamma} \end{aligned} \tag{2.3}$$

We can notice that this "energy gap" corresponds to the total input of the unit described Eq. 2.1. To determine the state of the unit, Hinton et al. (1984) used the analogy with a particle that has two energy states in statistical physics: at temperature T, the particle switches to one state or the other in an attempt to decrease the global energy. Each unit is in the 'on' state with probability:

$$p(s_i = 1 \mid \mathbf{s}_{-i}^{\alpha\gamma}) = \frac{1}{1 + e^{-\Delta E_i / T}} \tag{2.4}$$

[1] Remember that all the matrices are symmetric and have zeros in the diagonal.

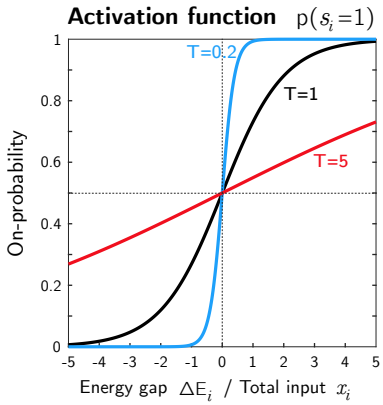


Figure 2.2: Sigmoid activation function. Transition probability of unit s_i.

where T is the temperature of the network, and $\mathbf{s}_{-i}^{\alpha\gamma}$ is the vector of states of the other units. Note that the updating rule is equivalent to applying the sigmoid function to the total input of the unit (Fig. 2.2). We see in Eq. 2.4 that if the energy gap is such that the system has a lower energy when $s_i = 1$ (i.e. $\Delta E_i > 0$), then the unit will most likely go into the 'on' state. The temperature T is a parameter that can be varied so as to scale the weights and alter the state transition probability. When this temperature is reduced to zero, the system becomes deterministic: unit i goes into the 'on' state if its total input is greater than zero, and into the 'off' state if its total input is less than zero. On the other hand, as the temperature increases the system becomes more noisy, until each unit has equal chance of being 'on' or 'off' regardless of the state of the other units. The bias term b_i shifts the activation probability of the unit: for instance, if the bias b_i of unit i is positive, then this unit is more likely in the 'on' state (see that when $\sum_{j\neq i} w_{ij} s_j = 0$ the energy gap in Eq. 2.3 is equal to $b_i > 0$).

2.1.2 Network memory and dynamics

In the previous subsection we have shown that each configuration of the network has an energy (Eq. 2.2), and that the transition probability of each unit is expressed as a function of the energy gap between its two states (Eq. 2.3). With this setting, when the network reaches thermal equilibrium it follows the Boltzmann-Gibbs distribution. Then BM then defines the joint probability distribution over all the configurations α of the visible units and configurations γ of the hidden units:

$$P(\mathbf{v}^{\alpha}, \mathbf{h}^{\gamma}) = \frac{e^{-E_{\alpha\gamma}/T}}{Z} \tag{2.5}$$

where $E_{\alpha\gamma} = E(\mathbf{v}^{\alpha}, \mathbf{h}^{\gamma})$ is the energy described Eq. 2.2, and $Z = \sum_{\beta\delta} e^{-E_{\beta\delta}/T}$ is the

normalisation factor and is called the *partition function*. However, we are ultimately interested in the distribution underlying the activity of the variables that are connected to the external world, which is defined by the marginal probability distribution:

$$P(\mathbf{v}^\alpha) = \sum_\gamma P(\mathbf{v}^\alpha, \mathbf{h}^\gamma) = \frac{1}{Z} \sum_\gamma e^{-E_{\alpha\gamma}/T} \tag{2.6}$$

We can see in Eq. 2.6 that the lower the energy, the higher the probability of the configuration. Therefore, given the constraints set by the network parameters (weights and biases), the BM has an "internal model" that allows to generate plausible/probable samples, to evaluate the probability of an observed configuration pattern, or to make inference (i.e. generate samples conditioned on a partial configuration). At high temperature T the energy gaps between configurations are reduced, and hence the probability distribution converges to a uniform distribution.

However, calculating $P(\mathbf{v}^\alpha)$ or sampling from it is not straightforward, primarily because of the partition function Z which sums over the 2^D configuration space. Instead of directly calculating it, we use the activity dynamics of the network to estimate the probabilities. The estimation process is generally referred to as Gibbs sampling (Fig. 2.3). The dynamics are governed by the energy function: the system gradually evolves so that the energy decreases or remains constant (strictly at zero temperature only). Starting from an initial global state, the units are asynchronously updated according to their transition probability (Eq. 2.4) in random order until the system finds a configuration of states that is a (local) minimum of the energy function. The asynchronous updating is required to prevent occurrence of cyclic states.

If the network has a deterministic update rule (see Eq. 2.4 when $T \to 0$), the system will always settle in the same stable state for a given initial state. Hopfield used this property to introduce the idea of associative memory, where the memory patterns correspond to local minima (attractors). While such system is useful to perform pattern completion, it can only retrieve specific stored configurations but cannot explore more combinations. For instance, going back to our toy example, if the network starts with {`elephant`, ?, `pink`} it would either settle on {`elephant`, `big`, `pink`} or {`elephant`, `small`, `pink`}, but cannot alternate between the two answers. In other words, one cannot represent probabilities[2].

By contrast, if the network uses the stochastic update rule (Eq. 2.4), the system has the ability to escape these attractors. Indeed, each unit will preferably go into the

[2]Stochastic versions of the Hopfield network are able to represent these probabilities.

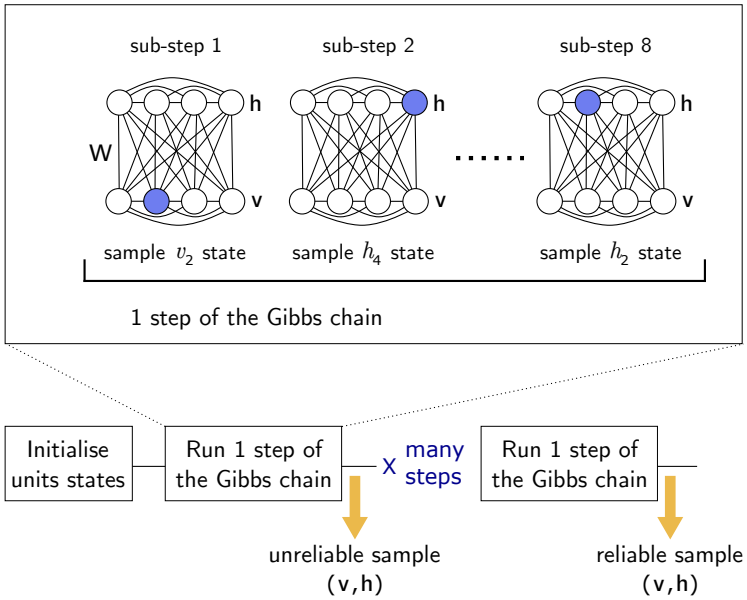


Figure 2.3: Gibbs sampling is a Markov Chain Monte Carlo (MCMC) algorithm to obtain samples from the model probability distribution P of the Boltzmann Machine. The variables of the network are initialised in a specific state $\mathbf{s} \in \{0,1\}^D$, where $\mathbf{s} = (\mathbf{v},\mathbf{h})$ with $\mathbf{v}$ the state of the visible units and $\mathbf{h}$ for the hidden units, and D is the total number of units in the network. The procedure constructs a chain of steps (Gibbs step) during which the state of all the variables is updated. Thus, each step produces a new sample $\mathbf{s}$, and longer chains lead to better samples as the distribution of the samples converges to the model distribution P (equilibrium distribution). For a Gibbs step in a BM each variable is updated sequentially (D sub-steps), using the conditional distribution given the state of the others units (Eq. 2.4).

state that decreases the global energy, but they can sometimes end up in the "wrong" state because the update is stochastic. The energy might temporary increase, but it allows the network to jump energy barriers before moving again towards other low-energy configurations. Therefore the system will continuously evolve and will keep on searching for interpretations. After reaching equilibrium, the distribution matches the Boltzmann distribution.

If we use this method from random starts, we can monitor the configurations visited and thus obtain samples from the distribution. Similarly, if we clamp a subset of the visible units (e.g. inference from a partial input from the environment), the remaining units adopt configurations that minimise the energy. The lower the energy, the more satisfactory the interpretation.

However, the downside of Gibbs sampling is that the system must reach thermal equilibrium before it starts generating configuration samples that correspond to the Boltzmann probability distribution (Eq. 2.5). While BMs are guaranteed to eventually reach equilibrium, this means that the system might have to run many update steps, which can be an issue for networks with large number of units since each variable has to be updated independently.

2.1.3 Learning

We have seen in the preceding part that a BM defines a probability distribution P over the state of the visible units, and we would like this distribution to reflect the underlying structure of the data. How does the network acquire this distribution? A BM is trained by observing examples of configurations that belong to the domain we wish to model, and the network adjusts its parameters until it can generate samples that match the probability distribution of the data. In most real-world applications there are many visible units (e.g. pixels of an image), and hence the network cannot exactly capture the probabilities of the 2^{D_v} configurations. Nonetheless, by learning the BM will approximate the unknown distribution by extracting the main features of the data.

The similarity between the model probability distribution P (at equilibrium) and the original distribution Q of the observed examples can be evaluated with the Kullback-Leibler divergence:

$$G = D_{KL}(Q \parallel P) = \sum_{\alpha} Q(\mathbf{v}^{\alpha}) \ln \frac{Q(\mathbf{v}^{\alpha})}{P(\mathbf{v}^{\alpha})} \tag{2.7}$$

where G measures the divergence of the distribution P from the distribution Q.

The divergence is null when the two distributions are a perfect match, otherwise it is positive. The aim is to minimise this divergence by modifying the weights and biases of the network. When the system is at thermal equilibrium, the partial derivative of the divergence G with respect to the weight w_{ij} connecting two units (visible or hidden) is (Hinton et al. 1984):

$$\begin{aligned}\frac{\partial G}{\partial w_{ij}} &= -\frac{1}{T}\left(\sum_{\alpha\gamma} Q(\mathbf{v}^{\alpha}) P(\mathbf{h}^{\gamma}|\mathbf{v}^{\alpha}) s_i^{\alpha\gamma} s_j^{\alpha\gamma} - \sum_{\beta\delta} P(\mathbf{v}^{\beta}, \mathbf{h}^{\delta}) s_i^{\beta\delta} s_j^{\beta\delta}\right) \\ &= -\frac{1}{T}\left(\langle s_i s_j\rangle_{data} - \langle s_i s_j\rangle_{model}\right)\end{aligned} \tag{2.8}$$

where α and β are configurations of the visible units, and γ and δ are configurations of the hidden units. The first term $\langle s_i s_j\rangle_{data}$ represents the average probability of the two units being 'on' when the visible states are clampled by the environment (observed examples). The second term $\langle s_i s_j\rangle_{model}$ represents the average probability of the two units being 'on' when the visible states are generated by the model distribution. Using gradient descent, the change of weight is then proportional to the difference between these two terms: $\Delta w_{ij} \propto \langle s_i s_j\rangle_{data} - \langle s_i s_j\rangle_{model}$. A similar learning rule is derived for the bias terms.

We can see that the weight update is local, as it is only determined by the co-activation probability of the two units it connects. For instance, if the gradient of G in the direction of w_{ij} is negative (Eq. 2.8), it means that the activations of units i and j are less correlated in the model than in the original distribution, and hence, the weight w_{ij} needs to be increased. The BM thus continuously adjusts its internal model. These properties, locality and error correction, resemble ideas of Hebbian and anti-Hebbian plasticity for synapses and learning by prediction error in the brain. However, despite its compelling character, it should also be stressed that it is unclear if and how the brain implements this two-phase learning rule.

We cannot calculate the gradient analytically since it is very hard to compute the probability distribution of the BM, as explained in the preceding Section 2.1.2, but instead we estimate it by again generating samples using Gibbs sampling (Fig. 2.3). The procedure has to be run twice, once for the "positive phase" to calculate $\langle s_i s_j\rangle_{data}$, and once for the "negative phase" to calculate $\langle s_i s_j\rangle_{model}$. During the "positive phase", we need to obtain samples of the hidden units activations when the visible states are provided externally, i.e. when the network observes examples from the unknown distribution Q. During the "negative phase", we let the system run freely so as to generate

samples of $(\mathbf{v}, \mathbf{h})$. However, we can only obtain samples after the system relaxes to equilibrium (at a fix temperature T), which can be quite long since the number of substeps in Gibbs sampling is equal to the number of units in the network. In addition, the procedure must be repeated for many examples from the external world so as to get a better approximation of the unknown distribution Q (sum over the configurations α in Eq. 2.8).

Techniques have been developed to improve the efficiency of the learning algorithm. For example, similar to physical systems, "simulated annealing" (Kirkpatrick et al. 1983) allows the system to reach equilibrium faster by starting the training at higher temperature, and then progressively reducing the temperature. Another popular technique is mean-field approximation (Peterson 1987, Welling & Hinton 2002). Recent work has intensively researched the use of variational approach and more advanced sampling methods to approximate the log-partition function or its upper bounds (Salakhutdinov 2008). Despite these methods, training takes usually very long, especially when the network gets larger. This has hindered the application of Boltzmann machines.

2.2 The Restricted Boltzmann Machines and the neocortex

BMs are powerful models, but we have highlighted in Section 2.1 that, in practice, the training is difficult and takes very long. A simplified variant of the network is generally used to circumvent this issue, originally introduced as Harmonium by Smolensky (1986) and now called Restricted Boltzmann Machine (RBM). As the name indicates, this network is a BM with restriction on the connectivity, where the visible-visible connections and hidden-hidden connections are removed (Fig. 2.4). As a result, the dependencies between the visible units are only captured by the hidden units. Despite the restriction of topology, a RBM is a universal approximator provided that the number of hidden units is sufficient (Le Roux & Bengio 2008), which means that any distribution can in principle be approximated with arbitrary precision. A complete description of RBMs can be found in Fischer & Igel (2014).

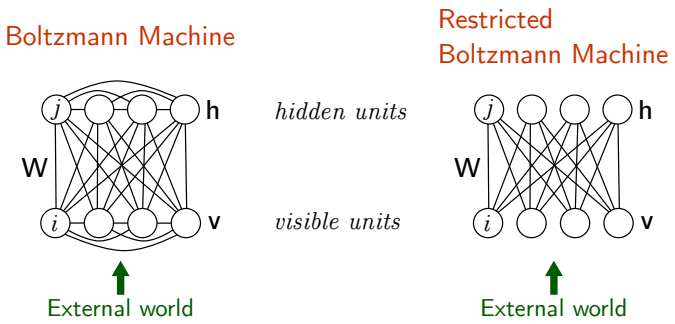


Figure 2.4: Boltzmann Machine and Restricted Boltzmann Machine.

Because of the simplification of connectivity, the energy function of Eq. 2.2 can be re-written for the RBM:

$$E_{\alpha\gamma} = -\left(\mathbf{v}^{\alpha}\right)^{\intercal} \mathbf{W}\mathbf{h}^{\gamma} \tag{2.9}$$

where $\mathbf{W}$ denotes the interaction terms between visible and hidden units (previously noted $\mathbf{W}_{vh}$ in Eq. 2.2). Note also that in the equation above we omitted the bias terms so as to simplify the notations, but also because we will not use them throughout this work. Indeed, the bias terms capture the average activity of each unit across the data set, and they are commonly used in applications where the RBM learns from large data sets and the visible units are active in many input patterns (e.g. pixels of images). However, in this work we considered very small data sets with sparse input patterns. In preliminary explorations we found that it was not crucial to use biases to monitor frequency of units activations, especially since we did not have constraints on the number of hidden units. Furthermore, because we had a small set of input data, our RBM network did not seem to learn properly the bias terms: for instance, for two visible units with same average on-probability, one unit could end up with a large positive bias while the other had a large negative bias. We wished to avoid this situation as our aim was to obtain meaningful cortical representation of knoweldge, so that the network could then use this knowledge to learn additional information.

Given the properties described in Section 2.1 above, we use the RBM to model associative areas in the neocortex, where the memories are consolidated. Similar to the interpretation of Káli & Dayan (2004), the bottom (visible) layer represents a vector of variables that receive input from higher sensory areas. We call it the "sensory cortex". As there are no lateral connections in the RBM, this layer can be divided into multiple pools of units representing different sensory modalities, or different features. For example, in Chapters 3 & 4 the visible layer has two pools of units, one to represent

flavours and the other to represent reward locations; in Chapter 5, the visible layer also has two pools of units, but this time to represent the polar coordinates of platforms in a circular watermaze.

2.2.1 Cortical recall

Although the energy function is simplified, the probability $P(\mathbf{v})$ (Eq. 2.6) is still not easy to compute because of the partition function. However, the modification of the connectivity leads to a useful property, namely the hidden units (resp. visible) are independent of each other and only depend on the state of the visible units (resp. hidden):

$$P(\mathbf{h}|\mathbf{v}) = \prod_{j=1}^{D_h} P\left(h_j|\mathbf{v}\right) \tag{2.10}$$

and

$$P(\mathbf{v}|\mathbf{h}) = \prod_{i=1}^{D_v} P(v_i|\mathbf{h}) \tag{2.11}$$

This factorisation greatly simplifies the process of Gibbs sampling, which we use to obtain samples from the model probability distribution as explained in Section 2.1.2. Indeed, we no longer need to evaluate successively the state of each unit given the state of the others, but instead we can sample the state of all the units belonging to the same layer simultaneously (Fig. 2.5). Therefore the number of sub-steps required for Gibbs sampling, which was equal to the number of units in the network for the BM, is now reduced to two sub-steps.

In our model we use Gibbs sampling to probe cortical memory when the network is presented with a partial input pattern. An example of pattern completion is given in Algorithm 2.1. Say a cue $\mathbf{v}$ is presented, and we wish to infer the rest of the visible units $\mathbf{u}$. The unknown part of the visible units $\mathbf{u}$ could be a noisy input that we would like to denoise (e.g. noisy version of an image), or it could simply be missing (i.e. $\mathbf{u} = \mathbf{0}$). In this work we use the latter during recall when testing memory performance, because we suppose that the neocortex receives cues from one sensory modality (e.g. a flavour) and must retrieve the activity in the second one (e.g. a reward location). The network performs several Gibbs sampling steps as shown Fig. 2.5, starting the chain from the cue. Note that, in practice, we do not need to compute the activity of the visible units $\mathbf{v}$ corresponding to the cue since we clamp this part at each iteration, but it can still be useful to check their reconstruction.

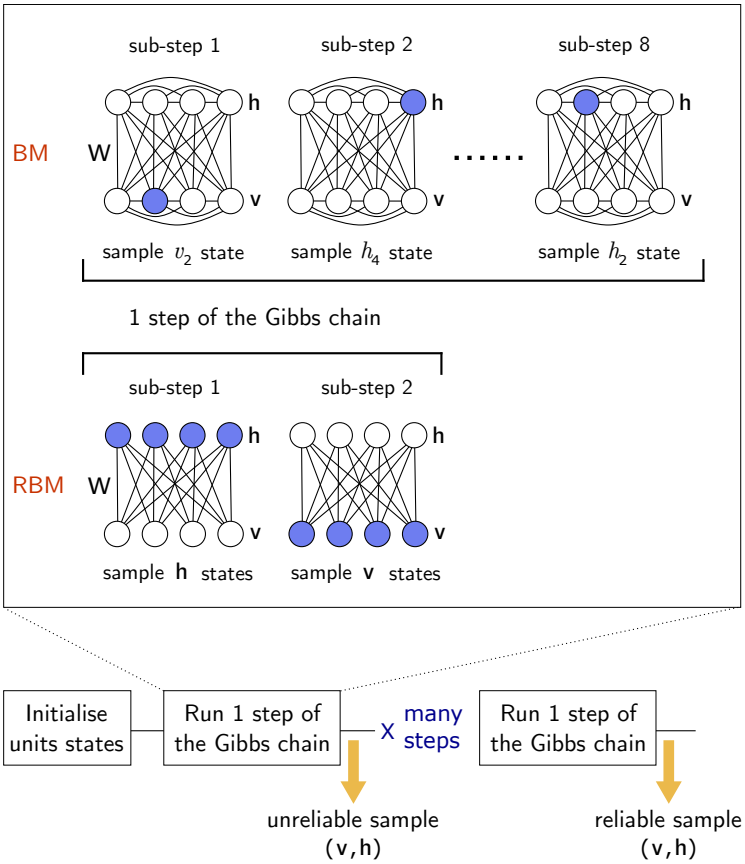


Figure 2.5: Gibbs sampling is a Markov Chain Monte Carlo (MCMC) algorithm to obtain samples from the model probability distribution P, for the Boltzmann Machine (BM) and Restricted Boltzmann Machine (RBM). The variables of the network are initialised in a specific state $\mathbf{s} \in \{0,1\}^D$, where $\mathbf{s} = (\mathbf{v}, \mathbf{h})$ with $\mathbf{v}$ the state of the visible units and $\mathbf{h}$ for the hidden units, and D is the total number of units in the network. The procedure constructs a chain of steps (Gibbs step) during which the state of all the variables is updated. Thus, each step produces a new sample $\mathbf{s}$, and longer chains lead to better samples as the distribution of the samples converges to the model distribution P (equilibrium distribution). Gibbs step in a BM: each variable is updated sequentially (D sub-steps), using the conditional distribution given the state of the others units (Eq. 2.4). For the RBM, each step of the chain only requires two sub-steps: sample $\mathbf{h}$ given the visible states $\mathbf{v}$, and then sample $\mathbf{v}$ given the hidden states $\mathbf{h}$ (Eq. 2.10&2.11).

Algorithm 2.1 Cued recall in a RBM

Notations:

$p(\mathbf{h}=1)$ is the real-valued vector with elements $p\left(h_j=1\right)$

$\sigma(\cdot)$ is the sigmoid activation function applied element-wise

Given the observed pattern $\mathbf{v}^0$ and noisy pattern $\mathbf{u}^0$ in the visible layer

1. Compute the activation probability of the hidden units

$$p\left(\mathbf{h}=1|\mathbf{v}^0,\mathbf{u}^0\right)=\sigma\left(\mathbf{W}^\intercal\mathbf{s}^0\right), \text{ where } \mathbf{s}^0=\begin{bmatrix}\mathbf{v}^0\\ \mathbf{u}^0\end{bmatrix}$$

2. Sample binary pattern $\mathbf{h}^0$ from $p\left(\mathbf{h}|\mathbf{v}^0,\mathbf{u}^0\right)$

   for n Gibbs sampling steps **do**

   a. Compute the activation probability of the visible units

$$p\left(\mathbf{s}=1|\mathbf{h}^{n-1}\right)=\sigma\left(\mathbf{W}\mathbf{h}^{n-1}\right), \text{ where } \mathbf{s}=\begin{bmatrix}\mathbf{v}\\ \mathbf{u}\end{bmatrix}$$

   b. Sample binary pattern $\mathbf{u}^n$ and clamp $\mathbf{v}\leftarrow\mathbf{v}^0$

   c. Compute the new activation probability of the hidden units

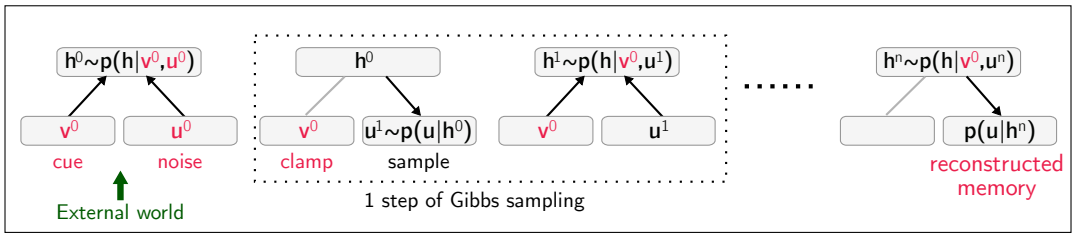
$$p\left(\mathbf{h}=1|\mathbf{v}^0,\mathbf{u}^n\right)=\sigma\left(\mathbf{W}^\intercal\mathbf{s}^n\right), \text{ where } \mathbf{s}^n=\begin{bmatrix}\mathbf{v}^0\\ \mathbf{u}^n\end{bmatrix}$$

   d. Sample binary pattern $\mathbf{h}^n$ from $p\left(\mathbf{h}|\mathbf{v}^0,\mathbf{u}^n\right)$

   end for

Given the last binary pattern $\mathbf{h}^n$, obtain the reconstruction of the memory pattern

$$p\left(\mathbf{s}=1|\mathbf{h}^n\right)=\sigma\left(\mathbf{W}\mathbf{h}^n\right), \text{ where } \mathbf{s}=\begin{bmatrix}\mathbf{v}\\ \mathbf{u}\end{bmatrix}$$



In Chapter 5, on the other hand, we model an experiment where rodents must swim in a pool of water to find a hidden platform so as to escape, and hence there is no cue presented to the network during recall. In this case, we start the chain from random states to probe memory.

Similar to Gibbs sampling in BMs, the network has to take many steps to ensure that we sample from the model equilibrium distribution. However, if we clamp a set of visible units to a training example pattern, the network then converges more rapidly. In addition our model learned small data sets and simple patterns, and thus required very few steps to correctly reconstruct the activities during recall. Biologically, Gibbs sampling resembles neural reverberation in the brain which has been proposed as a potential neural mechanism underlying recall.

2.2.2 Cortical learning

When training BMs, the weights are updated so as to mimise the divergence between the model distribution P and the true distribution of the data Q. This weight change is obtained by comparing the correlations when the visible units are clamped by the environment and the correlations predicted by the model, and thus requires to average over all possible state configurations of the visible patterns $\mathbf{v}^\alpha$ (Eq. 2.8). However, in practice we only observe a subset of these patterns. The standard procedure for training consists in dividing the training data into groups ("mini-batches"), and for each group we collect the statistics to estimate the weight change. Although it is usually more efficient to train the network on these mini-batches (Hinton 2010), in the current work we use online learning, i.e. we update the weights directly after we observe a single example. The motivation was that 1) we could afford online learning because we handled small data sets and simple tasks that did not require much computational power and 2) online learning seems more in line with the biology than having a positive phase cycling through training examples, then a negative phase cycling through model samples, and then finally update the weights.

When the RBM sees a single training pattern $\mathbf{v}^\alpha$, we compute the weight change by maximising the log-likelihood $\log P(\mathbf{v}^\alpha)$, where P is the marginal distribution over the visible states defined Eq. 2.6. To simplify the notations, from now on we assume that we operate at a temperature $T = 1$, and hence we omit this term in the equations. The derivative of the log-likelihood for the training example pattern $\mathbf{v}^\alpha$, w.r.t. the weight w_{ij} between visible unit i and hidden unit j is

$$\begin{aligned}
\frac{\partial \log P(\mathbf{v}^\alpha)}{\partial w_{ij}} &= \frac{\partial \log \sum_\gamma e^{-E_{\alpha\gamma}}}{\partial w_{ij}} - \frac{\partial \log \sum_{\beta\delta} e^{-E_{\beta\delta}}}{\partial w_{ij}} \\
&= -\frac{1}{\sum_\gamma e^{-E_{\alpha\gamma}}} \sum_\gamma e^{-E_{\alpha\gamma}} \frac{\partial E_{\alpha\gamma}}{\partial w_{ij}} + \frac{1}{\sum_{\beta\delta} e^{-E_{\beta\delta}}} \sum_{\beta\delta} e^{-E_{\beta\delta}} \frac{\partial E_{\beta\delta}}{\partial w_{ij}} \\
&= -\sum_\gamma p(\mathbf{h}^\gamma \mid \mathbf{v}^\alpha) \frac{\partial E_{\alpha\gamma}}{\partial w_{ij}} + \sum_{\beta\delta} p\left(\mathbf{v}^\beta, \mathbf{h}^\delta\right) \frac{\partial E_{\beta\delta}}{\partial w_{ij}} \\
&= \sum_\gamma p(\mathbf{h}^\gamma \mid \mathbf{v}^\alpha) v_i^\alpha h_j^\gamma - \sum_{\beta\delta} p\left(\mathbf{v}^\beta, \mathbf{h}^\delta\right) v_i^\beta h_j^\delta
\end{aligned} \tag{2.12}$$

We can notice that if we average Eq. 2.12 over a training set $\{\mathbf{v}^\alpha\}_\alpha$ we retrieve the learning rule[3] of Eq. 2.8: $\sum_\alpha \frac{\partial \log P(\mathbf{v}^\alpha)}{\partial w_{ij}} = \sum_{\alpha\gamma} p(\mathbf{h}^\gamma \mid \mathbf{v}^\alpha) v_i^\alpha h_j^\gamma - \sum_{\beta\delta} p\left(\mathbf{v}^\beta, \mathbf{h}^\delta\right) v_i^\beta h_j^\delta$. But since we do online learning, i.e. we update the weights after each example pattern presentation, we keep the gradient of Eq. 2.12 above. Because of the factorisation property of the RBM (Eq. 2.10), if we note $\mathbf{h}^\gamma_{-j}$ the pattern γ with all hidden units but unit j, this gradient can be written as

$$\begin{aligned}
\frac{\partial \log P(\mathbf{v}^\alpha)}{\partial w_{ij}} &= \sum_{h_j \in \{0,1\}} p\left(h_j \mid \mathbf{v}^\alpha\right) v_i^\alpha h_j \underbrace{\sum_\gamma p\left(\mathbf{h}^\gamma_{-j} \mid \mathbf{v}^\alpha\right)}_{=1} - \sum_\beta p\left(\mathbf{v}^\beta\right) \sum_\delta p\left(\mathbf{h}^\delta \mid \mathbf{v}^\beta\right) v_i^\beta h_j^\delta \\
&= p\left(h_j = 1 \mid \mathbf{v}^\alpha\right) v_i^\alpha - \sum_\beta p\left(\mathbf{v}^\beta\right) \sum_{h_j \in \{0,1\}} p\left(h_j \mid \mathbf{v}^\beta\right) v_i^\beta h_j \underbrace{\sum_\delta p\left(\mathbf{h}^\delta_{-j} \mid \mathbf{v}^\beta\right)}_{=1} \\
&= p\left(h_j = 1 \mid \mathbf{v}^\alpha\right) v_i^\alpha - \sum_\beta p\left(\mathbf{v}^\beta\right) p\left(h_j = 1 \mid \mathbf{v}^\beta\right) v_i^\beta
\end{aligned} \tag{2.13}$$

From Eq. 2.13 we can see that the first term of the gradient is easy to compute. On the other hand, the second term still requires that we collect the statistics for all configurations of the visible patterns $\mathbf{v}^\beta$. Once again, we use Gibbs sampling to approximate the log-likelihood gradient but, in contrast to the BM, it was shown that the RBM can be successfully trained by running short chains. For instance, the k-step Contrastive Divergence (CD-k)(Hinton 2002) is a common learning algorithms to train RBMs which works well with a few steps of Gibbs sampling. The main idea of CD-k is to start the chain from a training example pattern of activity and then run only k steps of Gibbs sampling (Fig. 2.5) to obtain a sample configuration $\left(\mathbf{v}^k, \mathbf{h}^k\right)$. The second term in

[3] Note that the gradients have opposite signs, but in the first case we minimise the divergence and hence the change of weight is proportional to the negative gradient, while here we maximise the log-likelihood, and hence the change of weight is proportional to the gradient.

the gradient is then approximated using this single sample instead of the average over all possible configurations (Algorithm 2.2 describes the case of CD-1). Surprisingly, theoretical results have demonstrated that 1-step Contrastive Divergence (CD-1) often performs well, albeit it introduces a bias in the gradient estimate since it works locally around training data points. This bias is reduced as the number of steps k increases. Although more advanced techniques have been developed since, we retained the CD-1 algorithm for its simplicity and efficiency but also because its interpretation suited the biology: the negative part of the gradient could be seen as the reactivation of activity patterns triggered by current sensory experience (i.e. model prediction).

For the calculation of the gradient we sampled both $\mathbf{v}$ and $\mathbf{h}$, while it is often advised to use $p(\mathbf{h} \mid \mathbf{v})$. The latter could be interpreted as the rate of the neurons, but since the input patterns were binary we wanted to be consistent between the visible and hidden units activity (during learning at least, as for recall we sometimes used the probabilities rather than the binary states...). This of course should be investigated in the future, especially for more complex data sets.

2.2.3 Constraints of RBMs as model of the neocortex

It should be stressed that we did not do a grid search to find optimal hyper-parameters for our models. Indeed, since we simulated behavioural experimental studies, we were not interested in achieving the best possible performance. Instead, the results of our simulations aimed to capture the tendency of the experimental data. A second reason is that most of methods to set the parameters require access to a batch of examples patterns (refer to Hinton 2010 for a list of training tips). However, this work assumes that our model is initially naive and trained online on sparse data. Thus, the idea was to set baseline parameters and to fix them throughout training rather than to fine-tune them.

Number of training epochs and number of cortical updates The number of training epochs represented the number of times the network was connected to and then disconnected from the external world, reflecting cycles of experience / sleep. We therefore limited this number according to the experimental task we simulated. During each epoch the patterns could be presented repeatedly. The patterns presented to the visible layer of the network ($\mathbf{v}^0$ in Algorithm 2.2) were either given by the external world (e.g. if observe $\mathbf{v}^\alpha$, set $\mathbf{v}^0 \leftarrow \mathbf{v}^\alpha$), or generated internally by the model (e.g. if the model

Algorithm 2.2 RBM training with 1-step Contrastive Divergence

Learning procedure for a Restricted Boltzmann Machine using CD-1

Notations:

$p(\mathbf{h}=1)$ is the real-valued vector with elements $p\left(h_j=1\right)$

$\sigma(\cdot)$ is the sigmoid activation function applied element-wise

η is learning rate for parameter update

Given the example pattern $\mathbf{v}^\alpha$ in the visible layer

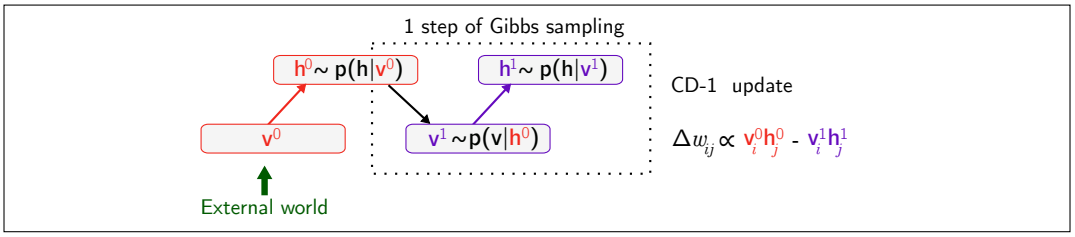
1. Initialise the positive chain with the example configuration $\mathbf{v}^0 \leftarrow \mathbf{v}^\alpha$
2. Compute the activation probability of the hidden units
   $p\left(\mathbf{h}=1|\mathbf{v}^0\right)=\sigma\left(\mathbf{W}^\intercal\mathbf{v}^0\right)$
3. Sample binary pattern $\mathbf{h}^0$ from $p\left(\mathbf{h}|\mathbf{v}^0\right)$

A. End of positive phase: compute positive update $\Delta\mathbf{W}^+=\mathbf{v}^0\left(\mathbf{h}^0\right)^\intercal$

4. Compute the activation probability of the visible units to start the negative chain
   $p\left(\mathbf{v}=1|\mathbf{h}^0\right)=\sigma\left(\mathbf{W}\mathbf{h}^0\right)$
5. Sample binary pattern $\mathbf{v}^1$ from $p\left(\mathbf{v}|\mathbf{h}^0\right)$
6. Compute the activation probability of the hidden units
   $p\left(\mathbf{h}=1|\mathbf{v}^1\right)=\sigma\left(\mathbf{W}^\intercal\mathbf{v}^1\right)$
7. Sample binary pattern $\mathbf{h}^1$ from $p\left(\mathbf{h}|\mathbf{v}^1\right)$

B. End of negative phase: compute negative update $\Delta\mathbf{W}^-=\mathbf{v}^1\left(\mathbf{h}^1\right)^\intercal$

8. Update the weights $\mathbf{W}\leftarrow\mathbf{W}+\eta\left(\Delta\mathbf{W}^+-\Delta\mathbf{W}^-\right)$



generates $\tilde{\mathbf{v}}$, set $\mathbf{v}^0 \leftarrow \tilde{\mathbf{v}}$; see sleep replay in Section 2.3). When learning was driven by an observed example $\mathbf{v}^\alpha$, we limited the number of repetitions, e.g. if in the experiment animals were exposed N times to a stimulus, we would present this stimulus N times to our network, with exception of the model in Section 3.2. On the other hand, we did not limit for the off-line repetitions, but we tuned this number to ensure that the network learned in a reasonable number of training epochs.

Number of hidden units An RBM can approximate any distribution with arbitrary precision if the number of hidden units is sufficient. Hence we set an arbitrary number of hidden units that was large enough to ensure that it would not hamper learning. In particular, we wanted to make sure that, once the model was trained on a particular task, it could learn new information.

Learning rate We aimed to set a fixed learning rate η to train the network with CD-1 learning (Algorithm 2.2). We did not implement a schedule to reduce the learning rate over time, nor did we include momentum and weight decay, as is commonly done in machine learning studies. However, we did have two learning rates: one when the model received input from the external world, and one when the model was disconnected from the external world. The former learning rate was usually higher than the latter because we had less training events driven by the environment. We come back to this in more details in Section 3.3.2 of result Chapter 3.

Model architecture

The architecture of the RBM is helpful as it greatly simplifies training algorithms, but imposes constraints that move the model away from neurobiology. Other architectures can be considered to relax this constraint, but the downside is that they usually increase the complexity of computations. For instance, the asumption that visible units (resp. hidden) activate independently is not compatible with the activity in the neocortex. We could consider the semi-restricted Boltzmann machine, which is similar to a RBM but allows visible-visible connections (Osindero & Hinton 2008). Similarly, hidden units could be connected with lateral inhibition in the hidden layer (similar to the radial basis Boltzmann machine, Kappen 1995)

Another criticism of RBMs from a biological point of view is that bottom-up and top-down information is processed by the same symmetric weight matrix. The Helmholtz machine (Dayan et al. 1995) is a network similar to the BM, but with two

sets of weights so that bottom-up recognition and top-down generative models are separated. The generative model is trained during a "wake" phase, and the recognition model during a "sleep" phase. Although the Helmholtz machine is an interesting framework to study the formation of knowledge and inference in the brain, research on this network did not follow-through because it is hard to train and it could not compete with simpler models that emerged at the same time.

Finally, our model could benefit from a deeper architecture if we simulated more complex data set, as numerous studies in machine learning show that networks with more hidden layers extract more useful representations. For instance, Salakhutdinov et al. (2013) implemented a deep Boltzmann machine and they showed that the first hidden layer extracted low-level features in the input, while the next layers discovered more abstract representations based on these low-level features. Thus, they could successfully train a hierarchical Dirichlet process prior over the top-level features, which learned the structure of knowledge and could subsequently generalise to new categories. More related to biology, Series et al. (2010) modelled the visual cortex with a deep Boltzmann machine. They introduced homeostatic mechanisms by manipulating the bias of the hidden units and investigated the emergence of hallucinations in Charles Bonnet Syndrome.

Such deeper architectures are therefore relevant to our research, and are promising models for studying memory processes in cortical networks; however, adding more layers means that the weights update require backpropagation, which is not obviously consistent with neurobiology.

2.3 Model for systems memory consolidation

The RBM is capable of learning and subsequently generating patterns that resemble those of the external world, and also making inference on new activity patterns it has never seen before. Knowledge in the RBM is represented as an "energy landscape" defined by the cortical weights (Hinton et al. 1984), and the network can explore it by stochastically updating the states of the neurons.

Káli & Dayan (2004) have exploited this idea to study systems memory consolidation, and they have developed a model of interactions between cortical areas, represented by the RBM, and the hippocampus (see Section 1.2.1). We have built our model on their work and we here introduce the main lines of this framework. However, more detail is provided in the results sections (Chapters 3 and 4).

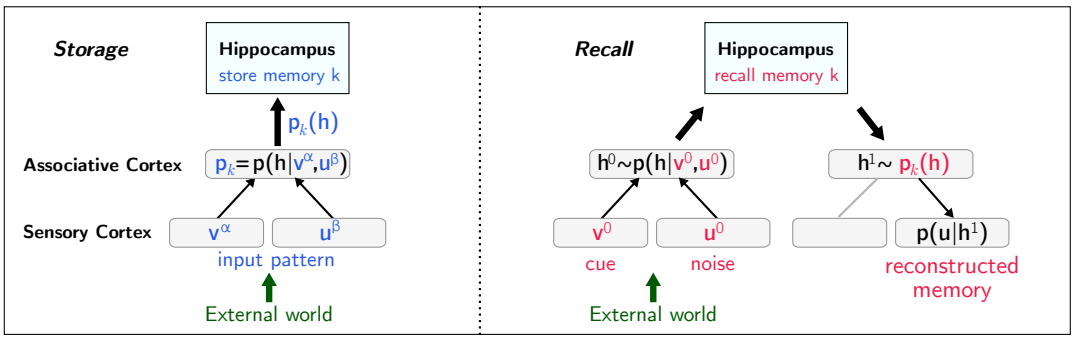


Figure 2.6: Cortico-Hippocampal interaction: memory storage (left) and recall (right).

2.3.1 Interaction with the hippocampus

The bottom layer of the RBM represents sensory areas in the cortex (higher areas, processed information), while the top layer integrates the information and relays it to the hippocampus (Fig. 2.6). Following the models of Alvarez & Squire (1994) and McClelland et al. (1995) (Section 1.2.1), the hippocampus has a temporary role, storing memories and mediating retrieval while these memories are slowly consolidated in the neocortex over time (days to years). Similar to the model of Káli & Dayan (2004), the hippocampus in our framework rapidly learns memory patterns by contrast with the neocortex which has an incremental, slow learning. In practice, the hippocampus takes instantaneous "snapshots" of the activity of the hidden, associative layer of the neocortex (Fig. 2.6). Memories in the hippocampus could be modelled with attractors in a Hopfield network, but we did not need this feature in our work and thus we retained a simple model to store memories, where the hippocampus acts as a temporary database.

The neocortex defines a probability distribution from which we can sample patterns of activity for the visible units. From Eq. 2.6 we can re-write this probability as

$$P\left(\mathbf{v}^{\alpha}\right) = \sum_{\gamma} P\left(\mathbf{v}^{\alpha} \mid \mathbf{h}^{\gamma}\right) P\left(\mathbf{h}^{\gamma}\right) \tag{2.14}$$

The hippocampus stores the conditional probability $P_k\left(\mathbf{h}\right) = P\left(\mathbf{h} \mid \mathbf{v}^k\right)$ of the activity of the hidden units in response to a specific sensory input pattern $\mathbf{v}^k$ in the visible layer. Thus, the activity in the visible layer given the reactivation of hippocampal memory k follows from

$$P\left(\mathbf{v}^{\alpha}|k\right) = \sum_{\gamma} P\left(\mathbf{v}^{\alpha} \mid \mathbf{h}^{\gamma}\right) P_k\left(\mathbf{h}^{\gamma}\right) \tag{2.15}$$

Therefore, the hippocampus can support recall in the neocortex by constraining the possible states of the hidden units. In particular, when the network is cued by a partial input pattern, the hippocampus recalls a corresponding memory k and reinstates the activity according to Eq. 2.15; thus, it allows to bypass the search in the neocortex (i.e. Gibbs sampling steps) that is described in Algorithm 2.1 and instead rapidly reconstruct memory patterns in the sensory layer (Fig. 2.6).

Critically, as we shall see in Chapter 3, memory storage and retrieval by the hippocampus depend on the knowledge in the neocortex, and hence the fact that memories are rapidly formed in the hippocampus does not imply that memory recall will be accurate.

Although we did not test this idea within our framework, the hippocampus could potentially be involved during cortical learning by initiating the negative chain to estimate the second term in the log-likelihood gradient (i.e. to generate $\mathbf{v}^1$ in Algorithm 2.2). Indeed, we mentioned in Section 2.2.2 that a common learning algorithm for RBM is Contrastive Divergence, where the chain starts from an observed training pattern (example in Algorithm 2.2). However, Tieleman (2008) showed that learning is usually better when the negative chain is initialised by persistent "fantasy particles", where the last sampled visible pattern is re-used to start the chain at $\mathbf{v}^1$, a technique called Persistent Contrastive Divergence. Thus, we could assume that the hippocampus mediates the temporary storage and reinstatement of these sensory patterns to compute the second, anti-Hebbian term of the gradient. This idea is consistent with experimental evidence of hippocampal memory replay during wake (during sharp wave-ripples, see Section 1.1.2), and studies have suggested that this replay could support consolidation in cortical areas already during wake (Carr et al. 2011).

Finally, in our model we were not concerned about memory storage capacity in the hippocampus, since we had small data sets that we assumed could all be stored. However, when we repeated the same patterns across training, old memories were overwritten with the newest ones. The motivation was the observation by Káli & Dayan (2004) that over training, as the cortical weights are updated, there is a shift in the representations of sensory stimuli in the hidden layer. Thus, when the hippocampus reinstated a memory pattern in the hidden layer, the reconstruction of the patterns of activity in the sensory cortices no longer matched the original stimulus. Káli & Dayan (2004) showed that off-line replay helped to maintain the correspondence: when a pattern was generated in a sensory cortical area following reactivation in the hippocampus, this pat-

tern was replaced with the closest match among the valid input patterns for this area; information was sent back to the hippocampus which then updated the memory representation. Nevertheless, in our model we assumed that the network did not have access to the set of valid input sensory patterns, and thus we never allowed the hippocampus to update representations during sleep.

2.3.2 Sleep replay and memory consolidation

In line with Káli & Dayan (2004), a training epoch has an experience phase followed by a a sleep replay phase. It is important to note that the experience and replay phases do not correspond to the two phases of the "wake-sleep" algorithm, which was developed to train models like the Helmholtz machine (Dayan et al. 1995), and which in contemporary notations usually refers to the two terms of the gradient when computing the weights change (Eq. 2.12).

The aim of the sleep replay phase is to allow the reactivation of memory patterns, and subsequently to strengthen the cortical weights. This idea was based on experimental studies which suggest that replay of experiences, in particular during sleep, promotes memory consolidation (see Section 1.1.2). In our framework, as in Káli & Dayan (2004), sleep replay is driven by the hippocampus which reinstates activity patterns in the hidden layer, followed by the generation of "dream" patterns in the sensory layer of the neocortex. During this phase, we trained the model on the "dream" patterns using the same method as during the experience phase (CD-1 learning; Algorithm 2.2). Consequently, during sleep the model had to generate training patterns that resembled those observed during experience. This operation was challenging as in most models we did not include "clean-up" connections within the sensory areas of the neocortex, and hence the model could in principle replay any pattern of activity.

We investigated two methods for generating the training patterns during sleep replay. The first method, which we primarily used, was similar to online training: we reactivated a memory in the hippocampus, reconstructed the corresponding activity in the sensory cortices, and the RBM was trained on this pattern. Next, we repeated the procedure by reactivating another memory in the hippocampus (Fig. 2.7, left panel).

The second method, which we used in Chapter 5, consisted in first generating a "batch" of training data, and then presenting in turn each pattern to update the weights (Fig. 2.7, right panel). We preferred the first method as it seemed more in line with biology. On the other hand, it necessitated to calibrate the right number of replay

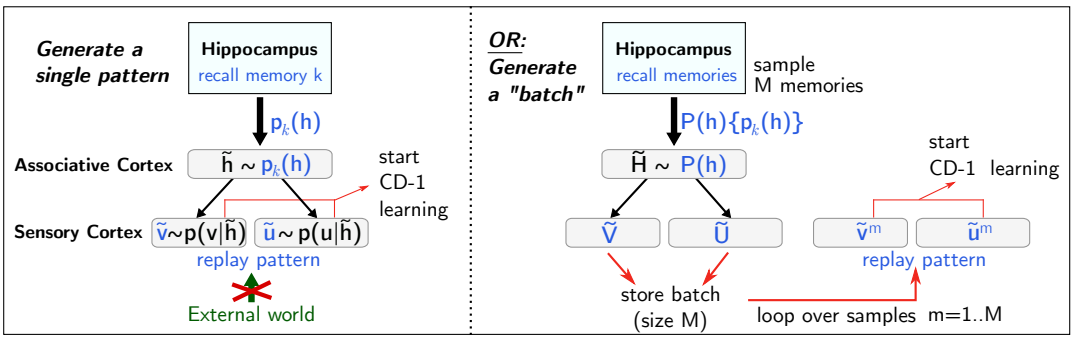


Figure 2.7: Replay of memories during sleep, driven by reactivation in the hippocampus. (Left) Online training during sleep replay: generate a pattern and update the weights. (Right) (1) Generate a "batch" of dream patterns: randomly reactivate multiple memories in the hippocampus, and reconstruct the corresponding activities in the sensory layer using the same set of weights. (2) Loop over the patterns generated to update the weights. Eventually generate a new batch after the update.

events along with the right learning rate so that the weight updates did not bias the generative model. The main downside of the second method was that learning tended to be "stuck", because the model was trained on patterns it could already generate, whereas the first method introduced some noise during training.

Although replay was initiated by the hippocampus in our model, it could easily be modified to start with the reactivation of a pattern in one sensory area of the neocortex. For instance, if replay was initiated with a partial visible pattern, the hippocampus could then recall a memory as depicted in Fig. 2.6 (right panel) in order to replay the associated memory. Such replay would be more compatible with recent experimental work suggesting that replay might start from cortical reactivations (Rothschild et al. 2017).

2.3.3 Evaluating "consistency" in the neocortex

Knowledge in the RBM is coded in the energy function, where plausible patterns of activity have a low energy, while less plausible patterns have a higher energy. We have exploited this property to define a measure of "consistency" of an activity pattern with the knowledge consolidated in the neocortex.

Given a visible activity pattern $\mathbf{v}$, the expected enery is $E(\mathbf{v}) = -\mathbf{x}^\intercal p(\mathbf{h} = 1|\mathbf{v})$, where $\mathbf{x} = \mathbf{v}^\intercal \mathbf{W}$ is the vector with the field (or total input) of the hidden units, and $p(\mathbf{h} = 1|\mathbf{v})$ is the vector with the on-probabilities of the hidden units given $\mathbf{v}$. We then normalised by $\sum_j p\left(h_j = 1 \mid \mathbf{v}\right)$ so that the final value could easily be transformed into a

probability by applying the sigmoid function. The resulting value provides information about the "consistency" of $\mathbf{v}$, given the current knowledge of the RBM. However, we generalised this to evaluate the "consistency" given an arbitrary expectation of the hidden units activity $\boldsymbol{\xi} = p(\mathbf{h} = 1)$:

$$C(\mathbf{x}; \boldsymbol{\xi}) = -\frac{1}{\sum_j \xi_j} \mathbf{x}^\intercal \boldsymbol{\xi} \tag{2.16}$$

In particular, we used Eq. 2.16 during recall mediated by the hippocampus (for instance, Eq. 3.6 in the second model implementation, Section 3.3.1). We also used this measure to track the consistency of the association between two patterns of activity in different sensory areas (Chapter 4).

We believe it is important to monitor this consistency, since new associations can involve new patterns of activity, or simply be novel combinations. Therefore, if the model can detect whether new information is new or familiar, consistent or conflicting with prior knowledge, it is then able to adapt its learning strategy. In our framework we assumed that the prefrontal cortex fulfilled this function.

Chapter 3

Modelling mental schemas

In this chapter we present exploratory work to study the acquisition and adaptation of semantic knowledge in the neocortex. We aim to gain a better understanding of the mechanisms underlying the development of "schemas", knowledge structures in the neocortex, and their influence on subsequent memory processing. To this end, we have implemented various computational frameworks, strongly inspired by Káli & Dayan (2004), and modelled the experiments of Tse et al. (2007), which investigated such schemas with rodents.

The experimental study of Tse et al. (2007) revealed in particular that, with an appropriate pre-existing schema, new memories could rapidly become independent of the hippocampus. These findings thus challenged the traditional belief that systems memory consolidation is always a slow process. Specifically, the discovery of Tse et al. (2007) seemed to challenge the Complementary Learning Systems Theory (CLST) (McClelland et al. 1995), which states that systems consolidation in the neocortex should be slow so as to prevent catastrophic interference with consolidated memories. McClelland (2013) argued that the experimental findings and CLST were actually compatible if one considered the consistency of the new memory with prior knowledge: he showed that neocortical learning in the CLST framework was fast and did not lead to catastrophic interference when learning a new consistent item, whereas slow learning was required when learning a new inconsistent item (see Section 1.2.3).

However, we argue that the consistency with prior knowledge advocated by McClelland (2013) corresponds to the problem of novelty versus familiarity with the environment in the experimental study (e.g. layout of the arena). Yet, the results reported by Tse et al. (2007) indicate that familiarity with the environment was not enough to trigger rapid learning and consolidation of new memories. Indeed, when rodents previ-

ously received information that was inconsistent over training, they did not exhibit fast learning during subsequent training even though they were in a familiar environment. Only when they initially learned consistent information rodents rapidly acquired new memories. This would suggest that the animals have developed an expectation about the reliability of the information in each environment. Thus, we believe the model of McClelland (2013) only partially explains the findings of Tse et al. (2007).

We propose that the first type of consistency, as described in the work of McClelland (2013), is related to *semantic schemas* in associative areas of the neocortex (knowledge), while the second type of consistency, as measure of information reliability, is related to *meta-schemas* (expectation about the structure of knowledge), which might be processed by the prefrontal cortex (Section 1.1.3). We predict that both are decisive for subsequent memory acquisition.

In Section 3.1 we summarise the experimental study of Tse et al. (2007) and describe our modelling approach. In Section 3.2 we show a simple toy model with a training protocol similar to Káli & Dayan (2004); we also introduce the prefrontal cortex module in our framework and the concept of the *meta-schema* in order to better define their relevance for the findings of Tse et al. (2007). Section 3.3 focuses on the formation of a *semantic schema* in the neocortex and its impact on the acquisition of new memories. For this study, we used a modified training protocol and a more suitable representation of the sensory data which allowed us to further investigate the role of sleep and the relationship between hippocampal memory and knowledge in the neocortex. We also investigate the consistency with prior knowledge of the new memories, in line with the work of McClelland (2013). These two models served as foundation for the final model (Chapter 4).

Besides the addition of the prefrontal cortex module, some studies presented here and in the next chapter differ from the work of Káli & Dayan (2004) since we are interested in the gradual formation of semantic knowledge, whereas that study implemented a semantic learning phase (pre-training) before the hippocampus was tasked to store episodic patterns.

3.1 Modelling the schema experiment

In this section we first summarise the experimental study of Tse et al. (2007), and we present our interpretation of their results that we wish to illustrate with our model. We then describe how we modelled the data and the task, and how we evaluated performance in our model. The specific setup used for the simulations will be outlined in each result section.

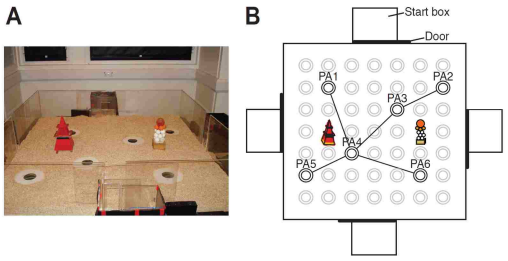
3.1.1 Summary of the experimental study and its significance

In the experiment of Tse et al. (2007), rats first acquired an associative schema by exploring an event arena with six baited sand-wells (Fig. 3.1a). Each sand-well was associated with a different flavour. The rats were trained each day concurrently on the six associations, and they had one trial per association. During each training trial, rats were placed in a randomly selected start box and were given a flavoured food (cue). They then had to fetch more food of that flavour hidden in one of the sand-wells. Slowly, over weeks of training sessions, the rats eventually learned to associate the six specific flavours with the six specific reward locations (Fig. 3.1b). That is, they went directly to the rewarded sand-well, as quantified by the performance index, and also did so during probe trials (denoted PT) when no reward was present.

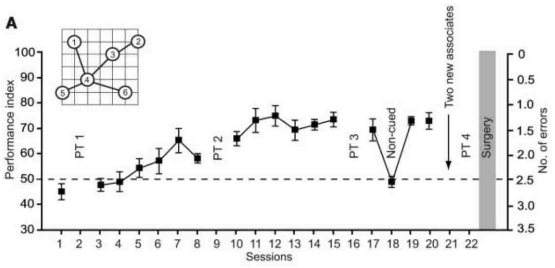
Once the schema had been acquired, two of the original sand-wells were moved to new, neighboring locations (Fig. 3.1c, inset left panel). In a single trial the rats successfully learned to associate two new flavours with each of these new locations (Fig. 3.1c, left panel 'new cued'). Furthermore, the rats still recalled the new associations after a hippocampal lesion only 48hrs after training, suggesting that the new memories were consolidated by then. This is much quicker than typical for cortical systems consolidation in rodents which is thought to take weeks for similar tasks (Section 1.1). Nevertheless, a lesion already 3hrs after training did impair the new memories, suggesting that these memories were initially hippocampal-dependent (Fig. 3.1c, right panel).

Importantly, these experiments revealed that the observed speed-up in learning was due to the learnt schema and not familiarity with the environment: when rats were initially trained in an inconsistent schema, where the associations flavour-place were randomly swapped every two sessions (Fig. 3.1d), single-trial acquisition failed (Fig. 3.1e).

