Dynamics of memory transformation:

Updating our knowledge of systems consolidation

Dissertation zur Erlangung des Doktorgrades an der Universität Hamburg

Fakultät für Psychologie und Bewegungswissenschaften,
Institut für Psychologie

vorgelegt von Lisa Catherine Dandolo
Hamburg, 2019

Tag der mündlichen Prüfung: 22.11.2019

Promotionsprüfungsausschuss

Disputationsvorsitz:	Prof. Dr. Jan Wacker
1. Dissertationsgutachten:	Prof. Dr. Lars Schwabe
2. Dissertationsgutachten:	Prof. Dr. Frank Rösler
1. Disputationsgutachten:	PD Dr. Kirsten Hötting
2. Disputationsgutachten:	Prof. Dr. Martin Spieß

Acknowledgements

Throughout the last years, whenever I was asked what I am doing, my answer has always been "still working on my thesis", so it feels quite surreal to finally write my last words today. It is time to thank all the people that supported me throughout this journey. First and foremost, I would like to thank all my colleagues in the lab in Hamburg: Prof. Dr. Lars Schwabe for his help and support throughout the years and for his willingness to accept my special liking for discussions; Lisa, Lisa and Mario for paving the way for me, proving that it is possible to finish a thesis; Nadine, Franzi, Anna, Gundi and Felix for bringing new life, laughter and "Sekt" to the lab; Conny and Susanne for the never-ending knowledge; as well as all the hard working students for their support during data collection. I also want to thank my colleagues at my new lab in Bremen for making me feel welcome and cheering me on in the last couple of months. Thank you, Lisa and Stella for your helpful comments on my thesis and especially thank you Nouna for spending all those long hours helping me to improve my English. A big thank you also to every single one of my friends who are too many to name individually; I have often neglected you over the last years and am looking forward to finally spending my free weekends and evenings with you. My biggest thank you and gratitude, however, goes to my family: to my parents for all your love and support, not only during the last years but also throughout all the long years of my studies and, in fact, all my life; to my big sister, for being my biggest role model in life, especially showing me how to stay positive no matter what comes; and of course to my husband for all your love and patience, accepting all my long hours of work and putting your life on hold to allow me to finish this thesis. Thank you all so much.

Summary

Memories are essential to everything in our lives and we constantly rely on them in all our actions and decisions. How memories are represented in our brain is therefore decisive to an overall better understanding of our brain functions, behaviour and cognition. However, a fundamental aspect of memory, i.e. the process of systems consolidation is still under vigorous debate. Different opinions exist on how memory representations change over time and how these changes are related to the transformation of memory quality. Especially the role of the hippocampus in representing detailed remote memories is still controversially discussed. This thesis contributes to this ongoing debate by examining different aspects of systems consolidation and relating them to the concept of memory dynamics. A review of different systems consolidation studies revealed, that varying time intervals have been used to test this process. For a more precise comparison of systems consolidation studies, I therefore propose a division into an early systems consolidation phase, comprising changes at brain systems level during the decisive first night of sleep, and a prolonged phase, examining consolidation during the following weeks. While investigating prolonged systems consolidation of declarative memories, we additionally illustrated the importance of dividing the hippocampus along its long axis. The transformation from detailed to gist-like memories across time was accompanied by changes along this axis, with anterior hippocampus involved in retrieving detailed memories and posterior hippocampus representing gist-like memories. We also investigated the prolonged systems consolidation process of nondeclarative motor memories, highlighting the involvement of the lateral prefrontal cortex in representing remote motor sequence memories. Finally, we demonstrated that memories which have already undergone an early systems consolidation phase can be dynamically updated by new information through a neural process orchestrated by the dorsolateral prefrontal cortex. Conclusively, this thesis adds valuable information to the present systems consolidation theories by focussing on more precise research questions regarding both brain structures and time intervals.

Contents

List of Figures

List of Abbreviations

CA Cornu Ammonis

DG Dentate Gyrus

dlPFC Dorsolateral Prefrontal Cortex

EEG Electroencephalography

FA False Alarm

fMRI Functional Magnetic Resonance Imaging

mPFC Medial Prefrontal Cortex

MTL Medial Temporal Lobe

MTT Multiple Trace Theory

RDM Representational Dissimilarity Matrix

ROIs Regions of Interest

RSA Representational Similarity Analysis

SCT Standard Consolidation Theory

SECPT Socially Evaluated Cold-Pressor Test

SRTT Serial Reaction Time Task

TTT Trace Transformation Theory

vmPFC Ventromedial Prefrontal Cortex

1 General introduction

Memories are crucially important in our daily lives. They influence our personalities, behaviours, decision making and future planning (Wimmer & Shohamy, 2012; Fellows, 2016 Schacter, Addis, & Buckner, 2007). All our knowledge about the world is based on memories of past experiences. When we lose our memories, we not only lose this knowledge about the world and information about our own past, but also part of what defines us as individuals and makes us function and survive in a complex society.

Therefore, understanding how we acquire, store, retrieve and use our memories has been an important area of research for many decades within different disciplines, including both cognitive psychological and neuroscience. Different lines of research have focussed on various aspects of memory and tried to describe different memory processes and memory systems. Throughout the years different memory theories and ideas have been developed and at times new findings and discoveries led to major revisions of original textbook views of memory. For example, over a long period of time the general view was that after an initial process of memory consolidation is complete, memories are stable and fixed, only to be disrupted by major brain injuries or degenerative diseases. However, this traditional textbook view had to make way for the more recent idea of memories being highly dynamic traces, that can be changed, transformed, updated or distorted each time a memory is retrieved or replayed (Nadel, Hupbach, Gomez, & Newman-Smith, 2012; Kroes & Fernandez, 2012; Dudai, Karni, & Born, 2015).

Thus, just as dynamic as memories are now thought to be, just as dynamically have theories about the process of memory consolidation changed and transformed throughout the years and are still updated whenever new information becomes available. However, this dynamic process has also led to the emergence of different, competing theories and, to date, there is no consensus amongst memory researchers on how exactly memory consolidation is performed within our brains (Squire & Bayley, 2007; Squire, Genzel, Wixted, & Morris, 2015; Nadel, Winocur, Ryan, & Moscovitch, 2007; Moscovitch, Cabeza, Winocur, & Nadel, 2016; Sekeres, Moscovitch, & Winocur, 2017).