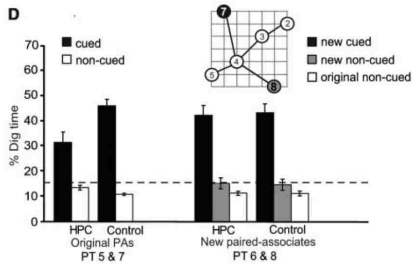
In a follow up study, Tse et al. (2011) investigated the role of the medial prefrontal



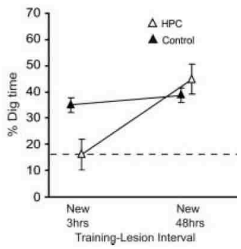
(a) Event arena.

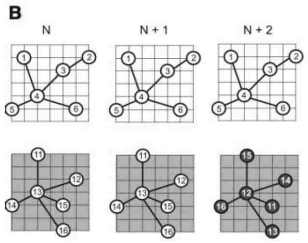


(b) Original training of six flavour-place associations.

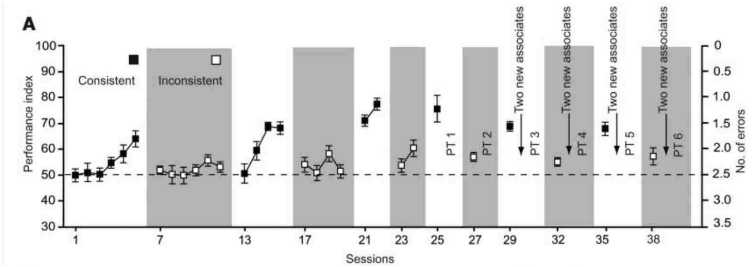


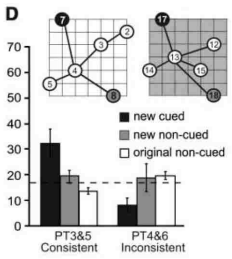
(c) Training two new associations.





(d) Consistent (top) & inconsistent (bottom) associations.





(e) Original training (left) and new training (right) in the consistent and inconsistent schemas.

Figure 3.1: Schema experiment of Tse et al. (2007). Abbreviations: PA=paired associates; PT=probe test; HPC=hippocampal lesions. (a) Photograph of the event arena with the six original sand-wells (left), and schematic of the spatial arrangement (right). (b) Acquisition of the six original associations over training sessions. Rats learned multiple flavour-place associations: after training, given a cue flavour in a start box, the animals recalled the location of the rewarded sand-well. (c) Acquisition of two new pairs of flavour-reward location. Animals with hippocampal lesions made 48 hours after new training recalled both original and new associations during probe tests, indicating that the memories were consolidated in the neocortex (left). Rats did not recall the new associations 14 days later if hippocampal lesions were made 3 hours after new training (right). (d) In the consistent schema, the associations between flavours and places were fixed across sessions, while in the inconsistent schema the associations between the six flavours and six locations were changed every two sessions. (e) Acquisition curve as the rats concurrently learned the consistent schema in one context and the inconsistent schema in another (left). Rats rapidly learned the new associations in the consistent schema context, but not in the inconsistent schema context (right). All panels taken from Tse et al. (2007).

cortex region (mPFC) for the acquisition of the new flavour-place associations. When rats were trained on the new associations in the consistent schema setting, it was accompanied with an increase of immediate early gene expression in the mPFC, suggesting a rapid encoding in this cortical region already at the time of learning. Furthermore, Tse et al. (2011) pharmacologically inactivated the mPFC region during learning, and animals in these conditions did not recall the new associations during a probe test 24hrs after training, indicating that memory consolidation had been disrupted. These findings imply that early encoding in the mPFC was necessary for the rapid assimilation of the new associations.

The two studies raise the question of what mechanisms could explain this fast learning and consolidation. We propose two influential factors.

The first factor is the base semantic knowledge to complete the task in the arena which is slowly consolidated in the associative areas of the neocortex over time. In both consistent and inconsistent arenas, the spatial arrangement of the reward locations, the cues, the room, and the task are stable across sessions, and thus animals become more and more familiar with the environment. We call this knowledge the *semantic schema*. For the example shown in Fig. 1.1, it would correspond to the set of colours and the set of animals categories. We suggest this schema allows rapid encoding in the hippocampus of new rewarded pairs, provided the new locations are compatible with the existing semantic layout.

The second factor is the high-level semantic knowledge underlying the associations, namely if there is a consistent relation between flavours and reward locations or not. We call this knowledge the *meta-schema*, because it captures abstract features that characterise a group of related memories. Note that in this particular example, the associations are characterised by a single *meta-parameter*, namely their consistency. For the example shown in Fig. 1.1, it would correspond to the variability of colours within a category. We suggest that the meta-schema gradually builds in the neocortex, and that its formation and use are supported by regions of the prefrontal cortex. We also suggest that the prefrontal cortex monitors the match/mismatch between an experience and the expectation of the meta-schema, and influences the acquisition of new memories accordingly. Therefore, we expect that if the system is first trained on consistent associations it will rapidly incorporate new associations, while if it is exposed to variable associations during initial training it will slowly incorporate new associations.

Our aim was to develop a computational framework to test these hypotheses and identify potential mechanisms underlying these processes.

3.1.2 Model of the data: flavours and reward locations

The model received inputs from the environment via higher sensory areas of the neocortex. As explained in Chapter 2, the sensory areas were represented by the visible layer of the Restricted Boltzmann Machine (RBM). In this case study, the visible layer had a pool of units that represented the flavours (odors/taste), and another pool of units that represented the reward locations in the event arena.

We worked with two representations for the flavours: 1) each flavour was represented by the activation of one out of N_F units, where N_F is the number of flavours, or 2) we attributed 100 units to the flavour sensory area, and each flavour was mapped to a random binary pattern, in which each unit turned on with probability 0.2.

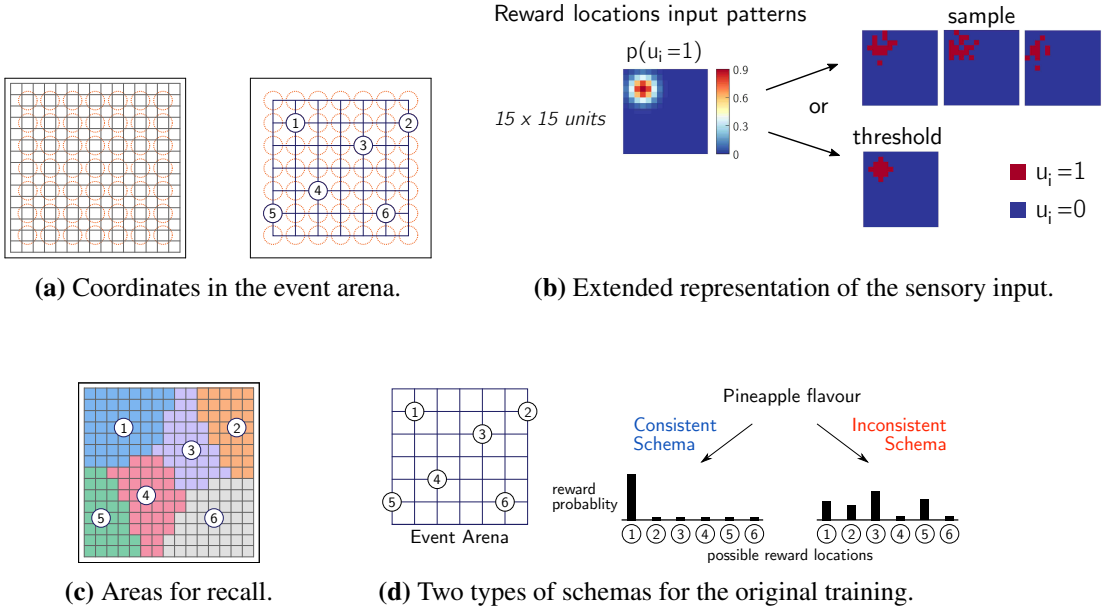
We also worked with two representations for the reward locations: 1) each reward location was represented by the activation of one out of N_L units, where N_L is the number of locations, or 2) we attributed 225 units to the reward locations sensory area, and each reward location was represented according to its coordinates in the arena. For the latter representation, each unit encoded a location with coordinates (x, y) in the arena (Fig. 3.2a). Similar to the experimental study, the arena had 7x7 possible locations for the sand-wells. We used a grid with a step of 0.5 to define the coordinates, which brought the total of reward locations units to 225 (15x15 units, the edges were not encoded). The units were active whenever there was a reward at the location they represented, or in the close neighbourhood. The activation was modelled by a Gaussian tuning curve, centered at the reward coordinates (x_0, y_0) :

$$f(x, y) = r_{max} e^{-\frac{1}{2}\left(\frac{(x-x_0)^2}{\sigma^2} + \frac{(y-y_0)^2}{\sigma^2}\right)}$$

where r_{max} is the maximum on-probability at the reward location, and σ denotes the spatial spread, which is identical in x and y directions (Fig. 3.2b). The 15x15 matrix was then rearranged into a vector containing the probability of activation of the 225 units. Finally, this vector was converted into a binary vector, since the RBM required binary inputs[1]. This conversion could be done in two ways, depending on the simulation (Fig. 3.2b):

1. Sample the activation probabilities ($r_{max} = 0.9$, $\sigma = 0.7$). In this case, the input data varied at each presentation.

[1] As explained in Chapter 2, RBMs can have real values units, but for our modelling purpose we chose to use binary units.



(a) Coordinates in the event arena.

(b) Extended representation of the sensory input.

(c) Areas for recall.

(d) Two types of schemas for the original training.

Figure 3.2: Model of the schema experiment of Tse et al. (2007). (a) Model of the event arena. (Left) Grid 15x15, where each square is encoded by a reward location unit in the sensory cortical layer. The orange circles are the potential sand-wells sites (the size of the circles is non-representative). (Right) Spatial arrangement of the six original sand-wells. Note: in the experimental study, the rodents entered the arena via one of the start boxes located at each side. The choice of the start box varied across trials to prevent procedural learning. However, in our modeling the point of entry did not matter and hence we did not take into account the start boxes. (b) Extended representation of the reward locations in the sensory layer of the neocortex. The left inset shows the probability of activation of the location units. To obtain the input binary patterns that we presented to the network, we either sampled or used a threshold. (c) With the extended representation of the locations shown in (b), the probability to go to a location is derived by taking into account the activity of the units that encode neighbouring coordinates. The figure shows the areas surrounding the six original sand-wells. For example, the activity of the units in the blue region contribute to the probability to go to the closest sand-well, which is no.1 in this case. (d) The two types of schemas for the original training. In the consistent schema, the flavour Pineapple is always associated with the sand-well no.1. In the inconsistent schema the reward location associated with flavour Pineapple is different every two training epochs.

2. Threshold the activation probabilities ($r_{max} = 1$, $\sigma = 1.5$, $threshold = 0.8$). In this case, the input data were identical at each presentation.

For the simulations in Section 3.2, we used the simplest representations for both flavours and reward locations. For the simulations in Section 3.3, we kept the simple representation for the flavours, but used the extended representation of the reward locations. For the final model and simulations in Chapter 4, we used the extended representations for all sensory inputs.

The advantage of the simple representations is that we can keep the size of the RBM small, and computations are easier. However, the extended representations are more realistic, as semantic memories are believed to have distributed representations in the neocortex rather than local representations.

Finally, the extended representation of the reward locations must not be confused with the place cells of the hippocampus. We are not modelling spatial navigation, but rather the specific locations of rewards in some representation of the event arena in the neocortex. Indeed, when Tse et al. (2007) lesioned the hippocampus at the end of training the rats correctly recalled the reward locations (Fig. 3.1), suggesting that the hippocampus was not required to find the locations, and, for instance, used the visual cues of the environment.

3.1.3 Memory tasks and training

Conform to the experimental study of Tse et al. (2007), the first task consisted of teaching the model to associate six different flavours with six sand-wells locations. We call this task the 'original training'. The six sand-wells were initially arranged in the arena as shown Fig. 3.2a. From the point of view of the neocortex (RBM), the task was to correlate the activity of the flavour units with the activity of the reward location units in the sensory layer.

In analogy with an experimental training + sleep/rest cycle we defined a 'training epoch', which comprised an *experience phase* followed by a *sleep replay phase*. During the *experience phase*, the model received inputs from the environment and was trained on the flavours and reward locations stimuli presented in the sensory cortex. The parameters of the associative cortex (RBM), the hippocampus and the prefrontal cortex were updated during that phase. During the *sleep replay phase*, the model was disconnected from the environment and generated its own sensory stimuli for training. Only the parameters of the associative cortex (RBM) were updated during that phase.

The exact training protocol was different for each model we implemented, and hence we describe the method in more details in the relevant sections.

We implemented two types of schemas for the original training, as in the experiment (Fig. 3.2d). In both conditions, we defined six flavours and six reward locations. In the first schema, called the 'consistent schema', the same flavour was associated with the same location over training. In the second schema, called the 'inconsistent schema', the associations were swapped over training every two epochs.

After the training of the six original associations, the model was presented with two new associations. We call this second memory task the 'new training'. Similar to the experiment, two original sand-wells, no.1 and no.6, were displaced and associated with two new flavours.

3.1.4 Memory performance

When cued with a flavour, the recall (explained below in Section 3.2.1) led to the activation of the location units in the sensory cortex, i.e. the on-probability of the units $p(u_i = 1)$ similar to Fig. 3.2b. We converted this activity into a probability distribution over the reward locations (sand-wells). For this conversion, we assumed that the probability to go to a certain location was proportional to the activity of the corresponding location unit.

The method for the conversion depended on the type of data representation (Section 3.1.2). For the simple representation, where each reward location was represented by only one unit in the sensory cortex, we simply normalised the on-probabilies of the location units. This gave us the probability to recall each reward location. For the extended representation, we assumed that the probability to visit a specific sand-well was equivalent to the probability to visit the region surrounding that sand-well (Fig. 3.2c). The probability to recall a reward location was then calculated by taking into account the activity of the all the units surrounding that sand-well, relative to the total activity in the arena. We evaluated memory performance in our model by measuring the average probability to recall the correct reward location associated to each flavour.

By default, memory was probed with the hippocampus active during recall, hence combining both the episodic and semantic contributions. To assess semantic memory, and track the time course of memory consolidation in the neocortex, we disabled the hippocampus during recall (mimicking the hippocampal lesions in experiments).

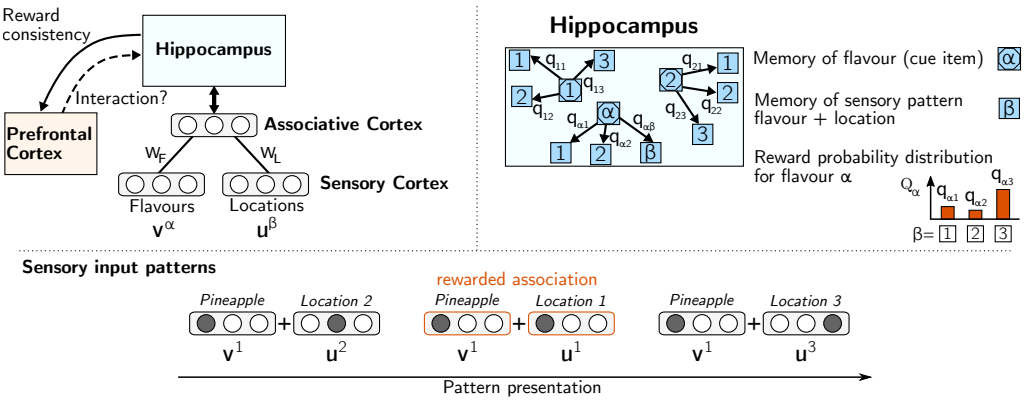


Figure 3.3: First version of the model. Global model architecture (left panel). The sensory cortex receives inputs from the environment. There is one unit per flavour, one unit per location, and one unit for the context (not shown). The hippocampus is connected to the associative layer of the neocortex. The prefrontal cortex learns the consistency of the rewarded associations, and our hypothesis is that it somehow influences learning in the hippocampus. The hippocampus (right panel) stores memories of all possible combinations $\left(\mathbf{v}^\alpha, \mathbf{u}^\beta\right)$ and the corresponding reward probabilities $q_{\alpha\beta}$.

3.2 The prefrontal cortex and the meta-schema

In this section we present our first attempt at modelling the experiment of Tse et al. (2007). In particular, we demonstrate why the abstract knowledge about the task is important to account for the experimental data. For this study, we used a simple toy model and we implemented a training protocol that was directly inspired by the original model of Káli & Dayan (2004). We also introduced the prefrontal cortex module, which was not present in the original model, to investigate its potential role.

3.2.1 Model setup

Model of the associative neocortex and sensory input data

For this first implementation we used a small network and the simple data representation: each potential reward location (sand-well), and each flavour, was represented by one unit in the sensory layer of the neocortex (visible layer of the RBM, Fig. 3.3). We added a context unit to indicate that the training took place in the same arena, and hence this unit was always in the 'on' state. The purpose of context units was to allow the network to learn multiple schemas in different contexts, but we never implemented

such multi-training, and thus we kept only one context unit[2]. We arbitrarily set the number of hidden units to the total number of possible flavour-location combinations used for the original training (36 units).

We note $\mathbf{v}^\alpha$ the vector of visible units $v_i \in \{0,1\}$ representing a flavour α, $\mathbf{u}^\beta$ the vector of visible units $u_i \in \{0,1\}$ representing a location β, $\mathbf{W}_F$ and $\mathbf{W}_L$ the weight matrices connecting the flavour and location units to the hidden layer. The vector with the state of the hidden units $h_j \in \{0,1\}$ is noted $\mathbf{h}$. In addition, we note $p(\mathbf{h}=1)$ the vector with as elements the probabilities $p\left(h_j = 1\right) \in [0 \ldots 1]$. Purely for convenience we assume that flavour 1 is associated to location no.1, and so on, so that the associations to be learned are simply given by $\alpha = \beta$.

Model of the hippocampus

During training the hippocampus took snapshots of the field of the hidden units $\mathbf{x}^{\alpha\beta}$ when the network was presented with a flavour pattern $\mathbf{v}^\alpha$ and a location pattern $\mathbf{u}^\beta$ (**step 1** in Fig. 3.5):

$$\mathbf{x}^{\alpha\beta} = \mathbf{W}_F^\intercal \mathbf{v}^\alpha + \mathbf{W}_L^\intercal \mathbf{u}^\beta \tag{3.1}$$

All flavour-place associations were memorised in the hippocampus, mimicking exploration. The hippocampus also created a memory for each flavour α by averaging the fields of the hidden units of the memories linked to that flavour:

$$\bar{\mathbf{x}}^\alpha = \langle \mathbf{x}^{\alpha\beta} \rangle_\beta \tag{3.2}$$

where $\langle\cdot\rangle_\beta$ is the average over the potential locations, computed element-wise. This flavour memory pattern was used during recall to recognise the flavour cue when we tested the network performance (procedure described below). The memory patterns $\mathbf{x}^{\alpha\beta}$ and $\bar{\mathbf{x}}^\alpha$ were overwritten at each epoch.

Finally, the hippocampus keeps track of the reward probability of each flavour-place association. We note $q_{\alpha\beta}$ the probability of reward at location β given the flavour α, and $Q_\alpha = \left\{q_{\alpha\beta}\right\}_\beta$ the reward probability distribution associated to the flavour memory α (Fig. 3.3, top right panel). As such, $\sum_\beta q_{\alpha\beta} = 1$. In contrast to the memory vectors $\mathbf{x}^{\alpha\beta}$ which were snapshots, the reward probabilities $q_{\alpha\beta}$ were increased by a fixed increment Δq over epochs (procedure described below).

[2]The context units are equivalent to the categories units in the data set used by McClelland (2013), such as “living thing”, “animals”, “plants”, etc...

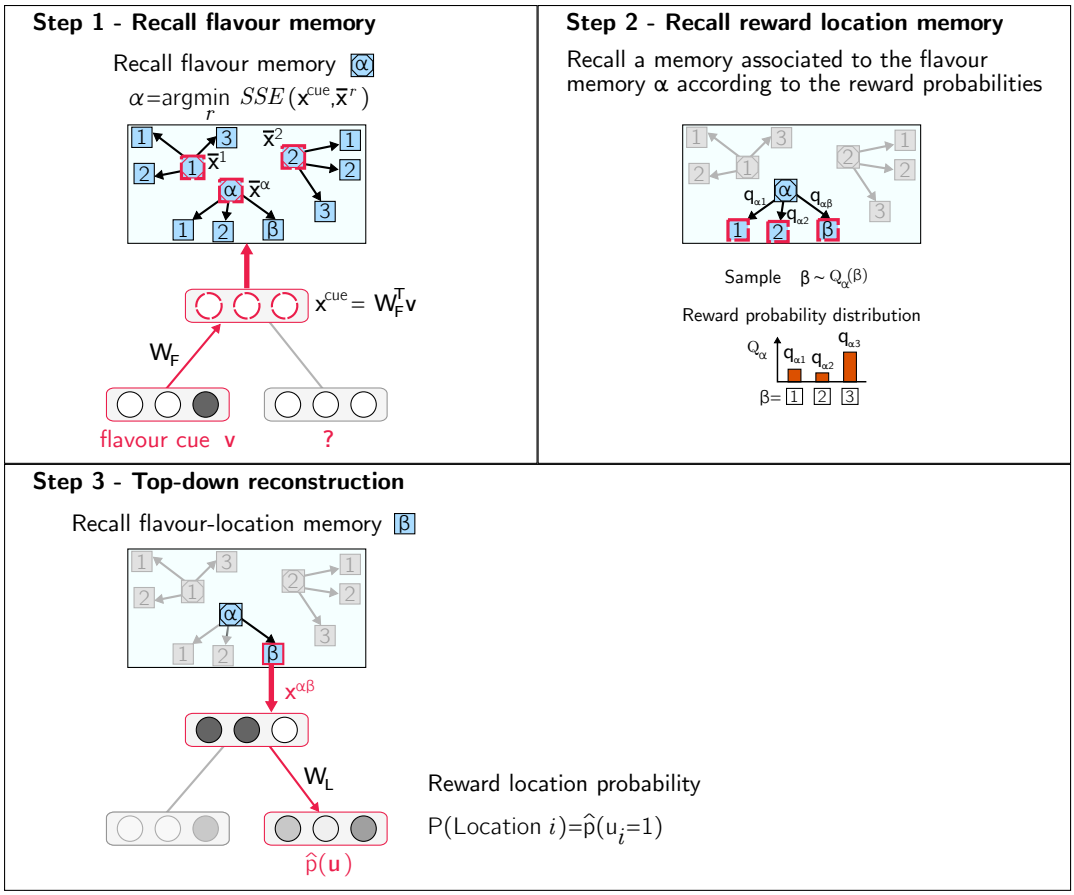


Figure 3.4: Recall mediated by the hippocampus. **Step 1**: A flavour cue is presented to the sensory cortex, and the activity propagates to the associative layer of the neocortex. The hippocampus computes the sum of squares errors between the resulting field vector $\mathbf{x}^{\text{cue}}$ and each of the stored field vectors $\bar{\mathbf{x}}^r$, the latter vectors corresponding the memories of the flavours. The memory that minimises the error is chosen. **Step 2**: Given the flavour memory α, the hippocampus can sample one of the associated flavour-place memories using the probability distribution of rewards Q_α. **Step 3**: Finally, the hippocampus reinstates the field $\mathbf{x}^{\alpha\beta}$ of the hidden units that corresponds to the sampled memory. The neocortex reconstructs the activity in the sensory layer via top-down connections to infer the reward location.

Model of the prefrontal cortex

We introduced a module that we identified as the prefrontal cortex. We attributed two main functions to this module. First, in line with experimental theories, we assumed that the prefrontal cortex could detect the context, and subsequently biased recall in the hippocampus towards memories relevant to the current context (Preston & Eichenbaum 2013). Although we did not implement the mechanisms whereby the prefrontal cortex biased hippocampal recall, we implemented two memory retrieval strategies in the hippocampus to mimic context-dependent or independent memory recall (next section). The second function of the prefrontal cortex was to extract the meta-schema. As explained in Section 3.1.1, the rewarded associations were characterised by their consistency across presentations. We defined a scalar variable ϕ^* as meta-parameter to capture the overall consistency of the associations. To compute the meta-parameter ϕ^* we simply took the expected reward probability, which we calculated by averaging the highest reward probabilities $q_{\alpha\beta}$ associated to each flavour memory α:

$$\phi^* = \langle \max_{\beta} \left(q_{\alpha\beta} \right) \rangle_{\alpha} \tag{3.3}$$

but in the last version of the model (Chapter 4) we introduce a different method to compute the meta-parameter.

The meta-parameter captures the shape of the distribution of rewards: for the task in the consistent schema the distribution would be tuned to a particular reward location, i.e. the reward expectation ϕ^* would be high, whereas the distribution would be broad for task in the inconsistent schema, i.e. expectation would be low.

Memory recall and performance

During recall, one of the flavours from the training set was presented as cue in the sensory layer of the neocortex. The network computed the field vector $\mathbf{x}^{\text{cue}} = \mathbf{W}_F^{\intercal}\mathbf{v}$ of the associative cortical layer (Fig. 3.4; note that we included the context unit with the flavour units to simplify the notations). The hippocampus compared the field vector $\mathbf{x}^{\text{cue}}$ with the stored flavour memory patterns $\bar{\mathbf{x}}^r$ and selected the flavour memory α that minimised the sum of squares error[3] (**Step 1** in Fig. 3.4):

$$\alpha = \underset{r}{\operatorname{argmin}} \|\mathbf{x}^{\text{cue}} - \bar{\mathbf{x}}^r\|_2^2$$

[3] On a more detailed level, such a computation could be done by a Hopfield network, in that case, the recalled flavour is the one the network converges to after being initialised with the cue.

Once the cue was identified, the hippocampus sampled one of the associated memories β according to the reward probability distribution Q_α (**Step 2** in Fig. 3.4). The hippocampus set the field of the hidden layer to $\mathbf{x}^{\alpha\beta}$ to reinstate the activity of the hidden units $p(\mathbf{h}=1)=\sigma\left(\mathbf{x}^{\alpha\beta}\right)$, where $\sigma(\cdot)$ is the sigmoid activation function ('on' probability) applied element-wise to the field vector $\mathbf{x}^{\alpha\beta}$. The neocortex then reconstructed the activity of the locations units $p(\mathbf{u}\mid\mathbf{h})$ in the sensory layer (**Step 3** in Fig. 3.4). We normalised the probabilities of activation to obtain the final 'on' probability $\hat{p}(u_i=1)$ of each location unit. Since each location unit represented a potential reward location, the probability to go to location no.1 for instance was $P(\textit{location 1})=\hat{p}(u_1=1)$.

As we mentioned above, we implemented a second recall pathway for the hippocampus, which we supposed was used when the prefrontal cortex did not recognise the context. We propose that the hippocampus in this case relied on semantic memory in the neocortex rather than episodic memory. For semantic-based recall, the model directly compared the activation of the hidden units $p(\mathbf{h}|\mathbf{v})$ resulting from the presentation of flavour $\mathbf{v}$, with the activations corresponding to memories of associations stored in the hippocampus (hence, replacing **steps 1** and **2** in Fig. 3.4). It then chose the memory pair $\alpha\beta$ that minimised the sum of squares error:

$$\{\alpha,\beta\}=\underset{rk}{\operatorname{argmin}}\left\|\mathbf{h}-\sigma\left(\mathbf{x}^{rk}\right)\right\|_2^2 \tag{3.4}$$

where $\mathbf{h}\sim p(\mathbf{h}|\mathbf{v})$ with $p(\mathbf{h}=1\mid\mathbf{v})=\sigma(\mathbf{x}^{\text{cue}})$, and $\mathbf{x}^{rk}$ is the field of the memory pair r,k in the hippocampus (flavour r and location k).

By default, recall was mediated by the hippocampus, and used the first, context-dependent method of recall. To test whether the memories were consolidated in the neocortex, we disabled the hippocampus during recall and let the RBM perform 5 Gibbs sampling steps to reconstruct the activity of the location units (see Cortical recall in Section 2.2.1).

Network training

Each epoch began with an **experience phase** during which the sensory layer of the neocortex received inputs from the environment, i.e. the flavours, the reward locations, and the context (Fig. 3.5). First, a flavour was randomly selected in the data set. Next, all the possible combinations of that flavour with all potential reward locations, even the non-rewarded combinations, were successively presented to the network in

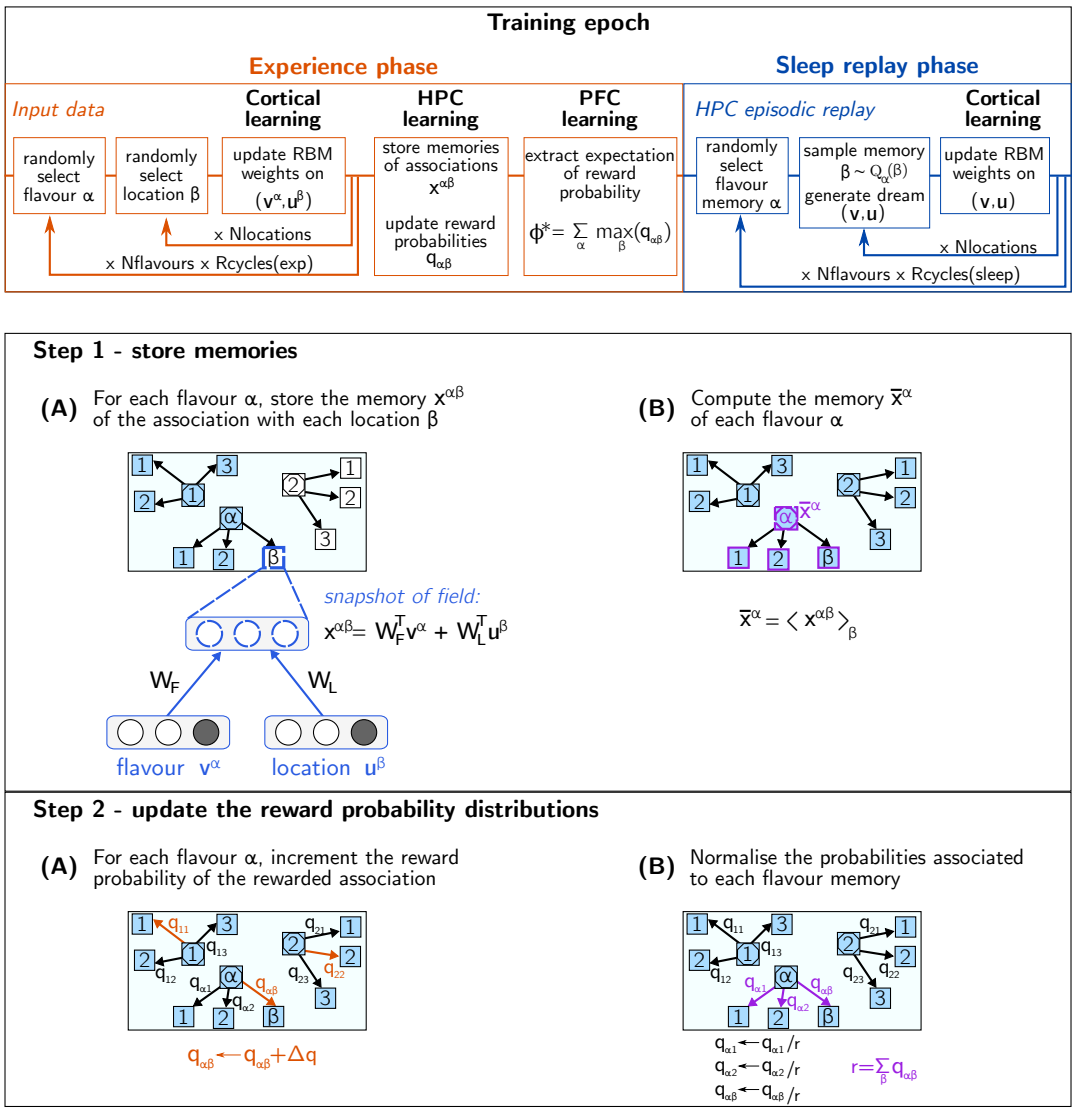


Figure 3.5: (Top panel) Training epoch. During the experience phase, the model sees all possible combinations of flavour pineapple with locations 1 to 6. The hippocampus stores the field $\mathbf{x}^{\alpha\beta}$ of the associative layer for each association $\left(\mathbf{v}^{\alpha}, \mathbf{u}^{\beta}\right)$ presented, and updates the reward probabilities (see lower panel for details). Lastly, the prefrontal cortex updates the meta-schema. During the sleep replay phase, memories are sampled in the hippocampus according to the reward probability distributions $Q_{\alpha} = \left\{q_{\alpha\beta}\right\}_{\beta}$. This memory reactivation leads to the replay of 'dreams' in the sensory cortices via the top-down connections. The hippocampus samples memories corresponding to the same flavour multiple times in a row (n being equal to the total number of potential reward locations) simply to mirror the experience phase. See Section 2.2 for details about cortical learning and sleep replay. (Lower panel) Hippocampal learning during the experience phase.

Training:		Original	New
reward locations		6	8
input patterns		6x6	2x8
memories in HPC		6x6	8x8
visible units		13	17
hidden units		36	36
cycles per epoch	EXP	50	150
	SLEEP	100	300

weights initialisation		mean 0 std 0.1
learning rate	EXP	0.05
	SLEEP	0.001
reward increment		0.1

Table 3.1: Model and training parameters. Notations: HPC = hippocampus; EXP = experience phase; SLEEP = sleep replay phase. The coloured numbers are parameters that will be modified during the simulations (see main text in the results sections). Number of input patterns during the experience phase / number of memory patterns in hippocampus: $n \times m$, where n is the number of flavours, and m the number of reward locations. For the cycles during the experience and sleep replay phases, refer to the diagram of the training epoch (Fig. 3.5).

a random order. This can be thought of the rat exploring the arena and visiting both rewarded and unrewarded sand-wells. The cortical weights were updated at each presentation (using CD-1 learning, see method in Section 2.2.2), even for non-rewarded experiences (we comment on this later). The model cycled 50 times[4] through all the flavours and applied the same protocol. The repetitions were necessary to update the cortical weights using a lower learning rate (see Table 3.1), which is common practice for training neural network so as to gradually incorporate information without erasing previous experience. However, the learning rate had to be large enough during the experience phase so that the network could later generate dream patterns that resembled the sensory input patterns (sleep replay phase).

After the cortex had updated its weights, the hippocampus memorised the representations of all the associations, again both rewarded and non-rewarded (we comment on this in the next paragraph). The hippocampus stored the field vector $\mathbf{x}^{\alpha\beta}$ of the hidden units (Eq. 3.1) corresponding to each flavour-place association $\left(\mathbf{v}^{\alpha}, \mathbf{u}^{\beta}\right)$ (**step 1** in lower panel of Fig. 3.5). Next, for each flavour memory α the hippocampus adjusted the reward probability distributions Q_{α} over the locations (**step 2** in lower panel of Fig. 3.5). To update the distribution, we simply incremented the probabilities $q_{\alpha\beta}$ corresponding to the rewarded associations ($\beta = \alpha$) by a fixed value Δq, and normalised.

[4]The rats had a limited number of trials to learn the associations at each session, but one could argue that some kind of replay took place, with working memory for example.

We chose an arbitrary, small value as the default reward increment (see Table 3.1).

For the last stage of the experience phase the prefrontal cortex extracted the meta-parameter ϕ^* (Eq. 3.3), which in this case represented the overall consistency of the rewarded associations.

Next, the model entered the **sleep replay phase**. The hippocampus randomly selected a flavour memory α, and sampled one of the associated memories according to the reward probability distribution Q_α. The hippocampus clamped the hidden layer to the selected memory pattern, and given this activation the cortex generated a 'dream' pattern in the sensory layer (Fig. 2.7). This procedure was the same as **steps 2** and **3** during recall, shown Fig. 3.4, except that it reconstructed the activity of all the visible units: context, flavours, and locations units. The final activity of the flavours and locations units was derived using a winner-take-all step. The cortical weights were trained on this 'dream' pattern (binary), using the same protocol as during the experience phase (CD-1 learning, see method in Section 2.2.2) but with a lower learning rate (Table 3.1). We decreased the learning rate during sleep replay to allow the gradual incorporation of memories and prevent catastrophic interference. The hippocampus cycled 100 times through all the memories of the flavours.

Modified experience phase. In the experience phase described above, the network updated the weights for all possible associations, rewarded and non-rewarded. During sleep on the other hand, the hippocampus preferentially replayed the memories with high reward probability. Therefore the consolidation of the correct associations was only possible because of sleep replay. However, we noticed that over the course of training, the cortical updates during the experience phase deteriorated what was consolidated during sleep: when we probed memory at the end of the experience phase, performance was lower than when we probed memory just after sleep. We could have compensated this effect by further increasing the number cortical updates during off-line replay, like Káli & Dayan (2004). However, instead we set a criterion to stop the update of the cortical weights during the experience phase at some point during training. The criterion was simply a recall test: learning stopped when all the rewarded locations had the highest probability of being recalled. The recall test was performed at the start of each experience phase. If the system fulfilled the stopping criterion, the experience phase only consisted of hippocampal learning and the update of the meta-parameter by the prefrontal cortex. Hippocampal learning was still necessary, because cortical plasticity during sleep replay altered the representations in the asso-

ciative layer of the sensory input; hence, the hippocampus had to re-store the field vectors $\mathbf{x}^{\alpha\beta}$ (Eq. 3.1) for each flavour-place association.

A different approach could have been to mimic the search behaviour of the animals by selecting the associations according to their reward probabilities, similar to the sleep replay phase. As the rat learns, it will presumable visit the unrewarded locations less and less, and hence the patterns presented during experience would reflect this choice.

Important remarks about training

In the current model the cortical weights were updated for both rewarded and non-rewarded events during the experience phase. This method was also implemented in the model of Káli & Dayan (2004). This update rule, however, is questionable as it seems at odds with the idea that memory persistence in the brain is strongly influenced by relevance (e.g. reward), which was shown experimentally by the action on LTP of the neurotransmitter dopamine (Lisman et al. 2011). This was the main reason for changing the training protocol in the next versions of the model. In the current model, if the network only learned the rewarded associations during the experience phase the neocortex could learn simple, consistent associations without sleep replay. However, in Section 3.3 we emphasise that sleep replay promotes the consolidation in the neocortex of more complex patterns of activity, and in Chapter 5 we propose that sleep replay is particulary necessary to incorporate multiple memories that interfere with each other.

Similarly, the choice of storing non-rewarded events in the hippocampus was also questionable, for the same reasons that we stated above about cortical learning. However, this method gave more flexibility for the hippocampus as it could recall different locations for a given flavour. It may seem like a waste of resources, but only when we a priori know that the associations are consistent. It is reasonable to assume that at the start of training, without prior knowledge, the rats do not know yet whether the information is reliable - for that matter, it might not even be relevant for the future. This is the case for the inconsistent schema where the associations are swapped over the course of training. In the next Chapter 4 we describe a different model for the hippocampal memory system which is a compromise between flexibility and efficiency.

3.2.2 Results: The acquisition of a consistent schema

We first investigated the acquisition and retrieval of the six original flavour-place associations. To check that there was not much variation in our results, we ran five simulations, each with a different random initialisation of the cortical weights (weights sampled from a Normal distribution, see parameters in Table 3.1). We initialised the reward probability distributions in the hippocampus as uniform distributions, i.e. $q_{\alpha\beta} = \frac{1}{6}$ for each flavour-location memory pair $\alpha\beta$ (right panel of Fig. 3.3). For all simulations the model stopped updating the cortical weights during the experience phase (coincidentally) after six epochs. After the sixth epoch, the cortical-cortical connectivity (RBM) was modified solely during sleep replay.

During training the probabilities of the rewarded association memories $\{q_{11}, \ldots, q_{66}\}$ in the hippocampus increased from the baseline value $\frac{1}{6}$ to a high value close to 1 (Fig. 3.6c). Recall performance of the correct reward locations improved in parallel over time (Fig. 3.6a, solid line). Furthermore, over time the rewarded associations memories in the hippocampus were more likely selected for replay during sleep, and thereby were gradually consolidated in the neocortex. When the hippocampus was disabled during recall, performance was initially lower but had less effect later in training, indicating a successful consolidation (Fig. 3.6a, dashed line).

At the end of the experience phase of each epoch, the prefrontal cortex averaged the highest reward probabilities $q_{\alpha\beta}$ of all flavour memories α in the hippocampus (Eq. 3.3). This average probability was then stored in the prefrontal cortex as the meta-parameter ϕ^*, and reflected the expectation about the consistency of the rewarded associations. Since the associations were consistent over training, the value of the meta-parameter ϕ^* increased and converged to 1 (Fig. 3.6b). On the other hand, if the associations were inconsistent over training, the meta-parameter ϕ^*, and hence the expected reward probability, would remain at the baseline value.

The recall probabilities of the rewarded locations that we evaluated from the activity in the sensory cortex were lower than the reward probabilities $q_{\alpha\beta}$ ($\beta = \alpha$) in the hippocampus (see Fig. 3.7a, compare solid blue line to semi-dashed black line). This effect could be the result of (i) the hippocampus recalling the wrong flavour memory, and/or (ii) the neocortex poorly reconstructing the activity of the location units in the sensory cortex (top-down reconstruction). We found that already after one epoch the recall of the flavour memories in the hippocampus was 100% accurate (not shown). Therefore recall performance with hippocampus was limited by the reconstruction of

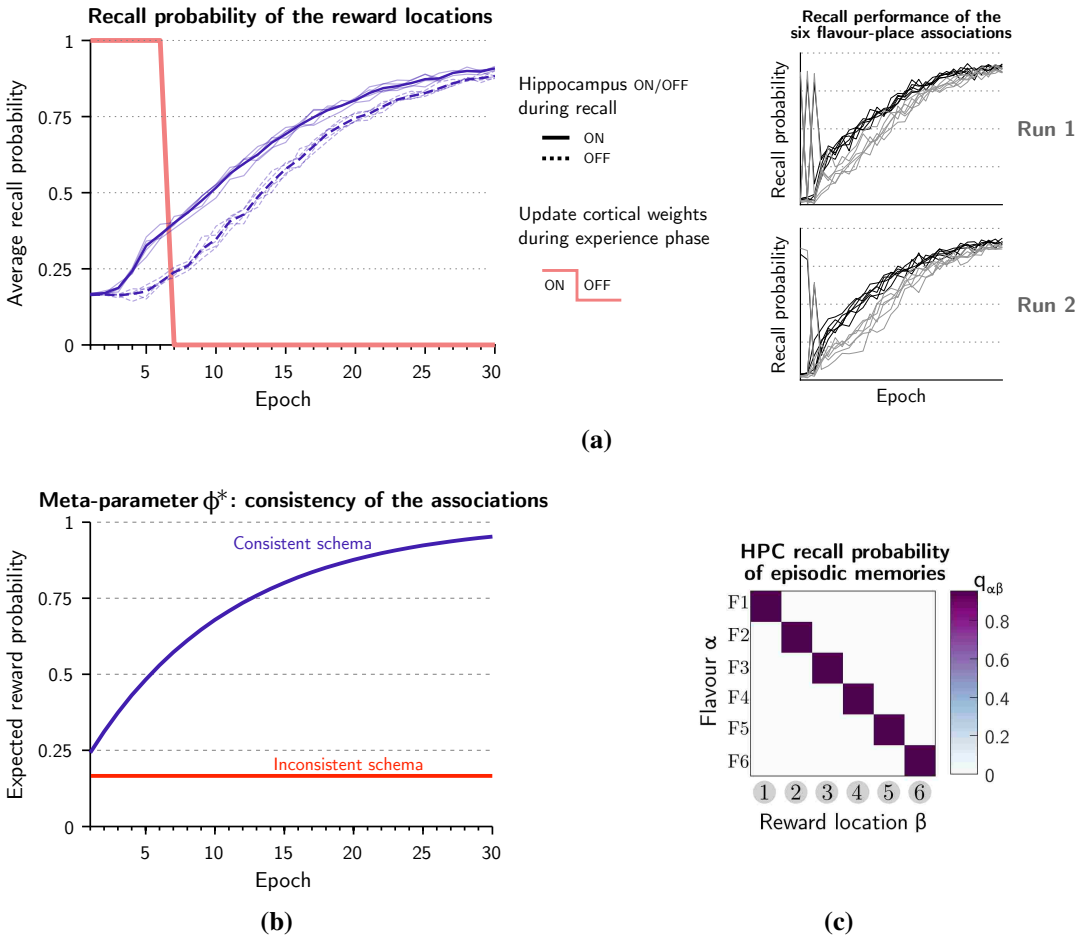


Figure 3.6: Training of the six original flavour-place associations in the consistent schema. (a) Recall performance over training epochs. Performance is the probability to recall, given a flavour, the correct rewarded location. The red line indicates the epochs when the cortical weights are updated during the experience phase. Solid blue lines: recall with hippocampus. Dashed blue lines: hippocampus disabled at recall (only). The thin lines are the average performance of each run (five RBM initialisations). The thick lines are the average over the five simulation runs. Note that a performance of 0.5 means that the correct locations are recalled on average with probability 0.5, and does not mean that on average 50% of the associations are recalled. This is illustrated in the two insets, which show the individual recalls of the six associations for two runs (black: hippocampus on; grey: hippocampus off). (b) The prefrontal cortex learns the meta-schema over time. The meta-schema tracks the consistency of the rewarded associations: the meta-parameter ϕ^* increases for the consistent schema but not for the inconsistent schema. (c) Reward probability distribution $Q_\alpha = \left\{q_{\alpha\beta}\right\}_\beta$ over the location memories β, for each flavour memory α in the hippocampus at the end of training. The hippocampus correctly recalls with high probability (diagonal $\beta = \alpha$).

the sensory patterns in the neocortex. As the flavours and reward locations were gradually consolidated in the neocortex, the difference was reduced towards the end of training. This observation also suggests that even if the probabilities $q_{\alpha\beta}$ of the rewarded associations ($\beta = \alpha$) in the hippocampus were equal to one from the start of training, the increase in recall performance would still be gradual while the cortex adjusts its weights.

Impact of blocking the replay of episodic memories during sleep

The current model stores both rewarded and unrewarded associations in the hippocampus. The reason that in the end the rewarded ones prevail is that the rewarded ones are replayed more often during sleep. To further demonstrate this effect we implemented a random replay where all associations stored in the hippocampus had equal chance of being selected, instead of a replay based on reward probabilities in the hippocampus.

As expected, when we disabled the hippocampus for recall, performance was at chance level throughout training (Fig. 3.7b), indicating that consolidation of the rewarded associations failed. Since the reward probabilities $q_{\alpha\beta}$ in the hippocampus were updated the same way as in the control condition, the probabilities of the rewarded associations had the same evolution as the semi-dashed black line Fig. 3.7a. Yet, the measured recall performance was even lower than the control condition (Fig. 3.7a). We checked that the hippocampus recalled the correct flavour memory, and found that this recall was perfect (not shown). Therefore this result further demonstrate the impact of the top-down cortical reconstruction of sensory patterns even when recall is mediated by the hippocampus.

Despite this limitation, the network still had more than 50% chance to recall the correct rewarded location by the end of training. This result would imply that cortical consolidation of the six specific flavour-place associations is not required to perform the task, as long as the episodic memories are stored in the hippocampus. Blocking the replay of rewarded experiences might be detrimental in the current task, yet we believe it could be helpful in situations where the consolidation of specific associations is not required, if for instance these associations are not particularly relevant for the future, while the hippocampus allows tempory storage and recall.

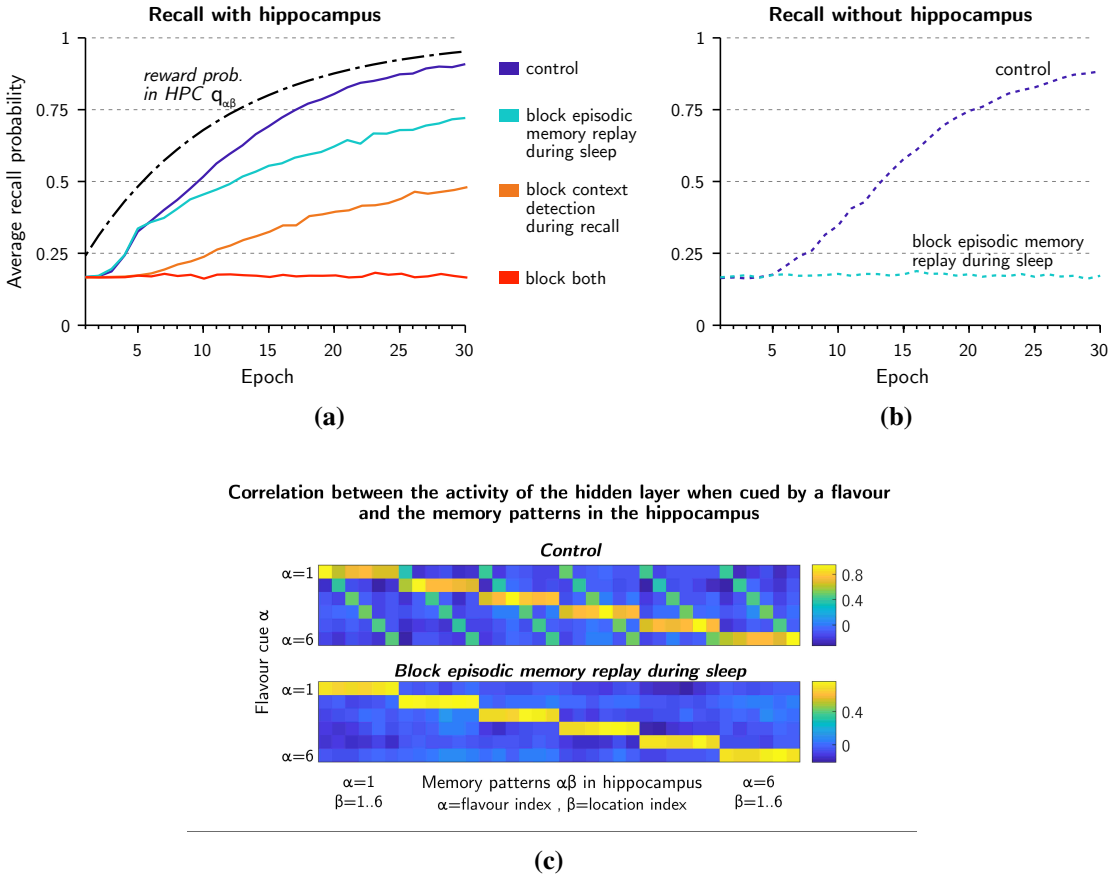


Figure 3.7: Impact on training of 1) blocking hippocampal replay of episodic memories during sleep (light blue lines), 2) blocking episodic memory recall, to mimic failure of context detection (orange line), or 3) blocking both (red line). The control condition (dark blue lines) corresponds to the results shown in Figure 3.6. (a) Performance when recall is mediated by the hippocampus. The semi-dashed black line shows the evolution of the reward probabilities $\{q_{11}, \ldots, q_{66}\}$ in the hippocampus corresponding to the correct associations (represented by one curve since these probabilities are all equal). The solid lines show the recall performance measured after the hippocampus recalled the memory and the neocortex reconstructed the sensory pattern. (b) Performance with the hippocampus disabled during recall to evaluate memory consolidation. (c) When the context detection fails (orange and red lines in (a)), the hippocampus relies on memories consolidated in the neocortex. To do so, it measures the correlation between the activities of the hidden units when cued by a flavour sensory pattern α, and the activities when the hippocampus reinstates stored memories.

Impact of blocking episodic memory retrieval in the hippocampus

Here we consider a new arena (say, with different walls colours) but with the same spatial arrangements of the sand-wells. Because it is a new context, the rule about the associations might have changed. Hence, when presented with one of the known flavours, it might be better to visit the closest potential reward location instead of going to the supposedly associated one at the other side of the arena. In our model, we modelled this flexibility of behaviour by allowing two recall pathways in the hippocampus (*recall* in Section 3.2.1).

When we blocked episodic memory recall, the hippocampus relied on the knowledge consolidated in the neocortex. Hence, we expected recall performance to be similar to the recall when we disabled the hippocampus (Fig. 3.6a, dashed line). However, this was not the case (Fig. 3.7a, orange line). Instead, performance increased at a low pace, and by the end of training, recall probability was lower than 0.5 when the hippocampus mediated recall, although recall probability without hippocampus was close to 0.9 since the associations were consolidated in the neocortex. The decrease of performance can be explained by looking at the correlations between the sensory-driven activities $p(\mathbf{h} = 1|\mathbf{v}^{\alpha})$, corresponding to the flavour α, and all the hippocampal-driven activities $p(\mathbf{h} = 1) = \sigma\left(\mathbf{x}^{\alpha\beta}\right)$, corresponding to the memories of associations flavour α and location β (Fig. 3.7c, upper panel). Despite the consolidation in the neocortex, when the network was cued by a flavour the hippocampus the resulting activity in the associative cortex was correlated with all the flavour-place memories in the hippocampus that corresponded to that flavour. Thus, the hippocampus had a chance to reactivate each memory β linked to the flavour memory α, but more likely reactivated the correct reward memory ($\beta = \alpha$) [5]. Therefore, the hippocampus could have two context-dependent recall strategies, one based on episodic memory which allowed the network to complete the task, and one based on semantic information which allowed the network to have a broader search.

On the other hand, if the network replayed random associations during sleep, hippocampal recall performance was flat throughout training when it relied on semantic memory (Fig. 3.7a, red line), because in this case the associations were not consolidated. This is further illustrated by the correlations between the activity in the associative cortex and the memories stored in the hippocampus (Fig. 3.7c, lower panel), which

[5] Note that if we used the activity $p(\mathbf{h}|\mathbf{v})$ instead of sampling $\mathbf{h}$ in Eq. 3.4, semantic-based recall performance in the hippocampus would be the same as episodic-based recall performance. Hence, the decrease in recall performance as observed in Fig. 3.7a (orange line) was only obtained with sampling.

show that when the network was cued with the flavour α, the hippocamus reactivated all the memory traces linked to that flavour equally.

3.2.3 Results: The acquisition of new associations

During the original training, the prefrontal cortex learned a meta-schema: it extracted the meta-parameter ϕ^* which captured the overall consistency of the flavour-place associations (Fig. 3.6b). However, the model has not used this information yet. Here we illustrate why this information is relevant for the data of Tse et al. (2007).

In line with the experimental paradigm, we trained a network on the task with the consistent schema and we examined how this network learned two new pairs of flavour-place associations (Fig. 3.2d). To check that there was not much variation in our results, we tested the acquisition of the new associations using ten networks that were trained on the original flavour-place associations. The training lasted 10 epochs instead of the 30 epochs used for the initial training.

To model the new flavours and new locations we simply added one unit for each in the visible layer of the RBM. The initial weights connecting the new units to the hidden layer were sampled from a Normal distribution with the same parameters as for the original training (Table 3.1). When we evaluated recall performance using the reconstruction of the activity of the location units (see Fig. 3.4), the reward probability distribution was defined over eight potential locations[6], the six original locations no.1 to no.6, and the two new locations no.7 and no.8.

The same training protocol was applied to the two new flavours (training epoch Fig. 3.5). All possible combinations were presented as input to the sensory layer of the neocortex, i.e. each new flavour was shown with both the new and original locations. The cortical weights were updated at each presentation, and the hippocampus stored each memory. During the first epoch, the reward probability distributions associated to the new flavour memories $Q_\alpha = \left\{q_{\alpha\beta}\right\}_\beta$ ($\alpha \in \{7,8\}$, $\beta \in \{1,\ldots,8\}$) were initialised as uniform distributions. We tripled the number of cycles during the experience phase (data set repetitions) compared to the original training (see Table 3.1). The aim was to boost the update of the cortical weights of the new units. The number of cycles

[6]Note that in the experiment of Tse et al. (2007) they closed the sand-wells no.1 and no.6 and opened neighbour sand-wells no.7 and no.8 (Fig. 3.1). However, in the current simulations the data representation did not take into account the spatial arrangement of the sand-wells. Hence we did not consider the proximity of the new locations to calculate performance. This of course will be an upgrade in the next models.

during sleep replay (memory reactivations) was increased accordingly in order to keep the same ratio with experience.

Despite the fact that we increased the number of cortical updates, the model learned the new associations only slowly with the baseline reward update parameter in the hippocampus (i.e. increment $\Delta q = 0.1$) (Fig. 3.8a, light green lines).

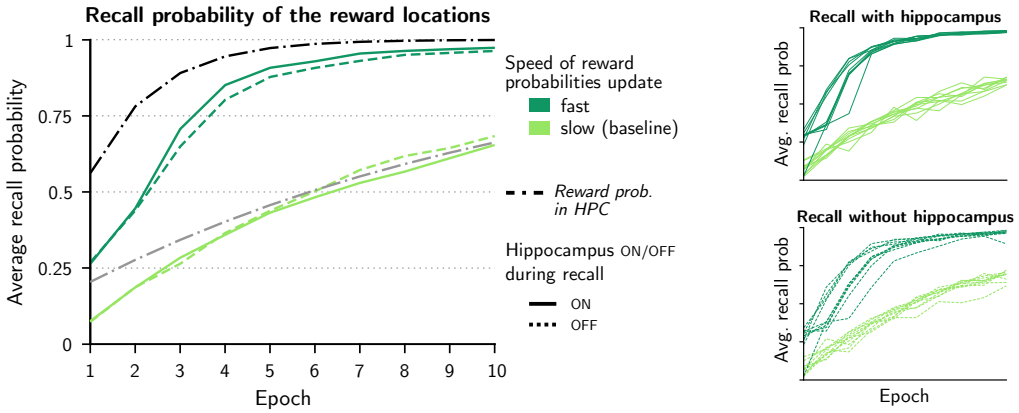
On the other hand, if the model could take the expectation about the consistency of associations into account, it might somehow adjust the rate of update of the reward probabilities in the hippocampus to generalise the prior belief to the new associations. For instance, for the consistent schema the learned reward probability expectation is high (Fig. 3.6b), and hence the model can increase Δq as it trusts the new associations. By contrast, for the inconsistent schema the learned reward probability expectation is low (Fig. 3.6b), and hence the model does not increase Δq as it considers that the new associations are unreliable.

In the current model we manually adjusted the parameter Δq for hippocampal learning at the start of the new training ($\Delta q = 0.1$ for baseline slow update, $\Delta q = 1$ for rapid update), but in the final model (Chapter 4) the prefrontal automatically modulated the learning in the hippocampus.

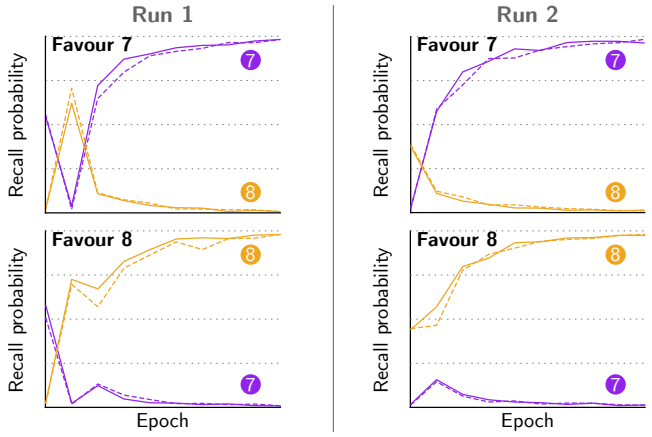
When the model had a fast reward update it reached more rapidly the same end-performance of the original training (Fig. 3.8a, dark green lines). After one epoch of training, the reward probabilities $q_{\alpha\beta}$ of the new associations in the hippocampus ($(\alpha, \beta) \in \{(7,7),(8,8)\}$) were higher than 0.5 (semi dashed black line); as a result, the new associations memories were more likely replayed during sleep, and hence the new associations were rapidly consolidated (Fig. 3.8a, dashed lines); note how it closely tracks the performance when recall is mediated by the hippocampus (solid curve).

As noted in Section 3.2.2, the observed recall performance was lower than the reward prediction in the hippocampus (semi dashed line). We found again that the limiting factor was not the identification of the flavour cue, but rather the top-down sensory reconstruction in the neocortex.

We further investigated why performance was low at the start in the case of the fast reward update. We found that one of new locations was learned at the expense of the other new location. For example, for the simulation run 1 in Fig. 3.8b, after one epoch of training the network more likely recalled the new location no.7 regardless of the flavour tested; in the next epoch, it was the opposite. Performance increased and fluctuations were reduced over training as the connectivity in the neocortex had undergone plasticity.



(a) Performance for fast versus slow reward update in the hippocampus.



(b) Performance of the two new flavours with fast reward update in the hippocampus (two runs).

Figure 3.8: Acquisition of two new associations (new flavours + new locations) with a consistent prior schema. (a) Recall performance of the new associations with the basic slow reward update in the hippocampus (light green lines), or with the fast update (dark green lines). The semi-dashed lines show the evolution of the reward probabilities $\{q_{77}, q_{88}\}$ in the hippocampus corresponding to the new associations. The two insets show the average performance of each simulation (10 runs, same legend as main plot), while the main plot shows the average over all runs. (b) Each panel shows the probability to recall the new locations no.7 or no.8 (purple and yellow lines respectively) when the network is cued with the new flavours, for two runs. The solid lines are for recall with hippocampus, the dashed lines are for recall without hippocampus.

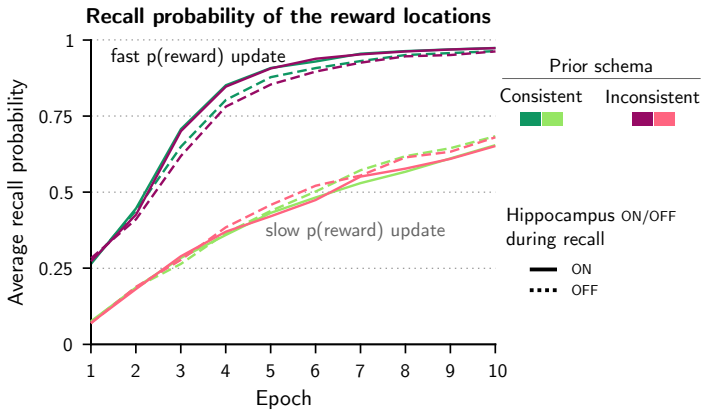
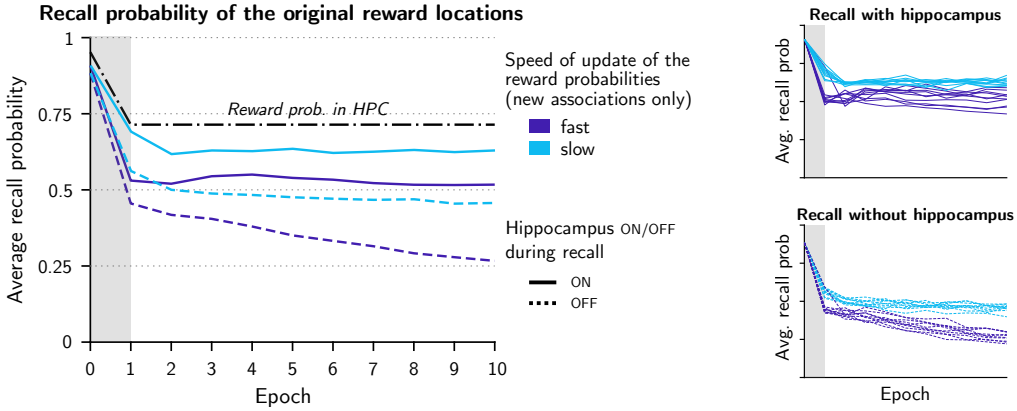


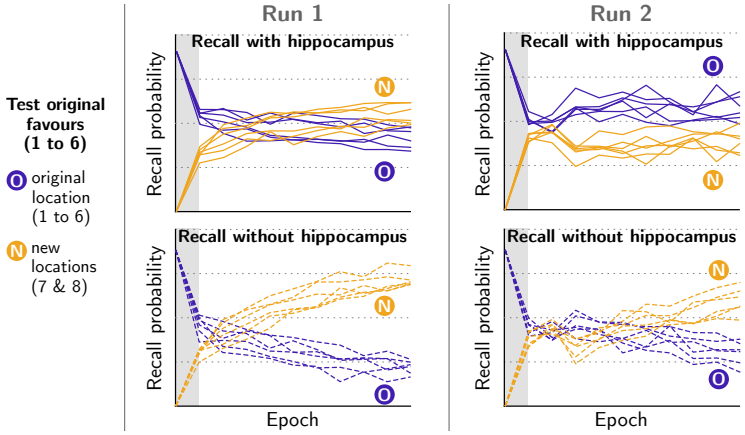
Figure 3.9: Comparison of the acquisition of two new associations with a consistent or inconsistent prior schema. The light colours show the case of the baseline slow update of the reward probabilities $q_{\alpha\beta}$ ($(\alpha,\beta) \in \{(7,7),(8,8)\}$) in the hippocampus. The dark colours show the case of a fast update. We suggest that the speed of the reward update in the hippocampus can be regulated by the prefrontal cortex to differentiate between consistent and inconsistent schemas.

What would happen in the inconsistent schema? To mimic the task in the inconsistent schema, we trained ten networks on the original flavours, but with random sleep replay. In this case, the original training looked like the training when we blocked episodic replay, described Section 3.2.2 and shown Fig. 3.7a&b (light blue lines). In addition, the reward probability distributions Q_α associated to the flavour memories in the hippocampus were uniform distributions.

We then trained the network on the new associations with the two types of reward update (baseline increment $\Delta q = 0.1$ or fast update $\Delta q = 1$). In contrast with experimental data of Tse et al. (2007), we found no difference between performance with consistent or inconsistent prior schemas (see Fig. 3.9 vs. Fig. 3.1d&e). Therefore, in our model, the knowledge consolidated in the associative neocortex cannot by itself account for the difference in the speed of acquisition of the new associations that was observed in experimental results. We thus propose that learning in the hippocampus should be modulated depending on prior consistency of associations. We suggest that the prefrontal cortex uses the meta-schema to decide whether hippocampal learning should be up-regulated or not.



(a) Interference with fast versus slow reward update in the hippocampus



(b) Example of two runs with fast reward update in the hippocampus.

Figure 3.10: Interference with the original flavour-place associations. Performance at epoch 0 corresponds to a test before the start of new training, indicated by the grey shaded area. (a) Recall performance of the original associations with the basic slow reward update in the hippocampus (light blue lines), or with the fast update (dark blue lines). The semi-dashed line show the evolution of the reward probabilities $\{q_{11}, \ldots, q_{66}\}$ in the hippocampus corresponding to the correct original associations (represented by one curve since these probabilities are all equal). These probabilities were updated once at the start of training to include the new locations memories. The two insets show the average performance of each simulation (10 runs, same legend as main plot), while the main plot show the average over all runs. (b) Each panel shows the probability to recall the original rewarded locations (dark blue) or the new locations (no.7 or no.8, yellow lines) when the network is cued with the six original flavours.

3.2.4 Results: Fate of the original associations

In the previous part we modulated the speed of acquisition of two new flavour-place associations by manually changing the increment during the update of the reward distributions in the hippocampus. Here we investigate the impact on the retention of the original associations.

During the first epoch of training the hippocampus also stored the memories of the associations $\mathbf{x}^{\alpha\beta}$ (Eq. 3.1) between the original flavours $\mathbf{v}^{\alpha}$ ($\alpha \in \{1,\ldots,6\}$) and the two new locations $\mathbf{u}^{\beta}$ ($\beta \in \{7,8\}$). However, when we presented these associations to the network, the cortical weights were not updated. In addition, the reward probabilities in the hippocampus corresponding to these associations were set to the baseline value $q_{\alpha\beta} = \frac{1}{6}$, and for each flavour memory α the reward probability distribution was normalised so that $\sum_{\beta=1}^{8} q_{\alpha\beta} = 1$. Thus, the reward probabilities of the original rewarded memories $\{q_{11},\ldots,q_{66}\}$ decreased slightly (semi-dashed line Fig. 3.10a; from $q \simeq 0.95$ to $q \simeq 0.71$). We chose to modify the reward expectation so that, given an old flavours, the model could select one of the new locations (mimicking novelty effect). In the absence of feedback for the old flavours the model had no reason to re-adjust its expectations. Thereafter the reward probability distributions associated with the original flavour memories were fixed for the rest of training.

Recall performance of the original associations deteriorated upon learning the new associations (Fig. 3.10a, solid lines). For instance, with the fast reward update, when cued by the original flavours the network progressively recalled the new locations instead of the old ones (Fig. 3.10b, solid lines in the upper row). However, performance was lower than the hippocampal reward prediction (semi dashed line Fig. 3.10a). This result implies that interference was not solely due to the modification of the reward probabilities $q_{\alpha\beta}$ in the hippocampus, but was also due to the sudden change of connectivity in the associative cortex to incorporate the new flavours and locations. Consequently we found a stronger effect for the fast update in the hippocampus of the new reward probabilities (dark blue lines).

We then probed the semantic memories of the original associations by disabling the hippocampus during recall. Recall performance also decreased upon learning the new associations (Fig. 3.10a, dashed lines), which further demonstrates that cortical plasticity was responsible for the interference observed when the hippocampus mediated recall. However, while recall with hippocampus stabilised during training, semantic knowledge continued declining over training for the fast reward update (Fig. 3.10a,

dark blue dashed lines). When cued with the original flavours after a few epochs of training, the network already had more chance to select the new locations than to select the original ones (Fig. 3.10b, dark blue dashed lines). This significant drop of supposedly consolidated memory was problematic. If anything, one would expect the opposite result: hippocampal memory is believed to be flexible so as to rapidly incorporate novel information, whereas semantic memory is thought to be gradually updated so as to preserve relevant information established in the neocortex.

Catastrophic forgetting is a well known issue in connectionist networks (McClelland et al. 1995, French 1999), and various computational models have shown that interleaved learning of new and old information (or a noisy version of old information) can prevent it. In line with this, we allowed the hippocampus to reinstate old patterns during sleep replay in order to compensate for interference, or at least minimise their effect (until now the model only replayed the memories related to the new flavours, i.e. the pairs flavours $\alpha \in \{7,8\}$ and locations $\beta \in \{1,\ldots,8\}$). We ran again ten simulations for the case of the fast reward update in the hippocampus. We initially kept the same number of cycles during sleep replay (300 cycles, see Table 3.1).

To assess the impact of replaying old memories we examined the recall performance at the end of training. We first examined recall with hippocampus, and we found that the performance for the original associations was closer to the reward probability predicted by the hippocampus (Fig. 3.11, recall with hippocampus, top right plot; the light blue histogram is closer to the semi-dashed line). In addition, this reactivation did not impair the recall of the new associations (Fig. 3.11, recall with hippocampus, top left plot; the only difference noted was a lower performance at the start of training, which was 0 instead of ~ 0.25 as in Fig. 3.8a, dark green).

We then checked the impact on memory consolidated in the neocortex by disabling the hippocampus during recall. We found that the interleaved replay during sleep prevented the new locations from completely overwriting the old ones (Fig. 3.11, recall without hippocampus, top right plot). However, this came at the expense of the consolidation speed of the new associations (Fig. 3.11, recall without hippocampus, top left plot). We increased the number of cycles (reactivations) during sleep replay in the hope to speed-up the reorganisation of neocortical connectivity, i.e. to more rapidly incorporate new information while preserving old knowledge. Increasing to 450 cycles helped the consolidation of the new flavour-place associations, but larger increase (600 cycles) proved to be ineffective (Fig. 3.11, recall without hippocampus, bottom row). This result is not very surprising, because when we increased the number of cycles we

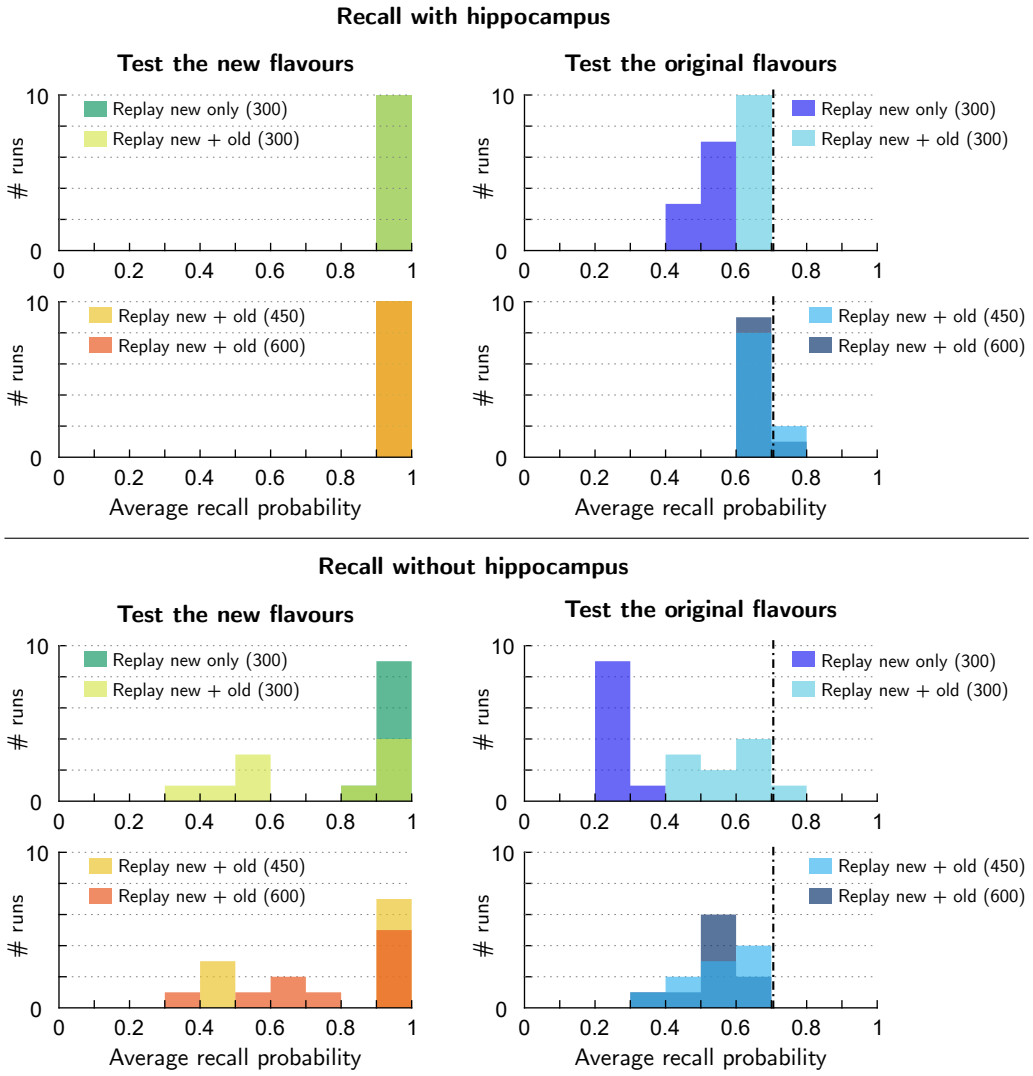


Figure 3.11: Interleaved learning with prior knowledge during sleep replay: impact on end-of-training performance for the new associations (plots on the left side) and for the original associations (plots on the right side). The baseline performances (dark green and dark blue) correspond to Fig. 3.8a (dark green lines) and Fig. 3.10a (dark blue lines). The plots in the bottom panel show the performance with the hippocampus disabled during recall to evaluate memory consolidation. Each histogram shows the distribution of the performances obtained with ten simulation runs. The colours correspond to various number of cycles used during sleep replay (300, 450 and 600). The semi-dashed line represents the updated reward probability of the original associations $q \simeq 0.71$ in the hippocampus.

did not alter the proportion of reactivations of new versus original memories. A replay schedule that prioritises the reactivation of new associations would probably be more effective.

On the other hand, a more surprising result in our model is that the hippocampus seems to be necessary to maintain the memories of both new and old associations, while the neocortex undergoes plasticity. This result implies that if the hippocampus is lesioned after new training, but before sleep, the old semantic memories would be impaired, and that only if lesions are made after sleep the old semantic memories would be preserved.

3.3 The neocortex and the semantic schema

In the preceding Section 3.2 we have presented the implementation of a simple network to model the experiment of Tse et al. (2007). We introduced the idea of the meta-schema to capture a global feature about the task, which in this particular case was the consistency of the rewarded flavour-place associations. We have shown that without the meta-schema, prior knowledge has no impact on learning new associations, in contrast with the experimental results of Tse et al. (2007).

However, we did not fully address the representation of knowledge about the flavours and locations encountered and its impact on the new training. This knowledge is stored in the connections between the sensory and associative cortices (RBM), and we call it semantic schema. In the previous simulations we could not explore this aspect as we implemented a simple representation of the sensory input that did not take into account the spatial arrangement of sand-wells in the arena. In this section, we chose a more elaborate representation of the reward locations so as to examine how the network reshapes its connectivity to incorporate new associations. In particular, this new representation allowed us 1) to investigate the role of replay for the formation of the semantic schema and the impact of cortical consolidation on recall mediated by the hippocampus (Section 3.3.2), 2) to investigate the impact of the prior semantic schema on the acquisition of new associations (Section 3.3.3), and 3) to investigate the impact of the consistency with prior knowledge of the new associations (Section 3.3.4).

In addition, in the first model the cortical weights were updated even for non-rewarded events; we have highlighted that such cortical update rule diverges from the "NeoHebbian" framework which predicts that the persistence of a memory depends on its relevance (Lisman et al. 2011). Thus, in this second computational model we

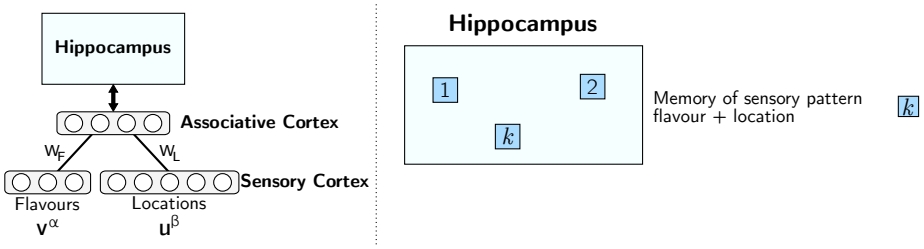


Figure 3.12: Second version of the model. Global model architecture (left panel). The sensory cortex receives inputs from the environment. There is one unit per flavour, while the location patterns are encoded by 225 units (Fig. 3.2b). The hippocampus is connected to the associative layer of the neocortex. This model did not include the prefrontal cortex module. In contrast with the first model (Fig. 3.3), only rewarded events are stored in the hippocampus (right panel).

modified the plasticity protocols so that both cortical weights update and hippocampal storage were restricted to rewarded events (Fig. 3.12).

3.3.1 Model setup

Note that in this second version of the model, we did not include the prefrontal cortex module as it did not play a role during the simulations. Nonetheless, we will re-introduce this module in Chapter 4.

Model of the associative neocortex and sensory input data

To model the spatial arrangement of the rewarded sand-wells we used the representation described in method section 3.1.2, where 225 location units encode the locations with coordinate (x,y) in the arena (Fig. 3.2b). The motivation of this change was to obtain a meaningful representation that could resemble semantic knowledge, where each location was defined in relation to the others. This representation was also more suitable to explore the adaptation of the semantic schema to include new reward locations, as these could be close or far from the original ones. We retained the simple representation for the flavours, where each flavour was encoded by one visible unit. The hidden layer of the cortical network had 200 units.

Since we only presented the rewarded flavour-place associations to the model, we changed the notations of indexation: $\mathbf{v}^k$ and $\mathbf{u}^k$ are the vectors of visible binary units representing the flavour and location of the kth rewarded association. Furthermore, to model the reward locations we used binary patterns $\mathbf{u}^k$ sampled from the activation

probabilities (Fig. 3.2b). Otherwise, we kept the same notation: $\mathbf{W}_F$ and $\mathbf{W}_L$ are the weight matrices connecting the flavour and location units to the hidden layer, $\mathbf{h}$ is the vector with the state of the hidden units $h_j \in \{0, 1\}$, and $p(\mathbf{h} = 1)$ is the vector with elements $p\left(h_j = 1\right)$.

Model of the hippocampus

The hippocampus only stored rewarded memories, as opposed to the first model where the hippocampus stored any incoming information regardless of its relevance (Section 3.2.1). Hence, the hippocampus no longer required to store the reward probabilities of each association to support recall and replay, as the only memories available were already the rewarded events.

The method to store memories in the hippocampus was similar to the method in the first model. The hippocampus took snapshots of the field of the hidden units when the network encountered a rewarded flavour-place association $\left(\mathbf{v}^k, \mathbf{u}^k\right)$ (Fig. 3.14):

$$\mathbf{x}^k = \mathbf{W}_F^\intercal \mathbf{v}^k + \lambda \mathbf{W}_L^\intercal \boldsymbol{\mu}^k \tag{3.5}$$

where $\boldsymbol{\mu}^k = p(\mathbf{u} = 1 \mid k)$ is the vector with the 'on' probabilities of the location units corresponding to the rewarded association k. We used the probabilities $\boldsymbol{\mu}^k$ as input instead of the binary samples $\mathbf{u}^k$ so that the field of the hidden units would reflect the average activity given a reward location. To compensate for the fact that we used real values between 0 and 1 for the location units, we weighted their contribution by a factor $\lambda = 2$. The memory patterns $\mathbf{x}^k$ were overwritten at each epoch.

Memory recall and performance

To test the network performance, we presented the flavours from the training set in the sensory layer of the neocortex. The network computed the on-probability of the hidden units $\boldsymbol{\xi}^{\text{cue}} = p(\mathbf{h} = 1 \mid \mathbf{v}) = \sigma\left(\mathbf{W}_F^\intercal \mathbf{v}\right)$ (**step 1** in Fig. 3.13). The hippocampus compared the activation $\boldsymbol{\xi}^{\text{cue}}$ of the associative (hidden) layer with the stored memory patterns $\mathbf{x}^k$ using the consistency measure $C(\mathbf{x}; \boldsymbol{\xi}) = -\frac{1}{\sum_j \xi_j} \mathbf{x}^\intercal \boldsymbol{\xi}$ (Section 2.3.3), and then sampled a memory k according to the softmax distribution:

$$p\left(\mathbf{x} = \mathbf{x}^k\right) = \frac{\exp\left(-C\left(\mathbf{x}^k; \boldsymbol{\xi}^{\text{cue}}\right)/T\right)}{\sum_r \exp\left(-C\left(\mathbf{x}^r; \boldsymbol{\xi}^{\text{cue}}\right)/T\right)} \tag{3.6}$$

where $T = 0.1$. We had to take a lower value for the temperature instead of $T = 1$ to obtain better recall performance. Next, the hippocampus clamped the hidden units to

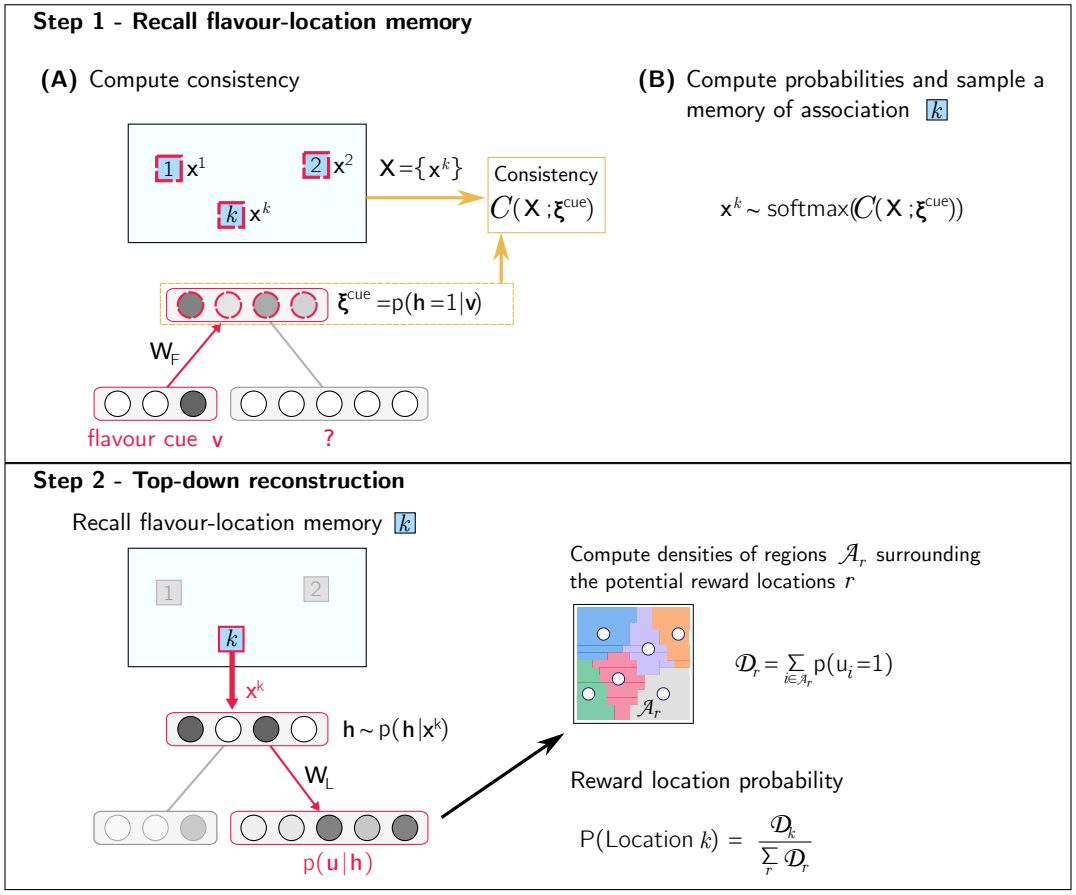


Figure 3.13: Recall mediated by the hippocampus. **Step 1**: A flavour cue is presented to the sensory cortex, and the activity propagates to the associative layer of the neocortex. Given the activity of the associative layer, the hippocampus computes the probability to recall each stored memory, and then samples a memory k. **Step** 2: The hippocampus reinstates the field $\mathbf{x}^k$ of the hidden units that corresponds to the sampled memory. The neocortex reconstructs the activity of the location units in the sensory layer via top-down connections. Finally, we derive the probability to go to each reward location from the activity of the location units.

the selected memory pattern $\mathbf{x}^k$, and the neocortex reconstructed the activity of the locations units $p(\mathbf{u} \mid \mathbf{h})$ in the sensory layer using the top-down cortical connections $\mathbf{W}_L$ (**step 2** in Fig. 3.13). As we explained in Section 3.1.2, we derived the probability to go to a location by summing the on-probabilities of the units surrounding it, and normalising by the total activity: if $\mathcal{A}_r$ is the region surrounding the potential reward location r, the probability to go to location no.1 for instance is $P(\textit{location 1}) = \frac{D_1}{\sum_r D_r}$, where $D_r = \sum_{i \in \mathcal{A}_r} p(u_i = 1 \mid \mathbf{h})$.

By default, recall was mediated by the hippocampus. To test whether the memories were consolidated in the neocortex, we disabled the hippocampus during recall and let the RBM perform 5 Gibbs sampling steps to reconstruct the activity of the location units (see Cortical recall in Section 2.2.1).

Network training

As in the previous model simulations (Section 3.2), the training was divided into epochs that simulated day and night cycles. However, we made two important modifications for the training protocol. The first modification was to restrict the number of presentations of each association during the experience phase, when the model receives input from the environment. The idea was to reflect that sensory experience is limited, as opposed to sleep replay which we suppose can have many reactivations. The second modification was to allow cortical plasticity and hippocampal storage only for rewarded events. This learning rule seemed more efficient and in line with biology compared to the learning implemented in the first model (Section 3.2).

An overview of the training epoch is given Fig. 3.14. Each epoch began with an **experience phase** during which the sensory layer of the neocortex received inputs from the environment. In contrast with the training inspired from Káli & Dayan (2004) which we used in the previous simulations, only the rewarded flavour-place associations were presented to the network during the experience phase[7], and each rewarded association was presented only three times[8]. The order of presentation was random at each epoch.

[7]Note that even if the network encountered the non-rewarded associations, to mimic the animal exploring other sand-wells, the result would be exactly the same because cortical plasticity and hippocampal learning were restricted to rewarded events.

[8]The number of presentations of each flavour-place association was chosen according to the experimental data. Once the rats found the correct sand-well, they collected the food pellet and returned to the starting box to eat. Since three food pellets were hidden in the sand-well, they came back and forth three times (without revisiting the other sand-wells).

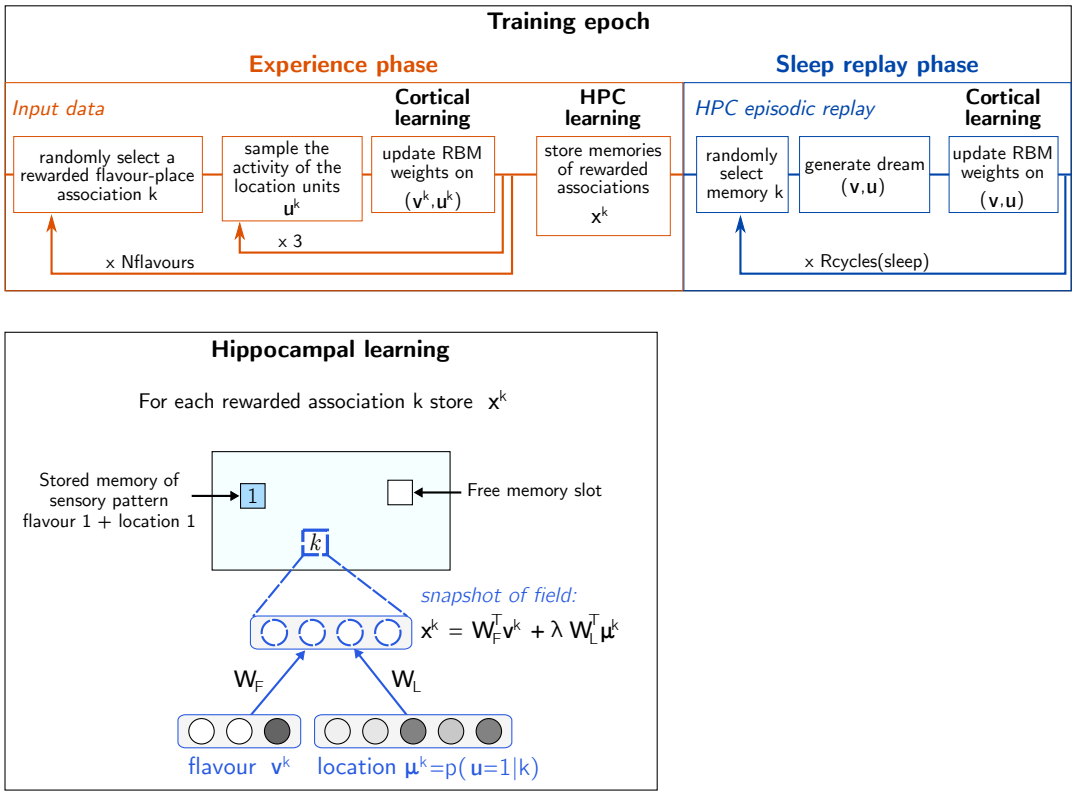


Figure 3.14: (Top panel) Training epoch. During the experience phase, the model sees the rewarded flavour-place associations. For each association, we sample the activation of the location units three times to obtain multiple binary input patterns (see Fig. 3.2b). The hippocampus stores the field $\mathbf{x}^k$ of the associative layer for each association presented (see lower panel). During the sleep replay phase, memories are randomly selected in the hippocampus. When a memory is selected, the hippocampus clamps the activity of the associative layer, leading to the replay of a 'dream' pattern in the sensory cortices via the top-down cortical connections $\mathbf{W}_F$ and $\mathbf{W}_L$. See Section 2.2 for details about cortical learning and sleep replay. (Lower panel) Hippocampal learning during the experience phase.

Training:	Original	New
reward locations	6	6
memories in HPC	6	2
visible units	231	233
hidden units	200	200
sleep cycles	600	600

weights initialisation		mean 0 std 0.1
learning rate	EXP	0.01
	SLEEP	0.01 (flavours) 0.0001 (locations)

Table 3.2: Model and training parameters. Notations: HPC = hippocampus; EXP = experience phase; SLEEP = sleep replay phase. The coloured numbers are parameters that will be modified during the training of new associations (see main text in the results sections). For the sleep cycles refer to the diagram of the training epoch Fig. 3.14.

Upon selecting a flavour-place association, we sampled the corresponding raw pattern representing the activation probabilities of the location units (Fig. 3.2b). We sampled three times so as to obtain three different binary input patterns per location. The cortical weights were updated for each rewarded stimulus presented (using CD1 learning, see method in Section 2.2.2).

After the cortex had updated its weights, the hippocampus stored the field vector $\mathbf{x}^k$ of the hidden units (Eq. 3.5) corresponding to each rewarded flavour-place association $\left(\mathbf{v}^k, \mathbf{u}^k\right)$ (Fig. 3.14, lower panel).

Next, the model started the **sleep replay phase**. The hippocampus randomly selected a flavour memory k, and clamped the hidden layer. Given the activation $\mathbf{h} \sim p\left(\mathbf{h} \mid \mathbf{x}^k\right)$ in the associative cortex, the network generated a "dream" pattern $(\mathbf{v}, \mathbf{u})$ in the sensory layer. This procedure was the same as the step 2 during recall, shown Fig. 3.13, except that it reconstructed the activity of all the visible units, the flavour units using the connection matrix $\mathbf{W}_F$, and the location units using the connection matrix $\mathbf{W}_L$. The cortical weights were trained on this "dream" pattern (binary), using the same protocol as during the experience phase (CD1 learning, see method in Section 2.2.2). The hippocampus cycled 600 times through all the stored memories in this way. Note that Girardeau et al. (2014) detected some 1000 ripple events per hour (which they then perturbed and observed impaired consolidation). Their detection method likely included false positives or task-irrelevant, and missed events. In any case our number of consolidation cycles seems reasonable.

We set a lower learning rate for the location units compared to the experience phase (Table 3.2). We justify how we set the cortical learning rates at the beginning of the results in Section 3.3.2.

3.3.2 Results: Sleep replay and consolidation in the neocortex

In the previous model in Section 3.2 sleep replay was needed because without it the correct associations could not be consolidated (Fig. 3.7b). By contrast, in this model consolidation was possible even without replay since we only presented the rewarded associations. Hence we first wished to study the potential role of sleep replay.

Trade-off between training time and performance without sleep replay

We initially trained the model on the six original associations, similar to the experimental paradigm of Tse et al. (2007), but without sleep replay. The cortical weights were updated during the 'experience phase' (Fig. 3.14) for 5000 epochs, with various cortical learning rates η between 0.001 and 0.1. For each learning rate η we ran 10 simulations with different random initialisation of the cortical weights (weights sampled from a Normal distribution, see parameters in Table 3.2). We tested the recall performance of the network at each epoch, and we assessed memory consolidation by blocking the hippocampus during recall.

To evaluate the training speed, we looked at the number of epochs the network required to reach an average recall performance of 0.5 (i.e. $\langle P(\text{Location } k)\rangle_{k=1..6} = 0.5$, where k is the rewarded location; see Fig. 3.13). As expected, we found that training was faster with higher learning rates (top left panel in Fig. 3.15). Next, we examined the patterns consolidated in the neocortex at the end of training by probing memories without hippocampus. We noticed that, for higher learning rates, the patterns reconstructed during recall did not reflect the average activity representing the locations, but rather the sparse sensory experience of the samples (see lowest panel Fig. 3.15; examples of samples are shown Fig. 3.2b).

To investigate this, we computed the 'reconstruction quality' of the activity of the location units $\hat{\mathbf{u}}$ in comparison with the original input $\mathbf{u}$ using the correlation measure:

$$rq(\hat{\mathbf{u}}, \mathbf{u}) = \frac{\langle (\hat{\mathbf{u}} - \langle \hat{\mathbf{u}} \rangle)^\intercal (\mathbf{u} - \langle \mathbf{u} \rangle) \rangle}{s(\hat{\mathbf{u}})\, s(\mathbf{u})} \tag{3.7}$$

where $\langle \cdot \rangle$ is the average over the location units, and s is the standard deviation. This measure is bounded between $[-1, 1]$, and $rq = 1$ means the observation is a perfect match with the input pattern. We monitored the reconstruction quality of the location units' activity when the network reached the target performance (top right panel in Fig. 3.15). For recall without hippocampus (pure cortical recall), we found that with higher cortical learning rate the quality was poor (cyan line). Recall quality with hip-

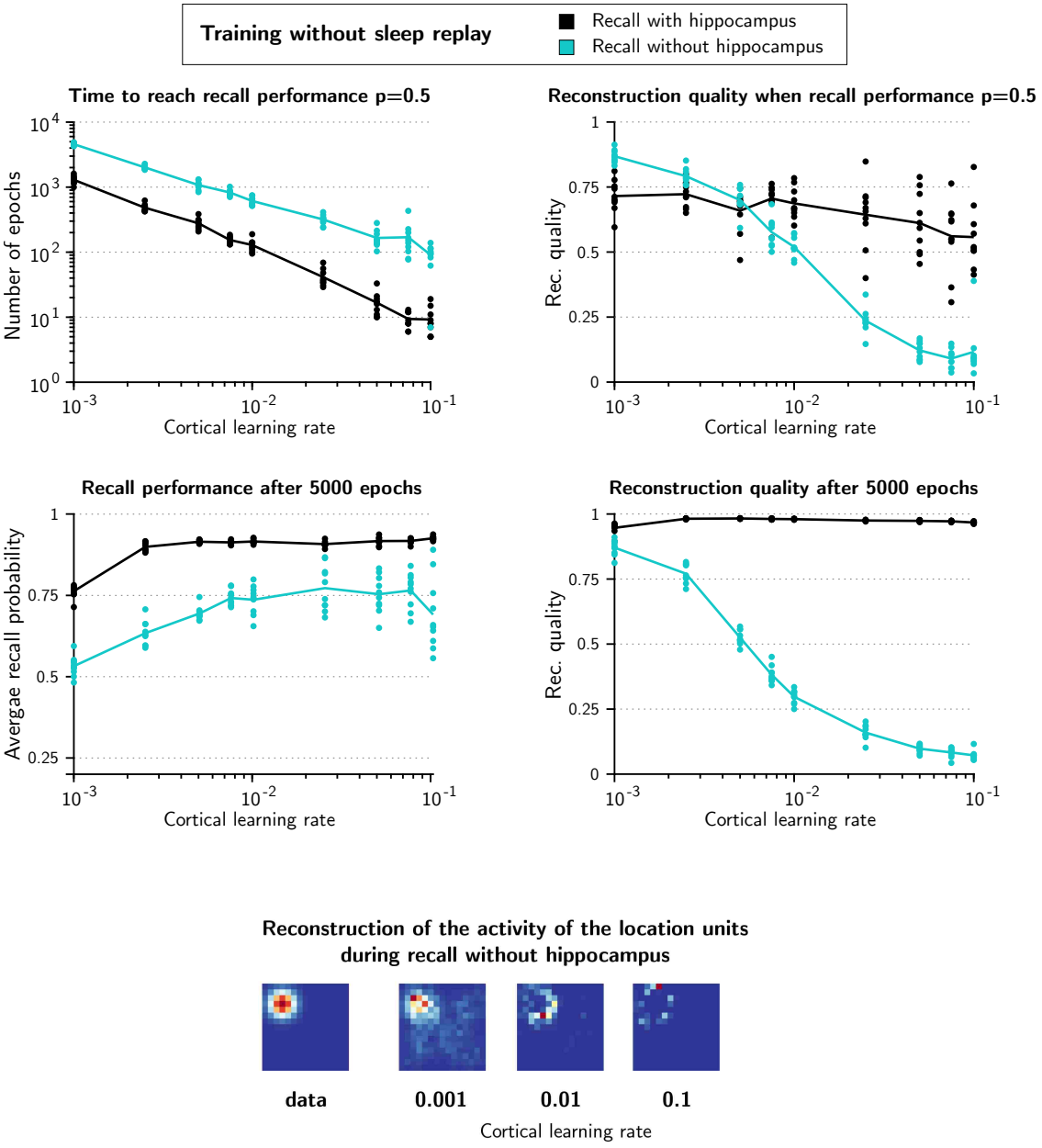


Figure 3.15: Training on the six original associations, without sleep replay, for various cortical learning rates. The reconstruction quality is a quantitative measure of the similarity between the true average activity patterns representing the locations and the activity patterns reconstructed in the sensory layer during recall. The lowest panel shows an example of the true pattern for location no.1 (data) and the reconstructed activities observed during recall at the end of training.

pocampus seemed more robust (black line), but we found more variability in the results for higher learning rates. At the end of training, quality of recall without hippocampus was even lower for high η (bottom right panel in Fig. 3.15, cyan line), which suggests that quality deteriorated over training. With the recall probability we could not see this effect as the performance measure was tricked by the activation of a few units.

The poor quality of the reconstructions indicated that the neocortex kept on overwriting past experiences instead of incorporating them over time. It memorised the specific activation patterns. We believed this aspect did not fit with cortical learning in the brain, and thus we concluded that the cortical learning should be slow.

Remarks. In the next parts, for all the simulations with similar data set we made sure the reconstruction quality was good during recall. In addition, it should be noted that when the location input patterns were not stochastically sampled, but always identical (Fig. 3.2b) we did not observe the effect described above. Instead, recall both with and without hippocampus were accurate over training, and training was faster with higher learning rates. We come back to the role of sleep replay for simple data versus more complex inputs in Section 5.2.1.

Sleep replay: a virtual training to promote consolidation

Sensory-driven experience, however, is limited. Hence, with small cortical learning rates the training would be extremely slow. We propose that replay during sleep could compensate by providing a 'virtual training'. However, to allow this 'virtual training' the network has to be able to replay sensory patterns that resemble the 'real' sensory experience. Thus, the learning rate should be large enough to allow the hippocampus to store memory patterns that can be reconstructed during subsequent sleep. This was possible because we saw that memory retrieval by the hippocampus was robust even for larger learning rates (right panel in Fig. 3.15, black line). Therefore, in theory, the hippocampus should be able to drive the replay of sensory patterns via the top-down neocortical connections.

For the remaining simulations, we found a trade-off (arbitrary) by setting the cortical learning rate to 0.01 during the experience phase, and to 0.0001 during the sleep replay phase.

We examined how sleep replay affected the learning and consolidation. We ran 5 simulations with different random initialisation of the cortical weights to check the variability of the results. Sleep replay improved recall performance (Fig. 3.16). The

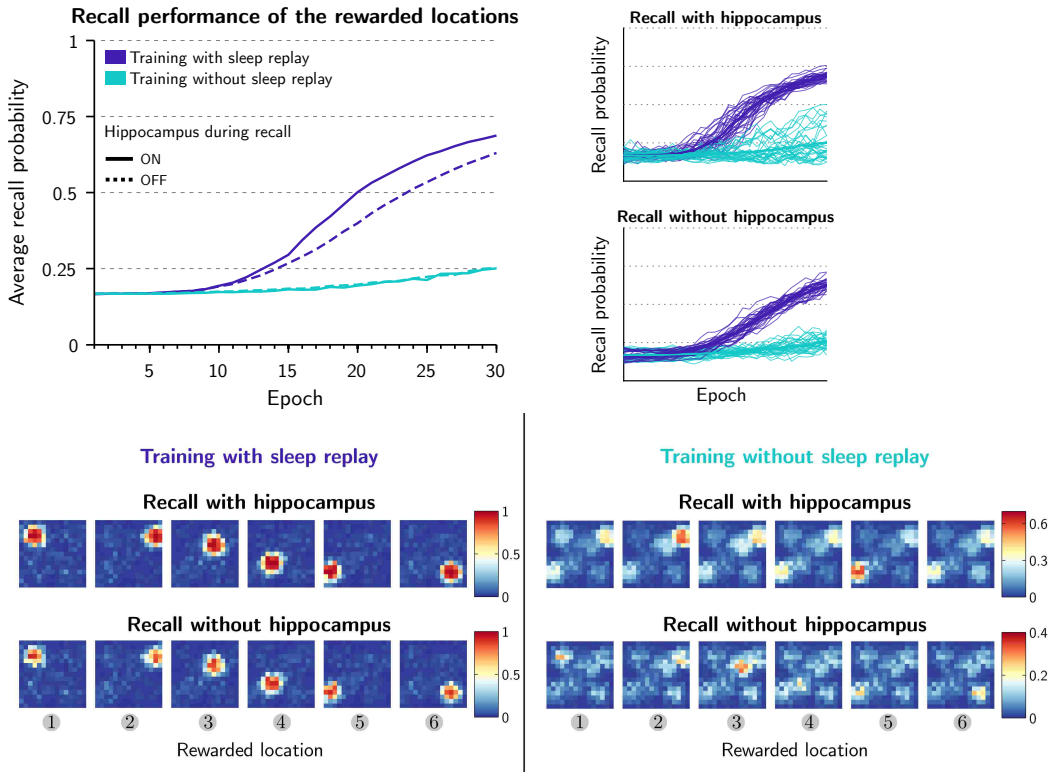


Figure 3.16: Training of the six original flavour-place associations, with (dark blue) or without (cyan) sleep replay. (Upper panel) Average recall performance over training epochs. Performance is the probability to recall, given a flavour, the correct rewarded location. Solid lines: recall with hippocampus. Dashed lines: hippocampus disabled at recall (only). The two insets show the individual recalls of the six associations for the five runs (thin lines), and the average for each run (thick lines). (Lower panel) Single example run of reconstruction of the activity of the locations units during recall at the end of training. The six insets in each row show the probability of activation of the 225 reward location units when the network is cued with one of the six flavours (numbers at the bottom). The probabilities are averaged over 100 trials.

training time required to reach the average performance of 0.5 was less than without sleep replay with the same learning rate $\eta = 0.01$ (about 20 epochs versus 100 epochs; see for comparison the top left panel in Fig. 3.15, black line). The associations were consolidated gradually over time in the neocortex, tracking the performance with hippocampus with smaller delay than without sleep replay (top left panel in Fig. 3.15).

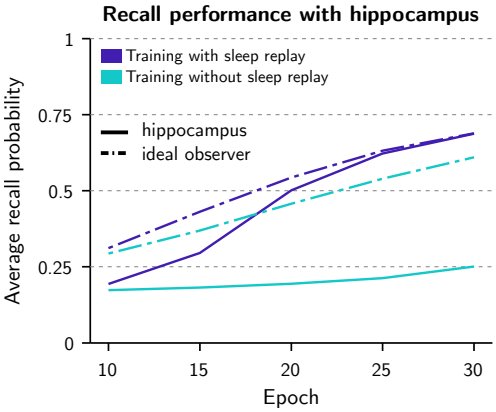
At the end of training, the network successfully recalled the six locations when cued with the corresponding flavours (see reconstructions of the locations in Fig. 3.16). Without sleep replay the reward locations were extracted, i.e. they were correlated with the input patterns, but the associations were weak and hence performance was low at the end of training. Note that without sleep replay, semantic recall appeared to be more accurate than hippocampal recall: the correct locations had the highest probability, whereas when recall was mediated by the hippocampus recall was biased towards locations no.2 and no.5 (Fig. 3.16, bottom right panel).

In summary, sleep acted as a virtual training for the neocortex, an opportunity to strengthen the connections that started to develop during the experience phase. Sleep replay thereby promoted the consolidation and stabilisation of the six associations in the neocortex.

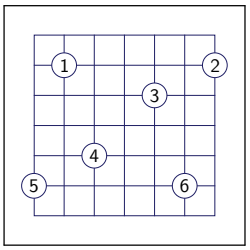
Recall mediated by the hippocampus is limited while memories are gradually consolidated in the neocortex

In the simulations with the previous model in Section 3.2.2, hippocampal performance increased gradually for two reasons: 1) it was regulated by the reward probabilities which were incremented over training epochs, and 2) the recall probability inferred from the sensory cortex activity was lower because of the top-down cortical reconstruction (Fig. 3.7a, solid blue line compared to semi-dashed black line). In the current model however, the hippocampus no longer relied on reward probabilities for recall, as it only stored rewarded memories. We then investigated two factors that could limit recall with hippocampus: (i) the problem of cue identification in the hippocampus (hidden layer activation, **step 1** in Fig. 3.13), and (ii) the quality of reconstruction in the sensory layer (from the hidden layer activation clamped by the hippocampus, **step 2** in Fig. 3.13). Both issues were due to the neocortical weights, the first one due to bottom-up stream, and the second one due to top-down stream.

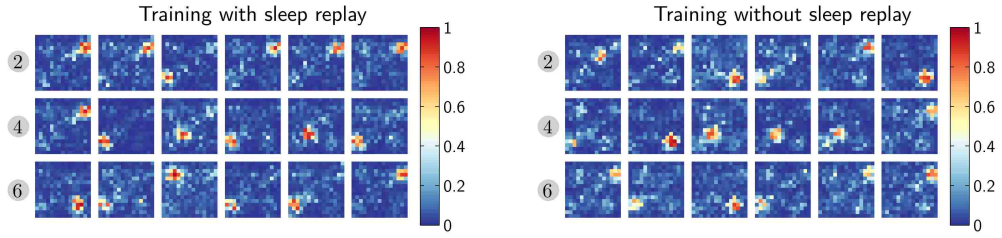
To separate the two contributions, we considered an 'ideal observer' which always correctly identified the cues; hence, recall consisted only of **step 2** in Fig. 3.13. As a



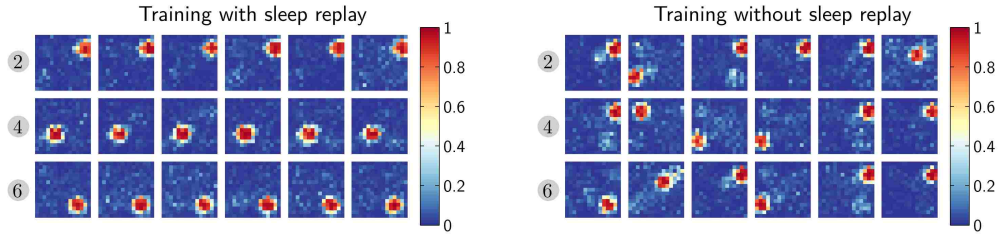
(a) HPC recall versus ideal observer.



(b) Reward locations.



(c) Test at epoch 15 (mid-training).



(d) Test at epoch 30 (end training).

Figure 3.17: Recall mediated by the hippocampus is limited by the gradual consolidation in the neocortex. (a) Recall performance (test every 5 epochs). The solid lines are the same results shown Fig. 3.16. The semi-dashed lines show the performance obtained when we bypassed memory recall in the hippocampus (**Step 1** in Fig. 3.13), and instead supposed that the hippocampus had perfect cue recognition ('ideal observer'). Hence, the performance measured is, in this case, only dependent on the top-down reconstruction in the neocortex (**Step 2** in Fig. 3.13). (b) Schematic of the sand-wells spatial arrangement for the original training, for reference for the next figures. (c&d) Reconstruction of the location units activities during recall of the associations no.2, no.4 and no.6. The insets in one row correspond to different trials when cued by the flavour indicated. The reconstructions shown were obtained with one of the five simulation runs.

result, if performance were low it would be due to the top-down reconstruction of the activities in the neocortex. We found that performance with the 'ideal observer' was low during the first half of training (Fig. 3.17, semi-dashed lines), which implies that the top-down cortical reconstruction was indeed a limiting factor in the performance when recall was mediated by the hippocampus. Over time, as the quality of the reconstructions improved, performance with the 'ideal observer' also improved. In addition, we noticed that the 'ideal observer' had a similar behaviour when the network was trained without sleep replay (Fig. 3.17, light blue semi-dashed line).

The actual recall performance was initially lower than the performance predicted by the 'ideal observer', but did catch up later on (Fig. 3.17a, dark blue solid line). This suggests that the cue identification in the hippocampus (**step 1** in Fig. 3.13) was also limiting performance. To illustrate this we looked at the reconstruction of the activity of the location units during recall of associations no.2, no.4 and no.6 (Fig. 3.17b&c, left panels): mid-training (epoch 15), the reconstructions were noisy as predicted, and we can see that the hippocampus wrongly identified the cues (mainly associations no.4 and no.6), but at the end of training recall was accurate.

When the network was trained without sleep replay, performance was always lower than the performance of the 'ideal observer' (Fig. 3.17a, light blue solid line). This suggests that the hippocampus never correctly identified the cues. When we tested the network mid-training (Fig. 3.17b, right panel), the reconstructions looked similar to those with sleep replay, but the identifications appeared more random. When tested at the end of training, (Fig. 3.17c, right panel), the network reinstated good quality sensory patterns corresponding to known locations, but it reinstated the wrong ones.

We further investigated the cue identification in the hippocampus during memory recall (**step 1** in Fig. 3.13). We found that the hippocampus preferentially picked the memories of associations no.2 and no.5 (Fig. 3.18). The only difference between these two associations and the other ones is that locations no.2 and no.5 are located near the edge, and thus have less noisy input patterns representations (less units with low probability of being active); however we do not know how this difference in sensory activation explains the bias observed during recall.

To conclude, recall mediated by the hippocampus was influenced by the gradual consolidation of memories in the neocortex. When the network was trained with sleep replay, recall performance was initially limited by both bottom-up and top-down streams. However, after a while the hippocampus correctly identified the cue and per-

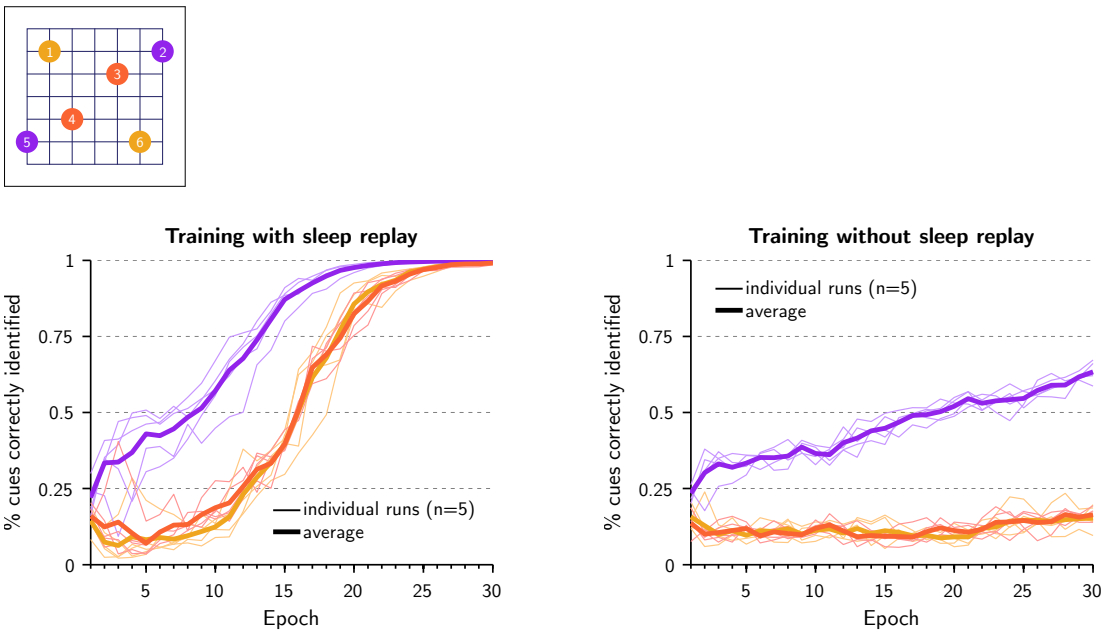


Figure 3.18: Identification of the memory cue in the hippocampus during recall, with or without sleep replay.

formance was only limited by the cortical reconstruction of the sensory patterns. On the other hand, when the network was trained without sleep replay the reconstructions of sensory patterns did increase in quality but the hippocampus constantly failed to recognise the memory cue coming from the bottom-up stream, and hence reinstated the wrong memories.