One of the primary goals of this thesis is to test current competing theories of memory consolidation. Yet, in order to understand the ongoing debate about memory consolidation, it is important to track the formation of these different theories throughout the history of memory research, as each new theory is not only based on new results and experimental data but is also strongly influenced by former beliefs and biases towards the one or other preceding theory. The first section of this general introduction therefore introduces the process of memory consolidation from a historical point of view, in order to better understand the emergence of the current memory consolidation theories.

1.1 Memory consolidation

Overall, memory functions can be divided into different processes, such as memory encoding, memory consolidation, and memory retrieval. Memory encoding and memory retrieval are quite easily defined: memory encoding is the process of initially perceiving and learning events (McDermott & Roediger, 2019), while memory retrieval is the process of accessing and reactivating internally stored information when needed (Dudai, 2002; McDermott & Roediger, 2019). But what exactly is the process of memory consolidation?

1.1.1 Historical background

The term 'consolidation' was introduced to memory research over 100 years ago by German scientists Müller and Pilzecker (1900) and stems from the Latin word *consolidare* that translates to 'make firm together' (Dudai, 2004; Dudai et al., 2015). Müller and Pilzecker (1900) had conducted a retroactive interference experiment in healthy participants showing that the presentation of new material, shortly after learning, disrupted the memory for the originally learned material. They suggested a perseveration-consolidation hypothesis, stating that memories initially exist in a fragile, transient state and only after some time *consolidate* into a stable version immune to disruption through new information (McGaugh, 2000; Hardt & Nadel, 2018). Two decades before these experiments Ribot (1882) had already described the phenomenon of a temporal gradient in retrograde amnesia in patients with traumatic head injury. This gradient, later named Ribot gradient implies that memories encoded shortly before the traumatic incident are lost while more remote memories are intact (Hardt & Nadel, 2018). This temporal gradient of retrograde amnesia was subsequently portrayed in support of Müller and Pilzecker (1900)'s idea of memory consolidation (Squire, 1992).

Thus, the idea of memory consolidation as a process of stabilizing initially fragile memories was introduced at the turn of the 20th century, a time that was later named the "Golden Age" of memory research (Rozin, 1976; Moscovitch et al., 2005). However, it was then forgotten for nearly 50 years until research in military veterans showing a Ribot gradient (Russell, 1946) and animal research by Hebb (1949) and others (Gerard, 1949; C. Duncan, 1949) initiated a renewed interest in the topic (McGaugh, 2000; Hardt & Nadel, 2018). Hebb (1949) formulated a biological dual-trace model for memory consolidation: a short-term memory trace exists of each experience in the form of reverberating neural activity within an assembly of neurons. If this activity reverberates long enough it can lead to

long-term traces in the form of structural changes at the synapses of the set of neurons. These newer studies therefore re-emphasized the time-dependency of memory consolidation, with time assumed to allow the stabilization of memories. However, up to this point it was completely unknown which parts of the brain were involved in this assumed stabilization process. In fact, it was commonly assumed that memory was distributed throughout the cortex (Lashley, 1929) with some accounts suggesting that different brain areas store different features of memory (Hebb, 1949), but no region was thought to be disproportionally involved in memory (Squire & Wixted, 2011). This drastically changed with the highly influential paper of Scoville and Milner (1957), the first researchers to discover, by coincidence, the importance of the medial temporal lobe (MTL) for memory. They described the case of the now famous patient H.M., who suffered from severe epilepsy leading to the decision to surgically remove his MTLs in both hemispheres. After the surgery H.M. showed severe anterograde and a temporally graded retrograde amnesia without any loss in intelligence and no perceptual disorders. This paper triggered many follow-up studies, leading to a special focus on the importance of one structure within the MTL, i.e. the hippocampus and its adjacent cortices (entorhinal, perirhinal, and parahippocampal) for memory (Squire & Zola-Morgan, 1991; Squire & Wixted, 2011).

When examining the time frame postulated for memory consolidation more precisely, it becomes evident that in some cases the described time interval is seconds to a few hours (e.g. Hebb, 1949; Müller & Pilzecker, 1900), while in others it is days, years and even decades (e.g. Ribot, 1882; Russell, 1946; Scoville & Milner, 1957). Thus, later accounts of memory consolidation divide these processes into two different categories that also depend on two different levels of brain structures: synaptic consolidation and systems consolidation (Dudai, 2004).

1.1.2 Synaptic consolidation and systems consolidation

Synaptic consolidation (or cellular or rapid consolidation) takes place within hours after acquisition and manifests itself in modifications of synapses (leading to the term synaptic consolidation), and other molecular changes at cellular level (Dudai, 2004; McGaugh, 2000). This process has been extensively studied throughout the years (for reviews see e.g. Kandel, 2001; Kandel, Dudai, & Mayford, 2014) and its existence is widely accepted among memory researchers (Dudai et al., 2015; Sekeres et al., 2017). Systems consolidation, on the other hand, refers to the reorganization of memory traces across large neuronal networks or brain circuits, thus changes at the brain systems level. This reorganization

can take days, months or even decades (Dudai, 2004).

The focus of this thesis is on systems consolidation and the ongoing debate about this process. Nevertheless, it is important to keep in mind, that synaptic and systems consolidation do belong to the same higher-level process and therefore naturally interact (Sekeres et al., 2017). In fact, Dudai and colleagues have recently argued that recurrent waves of synaptic consolidation are necessary for systems consolidation, thereby regarding synaptic consolidation processes as subroutines in systems consolidation (Dudai, 2012; Dudai et al., 2015).