3.3.3 Results: Impact of the cortical weights initialisation on the acquisition of new associations

At the end of the training on the six original associations, the network had learned a semantic schema: six flavours, six potential reward locations, and their associations. In line with the experimental study of Tse et al. (2007), the network was then concurrently trained on two new flavour-place associations. The aim was to investigate the impact of the prior semantic schema on the acquisition and consolidation of the new associations.

The new associations involved two new flavours, and we added one new unit per new flavour in the sensory layer of the neocortex (RBM). The weights connecting the new flavour units to the hidden units were set to the default values, i.e. they were sampled from a normal distribution with mean 0 and standard deviation 0.1 (see Table 3.2). The two new reward locations were in the neighbourhood of two original locations (no.7 replaced no.1, and no.8 replaced no.6, see inset at the top in Fig. 3.19). We

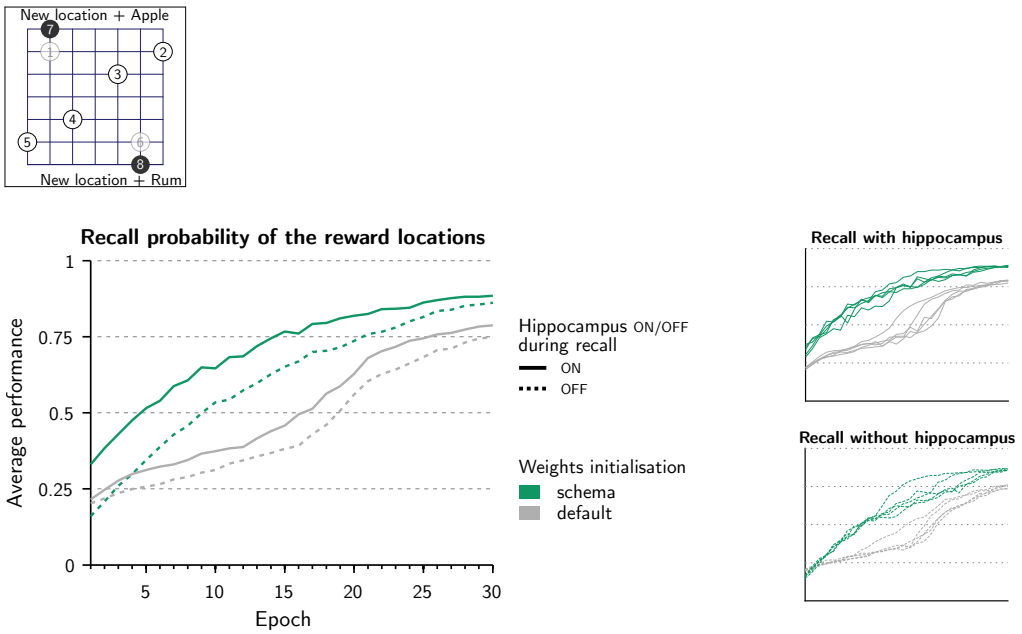


Figure 3.19: Acquisition of two new associations (new flavours + new locations) in a naive network (grey lines, weights with default initialisation) or in a network with a prior schema (green lines, trained weights). The two insets show the average performance of each simulation (5 runs per network type), while the main plot show the average over all runs. The axes of the insets are the same than the main plot. The solid lines represent recall performance with hippocampus, while the dashed lines show the performance with the hippocampus disabled during recall to evaluate memory consolidation.

applied the same training protocol described in Fig. 3.14, but only the new associations were presented to the network. The training parameters were the same as during the original training. The hippocampus was cleared at the start of training, and hence could only recall new memories.

To assess the impact of prior knowledge on the acquisition of the two new associations, we compared the performance of a **network with semantic schema**, i.e. which was previously trained on the six original associations, with the performance of a **naive network**, i.e. which started with the default weight initialisation for all units (Table 3.2). We ran five simulations for each network type. To evaluate recall performance we derived the probabilities to go to the new locations using the same procedure shown Fig. 3.13 (**step 2**), but replacing the locations no.1 and no.6 with the new locations.

The network with prior schema learned and consolidated the new flavour-place associations faster than the naive network (Fig. 3.19). With a prior schema, the hippocampus learned the new locations more rapidly because the network already knew

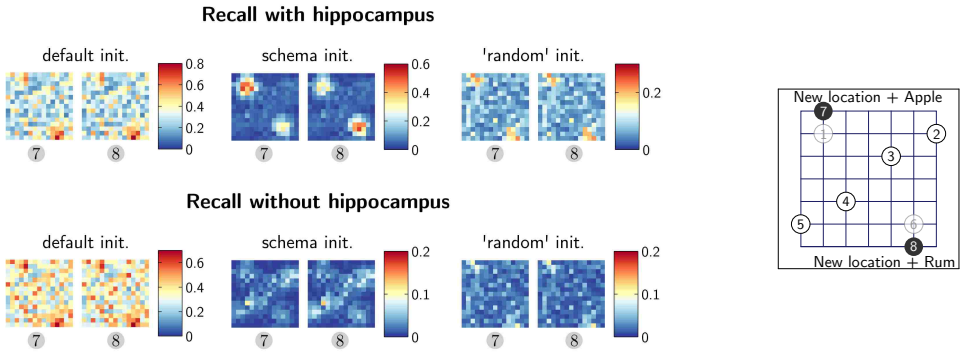


Figure 3.20: Recall of the two new flavour-place associations after one epoch of training with (1) a naive network (left column, default initialisation), (2) a network with prior schema (middle column, pre-trained with the six original associations), and (3) a network with random weight initialisation for the location units (right column, parameters extracted from the semantic schema, see Fig. 3.21). The images show the reconstruction of the activity of the location units during recall of the new associations no.7 and no.8. The reconstructions displayed were obtained with one of the five simulation runs. Note that for all the flavour units had default initialisation values.

the arena; this is illustrated by the reconstruction of the location units activities, shown in Fig. 3.20 (left and middle columns): after one epoch of training, the hippocampus could already reinstate the new locations, in contrast to the naive network for which reconstructions were noisy. Hence, while the naive network required more training epochs to adjust the cortical weights and reduce the background noise, the network with prior schema could be trained faster.

Next, we examined whether sleep replay was necessary for the rapid acquisition of new associations in a network with prior schema. We turned-off sleep replay during the new training, and we observed that both learning and consolidation were slower compared to the training with sleep replay (Fig. 3.22a, no markers). In the long run, the performance of the network with prior schema, but without sleep replay, was worse than the performance of the naive network (grey lines). However, when both networks were trained without sleep replay, performance with prior schema was better than performance of a naive network (no shown).

We suspected that learning without sleep replay might have been slow because the weights of the new flavour units had low initial values compared to the weights of the flavour units that represented the six original flavours. To check this, we ran new simulations where we initialised the weights of the two new flavour units using the statistics of the semantic schema. From now on we call this type of initialisation

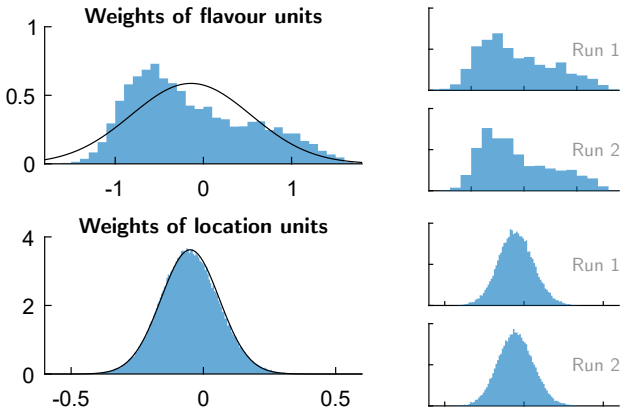
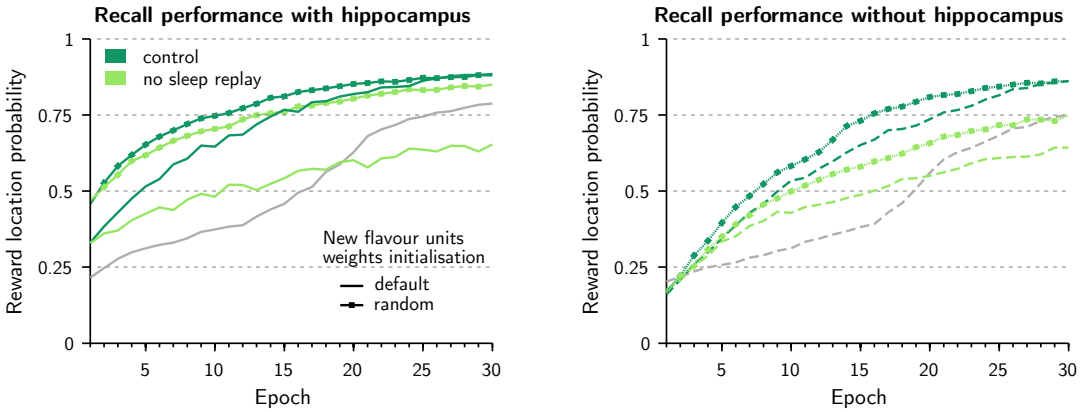


Figure 3.21: Statistics of the cortical semantic schema. The plots show the distributions of the cortical weights after training on the six original flavour-place associations. The main plots show the distributions for all the 5 simulation runs, and the small panels show the distribution for two simulation runs. Large panels show the distribution for all five simulations. Statistics of the weights of the flavour units: mean = [-0.14, -0.13, -0.13, -0.14, -0.14] ; std = [0.69, 0.68, 0.67, 0.68, 0.67]. Statistics of the weights of the location units: mean = [-0.05, -0.05, -0.05, -0.05, -0.05] ; std = [0.11, 0.11, 0.11, 0.11, 0.11]. The black lines represent the Normal distributions we used to sample the weights for the 'random' training condition (see main text) (Flavours: mean = -0.14 ; std = 0.68. Locations: mean = -0.05; std = 0.11).

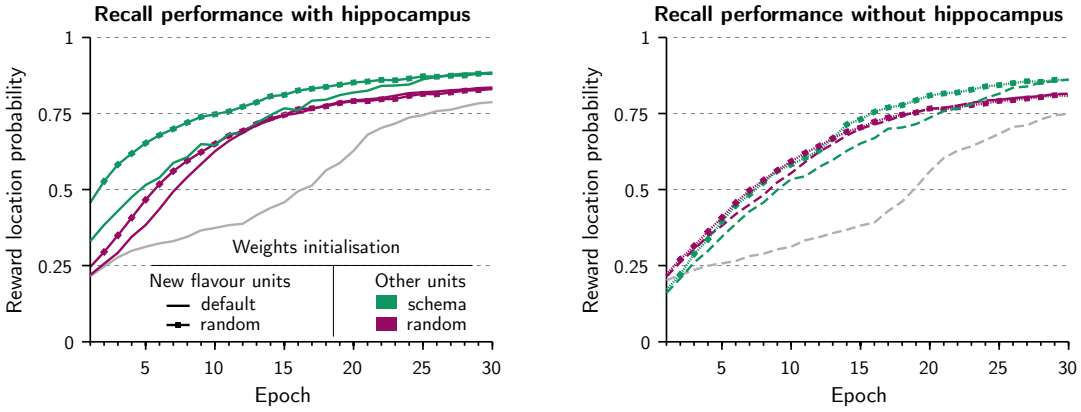
'random'. We extracted the mean and standard deviation of the flavour weights that were obtained after training the network on the original associations (Fig. 3.21, top row). We then initialised the weights from a broad normal distribution with mean $\bar{w} = -0.14$, and standard deviation std=0.68 (Fig. 3.21, black line).

With the 'random' initialisation, learning speed increased and consolidation speed slightly increased in networks trained with sleep replay (Fig. 3.22a, with markers). More importantly, performance of recall mediated by the hippocampus during training without sleep replay was almost identical to that with sleep replay (Fig. 3.22a left panel, light green line with markers). Thus for this particular task, a network with prior schema does not require sleep replay to learn the new associations when the weights of the new units are properly initialised.

Consolidation speed without sleep replay was also increased with the 'random' initialisation of the new flavour units, albeit in the long run, performance was better with sleep replay (Fig. 3.22a right panel, with markers). This suggests that consolidation still benefits from sleep replay.



(a) New training with replay versus without replay.



(b) Network with learned semantic schema versus 'random' initialisation.

Figure 3.22: Impact of the weight initialisation on the training of two new flavour-place associations (see Table 3.3). Notes: (1) The grey line shows the reference training of the naive network (default initialisation), and hence is the same in each plot. (2) The lines with markers represent performance of networks with the 'random' initialisation of the new flavour units (using the parameters extracted from the semantic schema). (3) The left panels show recall mediated by the hippocampus, and the right panels show the consolidated memories (hippocampus disabled during recall only). (a) Performance with or without sleep replay in a network with prior schema. (b) Performance with sleep replay, in a network with prior schema (green lines), or with 'random' weight initialisation for the location units and for the six original flavour units (purple lines, using the parameters extracted from the semantic schema).

Network	***Weight initialisation scheme***		
	original flavour & location units	new flavour units	
Naive	default	default	
Schema	end of original training	default	'random'
Random	'random'	default	'random'

Table 3.3: Networks trained on the new associations. Default refers to the initialisation we used for the original training (see Table 3.2). 'Random' initialisation uses the statistics of the semantic schema (see Fig. 3.21).

Since a 'random' initialisation of the new flavour units facilitated learning and consolidation, we considered training a naive network with similar 'random' initialisation for all the weights instead of the default initialisation. We extracted the parameters of the location weight distribution at the end of the original training (Fig. 3.21, lower panel). We then initialised the weights of the location units from a normal distribution with mean $\bar{w} = -0.05$, and standard deviation std=0.11 (Fig. 3.21, black line). Similarly, we initialised the weights of the six original flavour units from the distribution we mentioned earlier, but for now the units corresponding to the two new flavour units were still initialised from the default distribution. In the following paragraphs we refer to this "improved" naive network as **random network**. Table 3.3 summarises the different network initialisations implemented.

At the start of training, performance of the random network was lower than performance of the network with schema (Fig. 3.22b left panel, no marker). The reason was the same as for the naive network: it had no knowledge about the reward locations in the arena, and hence the reconstructions of the location units activities were noisy, as opposed to the network with the prior semantic schema (Fig. 3.20, middle and right columns). Nevertheless, the random network learned the two new associations much faster than the naive network, and consolidation was as fast as in the network with prior schema (Fig. 3.22b right panel, no marker).

Next, we repeated the simulations but we initialised the weights of the new flavour units from the normal distribution with the schema statistics (Fig. 3.21, upper panel). Learning speed and recall performance at the beginning of training did not improve as much as it did in the network with prior schema (Fig. 3.22b left panel, with markers). On the other hand, consolidation in the random network was again very similar to the one observed in the network with prior schema (Fig. 3.22b right panel, with markers).

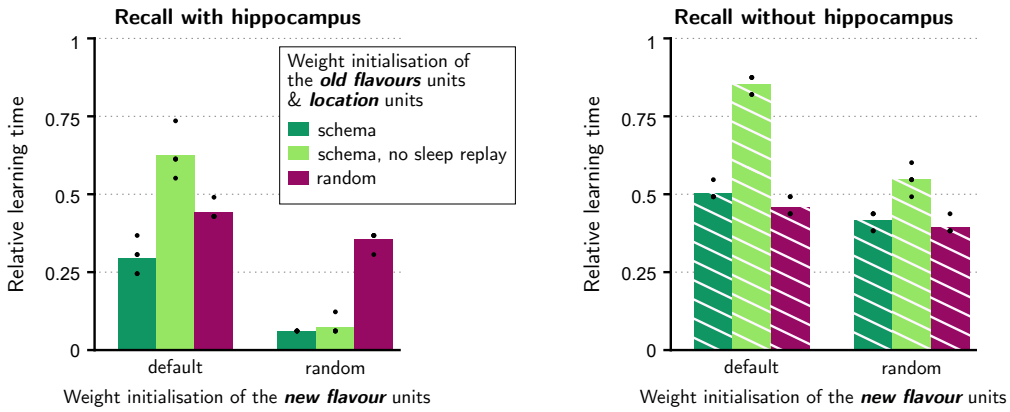
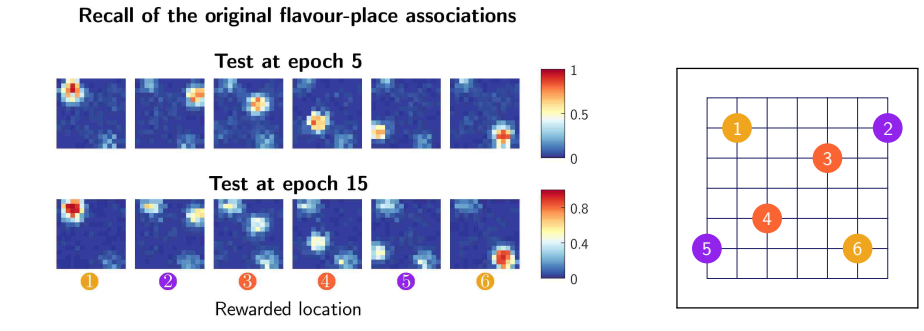


Figure 3.23: Relative training time for various weight initialisation schemes compared to a naive network (relative time is t/t_{ref}, where t_{ref} is the time for training a naive network; time was measured when performance reached p=0.5). The dots represent the results of individual runs. The left panel shows speedup in performance when recall is mediated by the hippocampus, and the right panel shows the consolidation speedup.

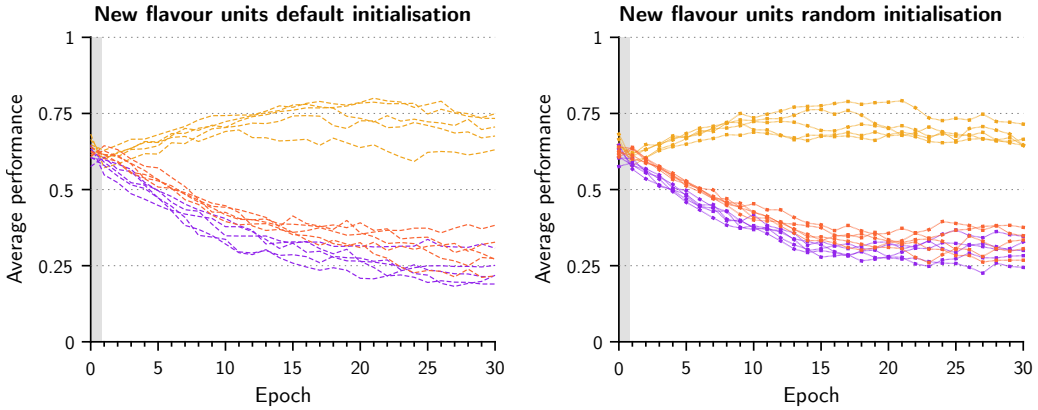
In summary, weight initialisation influenced how rapidly the network learned and consolidated the two new associations. The network with prior semantic schema assimilated the new memories more rapidly than a naive network (Fig. 3.23, in dark green), which is consistent with the data of Tse et al. (2007). In addition, we observed a speedup even if memories were not replayed during sleep (in light green); nonetheless, the speedup was greater with sleep replay, which is consistent with the fact that rats required 24-48hrs for the successful consolidation of the new associations. However, we also obtained rapid learning in a random network initialised from distributions with statistics similar to those of the semantic schema (in purple). In particular, consolidation was as fast as in the network with prior schema (Fig. 3.23, right panel).

Furthermore, sampling the initial weights of the new flavour units from a broader distribution promoted hippocampal memory acquisition, but only in a network with prior schema (Fig. 3.23, left panel). Importantly, the difference in learning with versus without sleep replay almost vanished. On the other hand, the consolidation speed with sleep replay was not much affected by the initialisation of the new units (right panel). Consolidation without sleep replay was faster in this case, yet performance was still overall lower than with sleep replay (Fig. 3.22b right panel, with markers).

All together, these results imply that the pre-training on the original associations facilitated the rapid acquisition of two new associations in the hippocampus, but this specific semantic schema did not benefit consolidation compared to a random initialisation with appropriate statistics.



(a)



(b) Cortical recall of the original associations (hippocampus disabled at recall).

Figure 3.24: Interference with the original associations during the new training.

Retention of the original associations

In order to assess the interference in the neocortex caused by the new training, we disabled the hippocampus at recall and we probed the memories of the on the original flavour-place associations. Note that the old memories were never replayed during sleep for the new training. We tested memory retention (i) for the default initialisation and (ii) for the 'random' initialisation of the two new flavour units (Fig. 3.24). Recall performance of the original associations with locations no.1 and no.6 increased because these locations were in the neighbourhood of the new locations. On the other hand, memories of the other four original associations (no.2 to no.5) were gradually overwritten with the new reward locations.

For both conditions, performance did not drop suddenly upon learning the new associations, in contrast with the interference observed with similar conditions (focused learning, old memories not reactivated) in the first model that used one unit per location coding (Fig. 3.10a). Not surprisingly, the distributed representation of reward location, was more robust to overwriting.

3.3.4 Results: impact of the consistency of the new associations with prior knowledge

In the Complementary Learning Systems Theory (CLST), McClelland (2013) highlighted that the consistency of new information with prior knowledge had a significant impact on the learning speed and on the extent of interference (see Section 1.2.3).

When we implemented the 'living things' data set with our RBM-hippocampus framework, we did not find that learning about penguin was an issue, but we did find similar results to those of McClelland (2013) when learning more alien items which had very unlikely combinations of features. However, unlike McClelland (2013), the consolidation of the 'penguin' was actually easier than the consolidation of the 'cardinal'. We assumed the novelty of the 'penguin' facilitated consolidation, while the 'cardinal' was too similar and hence consolidation was stuck (our guess was that there was barely a gradient during sleep replay).

We decided to further investigate this effect with the current simulations of the task of Tse et al. (2007). We repeated the training of two new associations 1) in a naive network (default weight initialisation), 2) in a network with prior schema with default initialisation of the new flavour units, or 3) in a network with prior schema with 'random' initialisation of the new flavour units, using the statistics shown Fig. 3.21. In the current setting, it was difficult to characterise what defined a consistent new association, hence we measured instead the similarity of the new locations with two original locations. We considered various potential locations in a squared area surrounding the original sand-wells no.1 and no.6 (Fig. 3.25, 'Possible new locations'). The two new locations coordinates mirrored each other: for instance, if the new location no.7 was at the original location no.1, then the new location no.8 was at the original location no.6 (locations indicated ***A*** in Fig. 3.25). In the results, we highlight four particular cases: identical locations ***A***, which we just mentioned; similar locations ***B***, which correspond to the previous simulation in Section 3.3.3; unknown locations ***C***, which challenge the old reward mapping; and finally conflicting locations ***D***, which are close to two original locations and close to each other.

Acquisition of the new associations

We first looked at recall performance with hippocampus after one epoch of training. We noticed in Fig. 3.22, which corresponds to the location ***B*** here, that this performance was already high because the network confused the new locations with the

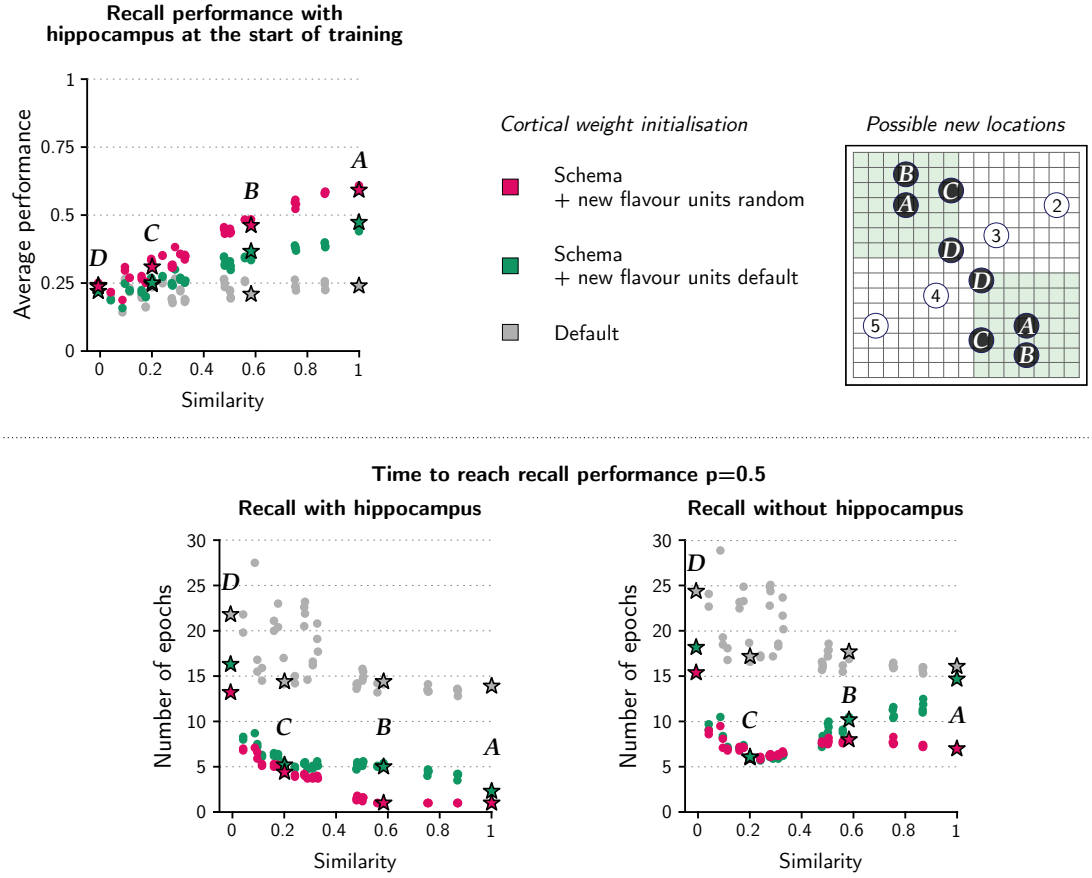


Figure 3.25: Acquisition of two new associations for different location similarity with prior knowledge. Similarity is measured by the reconstruction quality Eq. 3.7, between the activity vector of the location units corresponding to the new location no.7, and the activity vector of the location units corresponding to the original location no.1. For all plots, each dot represents the average result (10 runs) obtained for a given new location. There are 49 possible new locations that correspond to the small squares in the green areas surrounding the orginal locations no.1 and no.6 (top right panel). The results ***ABCD*** correspond to specific locations. The grey dot show the results obtained with a naive network (default weight initialisation). The green dots show the results obtained with a network with prior schema and default initialisation of the new flavour units, and the pink dots show the same but with the 'random' initialisation of the new flavour units. (Top left panel) Performance after one epoch of training the new associations, when recall is mediated by the hippocampus. (Lower panels) Training time to reach an average recall performance of 0.5, when recall is mediated by the hippocampus (left), or not (right).

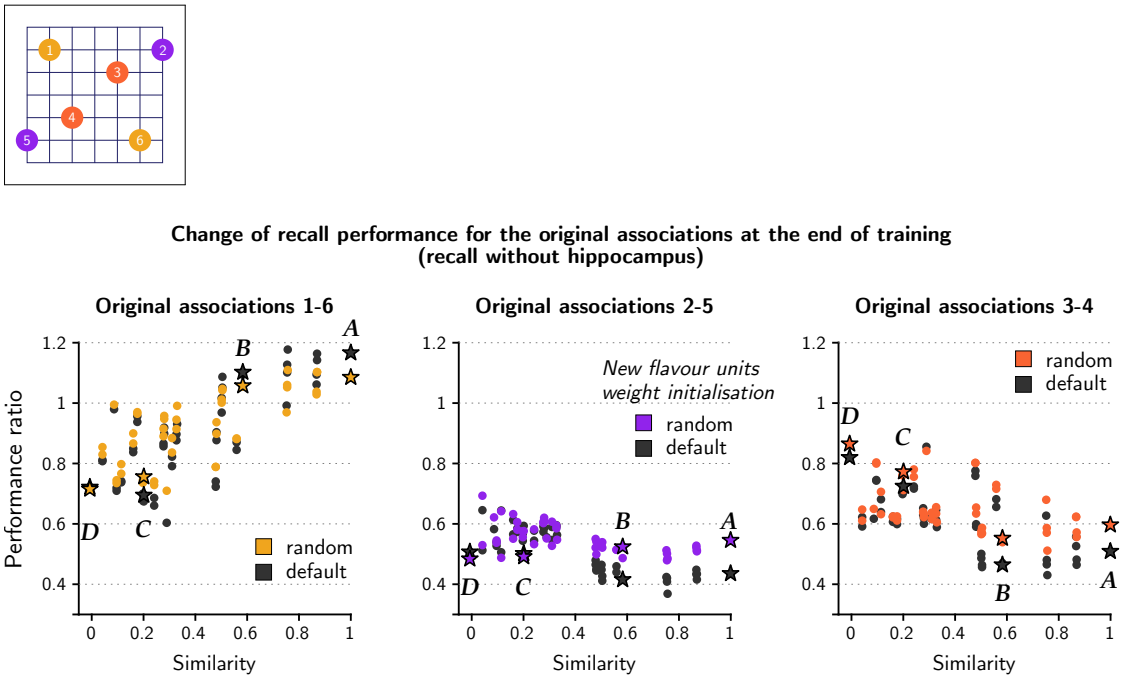


Figure 3.26: Interference with the six original flavour-place associations after 30 epochs of training two new associations for different location similarity with prior knowledge. For all plots, each dot represents the average result (10 runs) obtained for a given new location (49 in total) (see locations in Fig. 3.25). The dark dots show the results obtained with a network with prior schema and default initialisation of the new flavour units, and the dots in light colors show the same but with the 'random' initialisation of the new flavour units. We divided the results according to the placement of the original locations in the arena (see inset at the top). The performance ratio is $\frac{P_{new}}{P_{ref}}$, where P_{new} is the performance measured during recall without hippocampus at the end of the new training (for example, with location ***B*** see Fig. 3.24), and P_{ref} is the performance measured during recall without hippocampus at the end of the original training (Fig. 3.16, dark blue dashed line).

original ones: see for instance the reconstruction of the activities in Fig. 3.20, the network reactivates the original locations no.1 and no.6 rather than the new ones. This result was confirmed with the current simulations (top left panel in Fig. 3.25): when the new locations were closer to the original locations no.1 and no.6, recall performance improved in the network with prior schema (in green and pink), in contrast with the naive network (in grey). Performance further increased when we used the 'random' initialisation of the two new flavour units (in pink). However, when the networks with prior schema had to learn new locations further away from the original locations, they both lost their advantage. Nevertheless, we observed a speed up in training with prior schema for all new locations (bottom left panel in Fig. 3.25). The speedup was greater when the new locations were closer to the original ones. In addition, as anticipated, the conflicting locations ***D*** were the most difficult to train. Note that for the training with naive network (in grey), the training time was diverse for the locations of lower similarity: we found that the naive network learned with difficulty the rewards located at the edges, whereas the networks with prior schema did not have such problem.

The consolidation time had a more complex profile (bottom right panel in Fig. 3.25). With the new flavour default initialisation (in green), the consolidation took longer when the new locations were closer to the original ones: see for instance that locations ***A***, which perfectly overlap with the original locations, have similar consolidation speed than in the naive network. On the other hand, consolidation was faster for the unknown location ***C***. This effect was compensated with the 'random' weight initialisation of the new flavour units (pink dots).

Impact on the original associations

Next, we investigated the impact of the new training on the retention of the six original associations. We disabled the hippocampus during recall since 1) it only contained the new memories, and 2) we wished to evaluate if the neocortex had forgotten previously consolidated memories.

We checked recall performance at the end of the 30 epochs of training the new associations. Similar to the results in Fig. 3.24, we found that when the new locations were close to the original sites no.1 and no.6 (***AB***), performance of these two original associations improved or remained stable (Fig. 3.26, left panel). As the new associations were further away, performance decreased. For the two original locations no.3 and no.4 we saw a somewhat similar profile: as the two new locations were closer to them (***CD***), there was less interference (right panel). For the two original locations

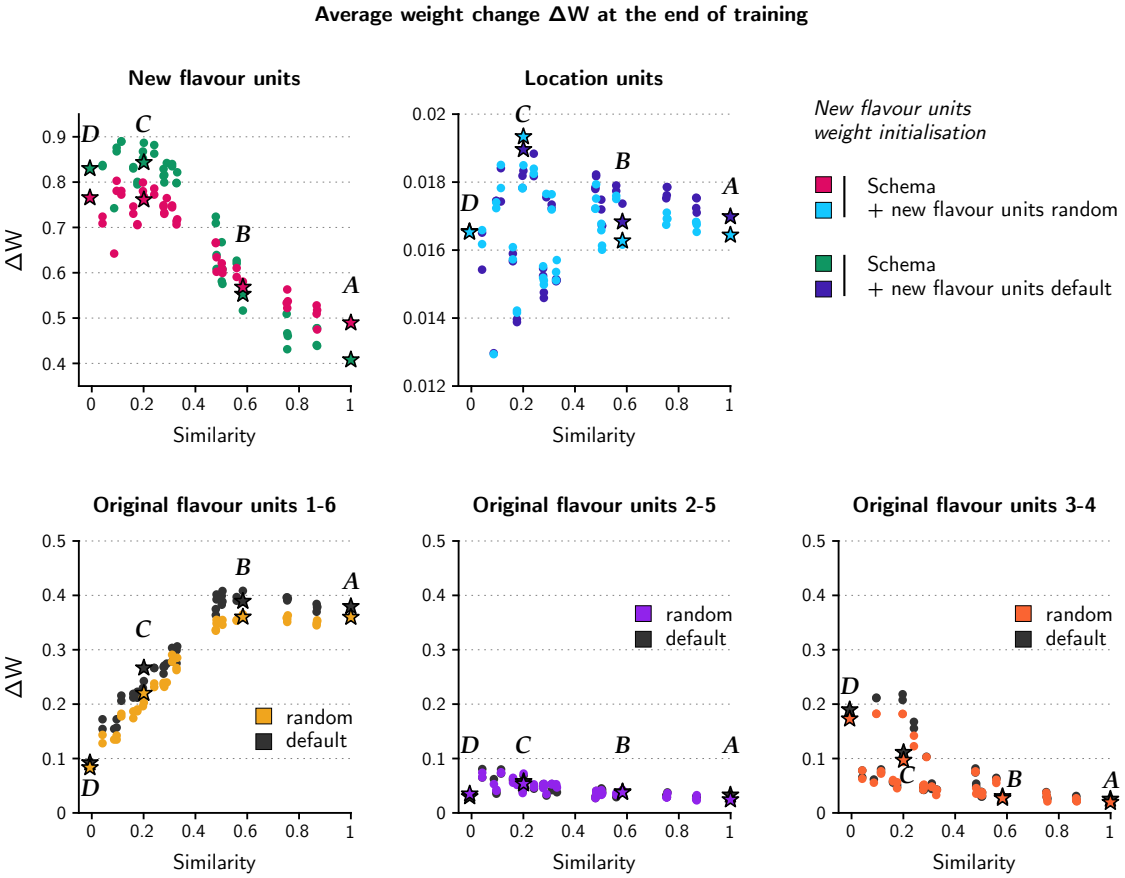


Figure 3.27: Total weight change in the neocortex after 30 epochs of training two new associations for different location similarity with prior knowledge. For all plots, each dot represents the average result (10 runs) obtained for a given new location (49 in total) (see locations in Fig. 3.25). The dark dots show the results obtained with a network with prior schema and default initialisation of the new flavour units, and the dots in light colors show the same but with the 'random' initialisation of the new flavour units. We divided the results according to the types of units. The total cortical plasticity was measured as the average absolute weight change for each group of units, between the end of training and before the start of training: $\Delta w = \langle \frac{1}{N_h} \sum_{j=1}^{N_h} |w_{ij}^{\text{end}} - w_{ij}^{\text{start}}| \rangle_i$, where N_h is the number of hidden units, and the index i represents the group of visible units considered. For example i can be the indices of the two original flavour units of the associations no.1 and no.6 (bottom left panel), the indices of the new flavour units (top left panel), or the indices of the 225 location units (top right panel).

no.2 and no.5, performance dropped to half for all new locations (middle panel).

Thus recall performance of old associations was less impaired when the new locations were closer to them. This observation is in line with the results of McClelland (2013), and we suggest that to prevent the forgetting of original locations that were further away from the new ones (e.g. locations no.2 and no.5) the network would have to replay them and do interleaved learning.

Note that when we probed the old memories at the start of training we did not find catastrophic interference upon presenting the new associations, regardless of the location of the new rewards. This result is not shown, but was similar to the example in Fig. 3.24, which corresponds to the training with location ***B***.

Remark on the performance of the two original locations no.1 and no.6. We already mentioned in Section 3.3.2 that a good performance did not imply that the memory reconstruction was of 'good quality', i.e. the reconstruction of the location units activity did not reflect the true sensory patterns (Fig. 3.15). In the current simulations, we suspect that the reconstruction quality of the sensory patterns for the locations no.1 and no.6 would be lower at the end of the new training, because it would capture the shift of the representations of the locations: the network did not reinstate the old locations per se but the new locations, as shown in Fig. 3.24a. However, we predict that this shift of representation would not affect performance at the behavioural level, which in our framework is modelled as the probability to recall the correct location, as long as the new locations are not too far from the two original reward sites.

Plasticity of the flavours and locations units

Finally, we monitored the total weight change after the new training for the different visible units (Fig. 3.27): the two new flavour units (top left panel), the original flavour units (the three lower panels), and the locations units (top right panel). We measured the plasticity as the average absolute weight change of the cortical connections.

For the location units (top right panel), we found more change for the unknown location (location ***C***), and less change for the locations in the edges (not highlighted in the figure). When the new locations were close to the two original (no.1 and no.6), the new flavour units required less plasticity (top left panel). The old flavour units 2 to 5 underwent little plasticity compared to the new flavour units[9] (middle and right

[9]We can compare the plasticity levels between the flavour units as we took groups of two visible units for each analysis; on the other hand, we cannot compare with the location units as the plasticity was averaged over the 225 units.

lower panels), which was surprising considering the drop of performance of the related associations (see Fig. 3.26, middle and right panels). This suggest that performance dropped because of the remapping of the location units activity, and not because the weights of the flavour units were modified. On the other hand, when the new locations overlapped with the two original no.1 and no.6 we found more connection change for the two original flavour units, almost as much as for the new flavour units (left lower panel, locations ***B*** and ***A***). We observed a similar effect, albeit less pronounced, when the new locations were closer to no.3 and no.4 (location ***D***).

3.4 Conclusion

In this chapter we modelled the experiment of Tse et al. (2007) in order to clarify the concept of schemas, the mechanisms of sleep replay, and the potential involvement of the prefrontal cortex.

In Section 3.2, we implemented a simple model to identify why our network required 1) the prefrontal cortex and 2) the meta-schema to replicate the findings of Tse et al. (2007).

First, we looked at the acquisition of a consistent schema by presenting constant flavour-place associations to the network (Section 3.2.2). We found that the hippocampus was important to select the relevant memories for replay during sleep, which allowed to consolidate the rewarded associations. Replay of random associations prevented consolidation, but it did not prevent hippocampal recall of rewarded memories as the network still acquired the body of knowledge (i.e. the flavours on one hand, and the locations on the other hand). We also suggested that the hippocampus might have two recall pathways to allow flexible behaviour depending on the context. We proposed that the hippocampus could either rely on recent, episodic memories, or probe memories consolidated in the neocortex. The selection of the retrieval pathway could perhaps be biased by the prefrontal cortex (Preston & Eichenbaum 2013).

Next, we trained the network on two new flavour-place associations (Section 3.2.3). Rapid acquisition was only possible if we up-regulated memory formation in the hippocampus. Yet, recall performance was low at the start of training as the network could not learn the two new associations simultaneously. We believe this competitive effect was a consequence of the novelty of the features since new units were added to the sensory layer, which made it difficult to replay them correctly during sleep. We then compared these results with the new training in an inconsistent schema. Tse

et al. (2007) showed that rats rapidly acquired the two new associations in the consistent but not in the inconsistent schema. However, our simulations indicated that there was no difference between the two schemas, implying that memory formation in the hippocampus had to be dynamically regulated. We suggested that the prefrontal cortex could monitor the expected consistency of the flavour-place associations during the original training, and then somehow use it to influence memory acquisition in the hippocampus.

Finally, we examined the retention of the original rewarded associations (Section 3.2.4). The new training did not impair the retention of remote memories in Tse et al. (2007), whereas in our model recall deteriorated significantly. In particular, consolidated memories were quickly overwritten, while recall mediated by the hippocampus was less impaired. In line with McClelland et al. (1995), reactivation of remote memories during sleep reduced interference but hindered the consolidation of the new memories. On the other hand, replaying old associations did not impair the retrieval of the new memories by the hippocampus. Increasing the number of replays helped the consolidation the new associations while preserving existing knowledge; however, the new memories were never fully consolidated, as opposed to when the system replayed only the new associations.

The second part of this chapter focused on the semantic schema in the neocortex during the training of consistent associations (Section 3.3). We implemented a different model as the previous one was too simplistic to explore the representation of knowledge in cortical networks.

We first studied the requirements for the successful acquisition of the flavours and the reward locations activity patterns (Section 3.3.2). We initially trained the network without sleep replay for different cortical learning rates, and we observed a trade-off between training time and quality of the memories. Since the network had limited access to training examples during experience, the learning rate had to be large enough to drive learning, but not too high otherwise the sensory patterns could not be properly reconstructed. Nevertheless, in these conditions learning would be too slow, and hence we suggested that replay during sleep could serve as a 'virtual' training to the network. We observed that, with a low learning rate and a large number of reactivations, sleep replay facilitated the consolidation and stabilisation of the memory patterns. Furthermore, our simulations revealed that cortical knowledge in the neocortex impacted hippocampal recall in two ways: first, during the bottom-up cue recognition, and second, during top-down reconstruction in the sensory cortices. Hence, hippocam-

pal recall performance was limited while memories were gradually consolidated in the neocortex.

The aim of the next simulations was to study the impact of the semantic schema on the acquisition of two new odour-place associations (Section 3.3.3). For comparison, we also trained a 'naive' network, i.e. with the same initial weights as prior to the original training. In other words, we wanted to evaluate whether knowing the six flavours and the six locations facilitated the assimilation of the new memories. In line with the findings of Tse et al. (2007), learning and consolidation were more rapid in a trained network compared to a naive network. However, the new memories could also be rapidly assimilated in a network with random weights sampled from distributions with similar statistics than the weights of a network trained on the six original associations. We tested whether any random weight initialisation would lead to the same outcome (not shown), but our preliminary results indicated that it only worked when we chose the statistics of the schema (e.g. it did not work with only the mean or the standard deviation).

Weight initialisation thus played an important role in how much more quickly the network learned the new information. Unfortunately, experimentally very little is known about this issue, and most experimental techniques are geared towards seeing changes in connections rather than measuring their strength before they are modified. In addition, weight initialisation in neural network is a hard problem as convergence is often unpredictable. For instance, the study by Kolen & Pollack (1991) revealed that a very small change of initial weights can have a significant impact on the convergence time even in a simple feedforward network and on a simple problem. Hence, it was interesting that we obtained similar speedup with the 'random' initialisation as with the prior semantic schema.

In the last part we further investigated the relationship between the new associations and prior knowledge (Section 3.3.4). The hippocampus assimilated more rapidly new associations when the new locations were closer to original ones. This effect was associated with less plasticity for the new flavours units. On the other hand, consolidation did not benefit from the similarity with old locations. If anything, consolidation of new flavours with old locations was slower. When the new locations overlapped with original locations, we found less interference for the corresponding original flavour-place associations, despite the fact that we did not replay them. This effect was associated with more plasticity for the concerned old flavours units. This could imply reconsolidation mechanisms.

Chapter 4

Interplay between hippocampus, prefrontal cortex and associative cortex

In this chapter, we present the final computational framework that we implemented to investigate the experimental results of Tse et al. (2007). Although this model was developed for this particular purpose, it can easily be extended to study similar memory tasks. This version of the model built upon the models introduced in the previous Chapter 3, and expands the implementation of the different modules: prefrontal cortex, associative cortex and hippocampus.

We noticed in Section 3.2 that a system consisting of only hippocampus and associative cortex was not able to learn and generalise the *structure of knowledge*. Indeed, a network that was trained in a consistent schema, where the associations flavour-place were fixed, learned new flavour-place associations the same way than a network that was trained in an inconsistent schema, where the associations flavour-place were swapped across training (see Figure 3.9 in Section 3.2). This result was at odds with the experiment of Tse et al. (2007), because in the experimental study the rodents distinguished the two types of schemas and learned rapidly new associations in the consistent, but not in the inconsistent schema (Fig. 3.1e). Thus, we suggested in Section 3.2 that the prefrontal cortex could extract the information about the *structure of knowledge*, which we called *meta-schema*, and was for this task characterised by a single meta-parameter that tracked the consistency of the rewarded associations. We also suggested that the prefrontal cortex could use this abstract knowledge for subsequent learning by modulating memory formation in the hippocampus. However, the model

in Section 3.2 did not address *how* the prefrontal cortex could mediate these operations. In the current work we propose an implementation that relies on the interaction between the prefrontal cortex, associative cortex and hippocampus.

Particularly, in the first model (Section 3.2), the prefrontal cortex did track the consistency of the associations but it required that the hippocampus explicitly stored the reward probabilities at all times (see model of the hippocampus Fig. 3.3 and update Fig. 3.5). In the current model, on the other hand, we will show that the network can directly compute the consistency of associations based on the knowledge consolidated in the neocortex (RBM).

In the first Section 4.1 of this chapter we describe the implementation of the framework. In the result Section 4.2 we investigate the interactions between the prefrontal cortex, associative cortex and hippocampus during the acquisition of six flavour-place associations. In line with the experiment of Tse et al. (2007), we considered two schemas to train the network[1]: a consistent schema, where the associations flavour-place were fixed, and an inconsistent schema, where the associations flavour-place were continually swapped across training (Section 4.2.1). For each training condition, we examine 1) the interaction between the prefrontal cortex and the hippocampus during episodic memory formation (Section 4.2.2), and 2) the off-line replay of episodic memories during sleep (Section 4.2.3).

In the second part of the results we investigate whether the acquisition of new associations is faster in a consistent schema than in an inconsistent schema (Section 4.3.1). In particular, we examine the impact of blocking the interaction between the prefrontal cortex and hippocampus (Section 4.3.2). Finally, we study other factors that could influence learning and consolidation, namely remote memories in the hippocampus (Section 4.3.3) and the number of reactivations during sleep (Section 4.3.4).

[1] Unlike the experimental study, we never trained simultaneously on the two types of schemas; however, this could be implemented by adding context units in the visible layer of the RBM.

4.1 Model setup

This version of the model incorporates all the modules introduced in Chapter 3: prefrontal cortex, associative cortex and hippocampus. In particular, we elaborated the model of the prefrontal cortex, introduced in Section 3.2, clarifying its interaction with the hippocampus and the semantic schema in the associative cortex. We also extended the model of the neocortex used in Section 3.3, and we implemented a new version of the hippocampus that combined elements of the models used in Sections 3.2&3.3. An overview of the model is shown Fig. 4.1.

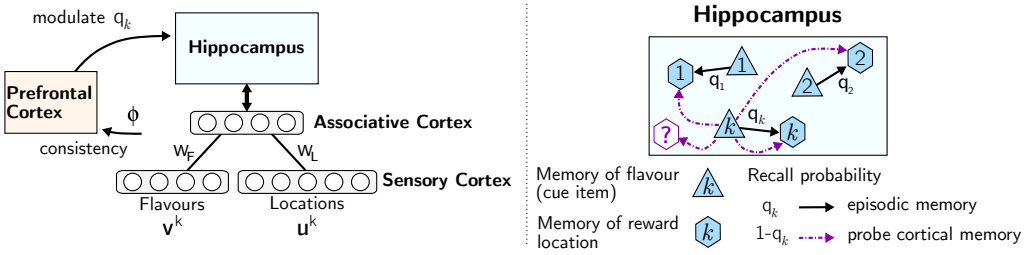


Figure 4.1: Overview of the last version of the model. Global model architecture (left panel). The sensory cortex receives inputs from the environment. The prefrontal cortex tracks the overall consistency of the associations over epochs. It can also detect novel or conflicting associations by interacting with the neocortex (Section 4.1.2), and subsequently influences the formation of episodic memories in the hippocampus. The hippocampus (right panel) is connected to the associative layer of the neocortex; it stores memories of rewarded associations, creating a memory for the flavour and a memory for the location, and connecting the two memories with an episodic link q_k. The value of q_k is modulated by the prefrontal cortex during the creation of the episodic memory (Section 4.1.3). The hippocampus mediates recall when the system is cued by a flavour (Section 4.1.4), but also recall during sleep replay and thus impacts memory consolidation in the neocortex (Section 4.1.5).

4.1.1 Model of the associative neocortex and measure of association consistency

The neocortical module was similar to the model in Section 3.3. The cortical network (RBM) had 325 visible units and 100 hidden units. We used the coordinate representation of the locations in the arena (225 units), as described in the method Section 3.1.2, but we used the fixed input activation patterns instead of sampling to obtain the input patterns (Fig. 3.2b, threshold method). Thus, we did not have to worry about the qual-

ity of sensory reconstructions during sleep replay and we could increase the learning rate during this phase of training compared to simulations in Section 3.3.

To represent flavours in the sensory layer, we defined random patterns of size 100 units with on-probability $p = 0.2$. This contrasted with the previous model in Section 3.3, where we used one visible unit to represent each flavour, and we added a new unit to represent a new flavour. Such change was motivated by two reasons: 1) sensory representations are believed to overlap in the cortex, and 2) we could define new flavour patterns without worrying about weight initialisation, which we saw was an influential parameter for new training (Section 3.3.3).

We used similar notations to the ones in Section 3.3: $\mathbf{v}^k$ and $\mathbf{u}^k$ are the vectors of visible binary units representing the flavour and location of the kth rewarded association. Their respective on-probabilities are noted $\boldsymbol{\nu}^k$ and $\boldsymbol{\mu}^k$. The matrices connecting the flavour units and location units to the hidden layer are noted $\mathbf{W}_F$ and $\mathbf{W}_L$, and $\mathbf{h}$ is the vector with the state of the hidden units $h_j \in \{0, 1\}$, and $p(\mathbf{h} = 1)$ is the vector with elements $p\left(h_j = 1\right)$.

We needed a method to infer the likelihood of reward for a given flavour-place association since the network did not explicitly keep track of the reward probabilities. We thus defined a probability ϕ to evaluate the consistency of an association between a flavour and a reward location $(\mathbf{v}, \mathbf{u})$:

$$\phi = \frac{1}{1 + \exp\left(\mathcal{G}_{\boldsymbol{\nu}}(\mathbf{u})\right)} \tag{4.1}$$

where $\mathcal{G}_{\boldsymbol{\nu}}(\mathbf{u})$ is an energy-based measure of the consistency between the reward location $\mathbf{u}$ and the flavour $\mathbf{v}$ (Algorithm 4.1). The method to evaluate consistency is based on cortical recall: starting from the flavour pattern $\mathbf{v}$, the neocortex infers the activity of the location units; then, the network compares the resulting activation of the hidden units $\boldsymbol{\xi}$ with the field of the hidden units $\mathbf{y}$ corresponding to the observed reward location $\mathbf{u}$. If $\mathcal{G}_{\boldsymbol{\nu}}(\mathbf{u})$ is low ($\mathcal{G}_{\boldsymbol{\nu}}(\mathbf{u}) < 0$) then the consistency is high, and if $\mathcal{G}_{\boldsymbol{\nu}}(\mathbf{u})$ is high ($\mathcal{G}_{\boldsymbol{\nu}}(\mathbf{u}) > 0$) then the consistency is low. Note that in Algorithm 4.1 to reduce the noise during the computations we used the real-valued vectors (probabilities) for the Gibbs sampling steps, and low temperature to infer the activity of the hidden units.

Although we refer to ϕ as the probability of an association, it must not be confused with the probability distribution $P(\mathbf{v}, \mathbf{u})$ over the all visible patterns of the Restricted Boltzmann Machine. We chose not use the probability distribution of the RBM because this distribution is rather hard to compute biologically because of the normalising fac-

Algorithm 4.1 Computation of consistency $\mathcal{G}_{\mathrm{v}}(\mathbf{u})$ of $\mathbf{u}$ given $\mathbf{v}$

Notations:

$p(\mathbf{h}=1)$ is the vector with elements $p\left(h_j=1\right)$

$\sigma(\cdot)$ is the sigmoid activation function applied element-wise

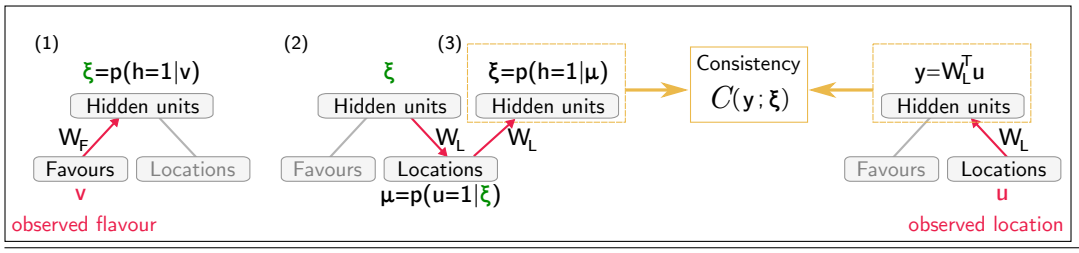
Given the observed flavour pattern $\mathbf{v}$

1. Compute the activation probability of the hidden units
   $\boldsymbol{\xi} = p(\mathbf{h}=1|\mathbf{v}) = \sigma\left(\mathbf{W}_F^{\intercal}\mathbf{v};T\right)$, temperature $T=0.5$
2. Compute the activation probability of the location units
   $\boldsymbol{\mu} = p(\mathbf{u}=1|\boldsymbol{\xi}) = \sigma(\mathbf{W}_L\boldsymbol{\xi};T)$, temperature $T=1$
3. Compute the new activation probability of the hidden units
   $\boldsymbol{\xi} = p(\mathbf{h}=1|\boldsymbol{\mu}) = \sigma\left(\mathbf{W}_L^{\intercal}\boldsymbol{\mu};T\right)$, temperature $T=0.5$

Given the observed reward location pattern $\mathbf{u}$

Compute the field of the hidden units $\mathbf{y} = \mathbf{W}_L^{\intercal}\mathbf{u}$

Compute consistency $\mathcal{G}_{\mathrm{v}}(\mathbf{u}) = C(\mathbf{y};\boldsymbol{\xi}) = -\frac{1}{\sum_j \xi_j}\mathbf{y}^{\intercal}\boldsymbol{\xi}$



tor (partition function) which sums over all possible visible patterns (see Chapter 2). One way of reducing the computational load is to restrict the visible patterns to the ones from the data set, i.e. the N flavours and N reward locations in our case. However, this solution would require that the system has full knowledge of these specific patterns, which is a constraint that we wished to avoid. Thus, we decided to derive a simple probability measure instead.

Another important remark is that ϕ should not be interpreted as the probability of location $\mathbf{u}$ given flavour $\mathbf{v}$, normalised by all possible locations; instead, it should be interpreted as the likelihood of the association: if $\phi \to 1$, then it is likely, if $\phi \to 0$ then it is unlikely. Since we measure ϕ by probing neocortical memory, we can also interpret it as measure of consolidation strength.

At the start of training, the initial cortical weights (RBM) were all zeros, so $\mathcal{G}_{\mathrm{v}}(\mathbf{u}) = 0$ and thus $\phi = 0.5$. This was the default value when the network had no prior knowl-

edge. Then, for a consistent association the probability ϕ will increase, indicating a strong and reliable association, whereas for an inconsistent association the probability ϕ will decrease, indicating a weak and unreliable association. In the next Section 4.1.2 we explain how the prefrontal cortex interpreted the probability ϕ of an association.

4.1.2 Model of the prefrontal cortex

The prefrontal cortex mediated the extraction of the meta-parameter ϕ^* which represented the expectation about the consistency (or variability) of the associations. Since there is no evidence about where such knowledge could be stored in the cortex, we supposed that the prefrontal somehow had access to this meta-parameter.

In the preceding Section 4.1.1 we defined a measure of consistency ϕ (Eq. 4.1), and hence the prefrontal cortex simply had to extract the overall consistency of the rewarded associations across epochs. As we mentioned earlier, prior to training the initial value od consistency was $\phi^* = 0.5$. Then the prefrontal cortex updated the meta-parameter at each epoch (see Section 4.1.5). If the flavours were always associated to the same locations, ϕ^* should increase above 0.5. This is the case of the 'Elephant' category in Fig. 1.1, and the system then will expect consistent associations. On the other hand, if the associations were swapped throughout training, ϕ^* should decrease below 0.5. This is the case of the 'Bird' category in Fig. 1.1, and the system will expect the associations to change in the future.

The meta-parameter set expectations for new incoming information, and thus if the expectations are not met the system knows something is amiss. We incorporated this function in our model so that the prefrontal cortex could detect whether a new experience was normal or surprising. The implementation was very simple: when presented with an association, the network computed the probability ϕ of the observed association, and then compared it to the model expectation ϕ^*. If $\phi \geq \phi^*$, the association met the expectation, but if $\phi \ll \phi^*$ then the system detected a surprising event.

However, the observed probability ϕ could be lower than the expectation for two reasons. In a first case, the association is conflicting with prior knowledge: for instance, in the current task, a known flavour that used to be always associated to a specific reward location is suddenly associated with a different reward location. In a second case, the association involves new elements: for instance, a new flavour is associated with a new reward location. To distinguish these two cases, the system estimated the consistency of the predicted reward location given the observed flavour. The

predicted consistency, noted $\hat{\phi}$, is defined for a given flavour $\mathbf{v}$ by:

$$\hat{\phi} = \frac{1}{1 + \exp(\mathcal{G}_v)} \tag{4.2}$$

Eq. 4.2 above is similar to Eq. 4.1, but it evaluates the average consistency of the recalled locations when the network is cued by the flavour $\mathbf{v}$, rather than evaluating the consistency of the recalled locations with the observed location $\mathbf{u}$ (Algorithm 4.2). Thus, in the conflict case the predicted consistency $\hat{\phi}$ is close to the model expectation, because the flavour was already consolidated with a different reward location. By contrast, in the novelty case the predicted consistency $\hat{\phi}$ is low because the new flavour was not associated to any particular location. The Fig. 4.2 describes conceptual examples of the different scenarios.

Algorithm 4.2 Computation of predicted association consistency $\mathcal{G}_\mathrm{v}$ given $\mathbf{v}$

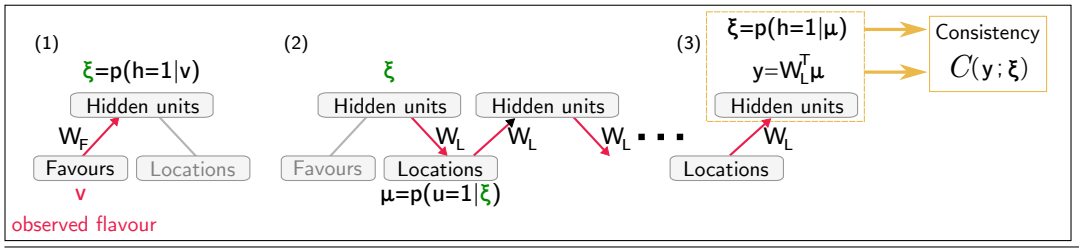
Notations:

$p(\mathbf{h}=1)$ is the vector with elements $p\left(h_j=1\right)$

$\sigma(\cdot)$ is the sigmoid activation function applied element-wise

Given the observed flavour pattern $\mathbf{v}$

1. Compute the activation probability of the hidden units
   $\boldsymbol{\xi} = p(\mathbf{h}=1|\mathbf{v}) = \sigma\left(\mathbf{W}_F^\intercal \mathbf{v}; T\right)$, temperature $T = 0.5$
2. **for** M iterations **do**
   2a. Compute the activation probability of the location units
   $\boldsymbol{\mu} = p(\mathbf{u}=1|\boldsymbol{\xi}) = \sigma(\mathbf{W}_L \boldsymbol{\xi}; T)$, temperature $T = 1$
   2b. Compute the field of the hidden units $\mathbf{y} = \mathbf{W}_L^\intercal \boldsymbol{\mu}$
   2c. Compute the new activation probability of the hidden units
   $\boldsymbol{\xi} = p(\mathbf{h}=1|\boldsymbol{\mu}) = \sigma(\mathbf{y}; T)$, temperature $T = 0.5$
   end for
3. Compute consistency $\mathcal{G}_\mathrm{v} = C(\mathbf{y}; \boldsymbol{\xi}) = -\frac{1}{\sum_j \xi_j} \mathbf{y}^\intercal \boldsymbol{\xi}$



In practice, to determine whether the observed probability ϕ or the predicted probability $\hat{\phi}$ were larger than the meta-parameter ϕ^* (expected probability), we set an arbitrary lower bound to $(1-\epsilon)\,\phi^*$, where $\epsilon = 0.15$.

In our model, the prefrontal cortex coordinated the detection of consistency and compliance with the meta-schema of incoming information, and changed its state accordingly (see Algorithm 4.3). In the next Section 4.1.3 we explain how the state of the prefrontal cortex influenced the creation of episodic memories in the hippocampus.

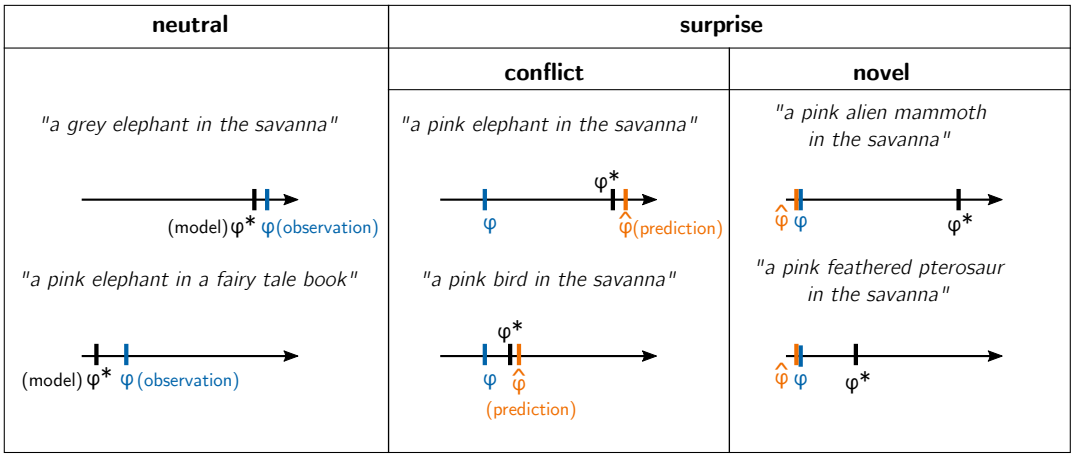


Figure 4.2: Conceptual examples to explain how the state of the prefrontal cortex is determined. The arrows represent the consistency or strength of the associations. (Left) In the context of the savanna, the model has a high expectation ϕ^* about the colour of the Elephant family (sharp distribution over colours, Fig. 1.1), whereas it does not in the context of a fairy tale. Thus, "a pink elephant", which has low probability ϕ (Eq. 4.1), is classified as *neutral* in the fairy tale context, but *surprising* in the savanna. (Middle) Since the model is confident about the grey color of the elephant in the savanna, measured by $\hat{\phi} \sim \phi^*$ (Eq. 4.2), "a pink elephant in the savana" is classified as *conflicting*. Similary, even though we expect more colour variability in the Bird category (Fig. 1.1), the model detects conflict because the colour "pink" was never seen before. (Right) Still in the savanna, if we see unknown animals, i.e. an "alien pink mammoth" or a "pink feathered pterosaur", we may assume that each belongs to the same super-category as either the Elephant or Bird family, and thus we use the corresponding expectation ϕ^* for comparison. Yet, we have no expectation for these new animals (low $\hat{\phi}$) and thus the episodes are classified as *novel*.

Algorithm 4.3 Setting the state of the prefrontal cortex

Notations:
#ϕ probability of the observed association (Eq. 4.1)
$\hat{\phi}$ probability of the predicted association (Eq. 4.2)
#ϕ^* model expectation (meta-parameter)
Threshold $\epsilon = 0.15$

if $\phi \geq (1-\epsilon)\,\phi^*$ **then** # observation conform to expectation
  PFC state = neutral
else
  Compute prediction $\hat{\phi}$
  if $\hat{\phi} \geq (1-\epsilon)\,\phi^*$ **then** # prediction conform to expectation
    PFC state = conflict
  else # prediction different from expectation
    PFC state = novelty
  end if
end if

Summary: surprise and the neocortex There are two different kind of surprise that should not be confused. (1) Whether an association is semantically odd (low ϕ) or likely (high ϕ), which is related to the consolidation in the neocortex. For example, "a pink elephant" is odd. (2) Whether an association is a mismatch ($\phi \ll \phi^*$) or is acceptable ($\phi \sim \phi^*$) relative to a specific model expectation ϕ^*. For example "a pink elephant in the savanna" is a mismatch, but "a pink elephant in a fairy tale book" is acceptable. We suggest this second type of surprise involves the prefrontal cortex, and necessitates a meta-schema linked to a context to define the meta-parameter ϕ^*.

4.1.3 Model of the hippocampus

Similar to the model in Section 3.3, the hippocampus only stored rewarded memories, and the memory patterns were overwritten at each epoch. However, unlike the previous model, here the hippocampus stored the elements of an episodic memory separately: when the network encountered a rewarded flavour-place association $\left(\mathbf{v}^k, \mathbf{u}^k\right)$, the hippocampus first took a snapshot of the field of the hidden units when the flavour pattern

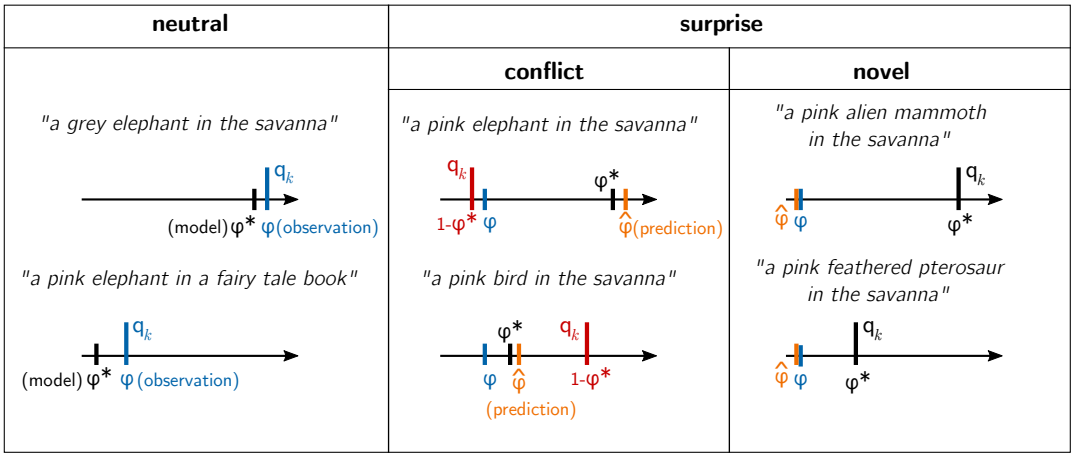


Figure 4.3: Conceptual examples to explain how the value of the episodic link in the hippocampus is determined according to the state of the prefrontal cortex. The arrows represent the consistency or strength of the associations. (Left) The two episodes are classified as *neutral* and hence the probability q_k of this episode is equal to ϕ, the probability of the association according to the knowledge consolidated in the neocortex (Eq. 4.1). (Middle) The two episodes are classified as *conflicting* and hence the probability q_k of the episodes are set to $1-\phi^*$, where ϕ^* is the expectation for each category. As a result, "pink bird" is more likely than "pink elephant", even though they are both semantically unlikely (low ϕ). (Right) The two episodes are classified as *novel* and hence the links q_k are set to the expectations ϕ^* of the known similar categories, generalising the structure to the new categories (Fig. 1.1).

$\mathbf{v}^k$ was presented, and then took a snapshot of the field of the hidden units when the location pattern $\mathbf{u}^k$ was presented (**step 1** in Fig. 4.5):

$$\begin{cases} \mathbf{x}^k &= \mathbf{W}_F^\intercal \mathbf{v}^k \\ \mathbf{y}^k &= \mathbf{W}_L^\intercal \mathbf{u}^k \end{cases} \tag{4.3}$$

These two memory patterns $\mathbf{x}^k$ and $\mathbf{y}^k$ were connected with an 'episodic link' q_k which represented the recall probability of the association. We assumed that the value of the episodic link q_k was determined by the state of the prefrontal cortex (Table 4.1). The procedure is illustrated in Fig. 4.3 for the conceptual examples mentioned in the Introduction of this thesis (Fig. 1.1). Briefly, if the prefrontal cortex was in the neutral state, the recall probability q_k of the episode was then equal to the probability ϕ of the association according to neocortical memory. If the prefrontal cortex detected conflict, the link q_k was then set to $1 - \phi^*$; hence, the higher the expectation ϕ^* (meta-parameter), the lower the value q_k of the episodic link. Finally, in case of novelty the episodic link was set to the value of the meta-parameter, that is $q_k = \phi^*$. The role of the episodic link q_k for recall and sleep replay is explained in the next Sections 4.1.4 and 4.1.5.

PFC state	Episodic link q_k
neutral	ϕ
conflict	$1 - \phi^*$
novelty	ϕ^*

Table 4.1: Value of episodic link q_k connecting a flavour and location memories in the hippocampus, according to the state of the prefrontal cortex. Notations: ϕ is the probability of the flavour-place association (Eq. 4.1), and ϕ^* is the model expectation (meta-parameter). For the states of the prefrontal cortex refer to Algorithm 4.3.

It is important to note that odd things usually tend to stick in our memory. Thus, we might actually remember all the conflicting examples given Fig. 4.3 vividly because they are extremely odd[2]. However, these memories will presumably be unique, long-term episodic-like memories, and hence fall out of the scope of this work as we are interested in memories that will be integrated within neocortical schemas.

We mentioned that the hippocampus stored separately the memory of the flavour and the memory of the location for each rewarded association; this method contrasted with previous model (Section 3.3), where the hippocampus directly stored the memory

[2]To our knowledge, there is no proof of existence of pink elephants...

of the associations (i.e. it stored the field $\mathbf{x}^k = \mathbf{W}_F^\intercal \mathbf{v}^k + \mathbf{W}_L^\intercal \mathbf{u}^k$). There were two reasons for this change. The first reason was that if the network learned novel or conflicting associations, it could not properly reconstruct the patterns in the sensory cortex when the hippocampus reinstated the new memories. For instance, if we swapped a flavour-place association at the end of training in the consistent schema, the memory could not be reconstructed using the top-down connections because the field of the hidden units $\mathbf{x}$ was mostly negative, as the combination of the two patterns $\mathbf{v}$ and $\mathbf{u}$ was implausible. This effect was particularly detrimental for sleep replay. To circumvent this issue we created two distinct memory patterns $\mathbf{x}$ and $\mathbf{y}$ in the hippocampus, and hence even if one pattern or the other was implausible, or if the association was implausible, the sensory patterns could nonetheless be reactivated. We show in the training Section 4.1.5 how we implemented this procedure during sleep replay.

The second reason for storing separate flavour and location memories was that the hippocampus was then able to reinstate associations that were not necessarily rewarded during the last epoch. We expand on this point in the following Section 4.1.4.

4.1.4 Memory recall

We decided to extend the hippocampal model of Section 3.3 to implement a model that was more flexible. Indeed, the former model only contained the memories of the last epoch, and thus if the network saw the association flavour PINEAPPLE + reward location no.1, the hippocampus could only reinstate this particular association and was unable to reinstate the flavour PINEAPPLE with another location. We believed that it was a limitation, because during training the animals did not know yet whether the associations were reliable or not. In particular, we wanted the hippocampus to have the ability to select not only the last rewarded location, but also any location that was rewarded in the past or that was rewarded with a different flavour.

The first hippocampal model in Section 3.2 took the uncertainty about the observed associations into account, but there the hippocampus was overloaded with unnecessary memories (it stored both rewarded and unrewarded events). Therefore, here we implemented an alternative solution where the hippocampus could recall episodic memories, to reinstate the last rewarded locations, or it could probe memory consolidated in the associative neocortex, to reinstate a location that could possibly be different. To compare the various models of the hippocampus, the reader can refer to Fig. 3.3&3.4 (first model), Fig. 3.12&3.13 (second model) and Fig. 4.1&4.4 (current model).

Standard recall was implemented as follows (same steps shown in Fig. 4.4):

Step 1 Given a flavour $\mathbf{v}$ in the sensory cortex (visible layer of the RBM), the activation propagated to the associative (hidden) layer; let $\mathbf{x}^{cue} = \mathbf{W}_F^{\intercal}\mathbf{v}$ be the vector with the field of the hidden units. The hippocampus selected the flavour memory vector $\mathbf{x}^k$ which had the highest correlation with $\mathbf{x}^{cue}$. We chose the maximum correlation because this method of recall was robust (correct identification), and hence we knew this first step of hippocampal recall would not influence memory performance. However, a more flexible method of recall would be preferable in the future (i.e. probabilistic recall), in particular to test weak or mixed flavours. In addition, if we tested old flavours that were consolidated in the neocortex, but no longer stored in the hippocampus, the current model was forced to select one of the newly stored flavour memories; instead, the model should have the option to switch to pure neocortical recall (i.e. in Fig. 4.4, the network would continue Gibbs sampling steps directly from $\mathbf{x}^{cue}$ instead of $\mathbf{x}^k$).

Step 2 Given the flavour memory k, the hippocampus could recall a location memory via to recall pathways: one direct recall pathway, recalling the association of the episodic memory (black arrow in Fig. 4.4), and one indirect recall pathway, probing the memory consolidated in the neocortex (purple dashed arrows in Fig. 4.4). The hippocampus recalled associated location memory vector $\mathbf{y}^k$ (episodic memory k) with probability q_k; otherwise, it switched to the cortical recall pathway, and could then recall any location memory r stored in the hippocampus with probability $\mathcal{P}_s\left(\mathbf{y} = \mathbf{y}^r | \mathbf{x}^k\right)$, where $\mathcal{P}_s$ is a recall probability defined by

$$\mathcal{P}_s\left(\mathbf{y} = \mathbf{y}^r | \mathbf{x}^k\right) = \frac{\exp\left(-C\left(\mathbf{y}^r; \boldsymbol{\xi}\right)\right)}{\sum_l \exp\left(-C\left(\mathbf{y}^l; \boldsymbol{\xi}\right)\right)} \tag{4.4}$$

The probability $\mathcal{P}_s$ was obtained by probing memory in the neocortex: the hippocampus clamped the hidden units to $\mathbf{x}^k$ (the flavour memory pattern recalled), and the neocortex did M Gibbs steps to reconstruct the activity $p\left(\hat{\mathbf{u}} \mid \mathbf{x}^k\right)$ of the location units in the sensory layer. Taking a last sample $\hat{\mathbf{u}}$, the network computed once again the activity of the hidden units $\boldsymbol{\xi} = p\left(\mathbf{h} = 1 \mid \hat{\mathbf{u}}\right)$. Finally, the system calculated the consistency $C\left(\mathbf{y}^r; \boldsymbol{\xi}\right) = -\frac{1}{\sum_j \xi_j}\left(\mathbf{y}^r\right)^{\intercal}\boldsymbol{\xi}$ (see Section 2.3.3 in Chapter 2) between $\boldsymbol{\xi}$ and each field vector $\mathbf{y}^r$ (location memory patterns) stored in the hippocampus. We then applied the softmax function to obtain the recall probabilities. We added a null-field vector $\mathbf{y}^0$ ($r = 0$) to take into account the case where no location memory was found to match the cue. Note that for the null vector $\mathbf{y}^0$ the consistency $C\left(\mathbf{y}^0; \boldsymbol{\xi}\right)$ is equal to zero.

Step 3 For the last step of recall the network inferred the final activity of the location units. If the hippocampus selected the null vector $\mathbf{y}^0$, the hippocampus did not reinstate a location memory and instead the network kept the activation $p\left(\hat{\mathbf{u}} \mid \mathbf{x}^k\right)$ of the location units obtained by cortical recall. Otherwise, the hippocampus clamped the hidden units to the selected location memory vector $\mathbf{y}^r$, and the neocortex reconstructed the activity of the locations units $p(\mathbf{u} \mid \mathbf{h})$ in the sensory layer using the top-down cortical connections $\mathbf{W}_L$. To calculate performance for the simulation of the task of Tse et al. (2007), the method was the same that the one we used in Section 3.3. We derived the probability to go to a location by summing the on-probabilities of the units surrounding it, and normalising by the total activity: if $\mathcal{A}_r$ is the region surrounding the potential reward location r, the probability to go to location no.1 for instance is $P(\textit{location 1}) = \frac{D_1}{\sum_r D_r}$, where $D_r = \sum_{i \in \mathcal{A}_r} p(u_i = 1)$.

Figure 4.4: (Next page.) Recall mediated by the hippocampus. **Step 1**: A flavour cue is presented to the sensory cortex, and the activity propagates to the associative layer of the neocortex. The hippocampus computes the correlation between the resulting field vector $\mathbf{x}^{\text{cue}}$ and each of the stored field vectors $\mathbf{x}^k$ corresponding to the memories of the flavours. The memory that maximises the correlation is chosen. **Step 2**: Given the selected flavour memory k, the hippocampus recalls the associated location memory pattern $\mathbf{y}^k$ with probability q_k (episodic link, black arrow), or the system probes memory consolidated in the neocortex with probability $(1 - q_k)$ (purple dashed arrows). If the hippocampus selects the cortical recall pathway, it may sample any of the stored location memory patterns $\mathbf{y}^r$ with probability $\mathcal{P}_s\left(\mathbf{y} = \mathbf{y}^r \mid \mathbf{x}^k\right)$. To compute this probability, we clamp the field of the hidden units to $\mathbf{x}^k$ (flavour memory pattern), and we let the cortical network do M Gibbs sampling steps to infer the activity of the location units $\hat{\mathbf{u}}$. We then calculate the new activity of the hidden units $\boldsymbol{\xi} = p(\mathbf{h} = 1 \mid \hat{\mathbf{u}})$ and compare it with each field vectors $\mathbf{y}^r$ (location memory patterns) and a null vector $\mathbf{y}^0$, using the consistency measure $C(\mathbf{y}; \boldsymbol{\xi}) = -\frac{1}{\sum_j \xi_j} (\mathbf{y}^r)^\intercal \boldsymbol{\xi}$. The probability $\mathcal{P}_s$ is then derived by taking the softmax. **Step 3**: If the hippocampus samples the null vector $\mathbf{y}^0$, the network keeps the activity $p\left(\hat{\mathbf{u}} \mid \mathbf{x}^k\right)$ reconstructed by the neocortex. Otherwise, the hippocampus reinstates the field $\mathbf{y}^k$ of the hidden units that corresponds to the sampled memory. The neocortex reconstructs the activity in the sensory layer via top-down connections, and we derive the probability to go to each reward location from the activity of the location units.

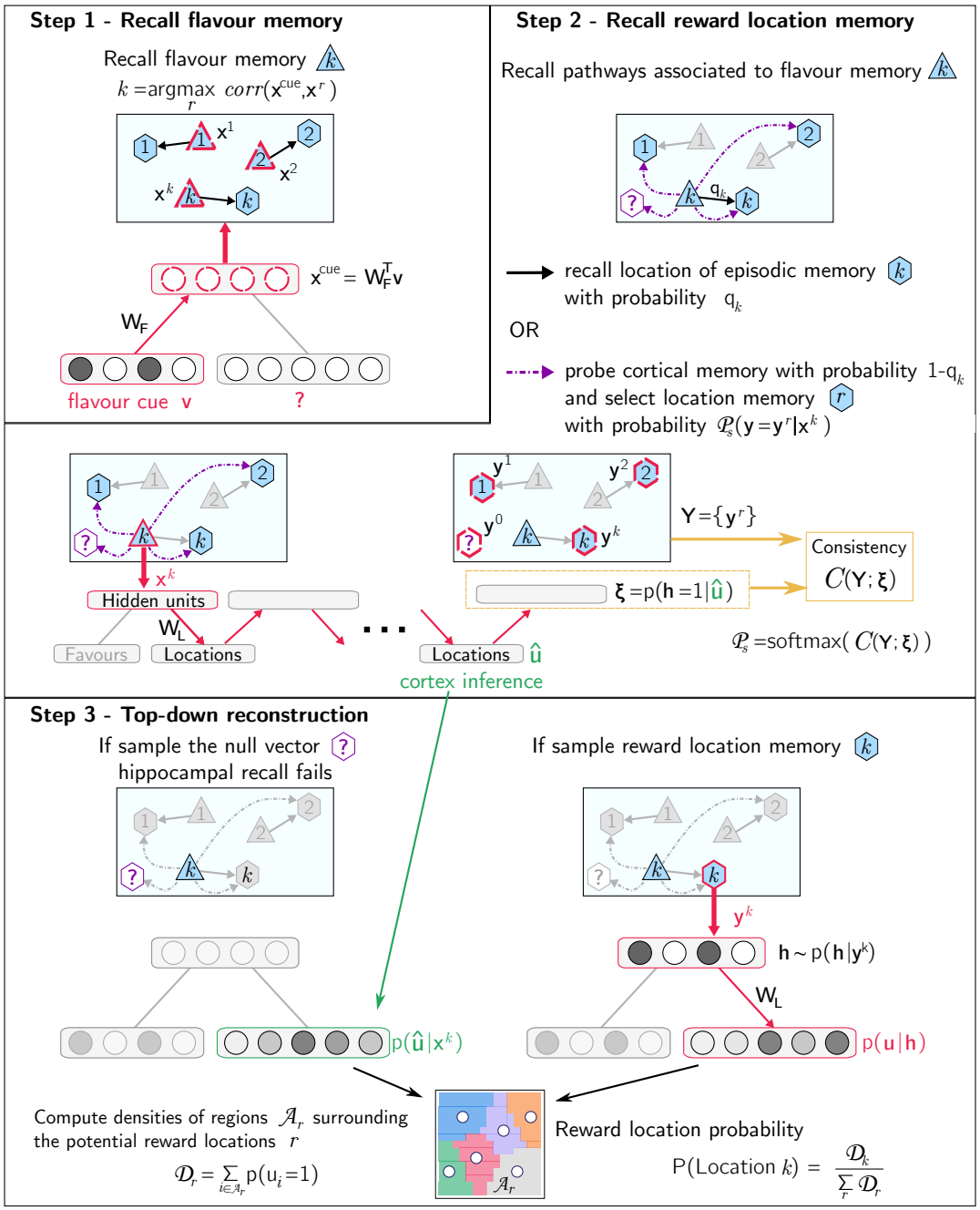


Figure 4.4: Recall mediated by the hippocampus. **(Continued on the following page.)**

Therefore, when the episodic link q_k was strong the hippocampus more likely recalled the episodic memory (direct pathway, black arrow in Fig. 4.4), and when the episodic link q_k was weak the hippocampus more likely probed cortical knowledge to infer the memory corresponding to the cue (indirect pathway, purple dashed arrows in Fig. 4.4). We believe it is useful for the hippocampus to have two pathways for recollection as it could potentially chose one and discard the other. For example, if the animal learn the task in one context, and subsequently are placed in a different context, they might discard episodic recall (direct pathway) and instead rely solely on memory consolidated in the neocortex (indirect pathway; i.e. do not sample q_k but sample $\mathcal{P}_s$ straight away). We have already suggested in Section 3.2 that, due to the context-dependency, the prefrontal cortex might be involved in the selection of the appropriate recall strategy in the hippocampus.

By default, recall was mediated by the hippocampus. To test if the memories were consolidated in the neocortex, we disabled the hippocampus during recall and let the RBM perform 2 Gibbs sampling steps to reconstruct the activity of the location units (see Cortical recall in Section 2.2.1). Note that this cortical recall differs from the "cortical recall" described above (**step 2**) when hippocampal recall failed: in the latter, the activity of the location units was inferred by Gibbs sampling starting from $\mathbf{x}^k$, the flavour memory recalled in the hippocampus, while for cortical recall non-mediated by the hippocampus the activity of the location units was inferred by Gibbs sampling starting directly from $\mathbf{x}^{cue}$.

In the next Section 4.1.5, we further develop on recall mediated by the hippocampus as we explain its role for sleep replay during training.

4.1.5 Network training

The cortical weights were initially set to zero, and the initial value of the meta-schema was $\phi^* = 0.5$ (see justification in the PFC model Section 4.1.2). Similar to the training in Chapter 3, each epoch was divided into an experience phase and a sleep replay phase. At the start of each epoch the hippocampus cleared out all its past memories. Similar to the training in Section 3.3, the network only learned rewarded associations and we limited the number of presentations of the sensory patterns during the experience phase at each epoch. An overview of the training epoch is given Fig. 4.5, and the parameters are in Table 4.2.

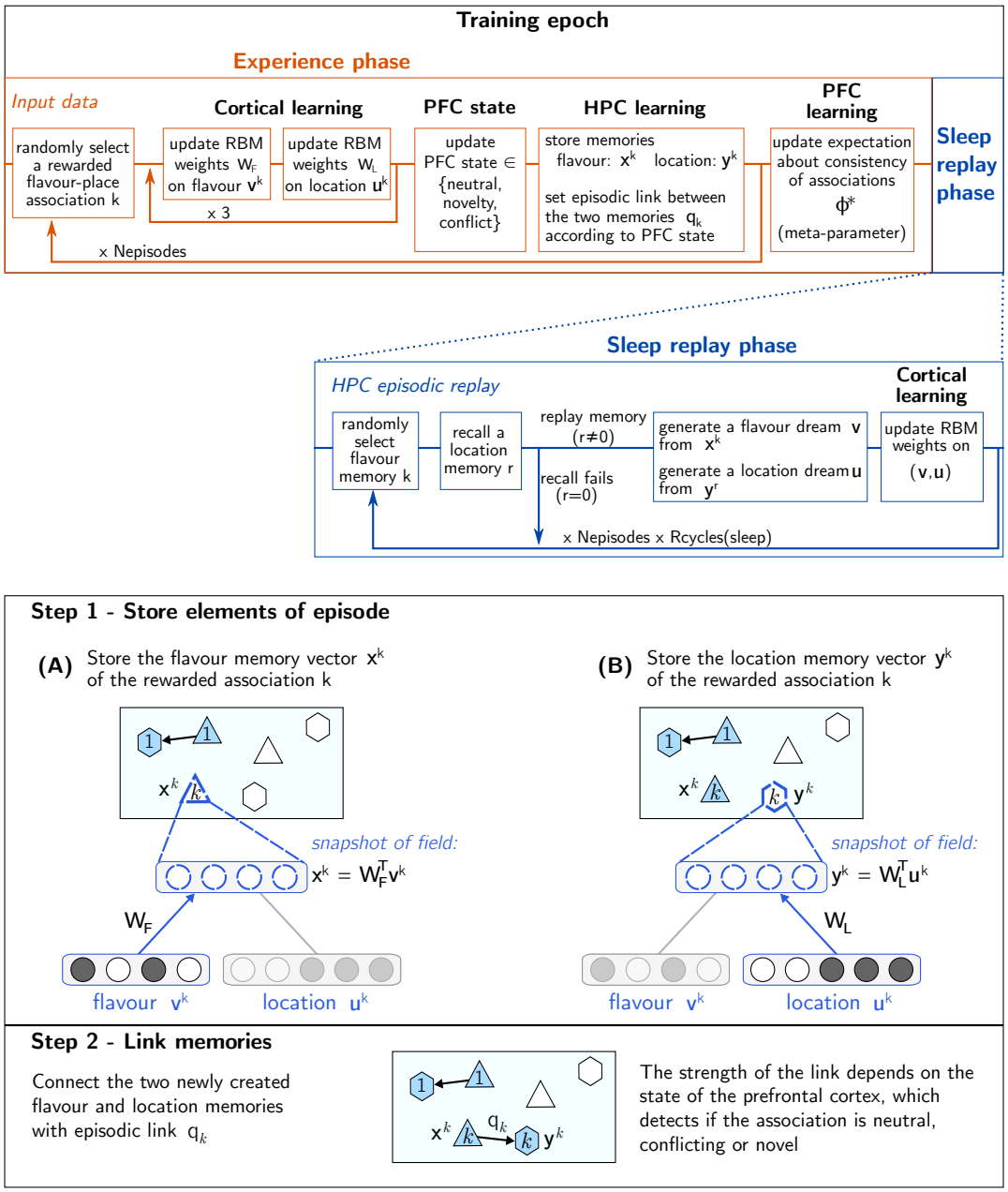


Figure 4.5: (Top panel) Training epoch. During the experience phase, the model sees the rewarded flavour-place associations in a random order. (Lower panel) Hippocampal learning during the experience phase. See Section 2.2.2 for details about cortical learning and Section 2.3.2 for sleep replay.

Training:	Original associations	New associations
reward locations	6	6
memories in HPC	6	2
visible units	325	
hidden units	100	
sleep cycles	6x100	2x100

weights init.		$w = 0$
learning rate	EXP	0.01
	SLEEP	0.001

Table 4.2: Model and training parameters. Notations: HPC = hippocampus; EXP = experience phase; SLEEP = sleep replay phase. The number of memories in HPC indicated is the total number of episodic links created per epoch. This number is then used to calculate the total number of replay events generated during sleep.

1 - Experience phase

The network cycled only once through the rewarded flavour-place associations during the experience phase, but each association was presented three times in a row to the sensory cortex. The order of presentation was random at each epoch.

Say the rewarded association k has been randomly selected (flavour k and reward location k in the arena). We first presented the flavour pattern $\mathbf{v}^k$ to the sensory cortex, and updated the cortical weights $\mathbf{W}_F$ connecting the flavour sensory area to the associative cortex (using CD-1 learning, see method in Section 2.2.2). Next, we presented the reward location pattern $\mathbf{u}^k$ to the sensory cortex, and updated the cortical weights $\mathbf{W}_L$ connecting the location sensory area to the associative cortex.

Following cortical update, we computed the probability ϕ (consistency, Eq. 4.1) of the flavour-place association $\left(\mathbf{v}^k, \mathbf{u}^k\right)$. The prefrontal cortex then compared the probability ϕ of the association with the current model expectation ϕ^* (meta-parameter, see Section 4.1.2), and detected if the association was neutral, novel or conflicting with prior knowledge; the prefrontal cortex set its state accordingly (Algorithm 4.3).

The hippocampus then stored the memory of the association (lower panel, Fig. 4.5). As explained in Section 4.1.3, it stored the field $\mathbf{x}^k$ of the hidden units when the flavour pattern was clamped in the sensory cortical area, and the hippocampus also stored the field $\mathbf{y}^k$ of the hidden units when the location pattern was clamped. The flavour and location memories in the hippocampus had to be connected so as to finalise the formation of the episodic memory. The value of episodic link q_k was set according to the state of the prefrontal cortex (see Section 4.1.3, Table 4.1).

The procedure was repeated for all rewarded flavour-place associations. At the end of the experience phase, the prefrontal cortex adjusted the meta-schema about the consistency of the associations so as to reduce the discrepancy between observation and expectation. It updated the meta-parameter ϕ^* by averaging the probabilities ϕ_k of the rewarded associations, but only considering the associations that were classified either as "neutral" or "conflicting":

$$\phi^* \leftarrow (1-\gamma)\phi^* + \gamma\langle\phi_k\rangle_k \quad \text{for } k \text{ where PFC state} \neq \text{novelty} \tag{4.5}$$

The rate of update was $\gamma = 0.2$. For this update we did not include the associations classified as "novel" because their probabilities ϕ_k were low, and thus if we included them it would wrongly decrease the value of the meta-parameter ϕ^*.

2 - Sleep replay phase

In the current version of the model, the cortical weights for the flavours $\mathbf{W}_F$ and the cortical weights for the locations $\mathbf{W}_L$ were adjusted independently during the experience phase. Hence, the flavours and the reward locations were consolidated, but their associations were not. The sleep replay phase allowed the consolidation of the associations. Importantly, only the associative part of the neocortex (RBM) was trained during sleep replay, whereas hippocampus and prefrontal cortex were not plastic and only supported the replay process.

Cortical learning during sleep was driven by the replay of flavour-place associations in the sensory cortex (Section 2.3.2). We assumed that the replay originated from reactivations in the hippocampus. At each epoch, the number of reactivations in the hippocampus was set to $K \times 100$, where K is the number of episodic memories stored in the hippocampus (memories were randomly selected, hence on average each memory was replayed 100 times).

A replay event started with the random reactivation of a hippocampal flavour memory pattern $\mathbf{x}^k$ (Algorithm 4.4). The hippocampus then recalled a reward location memory pattern $\mathbf{y}^r$ using the same method described Section 4.1.4. Note that we never started a replay event from a location memory stored in the hippocampus, but this could be an extension of the model.

If the hippocampus failed to recall a location memory, no pattern was generated in the sensory layer of the neocortex. In that case the replay event failed and was not replaced with a new recall attempt. Instead, the hippocampus proceeded to reactivate

another flavour memory.

If the hippocampus successfully recalled a location memory, the replay event was used to train the neocortex. The training had two stages. First, the network generated two dream sensory patterns (binary patterns), first a flavour and then a reward location, and for each the corresponding cortical weights were updated. Next, the two dream patterns were clamped simultaneously in the sensory cortex, and the cortical weights were updated again. This last update consolidated the association between flavour and reward location.

Algorithm 4.4 Sleep replay.

See Fig. 4.4 for recall mediated by the hippocampus.

for $t = 1$ to $K \times 100$ **do**

1. Randomly select a flavour memory pattern $\mathbf{x}^k$ in the hippocampus
2. Recall a reward location memory pattern $\mathbf{y}^r$ in the hippocampus

if $r = 0$ **then**

Replay event fails

else

Replay event is successful

2a. Generate a flavour dream $\mathbf{v}$ from $\mathbf{x}^k$, and update $\mathbf{W}_F$

2b. Generate a location dream $\mathbf{u}$ from $\mathbf{y}^r$, and update $\mathbf{W}_L$

2c. Clamp $(\mathbf{v}, \mathbf{u})$ in the sensory cortex, and update $\{\mathbf{W}_F, \mathbf{W}_L\}$

end if

end for

4.2 Results: Interplay between HPC, PFC and associative cortex during the acquisition of a schema

In this first part we ran five simulations for the training in the consistent schema, where the flavour-place associations are fixed, and five simulations for the training in the inconsistent schema, where the associations vary across epochs (Fig. 4.6a). For each run we defined a different set of six random patterns for the flavours. The initial weights of the cortical network (RBM) were all set to zero. The initial value of the meta-parameter was $\phi^* = 0.5$ (no expectation prior to training, see Section 4.1.2). For this part of the simulations, the network was trained for 50 epochs.

4.2.1 Training in a consistent or an inconsistent schema

For the consistent schema, recall performance improved over time (Fig. 4.6b, recall with hippocampus, blue solid line) and the flavour-place associations were gradually consolidated in the neocortex (Fig. 4.6b, recall without hippocampus, blue dashed line). At the end of training, the network successfully reconstructed the activity of the location units in the sensory cortex, where each rewarded location matched the corresponding flavour cue (Fig. 4.7a,b, left panels).

Performance in the inconsistent schema increased at a lower pace than performance in the consistent schema (Fig. 4.6b, recall with hippocampus, red solid line). Performance was measured as the probability to recall, for a given flavour, the last rewarded location. At the end of training, there was a larger variance of recall performance across the six associations compared to performance in the consistent schema (black dots in Fig. 4.7a, right panel). Indeed, the hippocampus correctly recalled some associations, but for others recall alternated between two locations, (i) the correct location and (ii) the consolidated location (Fig. 4.7b, upper right panel). The alternative behaviour arised because of the two recall pathways of hippocampal recall, direct episodic recall and indirect cortical-based recall (refer to Fig. 4.4). For example, for the flavours no.2, 3, 5 and 6, recall did not alternate because when the hippocampus probed cortical memory (indirect recall pathway), the distribution $\mathcal{P}_s$ over the stored location memories was uniform (Fig. 4.7d, right panel); this means that these flavours were not consolidated to any location, which is also illustrated by the reconstructions obtained when the hippocampus was disabled during recall (Fig. 4.7b, lower right panel). Note that performances (with hippocampus) for the flavours no.2, 5 and 6, were higher than

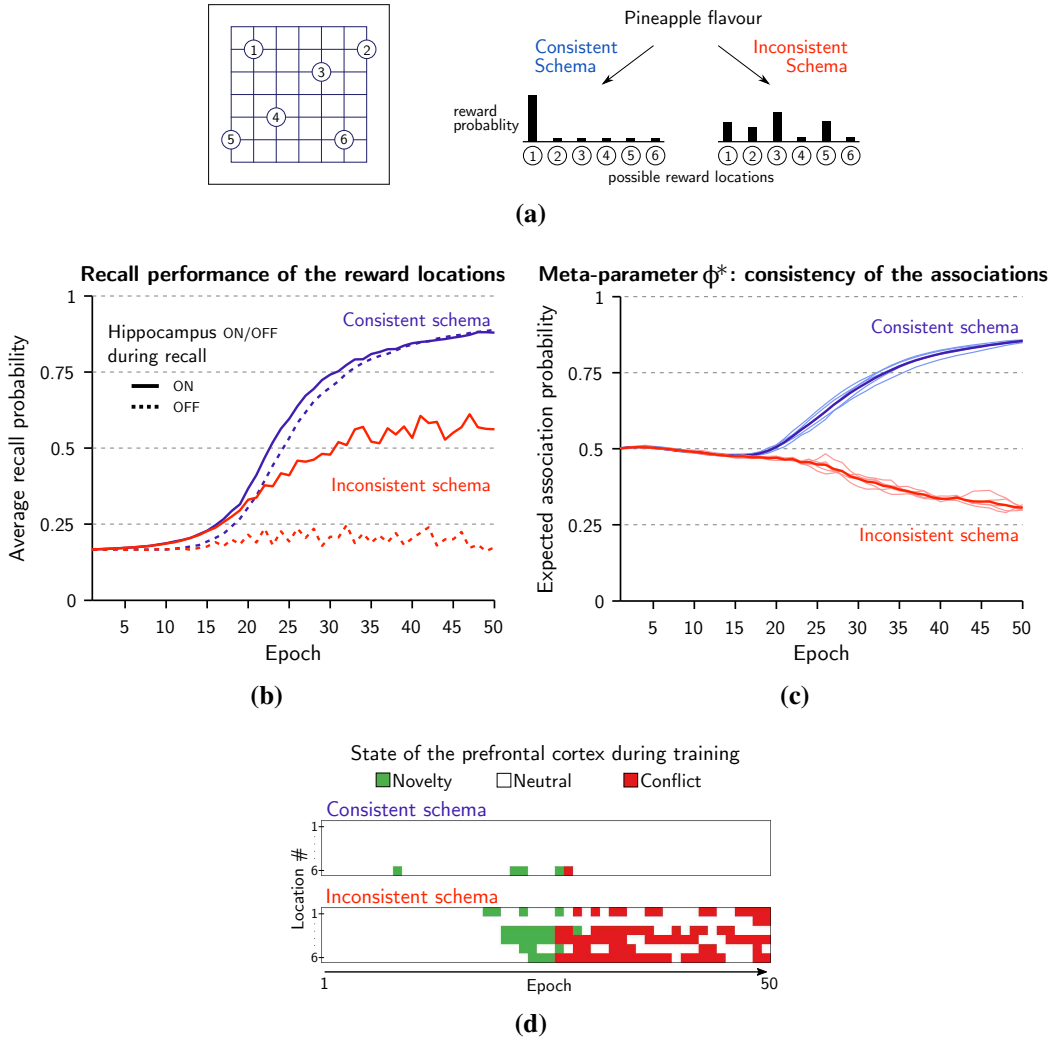


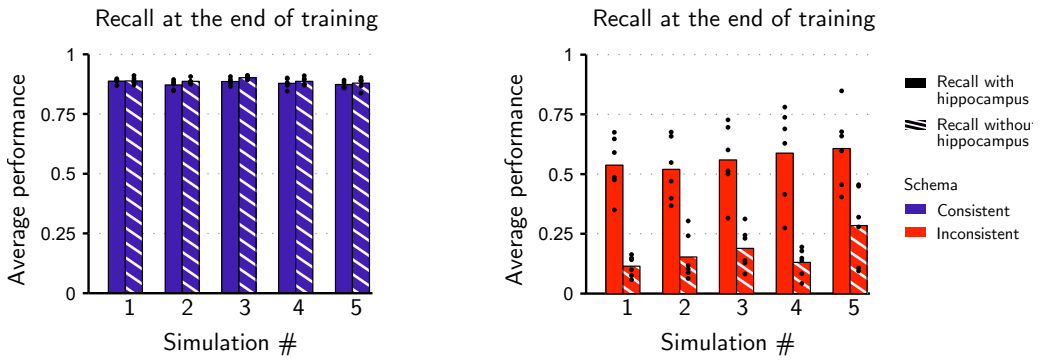
Figure 4.6: Original training in a consistent schema, where the associations are fixed, and an inconsistent schema, where the associations are shuffled every two epochs. (a) Example of associations in the two schemas. (b) Recall performance for the consistent schema (blue), and the inconsistent schema (red). Solid lines: recall with hippocampus intact; dashed lines: hippocampus blocked during recall only, to evaluate consolidated memory. (c) Acquisition of the meta-schema. The prefrontal cortex updates the expectation about the association probability between flavours and reward locations. This expectation is represented by the meta-parameter ϕ^*. The thin lines show the value of the meta-parameter for the different simulations. (d) State of the prefrontal cortex when the associations are presented at each training epoch. Each panel shows an example of simulation in the consistent and in the inconsistent schemas. Each row represents a flavour-place association, and the rows are ordered according to the location number. Note that for the training in the consistent schema we picked the result of a simulation where the prefrontal cortex did find novelty and conflict, but for the majority of the simulations the prefrontal cortex was always in the neutral state.

performance for flavour no.3; this is explained by the direct recall probability (episodic q_k), which was lower for the flavour no.3 and hence the hippocampus more likely probed cortical memory during recall (Fig. 4.7d, left panel). On the other hand, when cued with the flavours no.1 or no.4, the network reinstated both locations no.2 and no.5. The location no.2 (resp. no.5) corresponded to the correct reward location for flavour no.1 (resp. no.4) in the last epoch, and was recalled with a probability q_k (Fig. 4.7d, left panel). Since $q_k < 1$, the hippocampus had a probability $1 - q_k$ to probe cortical memory during recall. In contrast with flavours no.2, 3, 5 and 6, the distribution $\mathcal{P}_s$ over the stored location memories was peaked at one location for he flavours no.1 or no.4 (Fig. 4.7d, right panel); this means that these two flavours were consolidated with the two particular locations, which is also illustrated by the reconstructions obtained when the hippocampus was disabled during recall (Fig. 4.7b, lower right panel). In the case of flavour no.1, the episodic link q_k was larger than 0.5 and hence the network more likely recalled the correct location, while in the case of flavour no.4, the episodic link q_k was smaller than 0.5 and hence the network more likely recalled the incorrect, consolidated location.

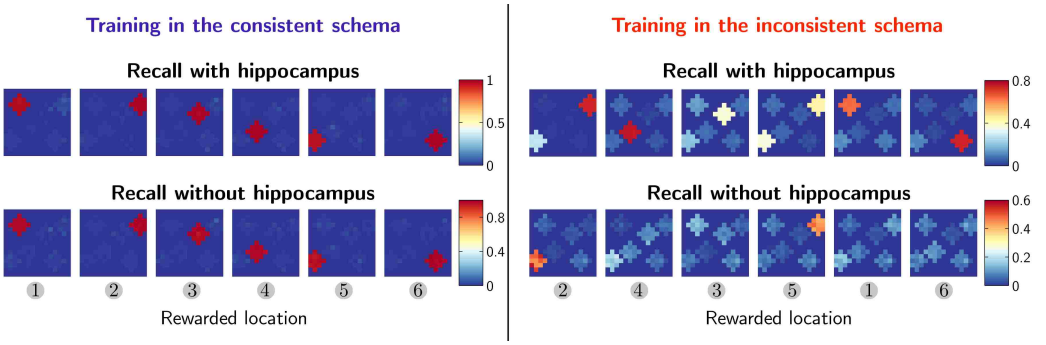
Despite the increase in hippocampal performance, semantic memory in the inconsistent schema performed poorly (Fig. 4.6b, recall without hippocampus, red dashed line). Indeed the associations were changed every two epochs and hence the network constantly overwrote the associations during sleep replay consolidation. Therefore, if the direct, episodic recall pathway in the hippocampus was blocked (black arrow in Fig. 4.4), performance would drop and would be similar to the recall performance without hippocampus.

By contrast, for the training in the consistent schema, all episodic links q_k were high (Fig. 4.7c, left panel), which means that the hippocampus almost always recalled location memories via the direct, episodic pathway. However, even if this recall pathway was blocked, the result would be the same since the associations were consolidated in the neocortex (Fig. 4.7b, left panel).

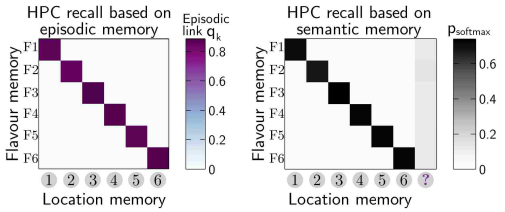
At each epoch, at the end of the experience phase, the prefrontal cortex averaged the probabilities ϕ (Eq. 4.1) of the associations encountered and updated the meta-parameter ϕ^* (Eq. 4.5). The meta-parameter monitored the consistency of the flavour-place associations across epochs. For the consistent schema, the value of the meta-parameter increased (Fig. 4.6c, blue lines), as the flavour-place associations were gradually consolidated in the neocortex. For the inconsistent schema, however, the meta-parameter decreased over time (Fig. 4.6c, red lines), as the flavour-place associ-



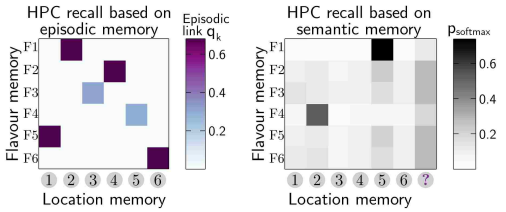
(a)



(b)



(c) Consistent schema.



(d) Inconsistent schema.

Figure 4.7: Recall of the six original flavour-place associations at the end of training in the two schemas. Except for (a), all figures show the results of one simulation run. (a) Recall probability of the reward locations for the five training simulations in the consistent schema (left) and inconsistent schema (right). The dots represent individual performances for the six associations. (b) Reconstruction of the activity of the location units in the sensory cortex (activities averaged over 100 trials). The six insets in each panel show the reconstruction for the six trained flavour-place associations, sorted by flavour tested. The correct reward locations are indicated below the insets (for the inconsistent schema, they correspond to the rewards allocation of the last epoch). (c,d) Hippocampal (HPC) recall probabilities (**step 2** in Fig. 4.4). When the hippocampus recalls a flavour memory, the associated location memory is recalled with a probability q_k (episodic link, left panel). If direct recall fails (probability $1-q_k$), the hippocampus probes cortical memory to compute the probability of each stored location memory (softmax, right panel; $\mathcal{P}_s$ in Fig. 4.4). The last column (?) is the 7th option in the softmax. When selected, the hippocampus does not reinstate a location memory and instead let the cortex infer the activity of the location units ($p(\hat{\mathbf{u}})$ in Fig. 4.4).

ations were, on average, only weakly consolidated in the neocortex.

We anticipated hippocampal performance to increase in the inconsistent schema because the neocortex learned over time the flavour and reward location patterns, and hence the quality of the reconstructions in the sensory cortex improved. However, performance was higher than expected, because the correct locations were recalled with high probability in the hippocampus (episodic links q_k, in Fig. 4.7d), while we expected these to be lower. Indeed, we expected the prefrontal cortex to be in the neutral state during the training of the six original associations in both consistent and inconsistent schema, because the meta-parameter ϕ^* (Eq. 4.5) is supposed to grow (or decrease) as the associations get progressively consolidated (or not) and the associations probabilities ϕ increase (resp. decrease). Thus if the prefrontal cortex was in the neutral state, the episodic link should have been set to $q_k = \phi$, where ϕ is the probability of the observed association; since we have seen earlier that the associations were not consolidated in the inconsistent schema, the probabilities q_k should have had low values (similar to the meta-parameter ϕ^*, Fig. 4.6c, red line). Therefore we hypothesised that the prefrontal might have been in a different state during the training in the inconsistent schema.

To investigate this behaviour, we monitored the state of the prefrontal cortex over training. While in the consistent schema the prefrontal cortex was, as expected, in the neutral state over training, for the training in the inconsistent schema the prefrontal cortex switched to the conflict state mid-training for some associations (Fig. 4.6d). This result explains why the episodic link had high value for training in the inconsistent schema: when the prefrontal cortex was in the conflict state, the episodic links of the associations were set to $q_k = 1 - \phi^*$ (Table 4.1), but the meta-parameter ϕ^* was low for the inconsistent schema (Fig. 4.6c, red lines).

Why did the network detect conflict in the inconsistent schema? Say flavour PINEAPPLE was associated with a reward at location no.4, and the probability ϕ_{P4} of the association had a value close to the meta-parameter ϕ^*. At the beginning of training this value was about 0.5, which means that flavour PINEAPPLE had more chance of being replayed with location no.4 than with any other location. Yet, after two epochs the associations were changed. Thus, flavour PINEAPPLE was then associated with a different location, say no.1. Hence, the new association had low probability ϕ_{P1} that was lower than the expectation set by the meta-parameter ϕ^*, but since $\hat{\phi} = \phi_{P4}$ was similar to ϕ^* the prefrontal cortex detected conflict. As an illustration, the process would look like to the "pink bird" example in Fig. 4.3: when the meta-parameter ϕ^* is low,

the episodic link q_k is high. Consequently, the new association will be more likely replayed during subsequent sleep, which will decrease ϕ_{P4} and increase ϕ_{P1}; but as the new association is consolidated, ϕ_{P1} might end up higher than the meta-parameter ϕ^*, and thus it might lead to conflict detection later when the associations are changed again. Hence we believe the conflict detection occurred because of the consolidation of associations during sleep replay.

4.2.2 Impact of blocking the PFC modulation of memory formation in the HPC

We have seen at the end of Section 4.2.1 that, in the case of the inconsistent schema, the prefrontal cortex detected conflicting associations over the course of training instead of remaining in the neutral state as we anticipated. By contrast, we did not observe such behaviour during the training in the consistent schema (Fig. 4.6d). Since the state of the prefrontal cortex modulated memory formation in the hippocampus during learning, here we investigate what would happen if we interfered with this modulation. In practice, it means that during **step 2** of hippocampal learning ("link memories" in Fig. 4.5), the strength q_k of the episodic link between the flavour memory and the location memory was set to ϕ, the probability of the association measured by the neocortex (Eq. 4.1), regardless of the state of the prefrontal cortex (while previously q_k was set according to Table 4.1). If we look at the examples in Fig. 4.3, this operation will not affect the outcome for the "neutral" associations, because in this case we already had $q_k = \phi$. This operation will not affect either the case of "pink elephant in the savanna", because $\phi \sim (1 - \phi^*)$, but it will affect the outcome for "pink bird in the savanna", since the episodic link will be set to a lower value than the value they would normally have after modulation by the prefrontal cortex. Note that during training the prefrontal cortex still monitored the consistency of the associations with prior knowledge.

For the consistent schema, blocking the modulation during the creation of the episodic link in the hippocampus had no impact on the performance (Fig. 4.8a&b, left panels, orange lines). This was expected, since we found in the control condition that the prefrontal cortex was almost always in the neutral state. This result is equivalent to our example of the "grey elephant in the savanna" in Fig. 4.3 ($\phi \sim \phi^*$ and $q_k = \phi$).

On the other hand, the performance in the inconsistent schema was impaired when the strength of the episodic link was not modulated by the state of the prefrontal cortex (Fig. 4.8a, right panel, orange line). Hippocampal recall randomly alternated be-

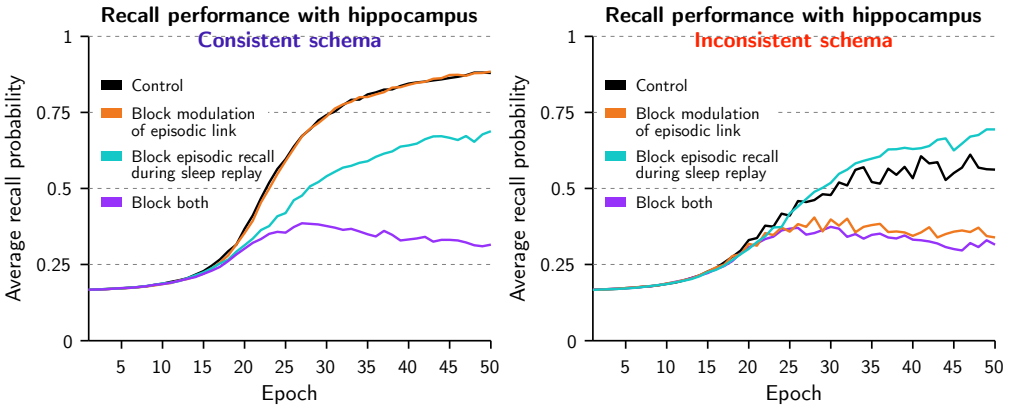


Figure 4.8: Impact on training in the consistent (left) and inconsistent schema (right) of: 1) blocking PFC modulation of the episodic link q_k during memory formation in the hippocampus (orange lines), 2) blocking hippocampal recall via the direct, episodic pathway (black arrow in Fig. 4.4) during sleep replay (light blue lines), or 3) blocking both (purple lines). The control condition (black lines) corresponds to the results shown in Fig. 4.6b.

tween the episodic memory of the last rewarded location, and the consolidated location (Fig. 4.9c, left panel). We noticed some of this effect already in the control condition, i.e. with prefrontal interaction intact (Fig. 4.7b, top right panel), but in the current condition the hippocampus mostly recalled the consolidated location. Indeed, the location with higher probability matched the location recalled without hippocampus (Fig. 4.9c, right panel), whereas the location with lower probability matched the correct location, indicated by the number below each inset.