1.1.3 Ongoing debate about systems consolidation

The classic, standard idea about systems consolidation is based on the above described reports of temporally graded retrograde amnesia in patients such as H.M. and describes a time-dependent reorganization of initially hippocampus-dependent memories. This Standard Consolidation Theory (SCT) thereby postulates that the components of memories are stored in different neocortical areas but initially the hippocampus is needed to bind those components together. During the systems consolidation process neocortical connections are strengthened until memories can be retrieved from the neocortex without involving the hippocampus (Squire, 1992; Alvarez & Squire, 1994; Squire & Alvarez, 1995). Lately, this standard view has been challenged by several new ideas and theories. Most prominent among them is the Multiple Trace Theory (MTT) introduced over 20 years ago (Nadel & Moscovitch, 1997; Nadel, Samsonovich, Ryan, & Moscovitch, 2000; Moscovitch et al., 2005) and a more recent extension of this theory, the Trace Transformation Theory (TTT; Winocur, Moscovitch, & Bontempi, 2010; Winocur & Moscovitch, 2011). At the centre of the debate are the questions whether remote detailed memories continue to rely on the hippocampus (MTT/TTT) or not (SCT) and whether memories that can be retrieved from the neocortex alone have the same qualities as hippocampus-dependent memories (SCT) or whether they have been qualitatively transformed (TTT)? Even after 20 years of research examining these questions, the debate about systems consolidation[1] is ongoing with proponents of both theories presenting studies favouring their ideas. I will introduce each of the systems consolidation theories in more detail in section 1.3 and will present a

[1] In this thesis the term *systems consolidation* is used as a general term to refer to time-dependent reorganization of memory traces across brain systems, in direct comparison to synaptic consolidation at cellular level, and not in the narrower sense described by the classic view of this process. To refer to this classic view I will use the term *Standard Consolidation Theory (SCT)*.

selection of animal studies, patient studies as well as neuroimaging studies in support of the respective theories.

At this point, it is essential to mention that sleep plays a decisive role in memory consolidation. I will describe the consolidation processes during sleep in more detail in section 1.5.1. For now, it is important to note that the first night of sleep after encoding is important for the fate of the memory traces and already leads to changes in the representation of these traces at brain systems level (Gais et al., 2007). I will use the term *early systems consolidation* to refer to these early changes at brain systems level after the first night of sleep. In contrast, I will refer to changes at brain systems level in the time frame of several days to weeks and months with the term *prolonged systems consolidation*. I will discuss this proposed differentiation between early and prolonged systems consolidation more thoroughly in the general discussion of this thesis in section 3.2.

Before continuing with a more detailed description of the opposing systems consolidation theories in section 1.3, I will introduce the concept of *multiple memory systems* in the following section 1.2, as it is important to understand to what type of memories the systems consolidation theories apply. Indeed, the ongoing debate between SCT and MTT/TTT focusses exclusively on systems consolidation of one type of memory, namely *declarative* memories. However, recently, researchers have also started to show interest in systems consolidation of so called *nondeclarative* memories (Dudai et al., 2015; see section 1.4). In line with this recent interest in this topic, one of the research questions of this thesis is whether systems consolidation of nondeclarative memories shows parallels to systems consolidation of declarative memories. In the following section I will therefore describe the differences between the declarative and nondeclarative memory systems, starting with an historical view of this division from the beginnings of memory research.

1.2 Multiple memory systems

1.2.1 Declarative and nondeclarative memory systems

The idea that more than one type of memory exists, was already expressed more than 100 years ago by William James (1890). In his book he wrote one chapter about memory and another about habits, thus expressing two essentially different ways of learning. Since then researchers have repeatedly suggested a division into two different forms of memory, thereby using different terminology, e.g. McDougall (1923) distinguished between explicit and implicit forms of recognition memory, and Ryle (1949) made a difference between knowing that and knowing how (Squire & Dede, 2015).

The start of experimental and biological research on multiple memory systems was again motivated by the famous patient H.M. (Scoville & Milner, 1957). After bi-hemispheric MTL resection H.M. showed severe memory impairments e.g. for events, facts and words but was, surprisingly, capable of learning a mirror drawing task, requiring hand-eye coordination, over multiple training sessions across days. Strikingly, he improved in the task, without having any memory of performing it the previous days (Milner, 1962). This encouraged researchers to distinguish between a cognitive MTL-dependent memory system and another system, responsible for learning motor skills, depending on extra-hippocampal structures.

With time, it was discovered that motor skills were not the only skills intact in patients with MTL damage, e.g. patients could also learn perceptual skills, like reading mirror-reversed words (Cohen & Squire, 1980). As a consequence, all preserved skills were classified as *procedural* knowledge, defined as knowledge in form of an acquired procedure, expressed only through performance. This *procedural* knowledge was contrasted to *declarative* knowledge, defined as knowledge characterized by conscious recollection and expressed in form of remembered materials including e.g. images, words, ideas, sensations, sounds (Cohen & Squire, 1980; see also Squire & Dede, 2015). Later, it was established that procedural knowledge is only one form of *nondeclarative* memory (Squire & Zola-Morgan, 1988; Squire & Knowlton, 1995), with other forms being identified such as priming (Tulving & Schacter, 1990), habit learning (Knowlton, Mangels, & Squire, 1996) and classical conditioning (Clark & Squire, 1998). Thus, the term *nondeclarative* should be seen as an umbrella term referring to multiple memory subsystems (Squire, 2004; Squire & Wixted, 2011).

Although this division of nondeclarative memories in many subsystems shows that there are more than two memory systems, much recent research has, nevertheless, focussed on

the distinction between a *cognitive* hippocampus-based memory system and a *habit* dorsal striatum-based system (White & McDonald, 2002; Schwabe, 2013). Questions arise on whether these systems work independently and in parallel (McDonald & White, 1994) or whether they interact cooperatively (McIntyre, Gold, & Marriott, 2003) or competitively (Poldrack et al., 2001). It has been suggested that the use of these different memory systems can depend on various factors, e.g. a shift from the cognitive to the habit system can be introduced by practice (Packard & McGaugh, 1996), distraction (Foerde, Knowlton, & Poldrack, 2006), or stress (Schwabe, 2013).

Declarative and nondeclarative memories are sometimes also termed *explicit* and *implicit* memories, respectively (Graf & Schacter, 1985). Explicit, declarative memories are defined as conscious memories, while implicit, nondeclarative memories are seen as unconscious memories, e.g. learning a motor skill without realizing that one has learned a procedure and without any recollection of actively learning it (Squire & Zola, 1996; Squire & Dede, 2015)[2].

Importantly, both declarative and nondeclarative memories are classified as *long-term memories*, thus memories that are stored for later retrieval. Note that another entirely different memory system exists, known as *working memory* (formerly also called short-term memory or immediate memory). This memory system has only a limited capacity, in other words, it can only maintain a limited amount of information at every point in time. On the other hand, this information can be actively "worked" with, thus it is used for ongoing cognitive functions such as reasoning (Baddeley, 2003; Shrager, Levy, Hopkins, & Squire, 2008).

As described above nondeclarative memory is an umbrella term for many subsystems (e.g. procedural memories, priming, habit learning), yet it is beyond the scope of this thesis to portray each of these systems in more detail. Importantly, however, declarative memories can also be divided into subsystems, and as this division is decisive for the debate between SCT and MTT/TTT, the next section will describe these declarative subsystems in more detail.