To understand this result we looked at hippocampal recall. We found that most episodic links q_k in the hippocampus were lower than 0.5 (Fig. 4.9e, left panel), in contrast with the control condition where most were higher than 0.5 (Fig. 4.7d, left

panel). As a result, in the current simulations the hippocampus most likely probed cortical memory (probability $1-q_k$) during recall rather than recalling the recent episodic memory; however, for four out of five simulations, the probability distribution $\mathcal{P}_s$ over the location memories stored in the hippocampus was peaked at one unique location memory (Fig. 4.9e, right panel). This location memory corresponded to the location consolidated (Fig. 4.9c, right panel). Accordingly, the episodic link of the association that involved this consolidated location memory was higher than the others (Fig. 4.9e, left panel; in this example, flavour no.5 was, by chance, associated with location no.4).

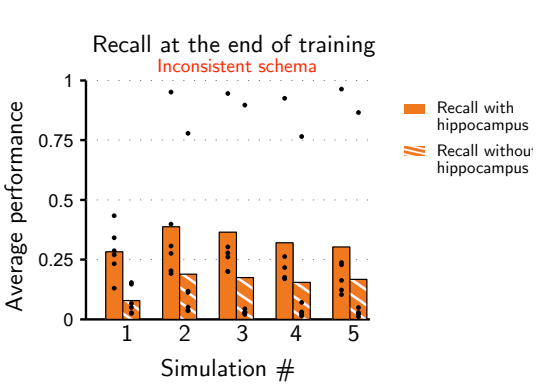
Therefore, for the majority of the simulations, we found that the network reinstated one rewarded location, regardless of the flavour tested (which location was dominantly consolidated was random across the different simulations).

Note about the consolidation in the inconsistent schema. To investigate why the network recalled only one reward location, we first checked whether the bias in recall was a consequence of the network having learned only one location. We tested memory recall with a random cue and found that the network reinstated all locations, suggesting that all locations were learned more or less equally (result not shown here, but this will be illustrated when we introduce the new associations in Section 4.3). Despite this, as we mentioned earlier, when the hippocampus probed cortical memory during recall the resulting probability distribution $\mathcal{P}_s$ over the location memories was peaked at one unique location memory (Fig. 4.9e, right panel), which implies that all flavours were consolidated with the same location. Indeed, if the flavours were not consolidated with any specific location, we would have observed that hippocampal recall had failed (option marked as (?) in Fig. 4.9e, right panel; refer to recall method in Fig. 4.4).

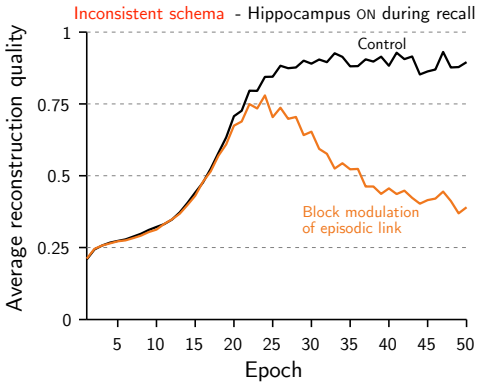
Next, we checked whether the consolidated location varied across training epochs. We monitored the state of the prefrontal cortex over time (Fig. 4.9d) and found that it switched to conflict mode for most associations presented. However, for one location the prefrontal cortex was always in the neutral state (for instance the location no.4 in Fig. 4.9b, top panel, corresponding to the simulation shown in panel (c) in the same figure). This means that, every epoch, any flavour presented with this specific location did not trigger conflict detection in the prefrontal cortex. This result implies that the consolidated location observed at the end of training was the same all the way.

In order to get further insight why such bias in recall occurred, we looked at the reconstruction quality[3] (Eq. 3.7) during recall with hippocampus. This measure reflects

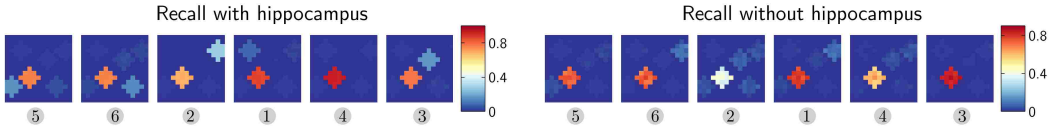
[3] We have introduced the reconstruction quality in Section 3.3.2 of Chapter 3 when we investigated the impact of cortical learning rate on the quality of the recalled sensory activations.



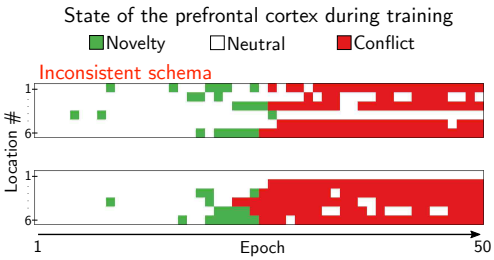
(a)



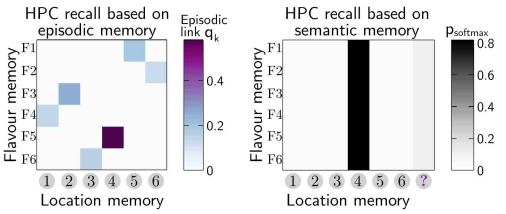
(b) Reconstruction quality.



(c) Recall in the inconsistent schema (simulation #4).



(d) State of the PFC (sim #4&5).



(e) Hippocampal recall (simulation #4).

Figure 4.9: Training in the inconsistent schema when we block modulation of the episodic link q_k by the prefrontal cortex during memory formation in the hippocampus. (a) Recall performance, at the end of training, for five simulations. The dots represent the individual performance of the six flavour-place associations. Notice that for four out of five simulations, one association has high recall performance while the others are low. (b) Quality of the reconstruction of the activity of the location units during recall over time. (c) Reconstruction of the activity of the location units when recall is mediated by the hippocampus (left) or or not (right). Notice that one location is preferably reinstated, which explains the performance shown in panel (a). (d) State of the prefrontal cortex when the associations are presented at each training epoch. Each panel shows an example of simulation. Each row represents a flavour-place association, and the rows are ordered according to the location number. (e) Hippocampal (HPC) recall probabilities (**step 2** in Fig. 4.4). When the hippocampus recalls a flavour memory, the associated location memory is recalled with a probability q_k (episodic link, left panel). If direct recall fails (probability $1-q_k$), the hippocampus probes cortical memory to compute the probability of each stored location memory (softmax, right panel; $\mathcal{P}_s$ in Fig. 4.4). The last column (?) is the 7th option in the softmax. When selected, the hippocampus does not reinstate a location memory and instead let the cortex infer the activity of the location units ($p(\hat{\mathbf{u}})$ in Fig. 4.4)

the correlation between the true sensory patterns and reconstructed ones; its evolution over training epochs is shown Fig. 4.9b. We noticed that the recall quality initially increased (orange curve) similar to the control condition (black curve). Yet, the recall quality started to decline mid-training while the reconstruction quality kept on increasing for the control condition. This suggests that mid-training, one location took over and, as it became more and more consolidated with all the flavours, the recall quality decreased over time.

To understand what led to the consolidation of a unique location, we need to go back to the example explained earlier in Section 4.2.1, when the network was trained in control conditions in the inconsistent schema. We saw that when the associations were swapped the prefrontal cortex detected conflict. We also saw that when flavour PINEAPPLE was suddenly associated with location no.1 instead of location no.4, with modulation by the prefrontal cortex the new memory in the hippocampus had a high episodic link (e.g. "pink bird" in Fig. 4.3). By contrast, the episodic link was here set to $q_k = \phi_{P1}$. But ϕ_{P1} was low, which means that when the memory of flavour PINEAPPLE was reactivated during subsequent sleep, the new location only had a low probability of being recalled, and hence the new association was less replayed and consequently it never got a chance to be consolidated. Instead, when the memory of flavour PINEAPPLE was reactivated, the hippocampus probed cortical memory with high probability $1 - q_k$, and thus recalled the initial location no.4 because the recall probability $\mathcal{P}_s$ was peaked at this location memory. Thus, the most consolidated association was again replayed, strengthened, and hence its probability ϕ_{P4} further increased. In other words, the "bird" will never turn pink in Fig. 4.3.

In the meantime, say location no.4 was then associated to flavour CHAMPAGNE. Say this flavour was not preferably consolidated to any particular location, and the episodic link $q_r = \phi_{C4}$ was lower than 0.5. Thus, like for the flavour PINEAPPLE, the hippocampus most likely probed cortical memory during recall during sleep replay. Since flavour CHAMPAGNE was not consolidated to a location, the recall probability $\mathcal{P}_s$ should have been uniform over the location memories in the hippocampus. However, we believe that as the flavour PINEAPPLE was more and more replayed with its preferred location no.4, it created a bias in memory retrieval for the flavour CHAMPAGNE when the hippocampus probed cortical memory (probability $\mathcal{P}_s$), leading to the memory of location no.4 being more sampled than the other location memories in the hippocampus. Therefore, location no.4 was often replayed with flavour CHAMPAGNE since it was predominantly recalled via both recall pathways in the hippocampus.

This created a snowball effect sustained by sleep replay, leading to the consolidation of a single location with all the flavours.

To conclude, blocking the hippocampal-prefrontal cortex interaction during the formation of memories in the hippocampus did not affect training in the consistent schema because there was no conflict detected during learning, and hence no need for modulating the episodic link (i.e. the grey elephant was already grey). On the other hand, blocking this interaction during training in the inconsistent schema did impair performance: when the modulation of the episodic link in the hippocampus was not allowed, memories were not correctly recalled during subsequent sleep and hence the network consolidated the wrong memories, which also impacted hippocampal recall.

4.2.3 Impact of blocking episodic memory recall in the hippocampus during sleep replay

In Section 4.2.2 above we saw that blocking the modulation of the episodic link did not impair performance in the consistent schema, but impaired performance in the inconsistent schema as memories were not correctly recalled during sleep. Furthermore, in most simulations in the inconsistent schema we found that one location was consolidated with all flavours, at the cost of the other locations. We suspected that if we suppressed the episodic link recall pathway during sleep replay for the training in the consistent schema we would find similar results. With this manipulation the hippocampus automatically recalled a location memory by probing cortical memory (see Fig. 4.10), and hence hippocampal replay never reflected experience but instead was based on prior knowledge. Since the cortical weights were initially set to zero, the neocortex had no prior knowledge and consequently the hippocampus should replay random combinations of flavour-place associations during sleep. We thus wondered whether we would observe a bias during the consolidation process (i.e. one location memory taking over the others), similar to the training in the inconsistent schema (Fig. 4.9c) or whether all locations would be equally consolidated with all flavours.

In the consistent schema, since the network replayed random flavour-location combinations we anticipated that consolidation would be impaired. We already reported this result in Section 3.2.2 of Chapter 3, where we prevented consolidation by blocking the replay of the rewarded associations (Fig. 3.7b). However, we also found that when the hippocampus mediated recall, performance was not greatly impaired by this procedure (Fig. 3.7a). Therefore we wanted to check whether the network would have

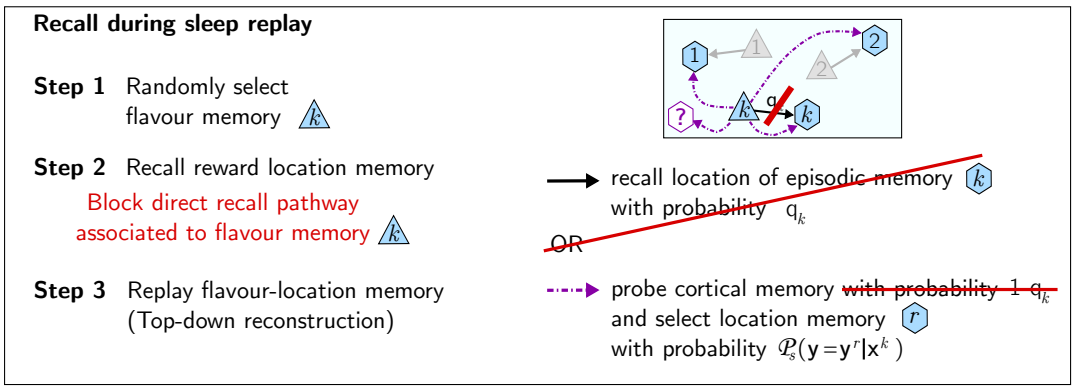


Figure 4.10: Blocking episodic memory recall in the hippocampus during sleep replay. Normal recall is described Fig. 4.4.

such behaviour in the current model.

We must stress that the two types of training - consistent and inconsistent - were equivalent when we blocked episodic-based recall during sleep replay. Indeed, in the control condition the difference between training in the consistent or inconsistent schemas was the replay of experiences during sleep that allowed consolidation of the associations in the consistent, but not in the inconsistent schema. This consolidation impacted the associations probabilities ϕ, and thus the update of the meta-parameter ϕ^* (Eq. 4.5), which in turn modulated the episodic link q_k during memory formation in the hippocampus. By contrast, if we block the replay of experiences it should prevent consolidation, and hence the meta-parameter ϕ^* should be the same in consistent and inconsistent training. Thus, the modulation of the episodic link q_k would also be the same, and as a result we expected similar hippocampal recall performance in the two schemas.

We found indeed similar recall performance for consistent and inconsistent schemas, as is illustrated in Fig. 4.8a&b (light blue lines). It is interesting to note that when we blocked the replay of experiences during sleep the network trained in the inconsistent schema performed best in this condition.

As expected, consolidation failed during training in the consistent schema (Fig. 4.8b, left panel, light blue dashed line). Indeed, as explained earlier, the consolidation of flavour-place associations only occurred during sleep replay, and since replay was now independent of the experience, the system never had the chance to correct the cortical knowledge over training. We noticed that the neocortex had consolidated a unique location at the end of training (Fig. 4.11c, right panel), similar to what we found for the inconsistent schema when the interaction with the prefrontal cortex was blocked dur-

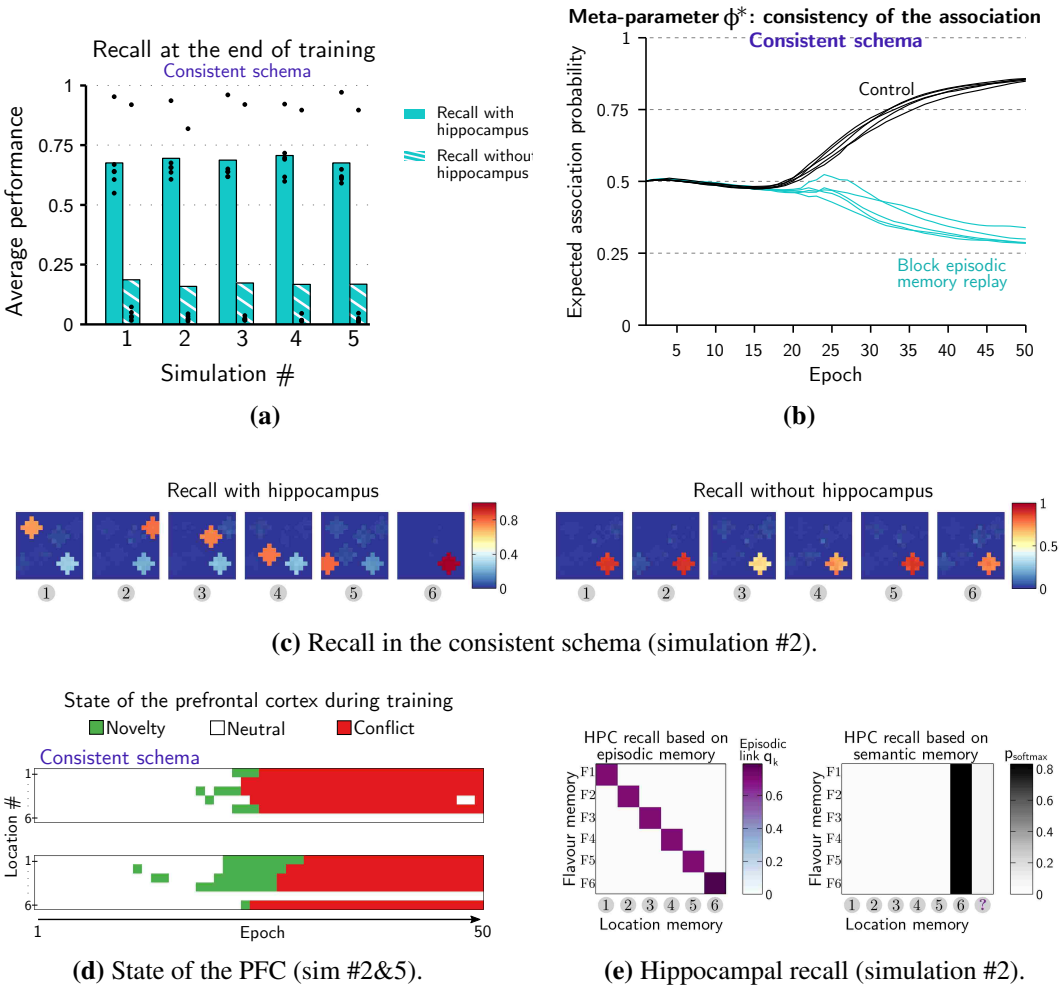


Figure 4.11: Training in the consistent schema when we block the hippocampal recall via the direct, episodic pathway during sleep replay (suppress episodic link q_k in Fig. 4.4). (a) Recall performance, at the end of training, for five simulations. The dots represent the individual performance of the six flavour-place associations. Notice that for all five simulations, one association has higher recall performance than the others. (b) Evolution of the meta-parameter ϕ^* extracted by the prefrontal cortex (expectation about the association probability between flavours and reward locations). The thin lines show the results for the different simulations. (c) Reconstruction of the activity of the location units when recall is mediated by the hippocampus (left) or or not (right). (d) State of the prefrontal cortex when the associations are presented at each training epoch. Each panel shows an example of simulation. Each row represents a flavour-place association, and the rows are ordered according to the location number. (e) Hippocampal (HPC) recall probabilities (**step 2** in Fig. 4.4). When the hippocampus recalls a flavour memory, the associated location memory is recalled with a probability q_k (episodic link, left panel). If direct recall fails (probability $1-q_k$), the hippocampus probes cortical memory to compute the probability of each stored location memory (softmax, right panel; $\mathcal{P}_s$ in Fig. 4.4). The last column (?) is the 7th option in the softmax. When selected, the hippocampus does not reinstate a location memory and instead let the cortex infer the activity of the location units ($p(\hat{\mathbf{u}})$ in Fig. 4.4)

ing the formation of episodic memories (Fig. 4.9c, right panel). The bias in recall was again due to a snowball effect during sleep replay: sleep replay was based on semantic recall, the more the associations were replayed the more they were consolidated, and the more they were consolidated the more they were recalled during replay. Eventually, one location took over the others. Since the neocortex did not consolidate the correct associations, the meta-parameter ϕ^* decreased as we predicted in the consistent schema (Fig. 4.11b, light blue lines).

In line with the results in Section 3.2.2, recall performance with the hippocampus was less impaired by the current manipulation (Fig. 4.8a, left panel, light blue solid line) than recall without hippocampus. The hippocampus mostly reinstated the correct location, but also reinstated the consolidated location (Fig. 4.11c, left panel, the consolidated location is the no.6 in the example displayed). We noticed that the prefrontal cortex switched to the conflict state for all but one association, which corresponded to the association with the consolidated location (Fig. 4.11d). This result is similar to what we observed in the inconsistent schema, when the network also consolidated a unique location (Figure 4.9d). However, in the current case the modulation of the episodic link by the prefrontal cortex was allowed, and thus for the five associations detected as conflicting this link was set to $q_k = 1 - \phi^*$ (see Table 4.1). Since ϕ^* was decreasing over time (Fig. 4.11b), q_k was increasing and thus the memories could be recalled by the hippocampus. In addition, the association corresponding to the consolidated location was labeled as neutral, and as such $q_k = \phi$, where the probability ϕ of this association was high, and hence this memory could also be recalled by the hippocampus. The values of the episodic links are shown for one simulation example Fig. 4.11e, left panel.

Thus, performance when recall was mediated by the hippocampus, which relied on the episodic link q_k of each association, appeared to be preserved thanks to the modulation by the prefrontal cortex during memory formation in the hippocampus. Indeed, without this modulation, the episodic links of the five associations that were detected as conflicting would be $q_k = \phi$, where the probability ϕ of each association was low. We confirmed this by blocking the modulation by the prefrontal cortex, and found that performance dropped as expected (Fig. 4.8a, left panel, purple solid line).

To conclude, as we predicted the consolidation of the associations in the consistent schema was impossible if we blocked the recall pathway in the hippocampus that allowed the replay of episodic memories during sleep. However, instead of replaying randomly each location with all the flavours, recall eventually became biased towards

one specific location memory mid-training, and hence this location was wrongly consolidated with all flavour memories. The network could not correct this bias despite the signal from the prefrontal cortex, which detected that associations were conflicting, because replay was independent of experience in the current simulations. On the other hand, performance when recall was mediated by the hippocampus was preserved, albeit it was lower than control performance.

The effect on hippocampal performance mirrors the results reported with the toy model when we blocked episodic memory replay (Section 3.2.2), but is interesting because this time the hippocampus did not need to explicitly store and update reward probabilities. Furthermore, in Section 3.2.2 we suggested that the model could decide to block episodic memory replay so as to prevent the consolidation of specific associations which might not be that relevant in the future. However, in the current model it seems better to simply not replay these associations during sleep, at least when the network has no prior knowledge, as otherwise recall during sleep becomes biased and leads to the replay and consolidation of false memories.

4.3 Results: Interplay between HPC, PFC and associative cortex during the acquisition of new associations

In Section 4.2 above we have trained our model to learn six flavour-place associations. These associations were either consistent or inconsistent across training. Such distinction is important as Tse et al. (2007) have shown that rodents could learn quickly two new associations in the consistent schema setting, but not in the inconsistent schema setting. However, in previous simulations in Section 3.2 our first model could not make this distinction, as the learning speed of new associations was identical in the consistent and inconsistent schemas (Fig. 3.9). Thus, we suggested that the prefrontal cortex regulated the learning in the hippocampus so as to influence the speed of acquisition of the new associations as a function of the prior schema. Yet, in our first model there was no connection between prefrontal cortex and hippocampus to allow such operation. By contrast, in the current model there is a direct link between the consistency of new information (neutral, conflicting or novel), the prior schema (meta-parameter ϕ^*) and memory formation in the hippocampus (Table 4.1). Therefore in this section we investigate how the current model learns the new associations with either a consistent

or inconsistent prior schema, and whether it can account for the experimental findings of Tse et al. (2007).

4.3.1 Acquisition of new associations with a consistent or inconsistent prior schema

We initialised the model using the parameters obtained in previous simulations of the original training (5 networks). We generated 5 sets of two random patterns to represent the two new flavours, and we trained each network on these sets (25 simulations in total). The training protocol was identical to the original training, but only the new associations were presented to the network (see training epoch Fig. 4.5). In addition, the hippocampus was cleared out before training and hence it only contained the memories of the two new flavours and the two new locations at each epoch. This also means that only the new memories could be replayed during sleep. Accordingly, we set the total number of reactivations during sleep to 200 (as opposed to 600 for the original training), so that each flavour memory was reactivated on average 100 times in the hippocampus (same as original training).

The network learned more rapidly the two new associations when it had been previously trained in the consistent schema than when trained in the inconsistent schema (Fig. 4.12a). From the start of the new training, recall performance with hippocampus was high in the consistent schema, while it was at chance level in the inconsistent schema (Fig. 4.12a, solid lines). Consolidation was also faster with the consistent prior schema than with the inconsistent prior schema (Fig. 4.12a, dashed lines).

The prefrontal cortex detected the novelty of the associations at the start of training for all simulations in the consistent schema, and for the majority of the simulations in the inconsistent schema (Fig. 4.13a, green line = % novelty detected). Indeed, the probability ϕ (Algorithm 4.1) of each new association was low, and it was lower than the expected association probability set by the meta-parameter ϕ^* in both consistent and inconsistent schemas (Fig. 4.13b, **epoch 1** column). The network thus detected a uprising event. Furthermore, to compute the predicted association probability $\hat{\phi}$ of the new flavours, the network had to infer the patterns of activity for the location units (Algorithm 4.2); yet, the flavour patterns of activity were new and consequently the network reconstructed improbable activation patterns (for illustration, similar to the reconstructions shown in Fig. 4.19, Epoch 1, recall without hippocampus). Thus, the predicted probability $\hat{\phi}$ was even lower than the observed probability ϕ Fig. 4.13b,

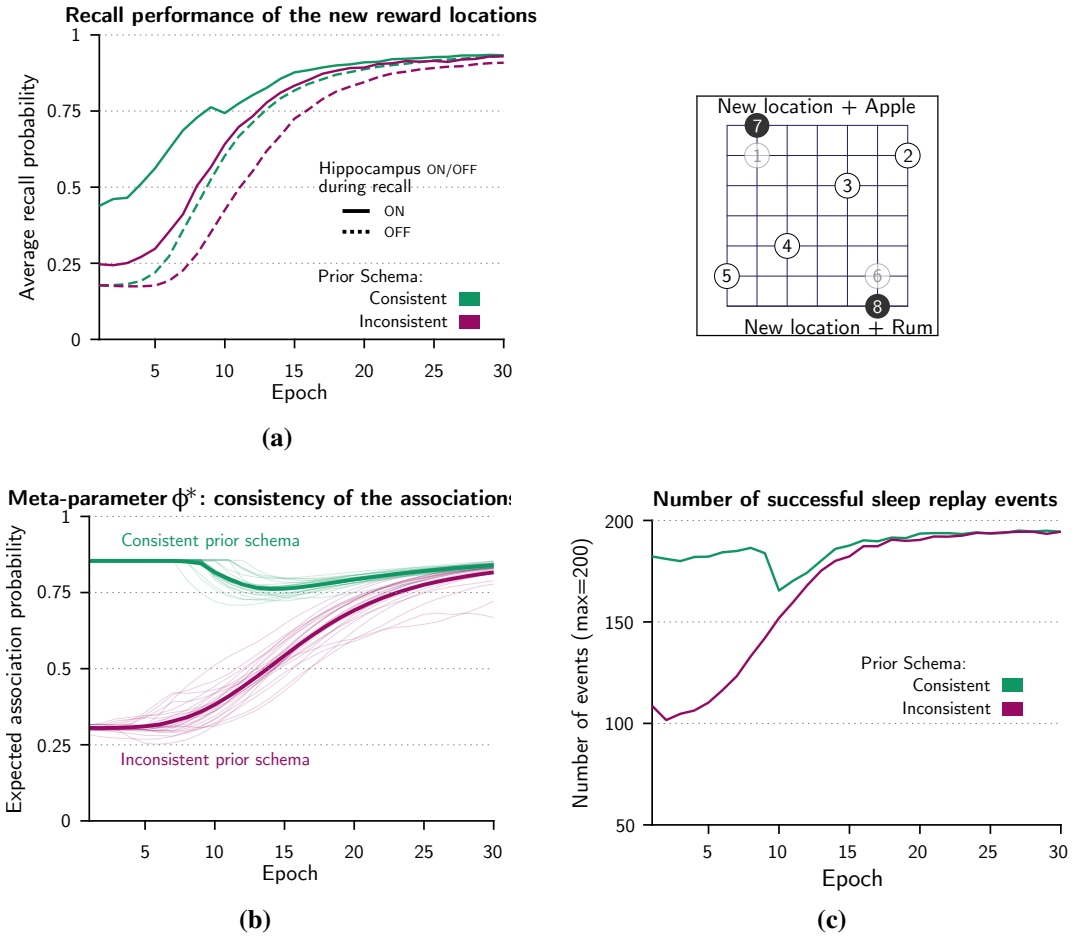
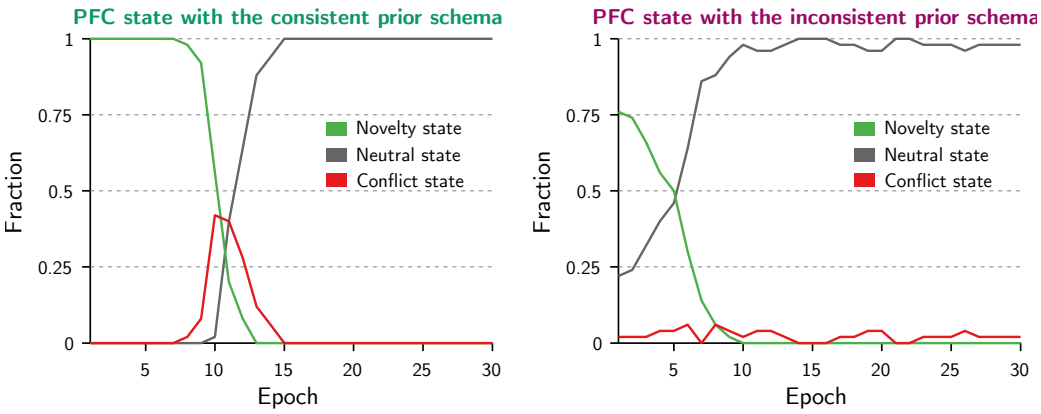
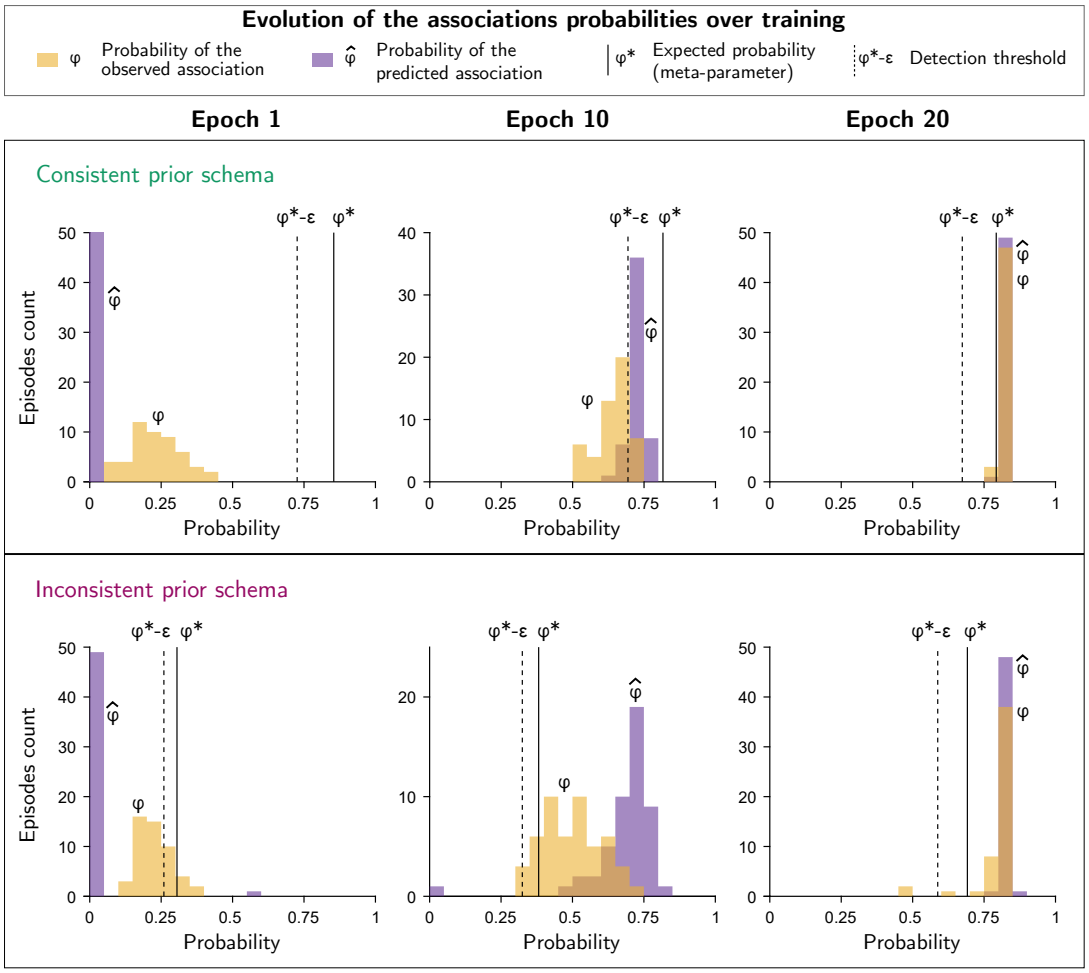


Figure 4.12: New training in the consistent schema (green) versus inconsistent schema (purple). (a) Recall performance during the training of the two new flavour-place associations. Solid lines: recall with hippocampus intact; dashed lines: hippocampus blocked during recall only, to evaluate consolidated memory. (b) Evolution of the meta-schema. The prefrontal cortex updates the expectation about the association probability between flavours and reward locations. This expectation is represented by the meta-parameter ϕ^*. The thin lines show the value of the meta-parameter for the different simulations (25 simulations, 5 pre-trained networks trained 5 times on two new flavour-place associations). For the consistent schema, the initial value of ϕ^* is high, while it low for the inconsistent schema (see Fig. 4.6c). (c) Sleep replay monitoring. Number of replay events that were used for training the network during sleep, out of the 200 reactivations in the hippocampus. Reactivations failed when the hippocampus could not recall a location memory associated with the flavour memory reactivated. See Algorithm 4.4 for sleep replay.



(a)



(b)

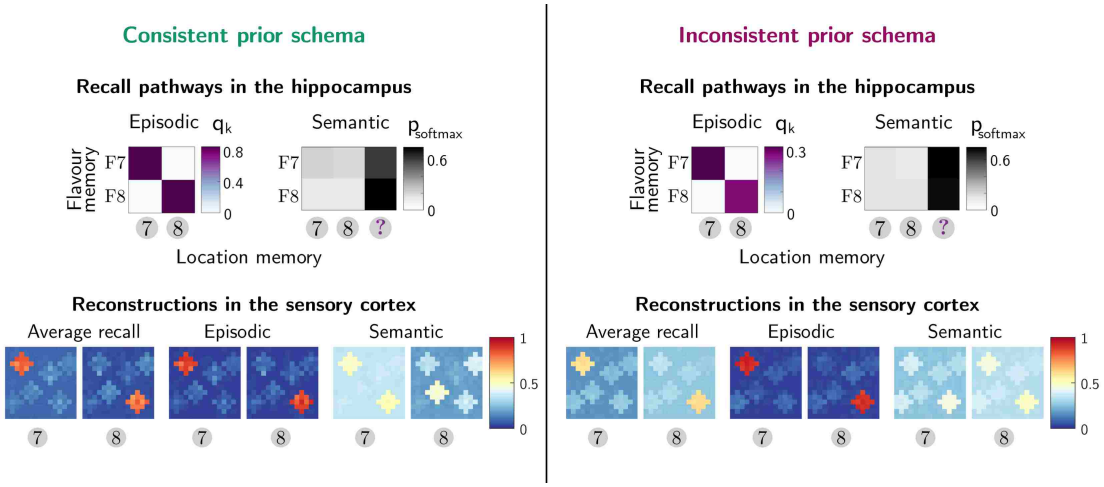
Figure 4.13: State of the prefrontal cortex during the new training in the consistent schema versus inconsistent schema. (a) State of the prefrontal cortex during the new training for all 25 simulations. Fraction = across all the simulations. (b) Examples of how the prefrontal cortex determines its state.

epoch 1 column). This case relates to our conceptual example of the "mammoth" depicted in Fig. 4.2, and similarly the network detected a novel event.

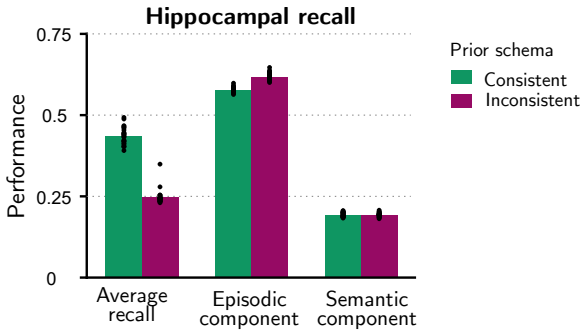
Over time, the new associations were consolidated in the neocortex, their respective probabilities ϕ increased (Fig. 4.13b, **epoch 10** column). Since the expected probability ϕ^* was low for the inconsistent schema, the prefrontal cortex rapidly switched back to the "neutral" state in this case, while it took longer in the consistent schema (Fig. 4.13a, grey line = % neutral detected). We also noticed in the consistent schema that during the transition of states from "novel" to "neutral" the prefrontal cortex detected conflicting associations (Fig. 4.13a, red line = % conflict detected), and we will expand on this later in Section 4.3.2.

New training with a consistent prior schema

Since the prefrontal cortex was in the novelty state at the beginning of training, the episodic link q_k in the hippocampus between a new flavour memory and its associated new location memory was set to the expected probability ϕ^* (Table 4.1). For the consistent schema, this expected probability ϕ^* was high, and hence the new episodic links in the hippocampus had the same high value $q_k \sim 0.8$ (episodic-based recall, Fig. 4.14a, top left panel). Consequently, when the network was cued with the new flavours during early training, the hippocampus already had a high probability of recalling the correct locations memories. When the hippocampus did recall the correct location memories, the reconstructions of the activity patterns in the sensory layer were very similar to the activity patterns of original locations no.1 and no.6 (episodic-based recall, Fig. 4.14a, lower left panel). This result was not surprising as the new rewarded sand-wells were located near the original ones (see schematic of the arena in Fig. 4.12). Despite this, the reconstructions were noisy as the other four original locations were also reactivated, and hence the corresponding performance was lower than expected, i.e. around 0.6 instead of 0.8 (Fig. 4.14b, episodic component, green bar). We had already reported with our previous models in Chapter 3 that recall performance was limited by the top-down reconstruction quality in the neocortex, and in particular in Section 3.3.4 we have highlighted that the similarity of the new locations with the original map (semantic schema) had a significant impact on hippocampal recall performance (Fig. 3.25). Accordingly, if the two new locations were the same as the original locations no.1 and no.6, early in training recall performance with hippocampus should be as high as the performance of



(a)



(b)

Figure 4.14: Recall mediated by the hippocampus after one epoch of training on the two new associations. For reference see Fig. 4.4 (**step 2**). Note that the hippocampus only contains the memories of the new flavours and locations. (a) Top panels: Hippocampal recall probabilities. When the hippocampus recalls a flavour memory, the associated location memory is recalled with a probability q_k (episodic pathway). If direct recall fails (probability $1 - q_k$), the hippocampus probes cortical memory (semantic pathway) to compute the probability of each stored location memory (softmax; $\mathcal{P}_s$ in Fig. 4.4). If the hippocampus selects the 3rd option (?) when probing cortical memory, it does not reinstate a location memory and instead let the cortex infer the activity of the location units ($p(\hat{\mathbf{u}})$ in Fig. 4.4). Lower panels: Reconstruction of the activity of the location units in the sensory cortex (activities averaged over 100 trials, results of one simulation run). The insets in each panel show the reconstruction for the two new flavour-place associations, sorted by flavour tested. The correct reward locations are indicated below the insets. The reconstructions were obtained either with the episodic recall pathway only (to evaluate this we set $q_k = 1$), or with the semantic recall pathway only (to evaluate this we set $q_k = 0$), or with the two recall pathways available (default, average recall). (b) Recall performance corresponding to each reconstruction shown in (a).

the original associations[4].

Since the new episodic links in the hippocampus were about 0.8, the hippocampus still had 20% chance $(1 - q_k)$ to switch to semantic-based recall, i.e. to probe neocortical memory in order to infer the location memory (see recall procedure in Fig. 4.4). In this case, the probability to recall the correct location memory was low since the new memories were not yet consolidated (semantic-based recall, Fig. 4.14a, top left panel). Thus, the reconstruction of the activity of the location units was poor, and the corresponding performance was very low (Fig. 4.14b, green bar, semantic component). Finally, the average recall performance reflected the performances of the two possible hippocampal recall pathways, considering that episodic recall was predominant (Fig. 4.14b, green bar, average recall).

However, the network required more epochs for the new associations to be consolidated in the neocortex. Indeed, performance was impaired if the hippocampus was disabled during recall up to about 10 epochs (Fig. 4.12a, dashed green line). We anticipated this result as we had already reported in Section 3.3.4 that consolidation was hampered by the novelty of the flavours and the overlap between the new and old locations. In addition, we believe that the difficulty was increased in the current model because the flavours were represented by random binary patterns (as opposed to one unit in previous models) and thus representations potentially overlapped. We will see later how we can speed-up consolidation with more sleep replay events (Section 4.3.4).

At the beginning of training the prefrontal cortex was in the novelty state, and as such meta-parameter ϕ^* was not updated (Fig. 4.12b, green lines). We have mentioned earlier that in the course of training the prefrontal cortex eventually switched to the conflict state before settling in the neutral state (Fig. 4.13a, left panel). While in these two states, the meta-parameter ϕ^* was updated with the values of the new associations probabilities ϕ (Eq. 4.5). Nevertheless, these probabilities ϕ had not yet caught up with the value of the meta-parameter ϕ^* (Fig. 4.13b, **epoch 10** in the consistent schema), because the probabilities ϕ gradually increased at similar pace than memory consolidation. This caused the meta-parameter ϕ^* to slightly drop and then increase again during training (Fig. 4.12b, green lines).

[4]In the current model, the representations of the flavours and the representations of the reward locations were stored as separate memories in the hippocampus. Hence, when the network stored the memories of the new locations, the representations in the hidden layer were not altered by the novelty of the flavours but solely by the novelty of the locations. Therefore if the new locations were the same as the original locations no.1 and no.6, the reconstructions should be of high quality.

New training with an inconsistent prior schema

For the inconsistent schema, we have seen that the prefrontal cortex was also mainly in the novelty state at the beginning of training. The episodic links q_k of the new associations in the hippocampus were respectively set to the expected probability ϕ^*, which in this case was low ($q_k \sim 0.3$, episodic-based recall, Fig. 4.14a, top right panel). Therefore, when the network was cued with the new flavours during early training, the hippocampus had a low probability of selecting the episodic memory recall pathway (i.e. recalling the locations memories that were associated to these flavour memories via the episodic link q_k; refer to **step 2** in Fig. 4.4). Instead, the hippocampus most likely switched to semantic-based recall with probability$1 - q_k$. In this case, the probability to recall the correct location memory was low since the new memories were not yet consolidated, and was the same as in the consistent schema (semantic-based recall, Fig. 4.14a, top right panel). Thus, similar to the consistent schema, the corresponding performance was very low (Fig. 4.14b, purple bar, semantic component). Nonetheless, when the hippocampus recalled the episodic memory, performance was the same with both consistent and inconsistent prior schemas (Fig. 4.14b, episodic component). This observation is important as it shows that if we set the episodic links of the new associations to the same values in the consistent and inconsistent schemas, we should obtain the same results in both conditions. We illustrate this point in next Section 4.3.2. But for now, in the inconsistent schema case the probability to recall the episodic memory was low and hence the average hippocampal recall performance was lower than in the consistent schema (Fig. 4.14b, purple bar, average recall).

Consolidation with the prior inconsistent schema was slower than with the consistent prior schema (Fig. 4.12a, dashed lines), and this effect was probably due to the model replaying the wrong associations during sleep. However this effect could also be a consequence of a reduction of the number of "dream" sensory patterns generated during sleep replay. Indeed, in our model replay during sleep was based on hippocampal recall, and we have seen above that the hippocampus most likely recalled location memories via the semantic recall pathway. Yet, with this recall the hippocampus most likely failed to retrieve a location memory (semantic-based recall, Fig. 4.14a, top right panel), and accordingly the replay event failed (Algorithm 4.4). Therefore we expected a reduction of the number of successful reactivations during sleep at the beginning of training. We found that this was the case, since out of the 200 replay attempts during sleep, almost half did not generate a "dream" sensory pattern (Fig. 4.12c).

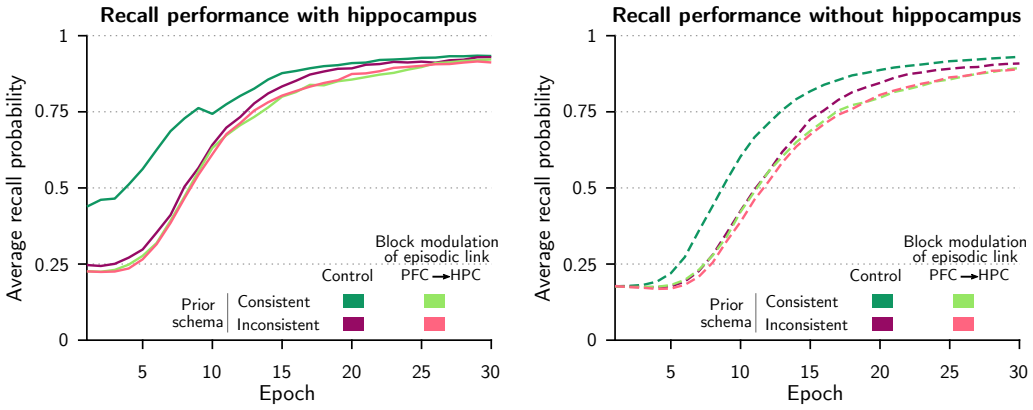
At the beginning of training the prefrontal cortex was in the novelty state, and as such meta-parameter ϕ^* was not updated (Fig. 4.12b, purple lines). When the prefrontal cortex eventually switched to the neutral state, the meta-parameter ϕ^* was then updated with the values of the new associations probabilities ϕ (Eq. 4.5). By contrast with the consistent schema, these probabilities ϕ were already similar to the meta-parameter ϕ^* of the inconsistent schema (Fig. 4.13b, **epoch 10**). Thus, probabilities ϕ and meta-parameter ϕ^* both gradually increased over time as the new memories were consolidated in the neocortex (Fig. 4.12b, purple lines).

4.3.2 Interaction between prefrontal cortex and hippocampus during episodic memory formation of the new associations

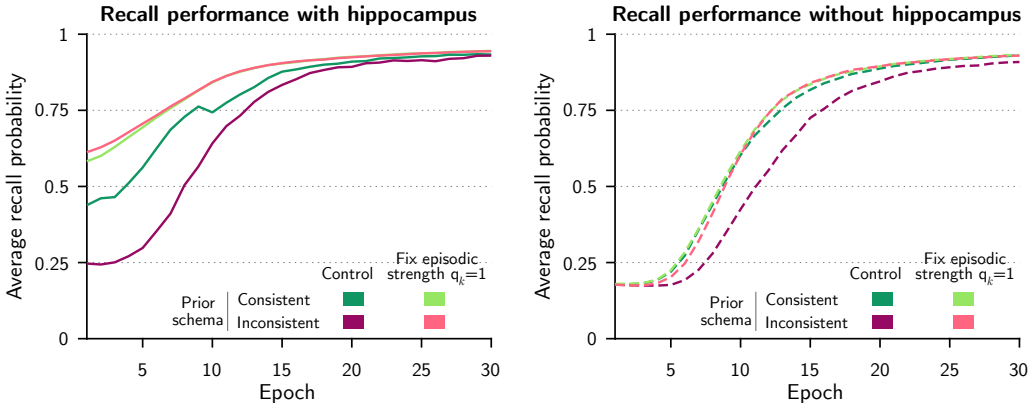
Of the importance of the prefrontal cortex to differentiate consistent and inconsistent schemas

In the Section 4.3.1 above we have seen that a network pre-trained in a consistent schema learned more rapidly two new associations than a network pre-trained in an inconsistent schema, in line with the experimental study of Tse et al. (2007). This distinction of behaviours was possible because the prefrontal cortex, which detected the novelty of the associations, transferred the *structure of knowledge*, i.e. the expectation about the consistency of associations to the new flavour-place memories. This transfer was done by modulating the formation of episodic memory in the hippocampus: the new episodic memories were strong with a consistent prior schema, while they were weak with an inconsistent prior schema. The strength of the episodic memories influenced how often they were replayed during subsequent sleep, which in turn regulated memory consolidation in the neocortex. To illustrate this, we blocked the interaction between the prefrontal cortex and hippocampus during episodic memory formation. As predicted, we found that learning was as slow with the consistent prior schema as with the inconsistent prior schema (Fig. 4.15a). This is consistent with the data of Tse et al. (2011) (Section 3.1.1).

When we interfered with the modulation of episodic memory formation (**step 2** in Fig. 4.5), the strength of the episodic link q_k between the flavour memory and the location memory in the hippocampus was independent of the state of the prefrontal cortex. Instead, it was set to $q_k = \phi$, where ϕ is the probability of the association measured by the neocortex. We have seen earlier that in both consistent and inconsistent schemas,



(a)



(b)

Figure 4.15: Blocking PFC modulation of the episodic link q_k during memory formation in the hippocampus: impact on new training with consistent and inconsistent prior schemas. (a) Blocking the modulation prevents rapid acquisition in the consistent schema. (b) While the modulation is blocked, if the episodic links q_k of the new associations are set to 1 in the hippocampus then learning becomes as rapid with the inconsistent prior schema as with the consistent prior schema.

the probabilities of the new associations were low at the beginning of training (ϕ in Fig. 4.13b, **epoch 1** column). Indeed, the new patterns of activity (new flavour and new location) were unknown according to the existing knowledge in neocortex (like the two new animals examples in Fig. 4.2).

With the inconsistent prior schema, at the beginning of training the new episodic links $q_k = \phi$ were very similar to the control condition, because when the interaction was allowed the episodic links were set to the low value of the meta-parameter ϕ^* (ϕ close to ϕ^* in Fig. 4.13b, inconsistent prior, **epoch 1** column). As such, blocking the interaction between prefrontal cortex and hippocampus did not affect the performance compared to the control condition (Fig. 4.15a, pink vs. purple lines). By contrast, in

the case of the consistent prior schema, the new episodic links q_k were much lower compared to the control condition, as these links were formerly set to the high value of the meta-parameter ϕ^* (Fig. 4.13b, consistent prior, **epoch 1** column). This prevented rapid episodic memory formation in the hippocampus, and thus performance of the new training with the consistent prior was reduced (Fig. 4.15a, light vs. dark green lines).

It is important to note that models pre-trained on a consistent and inconsistent schemas have similar knowledge in the neocortex, i.e. they both know the six original flavours and the six original reward sites. This explains why the new associations probabilities ϕ were similar with both prior schemas at the beginning of training, and why performance in the two conditions was identical when we blocked the interaction between prefrontal cortex and hippocampus. To further illustrate this, we ran a new training where we blocked again the modulation of episodic memory formation in the hippocampus, but the new episodic links q_k were set to a fixed value of 1. In this case, learning and consolidation were as fast with the inconsistent prior schema as with the consistent one (Fig. 4.15b).

Note about novelty and conflict detection

We have noticed earlier in Section 4.3.1 that in the course of training with the consistent prior schema, the network detected conflict around epoch 10 (Fig. 4.13a, left panel). This effect occurred when the probability ϕ of an association was lower than the expected probability ϕ^*, but the predicted association probability $\hat{\phi}$ was close or larger than ϕ^* (Fig. 4.13b, consistent prior, **epoch 10**). When the prefrontal cortex was in the conflict state, the episodic link was set to $q_k = 1 - \phi^*$, and thus both recall performance with hippocampus and the number of sleep replay events dropped (Fig. 4.12a&c). Yet, it did not impair consolidation (Fig. 4.12a) because the prefrontal cortex was only momentarily in this state. By contrast, we did not see such conflict detection for the new training with the inconsistent prior schema, and instead the prefrontal cortex was rapidly in the neutral state (Fig. 4.13a, right panel). Nonetheless, like in the consistent prior schema, we noted that the associations probabilities ϕ were lower than the predicted associations probabilities $\hat{\phi}$ around epoch 10 (Fig. 4.13b, inconsistent prior, **epoch 10**). Unlike in the consistent prior schema, the expected probability ϕ^* was low for the inconsistent schema and hence the observed associations probabilities ϕ caught up with ϕ^* before the prefrontal cortex could detect conflict.

Since the mismatch between observed and predicted associations probabilities occurred in both prior schemas, we suspected that this effect was a consequence of the displacement of the original reward locations (no.1 and no.6) to new neighbouring locations (no.7 and no.8 respectively). To verify this hypothesis we monitored the reconstruction of the activities of the location units over time (Fig. 4.19, "Control" panels). We noticed that at the beginning of the new training the network reinstated the original locations no.1 and no.6 when cued with the new flavours; only later on, around epoch 10, the neocortex adjusted the representations of the reward locations, which coincided with the conflict detection. Therefore we predict that if we did not move the original locations the network should not detect conflict; on the other hand we cannot yet predict the behaviour if we moved the locations further away, without overlapping with the original reward locations.

When we blocked conflict detection, the prefrontal cortex took longer to switch from the novelty to the neutral state, and recall performance with hippocampus was no longer impaired (Fig. 4.16).

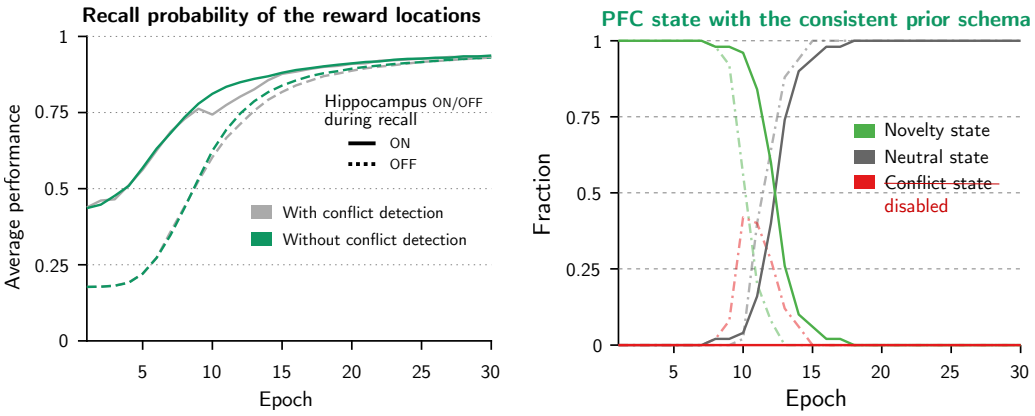


Figure 4.16: With the consistent prior schema, blocking conflict detection in PFC prevents the transient drop in recall performance with hippocampus (a) that was caused by the displacement of the two original reward locations to the two new ones. (b) When conflict detection is blocked, the prefrontal cortex stays longer in the novelty state (solid lines). The dashed lines represent the state of the prefrontal cortex when conflict detection was allowed.

4.3.3 Impact of remote location memories in the hippocampus

In the previous simulations, the hippocampus was cleared out prior to the new training, and thus the hippocampus could only recall the memories of the two new flavours and the two new locations throughout training. However, the six original memories were relevant to the task, and more specifically the memories of the reward locations. As such it might be interesting to keep them, and here we investigate how it would impact the acquisition of the two new flavour-place associations.

We kept the six location memory patterns $\mathbf{y}^k$ obtained at the end of the original training, and these memory patterns were never updated during the new training. In addition, we disabled the conflict detection in the prefrontal cortex to prevent the impact on performance of the location displacement, following the observations in Section 4.3.2 above; however in practice we found that this did not make any major difference.

When the hippocampus kept all location memories it did not impact the training with the consistent prior schema, whereas learning in the inconsistent prior schema was slowed down (Fig. 4.17a). To understand this result we monitored the recall probabilities of the location memories in the hippocampus. In particular, we focused on the semantic recall pathway of the hippocampus, because via this recall pathway the hippocampus could select any location memory according to the probability $\mathcal{P}_s$ (**step 2** in Fig. 4.4).

With the consistent prior schema, at the beginning of training the hippocampus selected mostly the original location memories with semantic-based recall (Fig. 4.18a, epoch 1), as opposed to the previous simulations where the hippocampus mostly selected the (?) option and hence the reconstructions were mainly obtained by cortical inference (Fig. 4.14a, upper left panel). However, this had no consequence on hippocampal recall since q_k was high and hence the hippocampus had a low probability $1 - q_k$ to use semantic-based recall (Fig. 4.18b, light blue line). If anything, we found a slight increase in performance for recall with hippocampus; indeed, when the hippocampus reinstated old location memory patterns $\mathbf{y}^k$ with semantic-based recall, the reconstructions of the sensory patterns were less noisy than during recall in the control condition[5]. Over time, the hippocampus reactivated the correct, new location memories along with their neighbouring original location memories (no.7 & no.1 and no.8

[5]For illustration of the difference, refer to the "semantic" reconstructions in Fig. 4.14a, compared to reconstructions in the panel "inconsistent prior schema, recall with hippocampus, old location memories in HPC" in Fig. 4.19. We actually found that this was caused by multiple steps of Gibbs sampling; when we allowed only one step of Gibbs sampling at the beginning of training, performance in controld condition improved.