[2] There are some memory researchers that would recommend the usage of the terms explicit and implicit instead of declarative and nondeclarative, as they claim that the term declarative is hard to define in a noncircular way (Nadel et al., 2007). On the other hand, other memory researchers have used the distinction declarative vs. nondeclarative in fairly recent reviews (e.g. Dudai et al., 2015; Squire & Dede, 2015), leading to the choice of these latter terms in this thesis.

1.2.2 Declarative memory subsystems

Episodic and semantic memory systems

Tulving (1972) subdivided declarative memories into two categories: episodic and semantic memories. Semantic memories incorporate facts and knowledge about the world, while episodic memories were defined by Tulving (1983) as re-experiencing specific events with respect to both space and time. Episodic memories rely on autonoetic consciousness, a sense of being aware of one's own existence in subjective time, also defined as self-knowing (Tulving, 1985), a requirement for recollecting experiences and mental time travel (Tulving, 2002). Semantic memories, on the other hand, require noetic consciousness, a sense of knowing about objects and events, thus having knowledge about the world (Tulving, 1985). Much of the original debate between SCT and MTT focussed on the question whether both semantic and episodic memories can be retrieved without the involvement of the hippocampus after systems consolidation (SCT; Alvarez & Squire, 1994), or whether the memories that can be retrieved from the neocortex are semantic in quality, and the retrieval of truly episodic memories requires an intact hippocampus (MTT; Nadel & Moscovitch, 1997).

Context-dependent and context-free memories

As much of the research on systems consolidation is conducted on animals, it is important to consider that, according to Tulving, episodic memories accompanied by autonoetic awareness and self-knowledge are unique to humans and cannot be found in other animals (Tulving, 2002). Nevertheless, researchers have tried to find homologues for episodic and semantic memories in animals. This is best accomplished by a distinction between *context-dependent* (episodic-like) and *context-free* (semantic-like) memories. Context-dependent memories are observed when animals express the acquired behaviour only in those retrieval situations that are set in the same context as the training situation was, but not in other contexts. Context-free memories are observed when animals also show the learned behaviour in other contexts, thus when they cannot distinguish between two similar contexts (Rosenbaum, Winocur, & Moscovitch, 2001; Sekeres et al., 2017).

Recollection and familiarity processes

An important concept closely related to the notion of remembering episodes and knowing about semantic facts, is the commonly used division of recognition processes into *recol-*

lection and *familiarity*. In recognition tasks, participants are presented with previously learned stimuli intermixed with new stimuli and are asked to indicate with 'yes' or 'no' whether they have seen the stimuli before. According to a dual-process model (Yonelinas, 1994; Yonelinas, 1999) this task can be solved using recollection, a process of actually remembering the learning of the respective stimuli during the encoding session, or alternatively using familiarity, a sense of knowing the stimuli without actually recollecting any details about where and when it was learned. A whole line of research (Yonelinas, 2002; Eichenbaum, Yonelinas, & Ranganath, 2007; Diana, Yonelinas, & Ranganath, 2007; Squire, Wixted, & Clark, 2007; Ranganath, 2010; Norman, 2010 Yonelinas, Aly, Wang, & Koen, 2010) has focussed on discriminating these two processes and identifying the brain systems involved in each process. Although this research line does not focus on remote memories per se, one can still relate it to the central questions in the systems consolidation debate, as I will outline in section 1.3.4.

Spatial memories

When considering subsystems of declarative memories, it is important to acknowledge that historically one subgroup of declarative memories, namely *spatial* memory, has been given a special status in memory research (O'Keefe & Nadel, 1978; O'Keefe, 1991; Nadel, 1991). Spatial memories are memories for places, environments, spatial layouts or the spatial contexts within an episode. Building on influential rodent studies, such as the discovery of hippocampal place cells (O'Keefe & Dostrovsky, 1971), O'Keefe and Nadel (1978) introduced the *cognitive map theory* suggesting that the main role of the hippocampus is to create and maintain allocentric (viewpoint-independent) spatial representations of the environment for as long as they exist. This sparked a debate about whether the hippocampus really does represent spatial memories in a unique way (O'Keefe, 1991; Nadel, 1995) or whether these spatial memories are just one good example of the more general category of declarative memories, which are all initially represented in the hippocampus (Squire, 1992; Zola-Morgan, Squire, & Ramus, 1994). Subsequently, researchers initially supporting the cognitive map theory proposed the newer MTT (Moscovitch, Nadel, Winocur, Gilboa, & Rosenbaum, 2006), and thereby acknowledged that spatial memories, as all other declarative memories, can be divided into two categories: detailed spatial memories (perceptually detailed representations of landmarks, houses, fields, etc.) and semantic spatial memories (schematic representation of topography, location of only the major landmarks, enough to navigate, but without detailed re-experiencing); with only the former depending on the

hippocampus at remote time points (Moscovitch et al., 2005; Rosenbaum et al., 2001).

Autobiographical memories

Another form of declarative memory that justifies closer consideration is *autobiographical memory*. This term is explicitly used for memories of events in everyday life in comparison to remembering stimuli in experimental memory tasks in the laboratory. Thus, autobiographical memories are events from one's own personal past. These are normally classified as episodic memories, as they involve detailed recollections of events and experiences of the past. Importantly, however, not all memories of one's personal past have episodic and recollective qualities. We also know facts about ourselves that fit better into the class of semantic memories (e.g. where we lived as a child, which school we attended). These memories have been termed *personal semantics* (Cermak & O'Connor, 1983; Nadel & Moscovitch, 1997; Moscovitch et al., 2005). When using autobiographical memory tasks to examine whether detailed episodic memories are dependent on the hippocampus or not, it is therefore important to make a distinction between truly episodic autobiographical memories and personal semantics.

Differences between semantic, schematic and gist-like memories

Finally, another important distinction within declarative memories to discuss is the difference between *semantic, schematic and gist-like memories*. In light of the TTT, the terms gist-like memory or schematic memory have at times been utilized as a synonym for the term semantic memory. However, most recently Sekeres, Winocur, and Moscovitch (2018) have suggested that these three terms should be defined more clearly and used for distinct concepts. Let us use Christmas Eve celebrations as an example: *semantic* memories are very general concepts, facts and world knowledge (e.g. what does Christmas Eve mean in general?, when is it celebrated?); *schematic* memories, or schemas are knowledge about certain types of events extracted from several similar occurrences of such events (e.g. what normally happens at Christmas Eve in my family, like eating a special type of meal, unwrapping presents); *gist-like* memories, however, are memories of single events (e.g. last year's Christmas Eve). In this sense, gist-like memories might be defined as a form of episodic memory, as they refer to only one single episode. Still, gist-like memories contain only the coarse-grained, global or central features of the event (e.g. last year on Christmas Eve our dog was very sick) and are therefore substantially different from the truly detailed episodic memories that include fine-grained spatial and temporal elements of the event

(e.g. I was sitting to the left of my sister at the table; Sekeres, Winocur, & Moscovitch, 2018). The differences between these three concepts have often been neglected in the past, but may be important to the systems consolidation debate. I will return to this issue in the general discussion of this thesis in section 3.1.

1.3 Systems consolidation of declarative memories

As outlined above there is a continuing debate about systems consolidation of declarative memories and one of the main aims of this thesis is to compare and test the competing systems consolidation theories. These theories will therefore be introduced in more detail below (section 1.3.1). Subsequently, I will discuss the key issue of the debate, i.e. the role of the hippocampus in representing remote episodic memories by introducing some of the experimental studies supporting the ideas of each theory (section 1.3.2).

1.3.1 Systems consolidation theories

Standard Consolidation Theory (SCT)

The ideas behind the SCT were derived from the early accounts of memory consolidation as portrayed in section 1.1.1 and were articulated in specific theories in the 1980s and 1990s (Squire, Cohen, & Nadel, 1984; Teyler & DiScenna, 1985; Teyler & DiScenna, 1986; Alvarez & Squire, 1994; McClelland, McNaughton, & O'Reilly, 1995). Previously, Marr (1971) had already acknowledged the special role of the hippocampus as a region that forms simple memory representations, but added that the hippocampal pyramid cells later transfer the information to the neocortex. This latter point subsequently changed to the idea, that single memory components are already represented in the neocortex from the point of memory acquisition with the hippocampus initially required as a pointer (Squire et al., 1984) or index (Teyler & DiScenna, 1986) to bind the different representations together. Thus, importantly, memories are not thought to be literally transferred from the hippocampus to the neocortex. Instead hippocampal activations during consolidation repeatedly coactivate the respective neocortical areas, thereby strengthening the neocortical connections till they can be coactivated without the involvement of the hippocampus. Alvarez and Squire (1994) added a simple computational model to these ideas and stressed that the MTL learns quickly but only has a limited capacity, whereas the neocortex learns slowly but has a larger capacity. Subsequently, McClelland et al. (1995) proposed the influential *complementary learning system* computational model. This SCT model proposes that the hippocampus and neocortex are the sites of two complementary learning systems: the hippocampus uses sparse coding and non-overlapping representations to store episodes or facts quickly without allowing interference from similar experiences (*pattern separation*, see in more detail in section 1.3.5). The hippocampus then initiates the slow learning in the cortical system, which allows overlapping representations for similar stimuli, storing

shared structures of events.

Importantly, within these models of SCT, it was stated that the memory consolidation process is identical for both episodic and semantic declarative memories (Squire, 1992; McClelland et al., 1995).

Multiple Trace Theory (MTT)

After reviewing all the available data about systems consolidation, Nadel and Moscovitch (1997) concluded that the SCT could not incorporate all the experimental results, necessitating a new theoretical approach. Nevertheless, Nadel and Moscovitch (1997) do agree with the SCT on the idea that the hippocampal complex rapidly encodes all information and acts as a pointer or index binding together neocortical areas that represent different components of the memory. They elaborate that the whole hippocampal-neocortical ensemble can be defined as the memory trace for the episode. They propose that each retrieval of the memory leads to a new memory trace that again has an index in the hippocampus which points to similar (partly the same) representations in the neocortex. Therefore, with time, multiple memory traces exist for each memory. These multiple traces facilitate the extraction of semantic information from the episode and can lead to an integration of this information with pre-existing semantic representations in the neocortex. Thus, MTT agrees with SCT that semantic memories can be retrieved independently from the hippocampal complex after systems consolidation. However, in clear distinction to the SCT, Nadel and Moscovitch (1997) suggest that the retrieval of detailed episodic memories continues to rely on the involvement of the hippocampal complex. They later also added a computational model to simulate the proposed mechanism of MTT (Nadel et al., 2000).

Trace Transformation Theory (TTT)

After a decade of research and thorough debate about the ideas of MTT, Moscovitch and colleagues slightly revised their initial ideas and introduced the TTT (Winocur et al., 2010). This theory explicitly states that, across time, the detailed hippocampus-dependent memory supports the evolving of a second version of the same memory in the neocortex. This neocortical memory trace, however, is a more gist-like, generic memory of the event, lacking the contextual details of the hippocampal memory. In other words, according to the TTT, the progression of memories from a hippocampus dependent representation to an exclusively neocortical representation involves a memory transformation from detailed to gist-like versions of the memory. Importantly this does not necessarily imply that the

detailed hippocampus-dependent version of the memory is automatically lost. On the contrary, both memories can co-exist and there can be a dynamic interplay between them. The task or circumstances of retrieval may then lead to the dominance of one version over the other (Winocur et al., 2010; Winocur & Moscovitch, 2011). Thus, TTT takes the dynamic aspect of memory into account and addresses the issue of qualitative changes of a particular, single episode, an aspect that was not yet addressed in the MTT (Sekeres, Winocur, & Moscovitch, 2018).

Alternative theories

Although the ongoing debate between the two main systems consolidation theories, i.e. SCT and MTT/TTT, is the main focus of this thesis, it should nevertheless be mentioned that some alternative theories have additionally emerged over the years.

For example, Sutherland, Lehmann and colleagues have introduced an *overshadowing theory* (Lehmann et al., 2009; Sutherland, Sparks, & Lehmann, 2010) and a *multiple storage site hypothesis* (Sutherland & Lehmann, 2011). In brief, they claim that memories can be acquired by different memory systems in parallel. But, under normal circumstances, the hippocampal system overshadows and dominates other systems. If multiple distributed learning episodes are performed the overshadowing is overcome by the slow parallel learning in other systems. Thus, memories acquired through multiple reiterations of the experiences can survive hippocampal damage (but see Winocur, Moscovitch, and Sekeres (2013) for arguments against this theory).