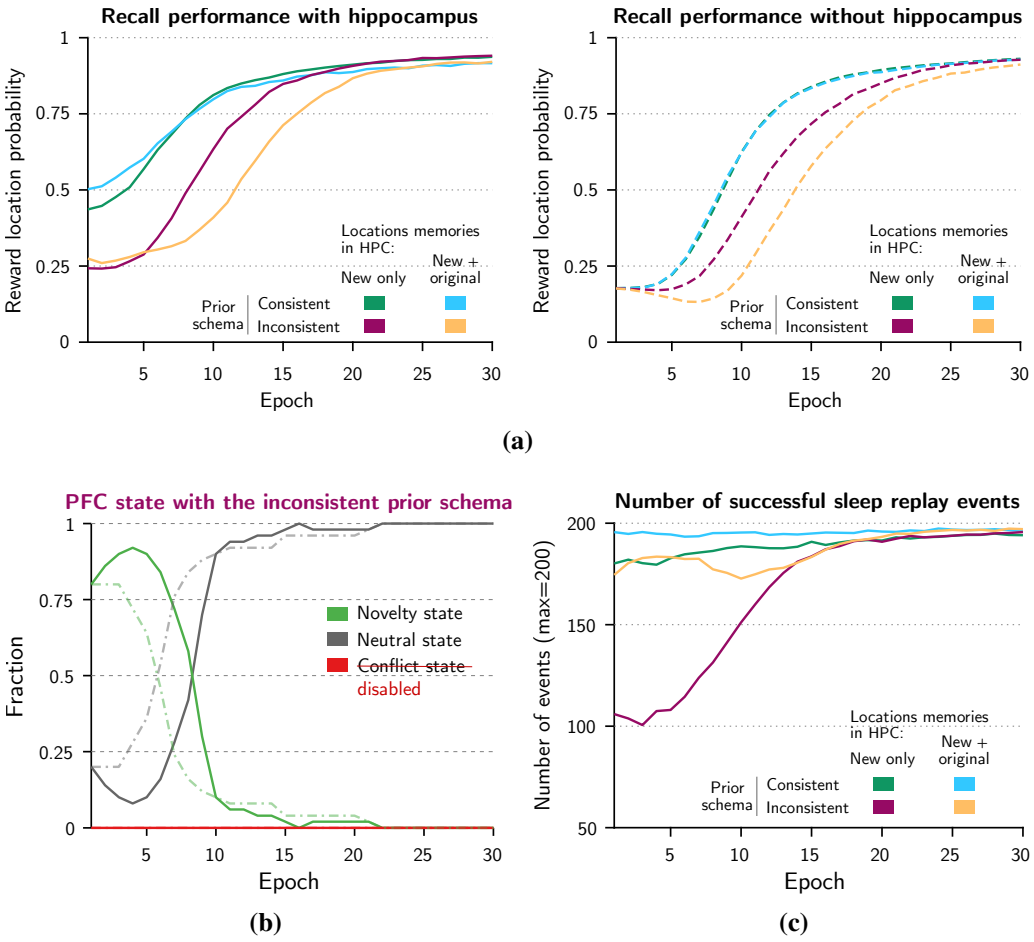


Figure 4.17: Impact of remote location memories in the hippocampus during the acquisition of two new associations. (a) Impact on recall performance mediated by the hippocampus and on consolidation. (b) Evolution of the state of the prefrontal cortex in the inconsistent schema. (There was no change in the consistent schema compared to control.) (c) Monitoring the number of successful replay events during sleep. See Algorithm 4.4 for sleep replay.

& no.6, Fig. 4.18a, epoch 10-20). Since hippocampal recall was not affected by the original location memories, sleep replay was unchanged and as a result consolidation was also the same as in the control condition (Fig. 4.17a).

With the inconsistent prior schema, the hippocampus most likely selected the remote location memories via semantic-based recall at the beginning of training, similar to the training in the consistent schema (Fig. 4.18a, epoch 1). However, in contrast with the consistent prior schema, q_k in this case was low and thus the hippocampus had a high probability $1 - q_k$ to recall the location memories via semantic-based recall (Fig. 4.18b, yellow line). We also noticed that (i) the hippocampus used the semantic recall pathway more often in the current simulations than in control condition

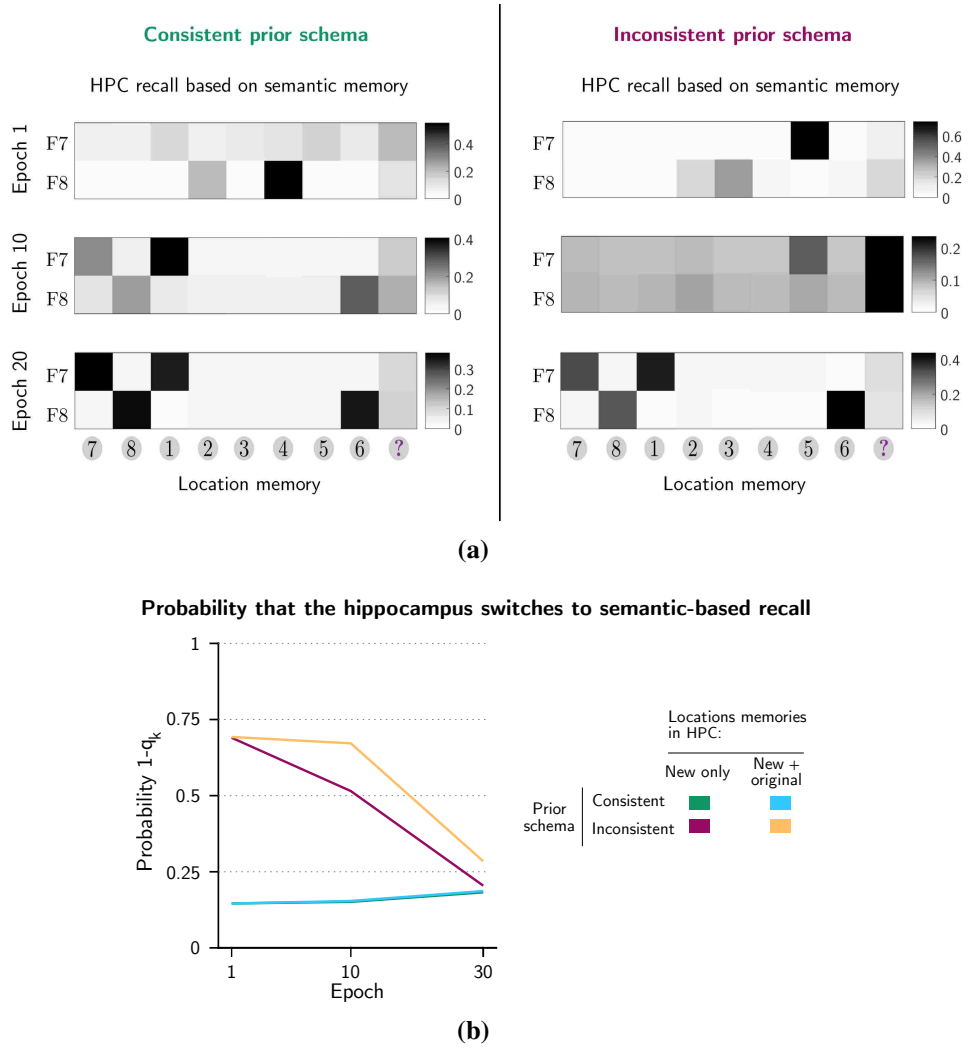


Figure 4.18: Recall mediated by the hippocampus when cued with the two new flavours. For reference see Fig. 4.4 (**step 2**). Note that the hippocampus only contains the memories of the two new flavours, but contains the memories of the two new and original six reward locations. (a) Top panels: hippocampal recall probabilities (1 simulation run). When the hippocampus recalls a flavour memory, the associated location memory is recalled with a probability q_k (episodic pathway). If direct recall fails (probability $1 - q_k$), the hippocampus probes cortical memory (semantic pathway) to compute the probability of each stored location memory (softmax; $\mathcal{P}_s$ in Fig. 4.4). If the hippocampus selects the 3rd option (?) when probing cortical memory, it does not reinstate a location memory and instead let the cortex infer the activity of the location units ($p(\hat{\mathbf{u}})$ in Fig. 4.4). (b) Probability $1 - q_k$ that the hippocampus probes cortical memory (semantic pathway) instead of recalling the episodic memory.

(Fig. 4.18b), and (ii) up to about 10 epochs of training the hippocampus randomly chose among the location memories via this recall pathway (Fig. 4.18a, epoch 10). All together, these results mean that, on average, the hippocampus most likely reinstated random location memory patterns $\mathbf{y}^k$ during early training. This is illustrated by the reconstruction of the activity of the location units at epochs 1 and 10 Fig. 4.19 (panel: inconsistent prior schema, recall with hippocampus, old location memories in HPC). Furthermore, since the hippocampus successfully recalled location memory patterns $\mathbf{y}^k$, albeit the wrong ones, almost all reactivations were successful during sleep replay, as opposed to the control condition where many reactivations failed (Fig. 4.17c). Yet, the hippocampus replayed the wrong associations, and this delayed consolidation of the new flavour-place memories (Fig. 4.17a).

4.3.4 Sleep replay and consolidation speed

With the consistent prior schema, the hippocampus could rapidly learn the two new flavour-place associations, but the consolidation in the neocortex was rather slow. While we anticipated this result as mentioned in Section 4.3.1, it was not comparable to the experimental results of Tse et al. (2007), as they showed that hippocampal lesions already 48 hours after new training did not impair the behaviour of the rodents (Fig. 3.1c).

In our model, we had multiple ways of influencing the consolidation speed. The first option was to increase the cortical learning rate (RBM), but this parameter was difficult to adjust in the current task, especially for the sleep replay phase. Indeed, if the learning rate was too large it could impair learning in the long run if not carefully adjusted over training epochs (e.g. risk of catastrophic interference). The second option was to allow consolidation between cortical areas already during the experience phase. However, as mentioned just before, it was difficult to calibrate how much plasticity could be allowed, and there was a risk that the network would no longer require sleep replay to consolidate the new associations. Since Tse et al. (2007) suggested that sleep was necessary for systems consolidation of the new associations (Fig. 3.1c), we decided against this option. The third option was to increase the number of reactivations during sleep replay; we considered this last option as it seemed the more appropriate and easiest to implement. Indeed, provided that the cortical learning rate was small enough, the only drawback with this hyper-parameter was that replay might become ineffective at a certain point during sleep (i.e. the patterns replayed would no

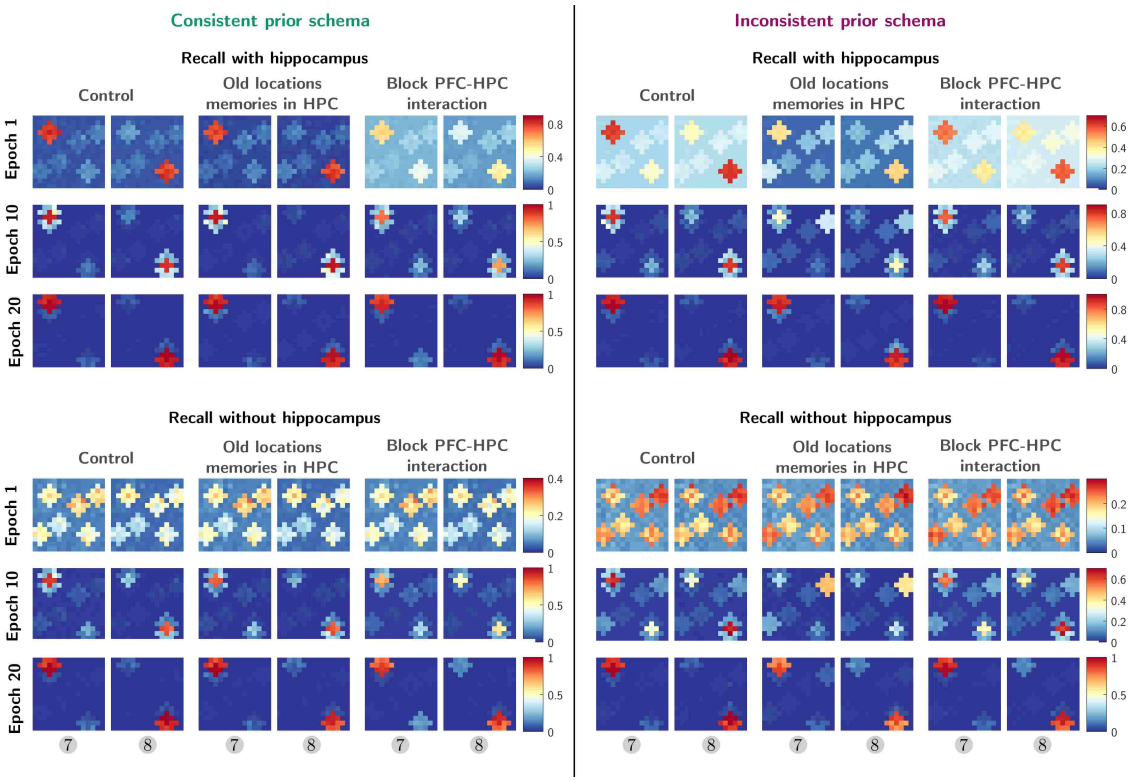


Figure 4.19: Reconstruction of the activity of the location units in the sensory cortex (activities averaged over 100 trials, 1 simulation run). Each group of two insets show the reconstruction for the two new flavour-place associations, sorted by flavour tested. The correct reward locations are indicated below the insets. Notations: HPC = hippocampus; PFC= prefrontal cortex. "Control" corresponds to the simulations of Section 4.3.1, "Block PFC-HPC interaction" corresponds to the simulations of Section 4.3.2, and "Old locations memories in HPC" corresponds to the simulations of Section 4.3.3. Note that for the reconstructions at epoch 1 without hippocampus, we did not allow iterations of Gibbs sampling.

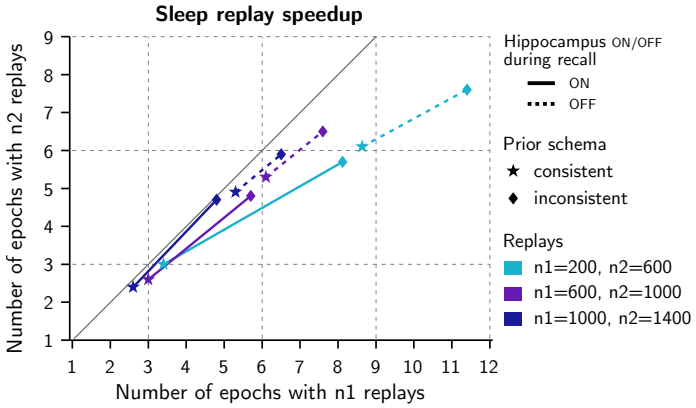


Figure 4.20: Influence of sleep replay on the speed of consolidation of the two new associations. The data points represent the number of epochs necessary to reach criterion performance (average recall probability p=0.5) when the network has n1 reactivations versus n2 reactivations during sleep replay. The data points below the black line (identity) indicate a speedup in learning. When a data point is below, and further away from the black line, it means that learning benefited from the increase in the number of reactivations during sleep (from n1 to n2).

longer drive learning).

We ran new simulations for the training of the new associations with both prior schemas. The conditions of training were the same as in control condition (Section 4.3.1). For each simulation, we increased the total number of reactivations during sleep replay (200, 600, 1000, or 1400). To assess the impact on the speed of acquisition (recall mediated by the hippocampus) and consolidation (recall not mediated by the hippocampus) we measured the number of training epochs required to reach an average performance $p = 0.5$ for the new associations. Increasing the number of reactivations to 1400 during sleep replay did not benefit acquisition in the consistent schema, but did reduce almost by half the time for acquisition in the inconsistent schema (Fig. 4.20 solid lines). Consolidation time in both schemas was also reduced almost by half (Fig. 4.20 dashed lines). However, we observed that with larger number of replay events the speedup saturated, although the consolidation time never reached memory acquisition time. We believe that memory consolidation was hindered by the novelty of the flavour sensory patterns, which could not be reconstructed properly during sleep. One possibility could be to have more cortical updates during experience. In this case it would imply that sleep replay cannot alone account for rapid consolidation of associations in the neocortex. Instead, consolidation of the new features should already be enhanced during the experience phase, so that the network is able to replay them dur-

ing subsequent sleep. This can be done by either increasing the cortical learning rate during experience, or by increasing the number of presentations of the input sensory patterns.

4.4 Conclusion

To summarise, our framework consisted of a closed-loop of information processing that involved the associative neocortex, prefrontal cortex, and hippocampus. The associative cortex maintained the representation of knowledge which allowed the prefrontal cortex to extract the expected consistency of associations, and to detect novel or conflicting events. In turn, the prefrontal cortex regulated the strength of the memory formed in the hippocampus based on the congruence of the event with prior knowledge. During subsequent sleep, the hippocampus replayed memories according to their strength, thus influencing memory consolidation and the update of knowledge in the neocortex.

We first tested this model on the acquisition of six consistent or six inconsistent flavour-place associations, as in the experiment of Tse et al. (2007). In both settings, the neocortex learned a set of features, i.e. the flavours and the reward sites. Following the process described above, the rewarded associations were consolidated in the consistent schema but not in the inconsistent schema. Recall mediated by the hippocampus was also impaired in the inconsistent schema, but the hippocampus nonetheless mostly reinstated the recent rewarded associations. However, hippocampal recall deteriorated in the inconsistent schema when we blocked the interaction between prefrontal cortex and hippocampus during episodic memory formation. This operation, on the other hand, had no effect on the acquisition of consistent associations. Blocking the replay of experiences during sleep prevented consolidation of associations in both consistent and inconsistent schemas, while the hippocampus was still able to recall recent rewarded association, albeit less accurately than in control condition.

We noticed that, when replay was mostly driven by prior knowledge rather than by experience, the network consolidated one location with all the flavours (Fig. 4.9&4.11). This effect occurred because replay was based on recall in the hippocampus, which was then biased by the consolidation in the neocortex. By contrast, in our first model in Section 3.2 we replayed random combinations in the hippocampus, and hence we did not observe this effect. Such result, however, is questionable: for instance, if

we blocked the replay of experiences, the system would then consolidate a memory "blue elephant" rather than not consolidating "elephant" with any colour. Yet, we implemented this replay method because we wanted our system to be able to reactivate prior knowledge in case of conflict so as to reduce interference.

What would happen if we disabled the semantic-based recall in the hippocampus ($\mathcal{P}_s$ in Fig. 4.4) during replay? We predict that, for training in normal conditions, learning and consolidation will be faster. However, in the inconsistent schema this will probably lead to fluctuations in cortical recall as the memories get consolidated and then overwritten every two epochs[6]. On the other hand, replay of prior knowledge allowed to keep the fluctuations low (on average).

How could we solve this problem? We thought about three solutions that could be explored in the future. The first is to find a better semantic-based recall (i.e. $\mathcal{P}_s$ Eq. 4.4), as here it led to biased recall. The second option would be to control whether the system should replay the associations, or the features alone. This would particularly be relevant in the case of the inconsistent schema, where there is no apparent gain in learning the associations, while the reward locations by themselves are relevant. The third option would be to delay the replay instead of reactivating memories during the sleep phase that immediately follows their acquisition. For instance, the hippocampus could accumulate experiences, and could replay all memories only after a few epochs.

Next, we investigated the impact on the acquisition of two new associations with either the consistent schema or inconsistent schema. In line with the results of Tse et al. (2007), learning and consolidation were faster in the consistent schema (Fig. 4.21, 'control'). Indeed, the prefrontal cortex detected novelty in both settings, but only in the consistent schema the system up-regulated memory strength in the hippocampus. Furthermore, when we blocked the modulation by the prefrontal cortex during the acquisition of the new associations, memories were slowly incorporated regardless of the prior schema (Fig. 4.21). This is in line with the experiments of Tse et al. (2011), where inhibition of the prefrontal cortex at the time of learning impaired memory acquisition. Thus, our results suggest that the consolidation of the six original associations in the neocortex is not enough by itself to account for the experimental findings, but that the system needs to extract and use abstract knowledge. This confirms our prediction in Section 3.2.

We further explored how to influence the speed of learning, and in particular we considered the impact of sleep replay. We ran simulations where the hippocampus had

[6]We found such behaviour in preliminary results, and we predict it will be the same here.

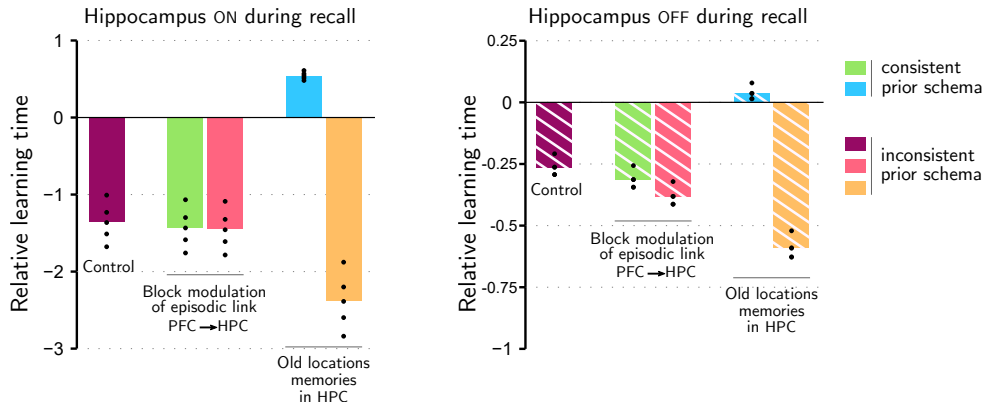


Figure 4.21: Summary of the learning time t of the new associations for various training conditions, compared to the time t_{ref} in control conditions with a prior consistent schema (relative time is $(t_{ref} - t)/t_{ref}$; time was measured when performance reached p=0.5). The dots represent the results of individual runs (five runs). The left panel shows speedup in performance when recall is mediated by the hippocampus, and the right panel shows the consolidation speedup. These results refer to Fig. 4.12, 4.15a&4.17.

access to the memories of the original reward locations; in this case it further delayed the acquisition of new memories in the inconsistent schema, but had no effect on the training in the consistent schema (Fig. 4.21). Similarly, we predict that if we kept the memories of the original flavours, these would be more rapidly associated with the new reward locations. Finally, we noticed that the consolidation speedup was not as strong as memory acquisition in the hippocampus (see the different time scales in Fig. 4.21). We increases the number of replays to boost the incorporation of the new memories into cortical networks. Consolidation was indeed faster, but it never reached the same speed as hippocampal learning. This suggest that other mechanisms are required to explain the rapid consolidation observed in Tse et al. (2007) and Tse et al. (2011).

Chapter 5

Sleep replay and plasticity: impact on episodic memory and consolidation

Internally driven (off-line) replay of declarative memories is believed to mediate their consolidation in the neocortex. An emerging hypothesis is that the salience of a memory (e.g. a rewarded experience) can modulate how often it will be reactivated during off-line replay (Stickgold & Walker 2013, Atherton et al. 2015, see Section 1.1.2). So far we have assumed that such selective replay was taking place in order to consolidate relevant memories in associative areas of the neocortex. For instance, in the first model in Section 3.2 and last model in Chapter 4, the hippocampus reactivated memories of flavour-place associations based on their reward probability. In the second model in Section 3.3, the selective replay was implicit as the hippocampus only stored rewarded memories, and hence could only reactivate these during subsequent sleep. Furthermore, computational studies have suggested that off-line replay of old memory patterns is a key mechanism to resolve the so-called 'stability - plasticity' problem (Grossberg 1987), allowing new information to be incorporated into cortical circuits while avoiding catastrophic interference with existing knowledge (McClelland et al. 1995, Robins 1995, Káli & Dayan 2004, Norman et al. 2005). Consistent with their findings, we have shown with our first model (Section 3.2) that replaying old memories during sleep helped preserve them, but as a consequence it hindered consolidation of new memories and so caused us to increase the total number of reactivations (Fig. 3.11). In addition, in our last model (Chapter 4) we could control the acquisition of new memories by manipulating the content available for replay (Fig. 4.17).

Therefore, our results support the view that off-line replay (e.g. sleep replay) is involved in controlling the entry of information into long-term memory, promoting the integration of relevant memories into cortical networks, while forgetting irrelevant

information. Such 'memory triage' has been proposed to underlie the development of knowledge (Stickgold & Walker 2013). In particular, we believe that the brain needs such selection when exposed to a new experience, as it must decide whether to reject it, incorporate it with existing knowledge or overwrite this knowledge. Importantly, Richards et al. (2014) have provided evidence that this decision balance changes over time: when animals received new information related to a current task, they seemingly integrated new and old memories, but if they received it after a longer delay, they quickly assimilated the new information at the expense of pre-extisting memories.

In this chapter we investigate this effect, and more specifically whether selective off-line (sleep) replay can account for the experiments of Richards et al. (2014). In Section 5.2.1, we modify the content of replay during sleep while the network discovers a schema by observing one example at each epoch of training. Importantly, we compare the relative gain provided by sleep replay when the schema is simple (the examples belong to the same distribution), or more complex (the examples belong to two distributions).

The Section 5.2.2 focuses on the second experiment of Richards et al. (2014), where they revealed that mice adopted different strategies over time to acquire new information, either learning slowly or rapidly. In our model, we had to increase cortical plasticity in order to rapidly assimilate new memories in one trial; however, this operation was likely to destabilise existing knowledge, and hence we examined the role of sleep replay to counterbalance the effect. In particular, we distinguished between the impact on the episodic memory (recall mediated by the hippocampus) and the impact on consolidation in the neocortex.

Finally, during all our experiments we could reactivate old memories, because we always assumed that the memories were still stored in the hippocampus. However, this goes against the prediction of experimental studies which suggest that once memories are consolidated they are no longer supported by the hippocampus. Thus, we modified sleep replay so that it could not only be driven by the hippocampus, which replays newly acquired memories, but was also driven by reactivations in the neocortex, so as to reactivate prior knowledge (Section 5.2.3).

The current work was done prior to the implementation of the last model introduced in Section 4, and hence the prefrontal cortex here is not explicitly implemented. As future work, these simulations could be done in the model with the prefrontal cortex. Another important remark is that we refer to a recall as "episodic memory" when the

hippocampus mediates it, and ”semantic memory” when it does not. The first should mostly reflect the memory of recent events, while the second should reflect the statistics across several episodes.

5.1 Model of the water maze experiment

5.1.1 Summary of the experimental study

Richards et al. (2014) investigated whether mice could extract regularities over time by integrating multiple related episodes. They trained the mice on a memory task in a water-maze. Each day mice had to find a circular platform (10 cm diameter) in the pool of water (120 cm in diameter and 50 cm in depth). The mice could not see the small platform as it was hidden just below the surface of the water, and hence they had to swim until they reached the platform in order to escape. The platform location was changed each day, but the locations had an underlying spatial distribution (Fig. 5.1a). Hence, if the mice did extract the regularities, their search path should reflect the mean of the distribution rather than retrace the specific locations of each platform they had encountered.

To test this, they trained mice to locate platforms drawn from a bimodal distribution[1] (2:1 weighting in the north vs. the east quadrant; see right panel in Fig. 5.1c). The mice were trained for 9 consecutive days, and each day the platform location was different (Fig. 5.1b). The animals had four trials to learn the platform location, and they started each trial from a different place in the pool that was randomly chosen from one of the cardinal coordinates (N, E, S, W). At the end of training, mice were tested on a probe trial that took place either one day after the training (1-d delay) or 30 days after (30-d delay) (Fig. 5.1b). For the probe test after 1-d delay (Fig. 5.1c, left panel), mice principally searched in the area of the platform they learned the day before, which was in this case in the East quadrant as indicated Fig. 5.1b. By contrast, a probe test after 30-d delay revealed that mice mostly returned to the North quadrant, which corresponded to the most likely location according to the underlying spatial distribution. The authors concluded that initially animals rely on memory of recent events, while over time they extract latent regularities across multiple memories.

Next, Richards et al. (2014) investigated whether this apparent memory transforma-

[1]They initially tested on a unimodal distribution. They did find evidence of regularity extraction, however it is difficult to conclude for such trivial distribution (Fig. 5.8). By contrast, a bimodal distribution should be more informative.

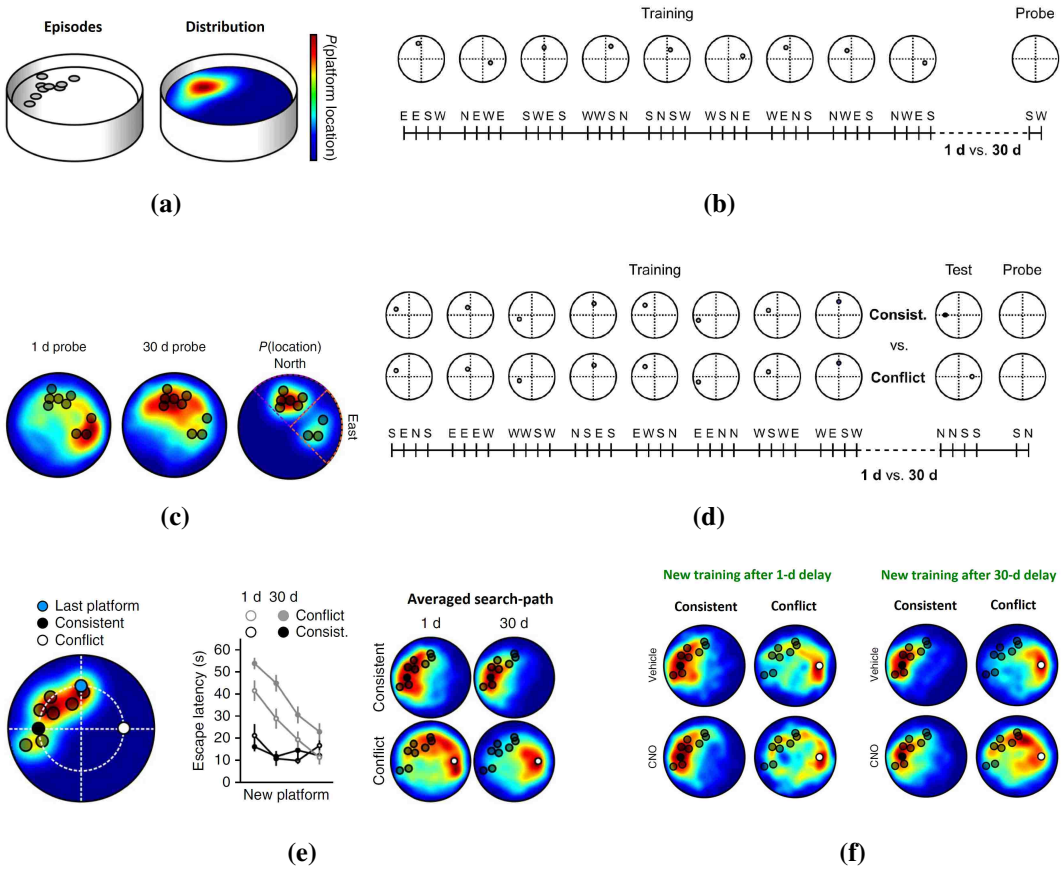


Figure 5.1: All panels adapted from Richards et al. (2014): Extraction of regularities underlying multiple episodes in a water-maze memory task. (a) Schematic of the water-maze. Mice searched for a platform that was submerged below water level so as to escape the pool. The specific location of this platform varied each day (left panel, gray circles). These locations were drawn from a probability distribution (right panel). (b) First experiment. Mice had 4 trials a day to learn the platform location, starting (randomly) from one of the four cardinal coordinates (N, E, S, W). After a delay of 1 day or 30 days following training, mice were tested on a probe test. (c) Averaged search paths during the probe test. Here, the platform locations alternated between two distributions (right panel). After 1-d delay (left panel), mice spent more time around the last platform they encountered whereas with 30-d delay (middle panel) the search was a better match to the actual distribution. (d) Second experiment. Same as (b), but mice were trained on a unimodal distribution, and after 1-d or 30-d delays mice were presented with a new platform location, either consistent (sampled from the original distribution) or conflicting (opposite side in the pool). Mice had a probe test 1 day after training. (e) Model (left panel), escape latencies during the 4 training trials (middle) and averaged search paths during probe test (right panel). When the new conflicting platform was presented after 1-d delay, mice searched in both old and new locations, while with a 30-d delay they focused on the new platform. (f) Averaged search paths during probe test (same protocol (d)), similar to (e) for control condition (vehicle). However, mPFC inhibition impairs learning of the new conflicting platform after 30-d delay.

tion had an impact on how animals acquired new conflicting information. They trained mice for 8 consecutive days on a uniform distribution in the North-West quadrant (left panel Fig. 5.1e). Following this training, the platform location was moved to a new location in the East quadrant, thus conflicting with the original spatial distribution. As control, they trained another group of mice on a new platform that was located near the original distribution, and hence was consistent with existing knowledge. To assess the influence of time on the acquisition of the new platform location, and in principle the influence of memory consolidation (as memory reorganisation), Richards et al. (2014) trained the mice on this new platform either after 1-d or 30-d delay (Fig. 5.1d).

During the training after both 1-d and 30-d delay on the consistent new platform, mice did find the platform quickly as they were used to swimming in this area (see escape latency in Fig. 5.1e, middle panel, black lines). By contrast, mice took longer to locate the new conflicting platform (Fig. 5.1e, middle panel, grey lines). Mice were quicker when the new conflicting platform was presented after 1-d delay compared to 30-d delay, however after four training trials in both conditions they eventually reached the same performance as the mice that learned the consistent platform. The main question then was whether mice would remember and go back to this new conflicting platform location on the following day, or whether they would revert to the original distribution. The study revealed that it depended on the time elapsed between original and new training. Indeed, when they were trained after 1-d delay, mice then searched in the area of the new conflicting location and also returned to the original area during the probe test on the next day (right panel in Fig. 5.1e, inset at the lower left). On the other hand, when they were trained after 30-d delay, mice searched primarily in the area of the new conflicting location (right panel in Fig. 5.1e, inset at the lower right). The result for 1-d delay was less pronounced in a second set of experiments (Fig. 5.1f, top row "Vehicle"), but still showed a broader averaged search path than the 30-d delay.

Finally, Richards et al. (2014) explored whether the medial prefrontal cortex (mPFC) was important to obtain this effect, since the mPFC is believed to play a role in the resolution of memory conflict (Preston & Eichenbaum 2013). They pharmacogenetically inhibited mPFC neurons during the new training, but the animals were drug-free during the probe test on the next day. The inhibition of mPFC did not alter the training after 1-d delay, but altered the training when it occurred 30 days later (Fig. 5.1f, "CNO"). In the latter case, the animals searched in both regions of the water-maze, instead of focusing on the new conflicting platform as they did before (Fig. 5.1f, "Vehicle"). Hence, their results indicate that the animals strategy to update knowledge evolves over time,

and the mPFC appears to play an important role in rapid update of search behaviour.

5.1.2 Model of the data

We used the same framework as in Section 3.3.1. In Chapters 3&4 the task of the network was to learn associations between a cue and a specific location of where to find a reward, while in the current task there is no cue. The visible layer only encodes probabilities of "reward" locations, which are in this case the platform locations to escape the water. To represent the platform locations we used polar coordinates, and thus the visible layer of the network had two vectors of units: one vector of units to encode the distance of the platform from the centre of the pool, and one vector of units to encode the angle.

We simplified the representations by allowing only discrete values for the distance and angle, as it seemed more likely that neurons encode an estimation of the distance and angle based on external cues rather than encoding their exact values. The distance units encoded discrete values of the distance d in the range $[0,60]$ with a step size $\delta d = 5$. We assumed that the probability of being at a distance d from the centre of the pool was represented by the on-probability of the corresponding visible unit. If the centre of the platform is located at the distance d_0 from the centre of the pool (Fig. 5.2, left panel), the activation of the unit encoding distance d was modelled by a Gaussian tuning curve:

$$f(d) = p_{max} e^{-\frac{(d-d_0)^2}{2\sigma^2}} \tag{5.1}$$

where $p_{max} = 1$ is the maximum on-probability at the platform location, $\sigma = r\sqrt{2}$ denotes the standard deviation and $r = 5$ is the radius of the platform. Similarly, the angle a of the platform took discrete values in the range $[0, 2\pi]$ with a step size $\delta a = 0.1\pi$, and we assumed that the probability of the platform being at an angle a was represented

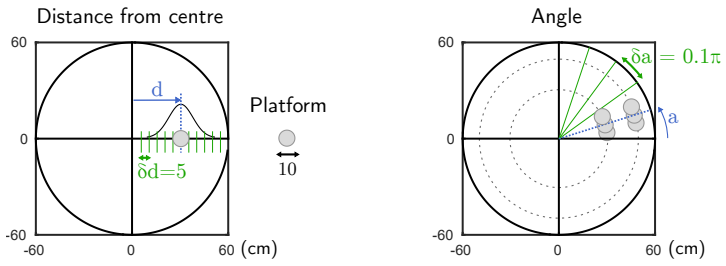


Figure 5.2: Model for the input data in the watermaze experiment.

by the on-probability of the corresponding visible unit. If the centre of the platform is located at the angle a_0 (Fig. 5.2, right panel), the activation of the angle unit encoding angle a was modelled by a Gaussian tuning curve:

$$f(a) = p_{max} e^{-\frac{(a-a_0)^2}{2\sigma^2}} \tag{5.2}$$

where $p_{max} = 1$ is the maximum on-probability at the platform location, and $\sigma = 0.05\pi$ denotes the standard deviation. However, each discrete angle value represented neighbouring platforms, and the closer these platforms were from the centre of the pool, the more likely they overlapped with several discrete angle values (Fig. 5.2, right panel). To take this effect into account, for a given distance d we calculated the angle with the edges of the platform: $a' = \arctan\left(\frac{2r}{d}\right)$, and if $a' \geq \delta a$ then we set $f(a) = p_{max}$ for all angles a such that $|a - a_0| \leq a'$.

5.1.3 Model setup

For the simulations we used our second model presented in Section 3.3.1. As stated in the introduction above, we did not include the model of the prefrontal cortex in the current simulations. Since we have already described the framework, we are succinct here and only highlight the differences.

The sensory layer of the **associative neocortex** (RBM) encoded the platform locations in the watermaze as described above in Section 5.1.2, and thus had a total of 33 units, 13 units to encode the distance of the platform from the centre of the pool, and 20 units to encode the angle. The associative layer had 30 hidden units. At each epoch, during the **experience phase** the network was presented with one platform location only. We took the on-probabilities of the visible units as input data to train the network (Eq. 5.1&5.2), and we limited the number of presentations of each input pattern to four trials per epoch, to mimic the four trials the mice had to learn the platform location in the experiment of Richards et al. (2014).

The **hippocampus** stored the memories of the platform locations using the method shown in Fig. 3.14. The hippocampus took snapshots of the field of the hidden units when the network learned a platform location, and to compute the field (Eq. 3.5) we used the on-probabilities of the visible units (Eq. 5.1&5.2). By contrast with previous models in Chapters 3&4, the hippocampus kept all episodic memories over epochs. Nonetheless, during the **sleep replay phase** the hippocampus either replayed the last stored memory only (default condition) or it reactivated all stored memories. Another

difference with previous models was the procedure to generate sensory patterns during sleep replay. Instead of generating patterns and subsequently updating the cortical weights, here we first generated a "batch" of dream sensory patterns, and then cycled through this batch to update the cortical weights (Fig. 2.7). We had to generate several batches (see Table 5.1) to increase the benefit of sleep replay. The number of reactivations per batch was quite large, and thus the total number of sleep reactivations seemed a bit excessive considering that we trained only one platform per epoch. We believe that this number can be reduced by adjusting other parameters such as the plasticity during sleep or the temperature to generate the samples.

In the experimental study of Richards et al. (2014) there was no specific cue provided for memory retrieval as mice were simply placed in the water and they swam until they found the hidden platform to escape the pool. To implement **memory recall** in our model we let the hippocampus reinstate the memories that were stored during the last epoch only. Since the network only learned one platform location per epoch, the hippocampus could only pick the corresponding memory. The method to reconstruct the activity of the distance and angle units was otherwise as described in Fig. 3.13 (step 2). By default, recall was mediated by the hippocampus. To assess semantic memory we disabled the hippocampus during recall: we set the value of the visible units to zero (i.e. the hidden units were randomly activated), and we let the RBM perform 5 Gibbs sampling steps to reconstruct the activity of the visible units (see Cortical recall in Section 2.2.1).

Note that we did not have a performance measure in this work. To examine the results we looked at the reconstruction of the activity of the sensory units in the neocortex. Since we assumed that the probability of the platform being at a distance d from the centre of the pool, and the probability of being at an angle a, were represented by the on-probabilities of the corresponding visible units, we derived the probability of being at the corresponding location with $p(\text{current location}) \propto p(d)\,p(a)$. To obtain a quantitative measure one could use the Kullback–Leibler divergence between this reconstructed probability and the empirical location distribution, as in the work of Richards et al. (2014).

Remarks about the role of the hippocampus for training and recall

In the experimental study mice started each trial from a random location (North, East, South or West quadrant). The random start was important to ensure that animals

# platform \ epoch	1
visible units	33
hidden units	30
sleep replay	10 batches 1000 replays \ batch

weights init.		$w = 0$
learning rate	η_{EXP}	0.05
	η_{SLEEP}	$\eta_{\text{EXP}}/10$

Table 5.1: Model and training parameters. Notations: HPC = hippocampus; EXP = experience phase; SLEEP = sleep replay phase. The coloured numbers are parameters that can be modified during the simulations.

learned the platform locations using the visual cues in the surroundings instead of memorising the path from a fixed starting point as a procedure, which does not require the hippocampus (Eichenbaum et al. 1990). In our simulations we did not implement navigation and hence we did not take into consideration where the animal started the trial from (similar to the start box in the arena that was not included in Chapters 3&4). However, we assumed that the hippocampus was somehow responsible for allowing the neocortex to develop a representation of relevant locations in space, i.e. to represent the distance and angle in the visible layer. Consequently, the hippocampus was required to learn the task, even though we did not implement its contribution explicitly.

This simplification of hippocampal contribution during memory processing can also explain another divergence between our framework and experimental data: the model could retrieve memories without the hippocampus as the memories of platform locations were consolidated in the neocortex, while studies in the water-maze have shown that after hippocampal damage remote spatial memories are impaired (e.g. Clark et al. 2005). However, we could suppose that when the hippocampus was disabled during recall in our model, it could still play a role in many ways: for instance, it could be required to interpret the activity in the neocortex to allow spatial navigation /orientation in the water-maze. Thus, when we “disable” the hippocampus during recall, we mean that the hippocampus does not retrieve a specific episodic memory to support recall, but we do not mean that it is not required to guide behaviour.

5.2 Results

5.2.1 Sleep replay and extraction of semantic information

Richards et al. (2014) showed that mice extracted the underlying spatial distribution of

platforms in a water-maze over time (Fig. 5.1a-c). Here we examined whether sleep replay helped the extraction of semantic information.

In Section 3.3 we have shown that sleep replay promoted consolidation of noisy input data (Fig. 3.16), because during sleep replay the network could generate many "virtual" training patterns to teach the RBM while keeping a low learning rate in the neocortex. In Chapter 4, consolidation of cross-modal memories was not possible without sleep replay as there was no plasticity between cortical areas during experience. However, in the current model we consider one sensory modality, and cortical plasticity was allowed during both experience and sleep replay. Therefore, since the RBM extracted the regularities of the activity patterns in the sensory area during both phases, we suspected that sleep replay might not be required for homogeneous input patterns, but might be important when the patterns were inconsistent.

To investigate this, we simulated the experiments of Richards et al. (2014). In a first experiment (**A**), the network was presented with platform locations drawn from a unimodal distribution (top panel in Fig. 5.3, similar to the original training in Fig. 5.1e, left panel). In the second experiment (**B**) the locations were drawn from a bimodal distribution (left panel in Fig. 5.4, similar to experimental paradigm in Fig. 5.1b&c). In line with the experimental study of Richards et al. (2014), the network was presented with one different platform location at each epoch.

In the first experiment (**A**) the platforms were located in the North-West quadrant (top panel in Fig. 5.3). The network was trained for 8 epochs, with or without sleep replay, and we tested memory recall before presenting a new platform at each epoch. The first important remark is that there was no difference between episodic and semantic memories, suggesting that memory was rapidly consolidated in the neocortex. Such a result was not too surprising as the sensory activity patterns were consistent across epochs (see examples of input patterns in the column **Data** in Fig. 5.3), and also because we had larger cortical learning rates and more sleep replay events than in Section 3.3 (e.g. here $\eta_{\text{EXP}} = 0.05$ compared to $\eta_{\text{EXP}} = 0.01$ in previous simulations). Hence, we will refer to the results simply as "memory recall".

At the end of training with sleep replay, memory recall reflected the platform distribution (last inset in the middle column in Fig. 5.3, compared to top panel). This means that hippocampal-mediated recall also reflected the statistics of the data rather than the exact last location (compare to last data in the first column in Fig. 5.3). This was compatible with the averaged search paths of the animals during a probe test one day after the training on the unimodal distribution (Fig. 5.8).

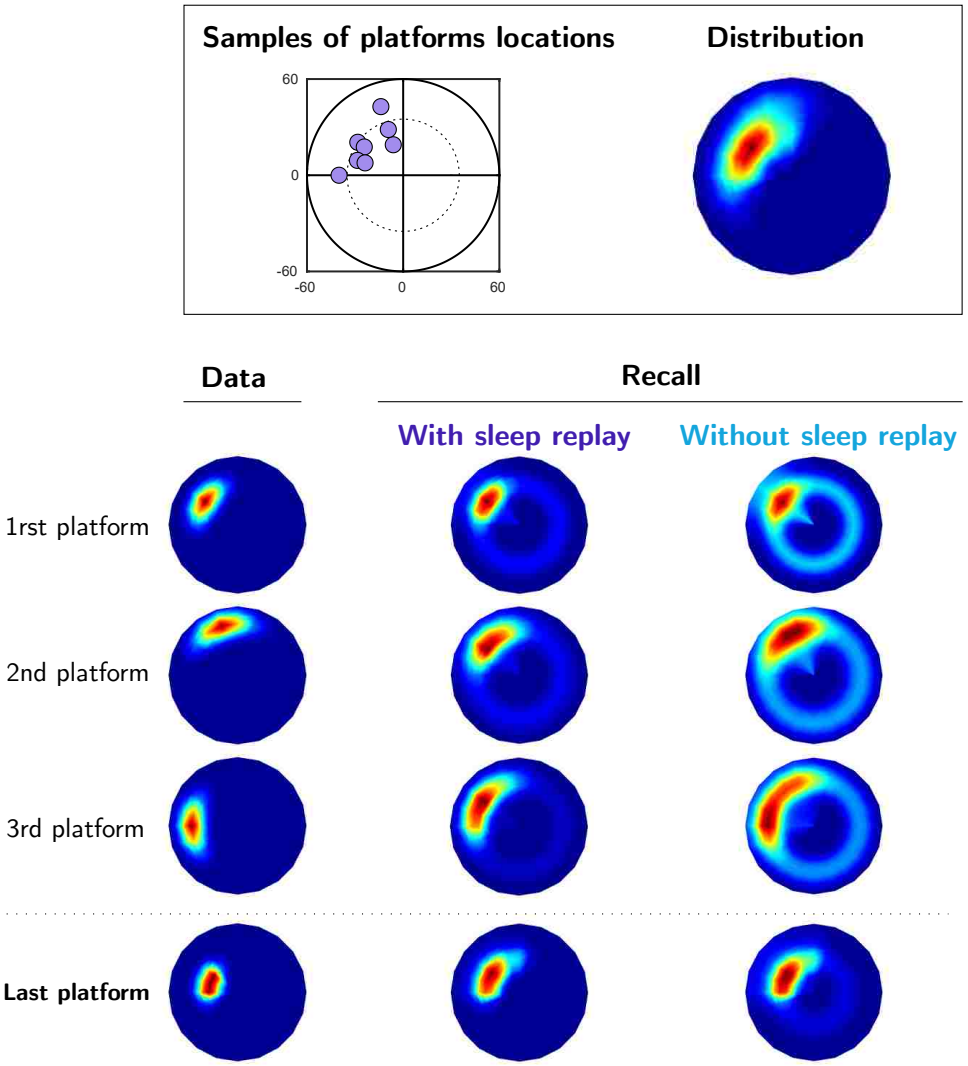


Figure 5.3: Experiment (**A**). Training where the platforms locations are drawn from a unimodal distribution. (Top panel) Schematic of the experiment (left) and distribution of the platform locations (right). The model learns one platform per epoch, and is trained in total for 8 epochs. (Main panel) The column on the left ('data') shows the input sensory patterns presented to the network (one platform each epoch). The middle and right columns show the reconstruction in the sensory cortex during recall, when the network was trained with or without sleep replay.

Importantly, the network also learned without sleep replay (last inset in the last column in Fig. 5.3). We suppose the reason was that the cortical learning rate during experience was large enough to allow cortical consolidation without sleep replay. Thus, the network seemingly does not require sleep replay when the input patterns are consistent. However, when we looked at memory recall during the first epochs of training, we noticed that with sleep replay the network quickly focused its search in the area where the last platform was seen, whereas without sleep replay the network had a broader search (Fig. 5.3).

We then investigated a more complex data set. In the experiment (**B**) the platforms were located in North or East sides, more likely in the North (top panel in Fig. 5.4). We followed the same protocol as in the experimental study (Fig. 5.1b), alternating the presentation of the platforms from each mode (6 in the North, and 3 in the East), and the last platform shown was in the East (top right inset in Fig. 5.4). In contrast to the unimodal distribution, we found that episodic and semantic memories differed (main panel in Fig. 5.4). Without sleep replay, semantic memory recall was biased towards the last platform location (first row in Fig. 5.4), contrary to experimental results (Fig. 5.1c, 30-d probe). A similar result was obtained when the network replayed only the last memory stored in the hippocampus (second row in Fig. 5.4). As we saw for the unimodal distribution, recall was similar at the end of training with or without sleep replay because the cortical learning rate was large enough during the experience phase. In contrast, when all the memories of the 9 platform locations were reactivated during sleep replay, semantic memory better matched the spatial distribution of the platforms (last row in Fig. 5.4). On the other hand, episodic memory always preferred the last location (left column in Fig. 5.4), consistent with the experimental findings (Fig. 5.1c, 1-d probe).

To conclude, we found that, given sufficient training, sleep replay was not necessary to learn the task when the sensory patterns of activity were drawn from a unimodal distribution. However, if data is limited, sleep replay is beneficial as it speeds up memory consolidation. On the other hand, when dealing with conflicting sensory patterns of activity, sleep replay allows the integration of multiple episodic memories into semantic cortical networks that reflect their underlying distribution.

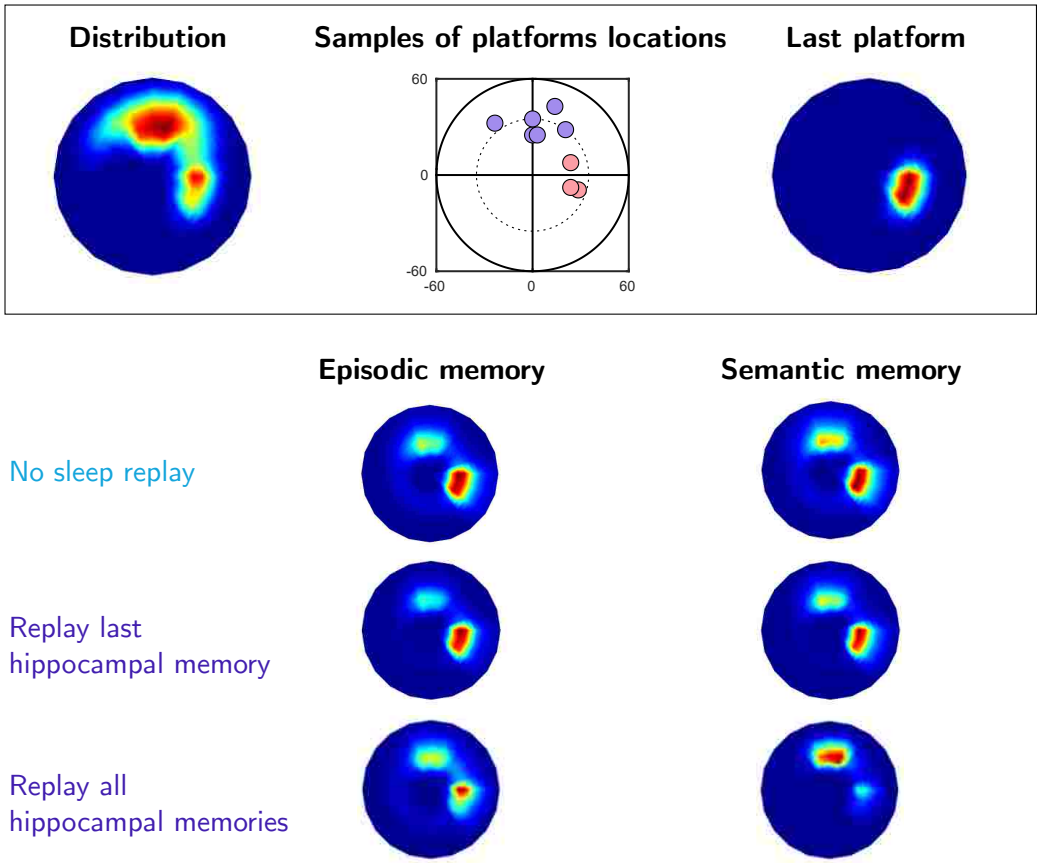


Figure 5.4: Experiment (**B**). Training where the platforms locations are drawn from a bimodal distribution (for reference, see Fig. 5.1b&c). (Top panel) Distribution of the platform locations (left), schematic of the experiment (middle) and last activity pattern shown (right). The model learns one platform per epoch, and is trained in total for 9 epochs. (Main panel) Each inset shows the reconstruction in the sensory cortex during recall. The first columns shows episodic memory, recall mediated by the hippocampus to mimick a probe test after 1-d delay, and the second column shows semantic memory, recall not mediated by the hippocampus to mimick a probe test after 30-d delay. Each row corresponds to a different replay protocol during sleep. We can see that selecting memories for off-line replay biases the extraction of semantic information, while episodic memories generally reflects recent events. Replaying all the experiences accumulated over training gives a better representation of the data statistics, consistent with the experimental results.

5.2.2 Interplay between sleep replay and plasticity during the modification of semantic information

Here we examined the requirements for incorporating new conflicting information.

Following the experiment of Richards et al. (2014), we first trained our model on the unimodal distribution as in Section 5.2.1, and then we presented the new conflicting location (Fig. 5.5). We investigated whether we could obtain the two learning strategies as in Fig. 5.1e (right panel) by manipulating the content of replay during sleep. Note that with our model we could monitor both episodic and semantic memories; however, in the study of Richards et al. (2014) the probe tests were done one day after the new training, which most likely corresponded to episodic memory. As such, our results about semantic memory are purely speculative and we do not have data to compare to.

The results are displayed in Fig. 5.6, and we assume in the following paragraph that we refer to this figure unless specified otherwise.

The network could not learn the new location in one shot when it was trained in the same conditions as the original training (insets 1a&b), which contradicts the experimental results. Thus, we increased cortical plasticity (baseline x2) and the network was then able to learn the new platform location, provided the network replayed the new memory (insets 2a&b). This shows again that sleep replay facilitates memory consolidation.

Episodic memory (inset 2a) was similar to the search behaviour of the animals when trained after 1-d delay (Fig. 5.1e, right panel). If we further increased plasticity (baseline x4), we obtained results that corresponded more to the search path of the animals when trained on the new conflicting platform after a delay of 30 days (episodic memory, inset 3a; for comparison, refer to Fig. 5.1e, right panel). However, with such a high level of plasticity the network erased previous semantic knowledge (semantic memory, inset 3a). On the other hand, without sleep replay episodic memory was also updated, but semantic memory was better preserved (insets 3b). We do not have experimental data about the update of semantic knowledge to decide which behaviour is correct. In fact, the two could be possible. Indeed, the results for the bimodal experiment would suggest that, since mice saw the new location only once, semantic memory would be gradually updated and not overwritten. On the other hand, the logic of survival would predict that mice rapidly consolidate the new location and forget the old, probably irrelevant ones, especially when the mice are trained 30 days after the original training.

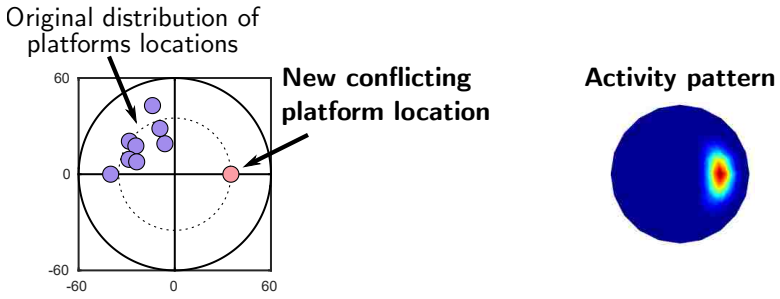


Figure 5.5: New conflicting platform location in the East side presented after the model was trained to locate the platforms in the North-West side of the pool.