Another line of work has compared the retrieval of episodic memories to the imagination of future events (Addis, Wong, & Schacter, 2007) or fictional experiences (Hassabis, Kumaran, & Maguire, 2007; Hassabis, Kumaran, Vann, & Maguire, 2007), proposing that similar brain areas are involved in all these tasks. This work led to the introduction of a *scene construction theory* (Hassabis & Maguire, 2007; Hassabis & Maguire, 2009; Maguire & Mullally, 2013), proposing that the hippocampus is central to the retrieval of vivid memories, imagination of experiences and spatial navigation tasks, as it is responsible for the construction of the scenes that allow the detailed events to be bound into a coherent spatial context. Importantly, *scene construction theory* agrees with MTT/TTT that the hippocampus is involved in the recall of vivid memories in perpetuity (Zeidman & Maguire, 2016).

1.3.2 The role of the hippocampus in remote episodic memory representation

While all the above described systems consolidation theories of declarative memories agree on both the importance of the hippocampus during encoding and retrieval of recent memories and the possibility of retrieving semantic remote memories without the involvement of the hippocampus, the key discussion point is the role of the hippocampus in retrieving remote detailed episodic memories. Different experimental approaches have been used to try to find answers to this key question and the following sections will describe some of the conducted experiments and their results, thereby also addressing the methodological advantages and problems each experimental approach entails.

Temporal gradients in human amnesic patients

The ideas of SCT are partly based on the observation of temporal gradients in retrograde amnesia in patients with MTL damage (Scoville & Milner, 1957; Russell, 1946). These gradients imply the loss of recent retrograde memories while remote memories are preserved, suggesting an independence of these remote memories from the hippocampus.
However, when introducing the MTT, Nadel and Moscovitch (1997) argued that a closer inspection of some of the patient studies showed that only semantic memories displayed such temporal gradients, while episodic memories were completely lost after MTL damage (Damasio, Eslinger, Damasio, & Van Hoesen, 1985; Tulving, Schacter, McLachland, & Moscovitch, 1988). To further test this hypothesis, Moscovitch, Yaschyshyn, Ziegler, and Nadel (1999) introduced a new scoring technique for autobiographical memories that allows the testing of detailed memories. Using this technique, they could show that amnesic patients had problems recalling truly detailed memories, even from early childhood (Moscovitch et al., 1999). The famous patient H.M. was also retested with a similar new scoring technique (Steinvorth, Levine, & Corkin, 2005), showing no detailed recollection of remote time periods. This study, however, was subsequently criticized by Squire and Wixted (2011), suggesting that these new deficits in H.M. could be related to new brain deficits that developed in other brain areas years after the surgery.
Proponents of the MTT found further support for their idea in meticulously described single case studies of patients. For example, patient V.C. showed a temporal gradient for semantic information, but episodic memory was lost across the whole life time (Cipolotti et al., 2001; Nadel & Moscovitch, 2001); patient K.C. showed intact semantic spatial memo-

ries, enough to navigate in his home town, but had lost all truly detailed spatial memories (Rosenbaum et al., 2000; Rosenbaum, McKinnon, Levine, & Moscovitch, 2004); and a former London taxi driver could still navigate along main roads after hippocampal damage but could not travel on less central roads (Maguire, Nannery, & Spiers, 2006). Yet, these studies were criticized by SCT supporters for being single case studies and for a lack of precision in the description of the respective brain damage (Squire & Bayley, 2007; Squire & Wixted, 2011). Bayley and colleagues had, on the other hand, tested patients with precisely described damage limited to the MTL in a series of group studies. They used either a detailed analysis of narrative content (Bayley, Hopkins, & Squire, 2003), a remember-know procedure and vividness ratings (Bayley, Gold, Hopkins, & Squire, 2005) or the Autobiographical Memory Interview (AMI, Bayley, Hopkins, & Squire, 2006; Kirwan, Bayley, Galvan, & Squire, 2008) and, in line with the SCT, reported no differences between the patients and control groups in the respective measures testing for the integrity of detailed remote memories (Squire & Bayley, 2007). However, MTT proponents countered that these studies by Bayley and colleagues had methodological problems of their own, as apparently the control groups in some of these studies seemed to perform much worse than control groups in other studies (Nadel et al., 2007).

In summary, both proponents of the SCT and proponents of the MTT/TTT reported patient studies supporting their theory and, at the same time, found ways to criticize the studies that did not align with their ideas.

Temporal gradients in animal lesion studies

In general, human patient studies will always face the problem of a lack of control over lesion site, lesion size and the timing between encoding of information and lesions. These problems can be, to a certain extent, overcome in animal studies. Here researchers can control both lesion site, size and lesion time in respect to the encoding session. Animal studies using e.g. contextual fear conditioning tasks (Kim & Fanselow, 1992; Anagnostaras, Maren, & Fanselow, 1999), socially acquired food preference tasks (Winocur, 1990), or trace eyeblink conditioning tasks (Kim, Clark, & Thompson, 1995; Takehara, Kawahara, & Kirino, 2003) showed temporal gradients after hippocampus damage with remote memories preserved while recent memories were lost, thus supporting the SCT. However, spatial tasks such as the water maze (e.g. Bolhuis, Stewart, & Forrest, 1994; Martin, de Hoz, & Morris, 2005; Clark, Broadbent, & Squire, 2005a) or cross maze (Winocur, Moscovitch, Caruana, & Binns, 2005) consistently showed no preserved memories after hippocampal lesion,

even for several months-old memories (Clark, Broadbent, & Squire, 2005b). While MTT proponents saw this as evidence that the hippocampus is permanently needed to retrieve complex spatial memories resembling episodic memories (Nadel et al., 2007), proponents of the SCT argued that these kinds of tasks may require the hippocampus because of their reliance on new performance-related learning and not because of the need to retrieve old memories (Squire & Bayley, 2007).

As TTT emerged, Winocur, Moscovitch, and Sekeres (2007) explicitly tested whether the memories that survived hippocampal damage at remote time points in the contextual fear conditioning task and the socially acquired food preference task were still context-specific (episodic-like) memories or whether the freezing response or food preference had generalized to other contexts, thus were context-free (semantic-like) memories. They found the later was true for remote memories and interpreted this as evidence of a memory transformation process from episodic-like to semantic-like memories (see Wiltgen & Silva, 2007 for same results using more than two time points and Richards et al., 2014 for a demonstration of memory generalization using a water maze task). Building on this, Winocur, Frankland, Sekeres, Fogel, and Moscovitch (2009) and later Einarsson, Pors, and Nader (2014) showed, that, as predicted by the TTT, both memories can coexist depending on the task demands and that reminding rats of the original conditioning environment before testing could reactivate the context-specific memory. Seemingly conflicting findings were reported by Wang, Teixeira, Wheeler, and Frankland (2009) using a context fear discrimination protocol that was shown to produce precise remote memories, thus memories that were not transformed. Interestingly, these precise, context-specific remote memories could be retrieved even after hippocampal lesions. However, Wang et al. (2009) also showed that these extra-hippocampal dependent precise remote memories were more fragile at later re-testing time points, thus some kind of difference between hippocampus-dependent and neocortex-dependent memories representations may still exist.

Hippocampus activity in healthy animal brains

Both the human patient studies and the animal lesion studies focussed on the question whether detailed remote memories can be retrieved without a functioning hippocampus. Yet, a slightly different but equally important question is whether the hippocampus is utilized to retrieve remote episodic memories in healthy functioning brains, or whether neocortical regions take over, even if the hippocampus is available.

In a series of studies by Bontempi and colleagues (for reviews see Frankland & Bontempi,

2005; Akers & Frankland, 2009), researchers tracked brain metabolic activity or the expression of activity-related genes (e.g. c-Fos or Zif268) in response to the retrieval of either recent or remote memories in healthy animal brains. Using either a spatial discrimination task (Bontempi, Laurent-Demir, Destrade, & Jaffard, 1999; Maviel, Durkin, Menzaghi, & Bontempi, 2004), a contextual fear conditioning task (Frankland, Bontempi, Talton, Kaczmarek, & Silva, 2004), or a socially transmitted food preference task (Ross & Eichenbaum, 2006), these studies showed that while the hippocampus was active during retrieval of recent memories, it was not active during retrieval of remote memories. The opposite pattern was observed in some neocortical areas and damage to the respective areas blocked the expression of remote but not recent memories (Akers & Frankland, 2009). Similar studies followed using other techniques, such as measuring dendritic spine growth (Restivo, Vetere, Bontempi, & Ammassari-Teule, 2009), or recording neural firing (Takehara-Nishiuchi & McNaughton, 2008), all suggesting a decrease in hippocampal involvement and paralleled increase in neocortical involvement in memory retrieval across time. These studies have therefore been reported in support of the SCT (Squire & Bayley, 2007, Squire et al., 2015). However, as noted in a review by Winocur et al. (2010), TTT agrees that cortical plasticity and increased involvement of neocortical areas is important for representations of remote memories. Yet, none of the above studies directly specified the nature and quality of the tested remote memories. Therefore, these studies cannot distinguish between SCT and TTT, as one cannot conclude from them that truly detailed memories depend less on the hippocampus at remote time points.

Lately, optogenetic methods and transgenic models, have given animal studies another new dimension. These techniques allow researchers to inhibit certain brain areas for short time intervals and even to tag and reactivate selectively those neurons activated during encoding. Thus, one can now try to tag and visualize the memory traces. Using optogenetic methods Goshen et al. (2011) showed that if the hippocampus is inactivated for a very short time remote memories cannot be retrieved, only if the hippocampus is deactivated for longer time periods (as is the case in lesion studies) the cortical version of the memory can be expressed, suggesting that the hippocampus is the default option in retrieving memories and it takes some time for the network to restructure to allow other brain areas to take over. Similar results were found by Denny et al. (2014) and they additionally showed that while memory traces in the hippocampus were initially more accurately reactivated in the conditioning context compared to a novel context, these differences disappeared at remote time points, suggesting a generalization of the memories. While these studies

favour the TTT, others have presented results in support of the SCT. Kitamura et al. (2017) demonstrated that with time prefrontal engram cells became capable of evoking a freezing response independently, while hippocampal engram cells gradually became silent. However, this study was promptly criticized by Hardt and Nadel (2018), as according to them the behavioural protocol of Kitamura et al. (2017) did not allow to establish a qualitative equivalence of the hippocampal and neocortical memory representations.
Thus, even animal studies with the newest techniques have not yet allowed researchers to draw a clear distinction between SCT and MTT/TTT. And, although animal studies allow the usage of modern techniques that cannot be applied in humans, it is important to remember that it is not possible to test truly episodic memories as defined by Tulving (2002) in animals.

Neuroimaging studies of autobiographical memories

In humans one can test hippocampal activity in normal functioning brains using neuroimaging techniques. Early neuroimaging studies in healthy humans comparing recent to remote autobiographical memories showed mixed results. Some studies reported a temporal gradient with less hippocampal activity for remote memories, in support of the SCT (Niki & Luo, 2002; Piefke, Weiss, Zilles, Markowitsch, & Fink, 2003), while others found no such gradient, in line with MTT/TTT (Ryan et al., 2001; Piolino et al., 2004; Maguire, Henson, Nummerg, & Frith, 2001).
Yet, the early studies showing temporal gradients were criticized for not adequately controlling for the vividness of the memories. Addis, Moscovitch, Crawley, and McAndrews (2004) showed that controlling for detail, emotionality, and personal significance of the memories actually eliminated or reduced the effect of time on hippocampal activity during autobiographical memory retrieval. On the other hand, studies in line with the MTT/TTT showing no difference in hippocampal activity for recent and remote memories were criticized for using pre-scan interviews to find detailed memories of each category, thus activity could reflect memories of the pre-scan interview instead of the remote time points. Therefore, other researchers explicitly avoided pre-scan interviews in their studies, using either personal photos chosen by family members (Gilboa, Winocur, Grady, Hevenor, & Moscovitch, 2004), commonplace questions not requiring pre-scan interviews (Rekkas & Constable, 2005) or information from interviews with close family members (Viard et al., 2007; Viard et al., 2010) to probe memory recall directly in the scanner. After controlling for vividness, all these studies still showed equivalent hippocampal activity for both recent and remote

memories, thus overcoming the prescan-interview argument. Another common argument against studies showing no difference between recent and remote memories is that one cannot exclude that hippocampal activity is due to new memory encoding and not memory retrieval. However, Gilboa et al. (2004) also controlled for this aspect, by using baseline trials that also prompt encoding as a control condition and they still found no difference between vivid remote and recent memories.