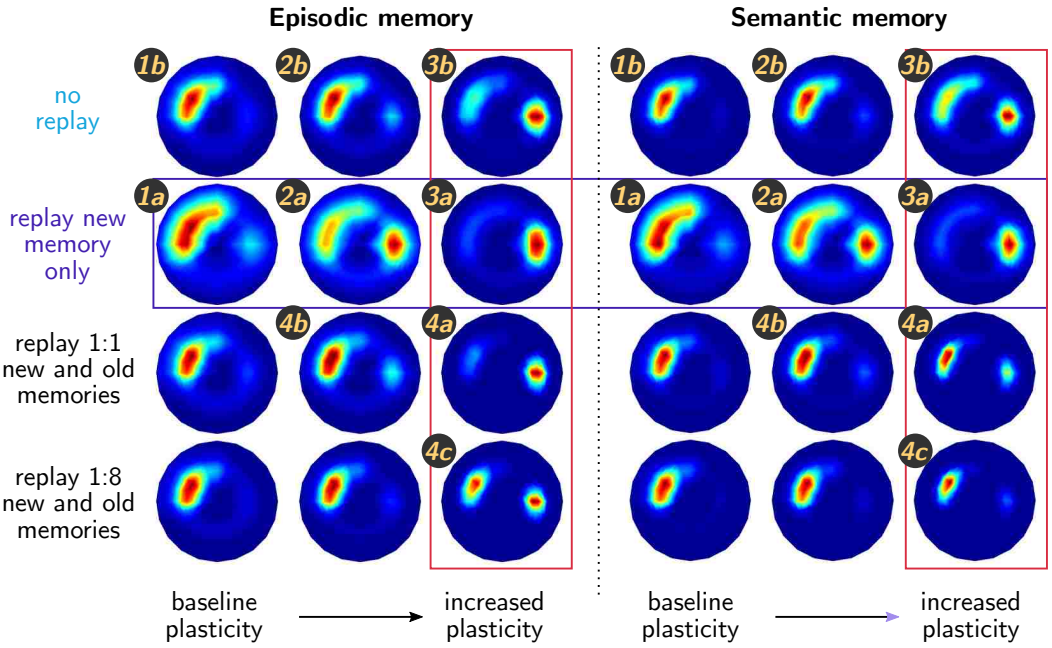


Figure 5.6: Interplay between sleep replay and plasticity during the training on the new conflicting platform location (schematic in Fig. 5.5). Sleep replay: 1) no off-line reactivation (learning during experience only), 2) control condition, the hippocampus replays the new location memory only, 3) equal chance to select the new or one of the old memories (8 original platforms in Fig. 5.5; reactivation probability of the new location is then $\frac{1}{2}$), 4) randomly pick a location memory (reactivation probability of the new location is then $\frac{1}{9}$). Plasticity: 1) baseline, 2) baseline x2, 3) baseline x4. For episodic memory, insets 2a and 4c could correspond to mice trained after 1-d delay, while insets 3a, 3b and 4a could correspond to mice trained after 30-d delay (Fig. 5.1e, right panel). Note that to obtain the latter results the network required higher plasticity.

The next question was, with the same high level of cortical plasticity (baseline x4), could we obtain a search pattern for episodic memory similar to the 1-d delay experiment (i.e. as in the inset 2a), while preserving semantic memory? In line with our simulations of the bimodal distribution, we let the hippocampus reactivate the 8 old memories of the original platforms along with the new memory during sleep replay (probability 0.5 to reactivate one of the old patterns, and probability 0.5 to reactivate the new one). In these conditions, we found that semantic memory was preserved, but the plasticity was too high and the reactivation of prior knowledge was not sufficient to slow down episodic memory update (insets 4a). Hence, we decreased the cortical plasticity (baseline x2) to the same level at which we managed to integrate new with old information, but we found in this case that reactivations of prior knowledge prevented the integration of the new location memory (inset 4b). Thus, we again increased cortical plasticity (baseline x4), and we tried instead to manipulate the content of replay: the hippocampus picked randomly a memory among the 9 stored memories (8 originals, and the new one), and as such it replayed the old memories more. As a result, episodic memory reinstated both old and new location memories, and semantic knowledge was completely preserved (insets 4c).

To conclude, our results show that the brain could have various strategies when learning new information that conflicts with established knowledge. We suggest that these strategies could rely on an interplay between plasticity and selective replay during off-line states such as sleep. In particular, our results demonstrate that episodic and semantic memories can be updated independently with this method, and also imply that the strategies might be chosen to target one or both memory systems.

5.2.3 Cortical-driven sleep replay

In Section 5.2.2 we could control the update of episodic and semantic information by reactivating the memories of the platform locations that were stored during the original training (the 8 locations indicated in Fig. 5.5). Yet, according to standard theories of memory consolidation, when animals are tested after a month memories are presumably consolidated in the neocortex, and as such they might have been removed from the hippocampus. Indeed, studies have shown that animals and amnesic patients with hippocampal lesions still recall remote spatial memories (Rosenbaum et al. 2000, Tse et al. 2007). Although this fact does not directly prove that memories are no longer stored in the hippocampus, having multiple versions of the same memory does

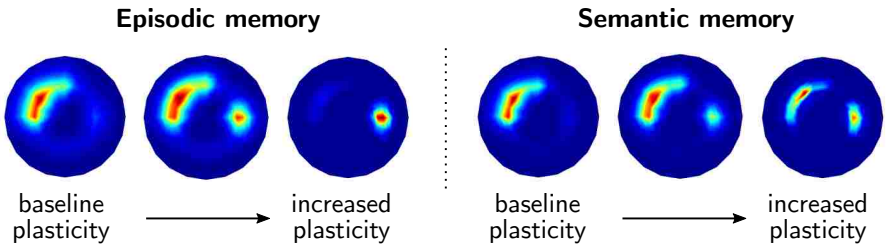


Figure 5.7: Prior knowledge can be reactivated by the neocortex during sleep replay when the hippocampus no longer stores the original memories. The network is trained on the new conflicting platform location (schematic in Fig. 5.5). Sleep replay alternates, with equal probability, between hippocampal reactivations of the new platform memory, and cortical reactivations of prior knowledge. Plasticity: 1) baseline, 2) baseline x2, 3) baseline x4. For episodic memory, the inset in the middle corresponds to mice trained after 1-d delay, while the inset on the right (higher plasticity) could correspond to mice trained after 30-d delay (Fig. 5.1e, right panel).

not seem very efficient. On the other hand, other theories stipulate that systems consolidation of semantic memory occurs in neocortical networks, but the hippocampus permanently stores contextual and spatial elements, and hence is always required to recall some remote memories (Multi trace theory, Nadel & Moscovitch 1997). In support of this, remote memories in a water-maze are usually impaired after hippocampal lesions (Clark et al. 2005). Nonetheless, since this question is still an open debate among neuroscientists, we decided to test whether we could obtain multiple learning strategies, like in Section 5.2.2, but without the original hippocampal memories. Instead, we implemented a cortical-driven replay during sleep, which alternated with traditional hippocampal-driven replay. When off-line replay was driven by the neocortex, the network randomly reactivated the hidden layer of the associative neocortex (i.e. $p\left(h_j = 1\right) \sim \text{Uniform}\,(0,1)$), and generated a dream pattern in the sensory layer of the neocortex. Hence, we expected to reactivate prior knowledge consolidated in the neocortex. On the other hand, the hippocampus could only replay the memory of the new location.

We initially trained the network on the uniform distribution (hippocampal-driven sleep replay only), and then we presented the new conflicting platform. Replay was split 50/50 between cortical and hippocampal-driven, thus we predicted comparable results as shown in the third row in Fig. 5.6. We found again that the network could rapidly assimilate the new location only if we increased the cortical plasticity, and we could replicate similar results than the ones observed in mice in the experimental

studies (Fig. 5.7). In contrast with our previous simulations, where sleep replay was only driven by the hippocampus, memory seemed to be updated more rapidly (see third row Fig. 5.6). We believe that the noisy cortical reactivations were better at driving learning during sleep.

Therefore, our results suggest that even if the memories are no longer stored in the hippocampus, the network can still have multiple strategies to incorporate new information. This is possible again by the interplay between plasticity and off-line replay, the hippocampus reactivating the new experiences and the cortex reactivating prior knowledge. In addition, we further support the idea that once memories are consolidated in the neocortex, there is no need to sustain them in the hippocampus.

5.3 Interpretation of the experimental data based on our model

We have seen in the preceding sections that we can control the entry of information into long-term memory by manipulating cortical plasticity and sleep replay. Hence, our model could select various strategies when exposed to a new event, and could decide to either reject it, incorporate it with existing knowledge, or overwrite prior information.

Such flexibility would suit the experiment of Richards et al. (2014), who have provided evidence that this decision balance changes over time (Fig. 5.1e, right panel). Furthermore, their data suggests that the prefrontal cortex plays a crucial role in learning new conflicting information. Indeed, 30 days after mice learned the distribution of the platform locations, mPFC inhibition during the training of a new conflicting platform prevented rapid acquisition (Fig. 5.1f).

Given our model of the prefrontal cortex described in Chapter 4, we interpret their findings when learning the conflicting platform after 30 days as follows: at the end of the training on the original platform distribution (Fig. 5.1e, left panel), the system has extracted the high consistencies of the locations in the pool. Next, when presented with the platform location, the prefrontal cortex module should detect conflict, and hence the new memory will be weak in the hippocampus (e.g. the “pink elephant in savanna” in Fig. 4.3). But this behaviour does not fit with the data, as learning would then be slow. Hence, we suggest that instead the prefrontal cortex overrides the strategy and generalises the expected consistency to the new platform location, similar to the

"novelty" case. We believe that such an override might be triggered by the relevance of this choice: while it does not make much sense to generalise the colour 'pink' to the Elephant family after seeing one example, in a water-maze there is the incentive to escape water. This would also fit the data when the mPFC is inhibited: in our model, if the prefrontal cortex cannot interact with the hippocampus during learning, the memory strength in the hippocampus is then set to its consolidation strength, which here would be low. Hence, we predict that we would observe similar results to those of Richards et al. (2014) when they trained the mice after 30-d delay, with or without mPFC inhibition (Fig. 5.1f, right panel).

The next question is why there is a difference of behaviour when the new training occurred 1 day or 30 days after the initial training. Richards et al. (2014) suggested that over time, mice extracted the statistics underlying multiple memories throughout consolidation, and thus had a better knowledge of the world after 30 days compared to after 1 day. They predicted that this effect was associated with an increase of sensitivity to conflict, and that the mPFC was therefore more engaged later on. Thus, they argued that learning was enhanced by the high prediction error that would be generated after 30 days.

There could be two interpretations with our model. The first would be that the memories of the 8 original platforms were consolidated later after training, and that the meta-parameter which tracked the variability of information would not be updated online over the course of training, but would rather be computed later on. Yet, with our current modelisation it would not work, as the memories of the original locations were directly consolidated during subsequent sleep (Fig. 5.3). Hence, regardless of when we train the new conflicting platform, its probability will be low and thus will trigger conflict detection. While this might be a limitation of our model, we believe that there could be an alternative explanation.

Specifically, we believe that the system could actually detect conflict both after 1 day and 30 days, but then would select different strategies. The results of Richards et al. (2014) would actually support this theory, as when they replicated the experiment mice seemingly returned to the new platform and less to the original locations even with 1-d delay (Fig. 5.1f). Furthermore, we were not convinced by the argument that the sensitivity to conflict increased over time after learning the unimodal distribution, as we did not find that pattern matching improved significantly enough to justify it (see Fig. 5.8).

According to our model, if the system detected conflict after 1-d the strength of

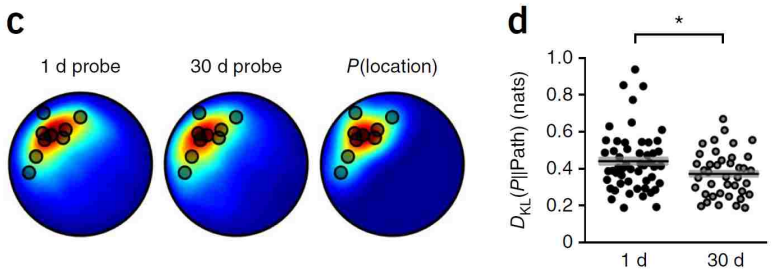


Figure 5.8: Adapted from Richards et al. (2014): Average search path when mice were trained on the unimodal platform distribution, indicated by P(location). Mice were either tested after 1 day (1 d probe) or after 30 days (30 d probe). The plot in the right panel shows the corresponding Kullback–Leibler divergence between the true distribution and the search path.

the memory in the hippocampus would be low (e.g. the "pink elephant in savanna" in Fig. 4.3); yet, if we blocked the prefrontal cortex during learning the outcome would be similar to the control condition (albeit we predict a slower learning), since the new location had a low probability. Hence, our results would explain why inhibition of the mPFC after 1 day delay did not affect the new training in the experiment of Richards et al. (2014) (Fig. 5.1f).

Why would the strategy be different then? We believe that, when mice are exposed to the new conflicting platform the following day, they might process this new information as part of the current task, i.e. they incorporate it into building the schema. This would correspond for instance to the training on the bimodal distribution (Fig. 5.1c) or to the original training in the inconsistent schema in Tse et al. (2007). By contrast, after 30-d delay, the task was presumably completed, and hence mice could use the learned "schema", i.e. high consistency of the locations, when exposed to the new platform.

In the end, our conclusion is not dissimilar to that of Richards et al. (2014), but only the interpretation of the underlying mechanisms.

To conclude, we suggest that when the prefrontal cortex detects conflicting events it can still choose different learning strategies: either slowly update knowledge, or rapidly assimilate the new information. We believe that neuromodulatory signals, triggered by relevant or emotional events for instance, might have a role in this selection. Furthermore, our results in this chapter suggest that the different learning strategies might involve adjusting plasticity levels and off-line reactivations.

Chapter 6

Discussion

6.1 Summary

Studies in neuroscience and psychology reveal that humans and animals are able to learn abstract knowledge from very few examples: in the experiment of Smith et al. (2002) children learned the shape bias given only four object categories with only two exemplar in each category; in the experiment of Tse et al. (2007) rats presumably learned the consistency of flavour-place associations given only six examples, albeit over weeks of training. Furthermore, in both experiments, children and animals were able generalise the abstract knowledge given a single new example.

While these concepts can be formalised at the cognitive level (Kemp et al. 2007), how the brain achieves this is unknown. Connectionist models such as McClelland (2013) have demonstrated that new information is rapidly assimilated when it fits with the established cortical mapping. However, we argue that these models do not explain how to integrate new information that is related to existing knowledge at an abstract level. Indeed, a new event can be consistent with existing knowledge but also unreliable (e.g. a new association in the inconsistent schema in Tse et al. 2007), and, although we did not test this directly, we believe that these models would nonetheless rapidly integrate it. Conversely, we might want to rapidly incorporate new information that might lead to catastrophic interference, but has strategic relevance. As an illustration, in the experiment of Richards et al. (2014), after mice learned that escape platforms were located in one area of the pool of water, they switched their search strategy to a different area after being exposed to a single new platform location. However, in the network of McClelland (2013) learning would be slow as the new platform is conflicting with prior knowledge.

6.1.1 Review of our results

We summarise the three models and results of Chapters 3 & 4 in Table 6.1 at the end of this chapter.

We proposed a definition of schemas in line with hierarchical Bayesian models (which Kemp et al. 2007 call overhypotheses), where a schema is actually composed of two schemas. The first is what we called the semantic schema, which defines a distribution over the features in play (e.g. the names and colours of animal categories in Fig. 1.1, the flavours and the locations in Tse et al. 2007, or simply the platform locations in Richards et al. 2014). The second is what we called the meta-schema, which represents the abstract knowledge; in our applications, it is characterised by a single meta-parameter (denoted ϕ^* in Fig. 1.1), which captures the variability, or consistency of associations between the features.

In chapter 2, we have introduced a computational model that was based on the framework of Káli & Dayan (2004). In this model, a neural network called Restricted Boltzmann Machine (RBM) represents associative cortical areas. A RBM is able to extract and reinstate patterns of activity, but is also able to evaluate the plausibility of events. Hence, it was suitable to our definition of schemas. We extended the model with a prefrontal cortex module to support operations related to the abstract knowledge. The neocortex (or RBM) interacts with a hippocampal module which can store and subsequently replay memories during sleep. Thus, this model was suitable to study on one hand episodic memory formation and recall mediated by the hippocampus, and on the other hand memory consolidation in the neocortex. Additionally, this framework allowed us to explore the role of sleep replay.

In chapter 3, we aimed to identify mechanisms that could support 1) the acquisition of the two schemas mentioned above, and 2) rapid learning and generalisation.

In Section 3.2, we implemented a first, simple model and simulated the associative task of Tse et al. (2007) (Fig. 3.1). We found that rapid assimilation of new information required up-regulation of hippocampal memory formation. In addition, if we up-regulated memory formation in the inconsistent schema the outcome was exactly the same as in the consistent schema. This contrasted with the experiment of Tse et al. (2007) where rodents did not learn rapidly in the inconsistent schema, and hence we suggested that memory formation in the hippocampus had to be dynamically regulated. Since the prefrontal cortex tracked the consistency of associations, we suggested that it could influence memory acquisition in the hippocampus. Furthermore, the hippocam-

pus was important to organise replay during sleep, not only to consolidate the relevant memories in the neocortex, but also to reconsolidate old memories that were damaged upon learning new information.

In Section 3.3, we focused on the semantic schema in the experiment of Tse et al. (2007). We first investigated its formation, and we found that there was a trade-off between plasticity and memory stability (i.e. reconstruction quality of activity patterns in sensory cortices). Given the limited exposure to training examples, sleep replay helped resolve this trade-off by providing a "virtual" training. Next, we examined the impact of knowledge consolidated in the neocortex on episodic memories in the hippocampus. Indeed, as sensory information is processed by associative cortices before it is relayed to the hippocampus, memory retrieval was affected twice: once by the bottom-up stream, which affected recognition in the hippocampus, and second by the top-down stream, which affected the reconstruction of the activity in the sensory cortex. As a result, the hippocampus rapidly recalled new associations if these were closely related to existing knowledge, which was compatible with the model of McClelland (2013). However, when we examined memories in the neocortex our results diverged from the model of McClelland (2013). First, we observed that consolidation was not facilitated by the similarity of the new reward location, in contrast to hippocampal memory (Fig. 3.25). Second, consolidation speed was not determined by the specific semantic schema, e.g. if animals had been trained in a different environment than the one of the new training, the outcome would have been the same (Fig. 3.23). This is not fully incompatible with the work of McClelland (2013), as the input activity patterns we considered were sparse and had no hierarchical structure, as opposed to the data set of McClelland (2013). Yet, these findings seem at odds with biology, and hence imply that other mechanisms might be in play to influence consolidation, mechanisms that perhaps can only be triggered when new information is related to a schema. Nonetheless, our results were overall in line with the findings of Tse et al. (2007), since a trained network learned and consolidated faster new associations compared to a naive network that was trained from scratch.

At this point, we had a better understanding of the meta-schema but we had to clarify the mechanisms by which it could impact the assimilation of new information. We addressed this in Chapter 4, where we defined a measure of the consistency of association (i.e. their variability or reliability) that was based on the strength of their consolidation in the neocortex. Thus, we could extract the meta-parameter that captured the overall consistency of the associations presented to the network. We assumed that the

prefrontal cortex mediated this operation. Hence, as consistent associations were consolidated over training the meta-parameter increased, and conversely decreased when the network was exposed to inconsistent associations. Next, we examined how the prefrontal cortex could influence the formation of episodic memories in the hippocampus. When an association was presented to the network, the prefrontal cortex detected if it was neutral, novel or conflicting by comparing the consistency of this association with the schema expectation (meta-parameter). The strength of the memory in the hippocampus was modulated accordingly, which in turn influenced memory consolidation as the hippocampus replayed memories during sleep based on their strength. When a new association was detected, the prefrontal cortex set the corresponding memory strength to the expected consistency. Hence, in the consistent schema it led to an up-regulation of the memory strength in the hippocampus, while it did not in the inconsistent schema. As a result, learning and consolidation were faster in the consistent schema compared to the inconsistent schema, similar to the experiment of Tse et al. (2007). Furthermore, if we blocked this modulation the distinction between consistent and inconsistent schemas was nullified, and memory acquisition in the consistent schema was as slow as in the inconsistent schema. This is in line with the data of Tse et al. (2011), where inhibition of the prefrontal cortex during training prevented the acquisition of the new associations in a single trial.

In Chapter 5 we further investigated this difference in our model between episodic memory in the hippocampus and memory consolidated in the neocortex. For this, we simulated the study of Richards et al. (2014) where animals had to quickly learn platform locations to escape water. In previous chapters we had considered the network intialisation and number of sleep replays to speedup learning, but had not yet manipulated plasticity. In this study we increased the cortical learning rate so that the network could learn rapidly, in line with the experimental data. Using the interplay between plasticity and sleep replay we obtained various strategies for memory acquisition, and we could separately control episodic memory and extraction of semantic knowledge. This interplay might underlie the different learning strategies that mice adopted in the experiment of Richards et al. (2014). Our results also suggest that plasticity and the content of replay sleep complement the upregulation of memory strength in the hippocampus described earlier, and together they might support the rapid assimilation of new information into existing knowledge.

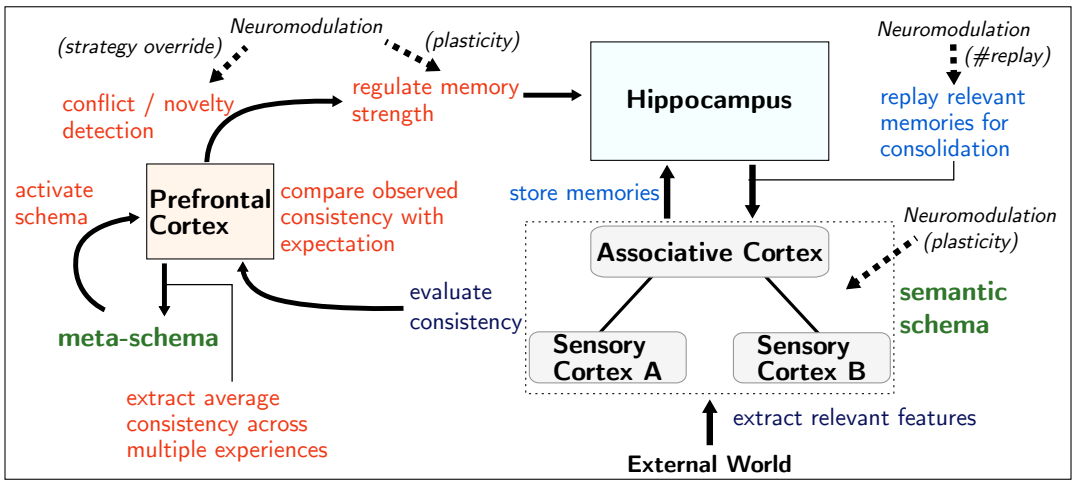


Figure 6.1: Overview of the theoretical framework.

6.2 Theoretical framework

Our framework, Fig. 6.1, attempts to address the three following open questions.

The first question is how the brain can learn mental schemas from limited data. The neocortex, i.e. the neural network of our model, discovers the features as it observes and slowly consolidates examples of associations (semantic schema, body of knowledge). These examples are stored in the hippocampus, which later recalls relevant memories, new and old, during sleep/rest. This recall might be triggered by spontaneous reactivations in the hippocampal region, or cued by reactivations in the neocortex (Rothschild et al. 2017). Such off-line replay facilitates the consolidation of memories in the neocortex, in particular when the number of training samples is limited (Section 3.3.2) and if there is a latent structure underlying the data (Section 5.2.1). The prefrontal cortex supports the extraction of the overall consistency of the associations (meta-schema, abstract knowledge), which we suggest can be done by evaluating the strength of the associations consolidated in the neocortex (Chapter 4).

The second question is *when* the hippocampus rapidly assimilates new information. In the model the hippocampus is not directly connected to the sensory cortices, but instead it stores the information relayed by associative cortices. Hence, a new event that is similar to existing knowledge can be accurately stored in the hippocampus, and when the hippocampus reinstates the memory, the neocortex is able to reconstruct the correct activation pattern in the sensory cortices (Section 3.3.4). Otherwise, plasticity in the neocortex may be increased to allow rapid memory acquisition in the hippocampus (Chapter 5). Next, the model decides whether it should rapidly learn the association or not. For this, both semantic and meta- schemas are fundamental (Chapter 4). Based on

the knowledge in the neocortex, the system can predict the consistency of the episode. The prefrontal cortex, presumably after activating a schema, compares the predicted and the expected consistency. This computation determines whether information is neutral, novel or conflicting. The prefrontal cortex modulates memory strength in the hippocampus accordingly, and for instance it may up-regulate the memory strength of a new memory in a consistent schema. However, we suggest that other neuromodulatory factors could be in play to override a learning strategy in the prefrontal cortex (Chapter 5), or to directly modulate memory strength in the hippocampus (e.g. dopamine).

The third question is how the neocortex rapidly consolidates new information. We already mentioned that the hippocampus replays memories according to their relevance during sleep, which we suppose is linked to their strength (Chapter 4). Nonetheless, the speed of consolidation will be determined by 1) the quality of the memories replayed in the sensory cortices, which is depends on hippocampal recall and top-down cortical reconstruction, 2) the number of reactivations, and 3) which memories can be reactivated (e.g. old memories) (Chapters 3-5).

As memories get consolidated in the neocortex, it will impact (1) the consistency of the associations, hence the meta-schema, and (2) the body of knowledge in the neocortex. Therefore, we have a closed-loop of information processing where a schema develops through the consolidation of multiple examples, and in turn the schema influences the acquisition of new memories.

Together, these mechanisms allow us to interpret the behavioural data observed in Richards et al. (2014) and the experiments of Tse et al. (2007) and Tse et al. (2011), but can also be extended to other experiments in the future.

6.2.1 Implication for experimental studies and open questions

Consistency, prefrontal cortex and interaction with hippocampus

Experimental studies have highlighted that the medial prefrontal cortex might help detect schema-congruency (van Kesteren et al. 2013) and conflict with prior experiences (Richards et al. 2014). In Chapter 4 we suggested that the prefrontal cortex is necessary to compare the consistency of an association with the schema expectation. We proposed that the consistency of an association could simply be the strength of its consolidation, which may be computed via recall in the neocortex.

Experimental findings of Tse et al. (2011) and Richards et al. (2014) have shown

that inhibition of the prefrontal cortex during training prevented the rapid acquisition of new memories. We proposed that the prefrontal cortex could modulate memory strength in the hippocampus during learning. We found in particular that this modulation was crucial to form a new memory in the consistent schema, since otherwise the memory would be too weak and rapid memory retrieval and consolidation would fail (Chapter 4). Such regulation of hippocampal activity would be compatible with the data of Bero et al. (2014) (Section 1.1.3).

Conversely, our model also suggests that, in principle, even in the inconsistent schema of the experiment of Tse et al. (2007) the rats should also be able to quickly learn two new associations (Fig. 4.15b). This would require an upregulation of the memory strength in the hippocampus. However, we believe that this modulation could be triggered by other signals that carry the relevance of information (say, the reward is really good!) that could override (or switch) the learning strategy set by the prefrontal cortex.

On the other hand, our model predicts that the modulation of memory strength in the hippocampus is not required during the formation of the consistent schema (original training), nor when learning a new association in the inconsistent schema (Chapter 4). Our hypothesis is that this interaction with the hippocampus is not required when there is no surprise (but the prefrontal cortex might still be important to relate the memories to a context for instance). The fact the prefrontal cortex has more than one function, like other regions in the brain, makes it difficult to test experimentally, unless we have a precise area or mechanism to target.

Schemas in the brain

We found an important role of prior knowledge in the neocortex for successful recall of the hippocampus. First, during the initial acquisition of consistent associations (Section 3.3.2), hippocampal recall performance was limited as memories were not yet consolidated in the neocortex (Section 3.3.4). Our results suggest that if hippocampal lesions were made at some point during training, recall performance would be similar to the performance prior lesions (see for instance Fig. 3.16). In support of this, in Tse et al. (2007) hippocampal lesions were made just after the new training, and the rats still correctly retrieved the old associations (Fig. 3.1b&c). But this result does not inform us about the time course of memory consolidation during training. However, if we delayed replay in our model hippocampal performance would no longer gradually increase as in Tse et al. (2007).

The second impact was during new memory acquisition, as memory was more rapidly assimilated by the hippocampus when it fitted prior knowledge (Section 3.3.4). Since replay in the sensory cortices relied on hippocampal-driven recall, we expected that the quality of the "dream" patterns would improve as the new memories were more consistent with prior knowledge, and thus we expected more benefit of sleep replay. Such result would be in line with the study of Groch et al. (2017) who found that only memories that were related to prior knowledge benefited from cued reactivation during sleep. However, with our models we did not obtain greater speedup of consolidation for these new memories (Section 3.3.4), but we believe that with a different replay method this could work (see next section).

In this work we argued that a schema also represented abstract knowledge, such as the expected consistency (or variability) of associations. However, there is no evidence about when this schema is extracted (although see discussion in 5.3), where it is stored, and how it is activated. Neuropsychological studies suggest that the prefrontal cortex primarily coordinates the use of schemas to process information (Fernández 2017). For instance, patients with damage to the medial prefrontal cortex are prone to confabulation (Ghosh & Gilboa 2014), but they are less prone to the "false memory effect" with the paradigm of Roediger and McDermott, where participants learn words that belong to a semantic schema (e.g., cold, blizzard, winter) and usually wrongly recall nonstudied word (e.g., snow) (Warren et al. 2014).

Replay during sleep (or rest) for consolidation and reconsolidation

In Chapter 3 we highlighted that learning new associations impaired the recall of consolidated memories, and our results suggest that replay is necessary to reconsolidate them. Indeed, when we disabled the hippocampus during recall, we found that performance of the original associations dropped after the new training but before sleep replay, and performance remained low, or further decreased, after sleep if we did not allow the replay of the original associations. In our model such replay occurs during sleep, but it might as well occur during wake (e.g. during immobility; see Section 1.1.2). Furthermore, we found that while memories are destabilised, the hippocampus can still support their retrieval (Fig. 3.11&3.15). To test these hypotheses, we could block the reactivation of old association patterns during sleep after the new training in Tse et al. (2007). If we do not observe interference afterwards, then it may be that (i) replay occurred during wake, and reconsolidated the memories, or (ii) that there was

no interference in the first place.

Another assumption of our model was that the network might have to increase the number of reactivations during sleep, for instance as a consequence of replaying remote memories (Section 3.2.4) or to boost the consolidation of a new memory (Section 4.3.4). Such operation would require a regulation of the number of replays as a function of consolidation requirements. Evidence of dynamic regulation was shown by Girardeau et al. (2014): when hippocampal sharp wave-ripples (SPW-Rs) were artificially suppressed during sleep after learning, the rate of occurrence of SPW-Rs increased.

6.2.2 Limitations of the model and future work

To test the hypotheses derived from the theoretical framework, the computational model has to be further developed. Here we highlight some features of the model and potential modifications which we believe are important to address.

Model of the hippocampus

The first important consideration about the model was memory storage of associations. In the two models in Chapter 3, the hippocampus stored directly the activity of the associative cortex given two patterns of activity in the sensory cortices. Thus, the representations of the sensory stimuli were integrated into one memory in the hippocampus. However, we found that this storage method did not allow to form memories of certain stimuli combinations. Indeed, when the hippocampus stored the memory of a known sensory stimulus, it could later reinstate the corresponding activity pattern in the sensory cortex. However, if the same stimulus was presented with an unknown stimulus in a different cortical area, the neocortex could no longer reconstruct the memory that was reinstated by the hippocampus. As mentioned earlier in the discussion, recall of memories stored in the hippocampus was strongly influenced by the knowledge in the neocortex. Thus, in the third model (Chapter 4) the hippocampus could form a memory for each sensory modality, and then bound these memories. This would be in line with experimental data (see Sections 1.1.1&1.1.2). This method was also interesting as it provided more flexibility for recall and replay (i.e. the elements of an episode could be reactivated independently).

However, our hippocampal model is currently very simplistic. In particular, we greatly simplified the model by assuming that the hippocampus cleared out past mem-

ories at the start of each training epoch, which does not seem plausible. Ideally, the hippocampus should keep the memories, and perhaps later overwrite or integrate them with new related information (Schlichting & Preston 2015). Furthermore, the hippocampus could determine whether it should keep a memory in storage or not: for instance, a memory could be forgotten if it is already consolidated or if is obsolete. We believe that a more sophisticated model of the hippocampus would give further insight into the mechanisms for schema formation (e.g. McKenzie et al. 2014) and sleep replay (Section 1.1.2).

Another important upgrade of the model would be to include other neuromodulatory system to control the formation of memories in the hippocampus. Indeed, iy is too restrictive to consider only the influence of the prefrontal cortex, as it is known that the persistence of long-term memory can be modulated with dopamine mechanisms in the hippocampus (Rossato et al. 2009).

Schemas

In this work we considered a very simple schema: only a few features, and only associations between two sensory modalities. This framework is suited to model many experimental tasks with rodents, as the animals usually learn simple associations. Nonetheless, it would be interesting to add more sensory elements in the associations. For example, in Fig. 1.1, the variability of both colour and size of the animals in a family could be each represented by a meta-parameter. Also, we assumed that the system knew which features were important - e.g. the flavours and locations in Tse et al. (2007), although the animals probably have to discover in the first place that these are the relevant stimuli among many others. In addition, the model should be extended to allow more layers of abstraction (i.e. over-overhypotheses in Kemp et al. 2007). For instance, in Fig. 1.1, even though the spectrum of colours in general is wide, it is limited to a subset of possible colours for the animals. We believe these modifications are important to model mental schemas and account for rapid generalisation seen in humans.

Furthermore, we did not explore how the model could have multiple schemas (e.g. “the pink elephant” either in the “savanna”, “fairy tale”, or “street”). This feature should be introduced in future work as it would help investigate how schemas are activated and updated in the brain, and how the prefrontal cortex would then coordinate their use. It would also allow to study behaviours such as confabulation.

Replay during sleep /rest

Sleep, or more accurately off-line replay, was important in all our models to facilitate memory consolidation and reconsolidation. However, we often observed that consolidation in the neocortex was rather slow despite replay (for example, see Fig. 4.21). Even though consolidation with a prior schema was faster than in a naive network, the speed we observed still contrasted with the rapid consolidation in rats in the study of Tse et al. (2007).

This issue may be solve by adjusting other parameters during sleep replay, in particular the temperature parameter (see Eq. 2.4). We found that this method usually worked for the top-down reconstruction of "dream" patterns in the sensory cortices (not shown), but in practice it was difficult to implement as the temperature had to be adjusted over the course of training. In addition, instead of having a fixed number of reactivations during sleep, the model should also be able to detect when the "dream" patterns generated are consolidated and hence stop replay.

More work should be done on the advantage of sleep replay, especially with a generative model similar to the RBM (or perhaps more complex... see Deep Dream Generator[1]). It would be interesting to explore more replay schemes that go beyond sensory experience (fantasy), which could be the basis of generalisation or gist extraction.

The organisation of replay should also be revised in the model. Indeed, we saw with our last model (Chapter 4) that the reactivation of prior memories could lead to the consolidation of false memories in the neocortex. We already discussed potential alternatives in the discussion about this model (Section 4.4). Cortical-driven replay should also be further studied, because it allows the replay of consolidated memories when the hippocampus can no longer support their reactivation. This is particularly important to preserve knowledge while incorporating new information. Robins (1995) for instance used noise (i.e. random patterns) in the network to internally reactivate "pseudopatterns". These patterns were similar to those observed during experience, and this technique reduced catastrophic interference when learning new patterns. We have implemented a similar, but somewhat too simplistic method in Section 5.2.3 which we believe only worked because the network had previously learned a uniform distribution.

[1] https://deepdreamgenerator.com/

Mechanisms for regulation of memory consolidation

Throughout this work we sometimes tuned cortical plasticity, increased the number of reactivation during sleep, and we also organised which memories were selected for replay. However, we yet have to define how these processes can be regulated. The findings of Girardeau et al. (2014) that we mentioned earlier suggest that the reactivations of memories might be regulated by consolidation requirements. The hypothesis of network "tagging" (Lesburguères et al. 2011), similar to synaptic tagging and capture process (Frey et al. 1997) but at the network level, should also be investigated as a regulatory mechanism to influence systems consolidation. For instance, Lesburguères et al. (2011) found that "tagging" in the orbitofrontal cortex (OFC) during encoding was fundamental for successful remote memory formation: inactivation of this cortical area during a food preference transmission task did not affect the acquisition since memory retrieval was intact 7 days later, but it impaired retrieval 30 days later, and also prevented the development of structural plasticity in OFC.

Model architecture

We mentioned in Section 2.2 that deeper architectures could be used instead of the RBM as models of cortical processing, in particular if we want to examine more complex data (Series et al. 2010, Salakhutdinov et al. 2013). Perhaps these models could also bring to the fore new hypotheses that we would miss with simpler structures. We also emphasised that such structures have constraints that are incompatible with biology (Chapter 2), although researchers actively investigate how to make the learning algorithms more biologically plausible (Bengio et al. 2015).

In addition, sleep replay had to be suitably parametrised so that the network was able to generate patterns related to experience. Káli & Dayan (2004) circumvented this issue by identifying the patterns reconstructed in the sensory cortices and correcting them. However, we did not want to memorise the input patterns, and instead we let the network reconstruct on its own. As a result, sleep replay was sometimes difficult to calibrate. We suggest as an alternative to use for instance a semi-Restricted Boltzmann machine which has connections within the visible layer (Osindero & Hinton 2008).

Our framework could also be extended with more modules to process the input / output which might facilitate the interpretation of behavioural data. For example Santoro et al. (2016) implemented a similar system to ours within a larger framework of reinforcement learning that modelled the interaction between episodic memory,

schematic memory (RBM) and decision making for navigation.

Table 6.1: (Two pages) Summary of results with the three models of the experiments of Tse et al (2007,2011). The main experimental results are numbered and referred to in the table. Notations: flavour-place association (paired associate, PA), hippocampus (HPC), medial prefrontal cortex (mPFC)

Key experimental results	(1) Rapid Acquisition of 2 new PAs in consistent schema (2) Rapid consolidation (<48hrs), but HPC lesions too early after new training prevent it (3) Rapid acquisition did not occur if rats were previously trained in an inconsistent schema (4) Early gene expression increases in mPFC compared to when rats (i) retrieve the 6 PAs or (ii) are exposed to 6 new PAs Inhibiting mPFC (5) during new training prevent rapid acquisition of the 2 new PAs (6) during probe test impair memory retrieval for both original and new PAs

Models		**Model1 (Chapter 3)**	**Model2 (Chapter 3)**	**Model3 (Chapter 4)**
Aim		- Show why a meta-schema is required to explain experimental data (3) - Show potential role for PFC (5) - A simple toy model as we do not focus on feature extraction	Focus on cortical connectivity and incorporation of novel information to study (1) and (2) - Impact of prior semantic knowledge - Impact of consistency with prior knowledge	Study how PFC and meta-schema could impact the assimilation of new information
Model Features	Cortex	- Each flavour / location is represented by one unit in the visible layer - Plasticity for all patterns presented	- Each flavour is represented by one unit in the visible layer - Distributed representation of locations (spatial encoding) - Plasticity only for rewarded events	- Each flavour is represented by a random pattern - Distributed representation of locations (spatial encoding) - Plasticity only for rewarded events
	HPC	- Store field of hidden units of input patterns - Store reward probabilities of each input pattern - Can reinstate all associations according to reward probabilities, during recall and replay	- Store field of hidden units of input patterns - Store rewarded events only - Can only reinstate associations that were rewarded during the last epoch - Replay all stored memories with equal probability	- Store field of hidden units of input patterns, separately for flavours and locations - Store rewarded events only, and reward probability - Can reinstate any possible association according to reward probabilities and consolidation strength, during recall and replay
	PFC	Store the average maximum reward probability for all flavours	Not implemented	- Evaluate consistency of association by measuring consolidation strength in the cortex - Detect neutral, conflicting or novel events using consistency of current association vs expected consistency (meta schema) - Modulate HPC memory strength during learning

Results	- Acquisition of new PAs with a consistent or with an inconsistent prior schema is identical, unless HPC learning is regulated - The PFC, by keeping track of overall associations consistencies, could mediate this operation	- Learning and consolidation of 2 new PAs is faster with a network pre-trained on the 6 original PAs than learning from scratch - When the new location is close to an original one HPC acquisition is faster but consolidation is slower - Consolidation speed is similar to a naive network with larger initial weights	- PFC detects novelty and up-regulates HPC learning for a consistent prior schema - Learning and consolidation of 2 new PAs are thus faster for a consistent schema than an inconsistent schema (3) - Blocking PFC modulation of HPC learning prevents the rapid new acquisition in the consistent schema (5)
Predictions	- Projections from regions of PFC to HPC are required - Up-regulation of memory formation in HPC is necessary for rapid acquisition in consistent schema	- Different learning acquisition speed depending on new reward location - Less interference for the original flavours associated to reward locations close to the new one	- Blocking PFC modulation of HPC learning has no impact (i) during the training of 6 original PAs nor (ii) during the new training in the inconsistent schema - PFC inhibition during the original training should prevent the update of the meta-schema, and hence should impair the new training - With an inconsistent prior schema, up-regulating HPC learning should lead to the same results as with a consistent schema
Limitations	- Not suitable to study semantic knowledge and impact on acquisition of new information - Missing PFC-HPC regulation mechanisms, consistency and novelty detection - Learning rule: same cortical update for rewarded and non-rewarded events	- Consolidation is slow compared to experimental data - No PFC, no distinction new training in consistent vs inconsistent schemas	- Consolidation is still slow compared to experimental data - Model does not provide detailed description of the neural mechanisms - Replay in the model leads to the consolidation of false memories
What no model explains	- (2) Reasons: difficulty to encode novel features and replay them during sleep - (4)(ii) We did not test training 6 new PAs - (6) PFC action during recall not implemented		
How to address shortcomings	- PFC model: allow to deal with multiple schemas; role during recall, interaction with associative cortical areas and HPC - HPC model: too simplistic (e.g. clear out all past memories at each epoch); recall and replay should also be revised - Improve the impact of sleep replay on consolidation: e.g. parameters scheduling, attractors in visible layer (cite KD), - Cortical connectivity: other network architecture (e.g. semi-RBM) - Online tracking of consolidation requirements (Girardeau et al. (2014))		

Bibliography

Alvarez, P. & Squire, L. R. (1994), 'Memory consolidation and the medial temporal lobe: a simple network model', *Proceedings of the National Academy of Sciences* **91**(15), 7041–7045.

Atherton, L. A., Dupret, D. & Mellor, J. R. (2015), 'Memory trace replay: the shaping of memory consolidation by neuromodulation', *Trends in neurosciences* **38**(9), 560–570.

Bayley, P. J., Hopkins, R. O. & Squire, L. R. (2003), 'Successful recollection of remote autobiographical memories by amnesic patients with medial temporal lobe lesions', *Neuron* **38**(1), 135–144.

Bendor, D. & Wilson, M. A. (2012), 'Biasing the content of hippocampal replay during sleep', *Nature neuroscience* **15**(10), 1439.

Bengio, Y., Lee, D.-H., Bornschein, J., Mesnard, T. & Lin, Z. (2015), 'Towards biologically plausible deep learning', *arXiv preprint arXiv:1502.04156* .

Bero, A. W., Meng, J., Cho, S., Shen, A. H., Canter, R. G., Ericsson, M. & Tsai, L.-H. (2014), 'Early remodeling of the neocortex upon episodic memory encoding', *Proceedings of the National Academy of Sciences* **111**(32), 11852–11857.

Buzsáki, G. (2015), 'Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning', *Hippocampus* **25**(10), 1073–1188.

Carr, M. F., Jadhav, S. P. & Frank, L. M. (2011), 'Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval', *Nature neuroscience* **14**(2), 147.

Chadwick, A., van Rossum, M. C. & Nolan, M. F. (2015), 'Independent theta phase coding accounts for ca1 population sequences and enables flexible remapping', *Elife* **4**, e03542.

Clark, R. E., Broadbent, N. J. & Squire, L. R. (2005), 'Hippocampus and remote spatial memory in rats', *Hippocampus* **15**(2), 260–272.

Cowansage, K. K., Shuman, T., Dillingham, B. C., Chang, A., Golshani, P. & Mayford, M. (2014), 'Direct reactivation of a coherent neocortical memory of context', *Neuron* **84**(2), 432–441.

Dayan, P., Hinton, G. E., Neal, R. M. & Zemel, R. S. (1995), 'The helmholtz machine', *Neural computation* **7**(5), 889–904.

Ego-Stengel, V. & Wilson, M. A. (2010), 'Disruption of ripple-associated hippocampal activity during rest impairs spatial learning in the rat', *Hippocampus* **20**(1), 1–10.

Eichenbaum, H., Stewart, C. & Morris, R. (1990), 'Hippocampal representation in place learning', *Journal of Neuroscience* **10**(11), 3531–3542.

Euston, D. R., Tatsuno, M. & McNaughton, B. L. (2007), 'Fast-forward playback of recent memory sequences in prefrontal cortex during sleep', *science* **318**(5853), 1147–1150.

Fernández, G. (2017), The medial prefrontal cortex is a critical hub in the declarative memory system, *in* 'Cognitive Neuroscience of Memory Consolidation', Springer, pp. 45–56.

Fiebig, F. & Lansner, A. (2014), 'Memory consolidation from seconds to weeks: a three-stage neural network model with autonomous reinstatement dynamics', *Frontiers in computational neuroscience* **8**, 64.

Fischer, A. & Igel, C. (2014), 'Training restricted boltzmann machines: An introduction', *Pattern Recognition* **47**(1), 25–39.

Frankland, P. W. & Bontempi, B. (2005), 'The organization of recent and remote memories', *Nature Reviews Neuroscience* **6**(2), 119–130.

French, R. M. (1999), 'Catastrophic forgetting in connectionist networks', *Trends in cognitive sciences* **3**(4), 128–135.

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

Genzel, L., Kroes, M. C., Dresler, M. & Battaglia, F. P. (2014), 'Light sleep versus slow wave sleep in memory consolidation: a question of global versus local processes?', *Trends in neurosciences* **37**(1), 10–19.

Genzel, L. & Wixted, J. T. (2017), Cellular and systems consolidation of declarative memory, *in* 'Cognitive Neuroscience of Memory Consolidation', Springer, pp. 3–16.

Ghosh, V. E. & Gilboa, A. (2014), 'What is a memory schema? a historical perspective on current neuroscience literature', *Neuropsychologia* **53**, 104–114.

Girardeau, G., Benchenane, K., Wiener, S. I., Buzsáki, G. & Zugaro, M. B. (2009), 'Selective suppression of hippocampal ripples impairs spatial memory', *Nature neuroscience* **12**(10), 1222.

Girardeau, G., Cei, A. & Zugaro, M. (2014), 'Learning-induced plasticity regulates hippocampal sharp wave-ripple drive', *Journal of Neuroscience* **34**(15), 5176–5183.

Groch, S., Schreiner, T., Rasch, B., Huber, R. & Wilhelm, I. (2017), 'Prior knowledge is essential for the beneficial effect of targeted memory reactivation during sleep', *Scientific reports* **7**, 39763.

Grossberg, S. (1987), 'Competitive learning: From interactive activation to adaptive resonance', *Cognitive science* **11**(1), 23–63.

Gupta, A. S., van der Meer, M. A., Touretzky, D. S. & Redish, A. D. (2010), 'Hippocampal replay is not a simple function of experience', *Neuron* **65**(5), 695–705.

Hennies, N., Ralph, M. A. L., Kempkes, M., Cousins, J. N. & Lewis, P. A. (2016), 'Sleep spindle density predicts the effect of prior knowledge on memory consolidation', *Journal of Neuroscience* **36**(13), 3799–3810.

Hinton, G. (2010), 'A practical guide to training restricted boltzmann machines', *Momentum* **9**(1), 926.

Hinton, G. E. (2002), 'Training products of experts by minimizing contrastive divergence', *Neural computation* **14**(8), 1771–1800.

Hinton, G. E., Sejnowski, T. J. & Ackley, D. H. (1984), *Boltzmann machines: Constraint satisfaction networks that learn*, Carnegie-Mellon University, Department of Computer Science Pittsburgh, PA.

Hopfield, J. J. (1982), 'Neural networks and physical systems with emergent collective computational abilities', *Proceedings of the national academy of sciences* **79**(8), 2554–2558.

Igloi, K., Gaggioni, G., Sterpenich, V. & Schwartz, S. (2015), 'A nap to recap or how reward regulates hippocampal-prefrontal memory networks during daytime sleep in humans', *eLife* .

Jadhav, S. P., Kemere, C., German, P. W. & Frank, L. M. (2012), 'Awake hippocampal sharp-wave ripples support spatial memory', *Science* p. 1217230.

Ji, D. & Wilson, M. A. (2007), 'Coordinated memory replay in the visual cortex and hippocampus during sleep', *Nature neuroscience* **10**(1), 100.

Káli, S. & Dayan, P. (2004), 'Off-line replay maintains declarative memories in a model of hippocampal-neocortical interactions', *Nature neuroscience* **7**(3), 286–294.

Kappen, H. J. (1995), 'Deterministic learning rules for boltzmann machines', *Neural Networks* **8**(4), 537–548.

Kemp, C., Perfors, A. & Tenenbaum, J. B. (2007), 'Learning overhypotheses with hierarchical bayesian models', *Developmental science* **10**(3), 307–321.

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

Kolen, J. F. & Pollack, J. B. (1991), Back propagation is sensitive to initial conditions, *in* 'Advances in neural information processing systems', pp. 860–867.

Lavenex, P. & Amaral, D. G. (2000), 'Hippocampal-neocortical interaction: a hierarchy of associativity', *Hippocampus* **10**(4), 420–430.

Le Roux, N. & Bengio, Y. (2008), 'Representational power of restricted boltzmann machines and deep belief networks', *Neural computation* **20**(6), 1631–1649.

Lesburguères, E., Gobbo, O. L., Alaux-Cantin, S., Hambucken, A., Trifilieff, P. & Bontempi, B. (2011), 'Early tagging of cortical networks is required for the formation of enduring associative memory', *Science* **331**(6019), 924–928.

Lewis, P. A. & Durrant, S. J. (2011), 'Overlapping memory replay during sleep builds cognitive schemata', *Trends in cognitive sciences* **15**(8), 343–351.

Lisman, J., Grace, A. A. & Duzel, E. (2011), 'A neohebbian framework for episodic memory; role of dopamine-dependent late ltp', *Trends in neurosciences* **34**(10), 536–547.

Logothetis, N. K., Eschenko, O., Murayama, Y., Augath, M., Steudel, T., Evrard, H., Besserve, M. & Oeltermann, A. (2012), 'Hippocampal–cortical interaction during periods of subcortical silence', *Nature* **491**(7425), 547.

Maingret, N., Girardeau, G., Todorova, R., Goutierre, M. & Zugaro, M. (2016), 'Hippocampo-cortical coupling mediates memory consolidation during sleep', *Nature neuroscience* **19**(7), 959.

Martin, S. & Morris, R. (2002), 'New life in an old idea: the synaptic plasticity and memory hypothesis revisited', *Hippocampus* **12**(5), 609–636.

McClelland, J. L. (2013), 'Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory.', *Journal of Experimental Psychology: General* **142**(4), 1190.

McClelland, J. L., McNaughton, B. L. & O'reilly, R. C. (1995), 'Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory.', *Psychological review* **102**(3), 419.

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

McKenzie, S., Frank, A. J., Kinsky, N. R., Porter, B., Rivière, P. D. & Eichenbaum, H. (2014), 'Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas', *Neuron* **83**(1), 202–215.

McNamara, C. G., Tejero-Cantero, Á., Trouche, S., Campo-Urriza, N. & Dupret, D. (2014), 'Dopaminergic neurons promote hippocampal reactivation and spatial memory persistence', *Nature neuroscience* **17**(12), 1658.

Milner, B., Corkin, S. & Teuber, H.-L. (1968), 'Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of hm', *Neuropsychologia* **6**(3), 215–234.

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

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

Norman, K. A., Newman, E. L. & Perotte, A. J. (2005), 'Methods for reducing interference in the complementary learning systems model: oscillating inhibition and autonomous memory rehearsal', *Neural Networks* **18**(9), 1212–1228.

Norman, K. A. & O'Reilly, R. C. (2003), 'Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach.', *Psychological review* **110**(4), 611.

Osindero, S. & Hinton, G. E. (2008), Modeling image patches with a directed hierarchy of markov random fields, *in* 'Advances in neural information processing systems', pp. 1121–1128.

Peterson, C. (1987), 'A mean fied theory learning algorithm for neural networks', *Complex System* **1**, 995–1019.

Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S. I. & Battaglia, F. P. (2009), 'Replay of rule-learning related neural patterns in the prefrontal cortex during sleep', *Nature neuroscience* **12**(7), 919.

Pfeiffer, B. E. & Foster, D. J. (2013), 'Hippocampal place-cell sequences depict future paths to remembered goals', *Nature* **497**(7447), 74.

Preston, A. R. & Eichenbaum, H. (2013), 'Interplay of hippocampus and prefrontal cortex in memory', *Current Biology* **23**(17), R764–R773.

Richards, B. A., Xia, F., Santoro, A., Husse, J., Woodin, M. A., Josselyn, S. A. & Frankland, P. W. (2014), 'Patterns across multiple memories are identified over time', *Nature Neuroscience* .

Robins, A. (1995), 'Catastrophic forgetting, rehearsal and pseudorehearsal', *Connection Science* **7**(2), 123–146.

Rosenbaum, R. S., Priselac, S., Köhler, S., Black, S. E., Gao, F., Nadel, L. & Moscovitch, M. (2000), 'Remote spatial memory in an amnesic person with extensive bilateral hippocampal lesions', *Nature neuroscience* **3**(10), 1044.

Rossato, J. I., Bevilaqua, L. R., Izquierdo, I., Medina, J. H. & Cammarota, M. (2009), 'Dopamine controls persistence of long-term memory storage', *Science* **325**(5943), 1017–1020.

Rothschild, G., Eban, E. & Frank, L. M. (2017), 'A cortical–hippocampal–cortical loop of information processing during memory consolidation', *Nature neuroscience* **20**(2), 251.

Roux, L., Hu, B., Eichler, R., Stark, E. & Buzsáki, G. (2017), 'Sharp wave ripples during learning stabilize the hippocampal spatial map', *Nature neuroscience* **20**(6), 845.

Salakhutdinov, R. (2008), 'Learning and evaluating boltzmann machines', *Tech. Rep., Technical Report UTML TR 2008-002, Department of Computer Science, University of Toronto* .

Salakhutdinov, R., Tenenbaum, J. B. & Torralba, A. (2013), 'Learning with hierarchical-deep models', *IEEE transactions on pattern analysis and machine intelligence* **35**(8), 1958–1971.

Santoro, A., Frankland, P. W. & Richards, B. A. (2016), 'Memory transformation enhances reinforcement learning in dynamic environments', *Journal of Neuroscience* **36**(48), 12228–12242.

Schlichting, M. L. & Preston, A. R. (2015), 'Memory integration: neural mechanisms and implications for behavior', *Current opinion in behavioral sciences* **1**, 1–8.

Series, P., Reichert, D. P. & Storkey, A. J. (2010), Hallucinations in charles bonnet syndrome induced by homeostasis: a deep boltzmann machine model, *in* 'Advances in Neural Information Processing Systems', pp. 2020–2028.

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L. & Samuelson, L. (2002), 'Object name learning provides on-the-job training for attention', *Psychological Science* **13**(1), 13–19.

Smolensky, P. (1986), 'Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1. chapter information processing in dynamical systems: Foundations of harmony theory', *MIT Press, Cambridge, MA, USA* **15**, 18.

Stickgold, R. & Walker, M. P. (2013), 'Sleep-dependent memory triage: evolving generalization through selective processing', *Nature neuroscience* **16**(2), 139.

Tamminen, J., Ralph, M. A. L. & Lewis, P. A. (2013), 'The role of sleep spindles and slow-wave activity in integrating new information in semantic memory', *Journal of Neuroscience* **33**(39), 15376–15381.

Tanaka, K. Z., Pevzner, A., Hamidi, A. B., Nakazawa, Y., Graham, J. & Wiltgen, B. J. (2014), 'Cortical representations are reinstated by the hippocampus during memory retrieval', *Neuron* **84**(2), 347–354.

Tang, W. & Jadhav, S. P. (2018), 'Sharp-wave ripples as a signature of hippocampal-prefrontal reactivation for memory during sleep and waking states', *Neurobiology of learning and memory* .

Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. (2011), 'How to grow a mind: Statistics, structure, and abstraction', *science* **331**(6022), 1279–1285.

Tieleman, T. (2008), Training restricted boltzmann machines using approximations to the likelihood gradient, *in* 'Proceedings of the 25th international conference on Machine learning', ACM, pp. 1064–1071.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., Witter, M. P. & Morris, R. G. (2007), 'Schemas and memory consolidation', *Science* **316**(5821), 76–82.

Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., Bito, H. & Morris, R. G. (2011), 'Schema-dependent gene activation and memory encoding in neocortex', *Science* **333**(6044), 891–895.

Tulving, E. (1972), 'Episodic and semantic memory 1', *Organization of Memory. London: Academic* **381**(4), 382–404.

van Kesteren, M. T., Beul, S. F., Takashima, A., Henson, R. N., Ruiter, D. J. & Fernández, G. (2013), 'Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: from congruent to incongruent', *Neuropsychologia* **51**(12), 2352–2359.

van Kesteren, M. T., Ruiter, D. J., Fernández, G. & Henson, R. N. (2012), 'How schema and novelty augment memory formation', *Trends in neurosciences* **35**(4), 211–219.

Warren, D. E., Jones, S. H., Duff, M. C. & Tranel, D. (2014), 'False recall is reduced by damage to the ventromedial prefrontal cortex: implications for understanding the neural correlates of schematic memory', *Journal of Neuroscience* **34**(22), 7677–7682.

Welling, M. & Hinton, G. E. (2002), A new learning algorithm for mean field boltzmann machines, *in* 'International Conference on Artificial Neural Networks', Springer, pp. 351–357.

Wiltgen, B. J., Zhou, M., Cai, Y., Balaji, J., Karlsson, M. G., Parivash, S. N., Li, W. & Silva, A. J. (2010), 'The hippocampus plays a selective role in the retrieval of detailed contextual memories', *Current Biology* **20**(15), 1336–1344.

Winocur, G. (1990), 'Anterograde and retrograde amnesia in rats with dorsal hippocampal or dorsomedial thalamic lesions', *Behavioural brain research* **38**(2), 145–154.

Winocur, G. & Moscovitch, M. (2011), 'Memory transformation and systems consolidation', *Journal of the International Neuropsychological Society* **17**(5), 766–780.

Zola-Morgan, S. M. & Squire, L. R. (1990), 'The primate hippocampal formation: evidence for a time-limited role in memory storage', *Science* **250**(4978), 288–290.