Prospective neuroimaging studies

Although autobiographical memories have the advantage of being truly episodic and rich in details and spatio-temporal experiencing, they do tend to be hard to fully control and match in attributes such as details, emotionality, times of subsequent retrieval and factual correctness (see false memory literature; Schacter, 1999). Therefore, conducting prospective studies in humans under controlled conditions can bring additional insights. In such studies participants encode information in the laboratory such as pictures (Takashima et al., 2006; Harand et al., 2012), stimuli-pair associations (Bosshardt, Degonda, et al., 2005; Takashima et al., 2009; Viskontas, Carr, Engel, & Knowlton, 2009; Yamashita et al., 2009) or film clips (Furman, Mendelsohn, & Dudai, 2012; Sekeres, Winocur, Moscovitch, Anderson, et al., 2018) before returning to the lab for memory retrieval after a certain time interval has passed. In a functional magnetic resonance imaging (fMRI) study of Takashima et al. (2006) participants came to four recognition tests in the scanner on day 1, day 2, day 30 and day 90, and were presented with a quarter of the picture stimuli learned at encoding in each session. In line with the SCT, they found a decrease in hippocampus activity for confident hits across time and a parallel increase in activity in the ventromedial prefrontal cortex (vmPFC). The same group of researchers later reported a decrease in hippocampal-neocortical connectivity and a parallel increase in cortico-cortical connectivity when comparing 1 day old retrieval to immediate (15 min) retrieval, again supporting the SCT (Takashima et al., 2009). Similarly, Yamashita et al. (2009) reported a decrease in (right) hippocampal activity and an increase in temporal neocortex activity when comparing 8 week old pair association memory to immediate retrieval. Completely contrary to these findings, hippocampal activity actually increased over time in other early prospective studies (Bosshardt, Degonda, et al., 2005; Bosshardt, Schmidt, et al., 2005). Other researchers, concentrating on differentiating between recognition memory based on recollection and memory based on familiarity, used prospective studies and the 'remember/know' procedure to track the evolution of memories (Viskontas et al., 2009; Harand

et al., 2012). By presenting the same stimuli at two time points (10 min and 1 week in Viskontas et al. (2009); 3 days and 3 months in Harand et al. (2012)), they could classify memories as being consistently episodic ('remember' answers at both time points) or being initially episodic ('remember' answer) and later transformed to familiar ('know' answer). In line with MTT/TTT, both studies showed that for those memories that were based on recollection at both time points hippocampal activity stayed stable, while activity in the hippocampus decreased for those memories that were based on familiarity at the later delay.

Furman et al. (2012) used a documentary movie as memoranda, as this is closer to episodic memories outside the laboratory. In a between subject design they compared 3 hours old memories to 3 week old and 3 month old memories in both recall and recognition tasks. They showed that during the recall task hippocampal engagement was similar at all time points, while hippocampal activity was lower in the recognition task at the 3 month interval compared to the two other time points. The authors suggested that these results are in line with the TTT view and note that hippocampal engagement seems to critically depend on the retrieval demands of the task. Thus, hippocampus-dependent episodic memories are necessary for recall tasks, but transformed, gist-like versions of the memories can be used to perform a recognition task at longer time intervals.

In summary, reviewing prospective neuroimaging studies again revealed studies in support of either the SCT (Takashima et al., 2006; Takashima et al., 2009; Yamashita et al., 2009) or the MTT/TTT (Viskontas et al., 2009; Harand et al., 2012; Furman et al., 2012), emphasizing the need for further research in this area.

1.3.3 Memory representation in neocortex

The last section has concentrated on the key question, whether the hippocampus is involved in retrieving remote detailed declarative memories, or whether these memories can be retrieved from the *neocortex* alone. The systems consolidation theories thereby mostly refer to the *neocortex* in general. However, there are other lines of research that investigate which neocortical areas and networks are involved in storing and retrieving declarative memories. In the next sections, different memory networks will be introduced briefly to provide a short but concise overview about neocortical areas important in representing different types of declarative memories.

Autobiographical memory network

Using a meta-analysis, Svoboda, McKinnon, and Levine (2006) have identified a network of brain areas involved in representing autobiographical memories including medial and ventrolateral prefrontal cortices, retrosplenial/posterior cingulate cortices, medial and lateral temporal cortices, the temporoparietal junction and the cerebellum. Cabeza and St Jacques (2007) report similar regions within their idea of an autobiographical memory retrieval network and suggest that, within this network, the lateral prefrontal cortex is responsible for search and control processes, the medial prefrontal cortex (mPFC) is involved in self-referential processes, the hippocampus and retrosplenial cortex control recollection processes, while the amygdala is responsible for emotional processing and the occipital and cuneus/precuneus regions are responsible for visual imagery and, finally, the vmPFC is involved in a feeling-of-rightness monitoring. However, McDermott, Szpunar, and Christ (2009) showed that the autobiographical memory network and the network for laboratory-based memories do not necessarily overlap, thus this autobiographical memory network cannot be generalized to all forms of declarative memory.

Semantic memory network

All the systems consolidation theories agree that remote semantic memories can be retrieved from neocortical areas. In a meta-analysis of 120 fMRI studies on semantic processing, activity was reliably found in middle temporal gyrus, posterior inferior parietal lobe, fusiform and parahippocampal gyri, inferior frontal gyrus, dorsomedial prefrontal cortex, posterior cingulate gyrus and vmPFC (Binder, Desai, Graves, & Conant, 2009). A theory based on clinical studies of semantic dementia, has suggested that the anterior temporal lobe plays a critical role in representing semantic concepts and memories (Lambon Ralph & Patterson, 2008; Patterson, Nestor, & Rogers, 2007; Lambon Ralph, Jefferies, Patterson, & Rogers, 2017, see also Nieuwenhuis et al., 2012), while others have suggested that the angular gyrus plays a key role in combining semantic memories (Price, Bonner, Peelle, & Grossman, 2015).

Schema memory network

Another line of research has focussed on the emergence of schemas from commonalities across multiple experiences over time and the neocortical representation of these schemas (Tse et al., 2007; Tse et al., 2011; van Kesteren, Fernandez, Norris, & Hermans, 2010;

van Kesteren, Ruiter, Fernandez, & Henson, 2012; Ghosh & Gilboa, 2014; Sommer, 2017; Gilboa & Marlatte, 2017). Original rodent studies had shown that memories that can be integrated within an existing schema are more rapidly represented in neocortical areas and rely less on the hippocampus than completely new information (Tse et al., 2007; Tse et al., 2011). This finding has recently also been shown in humans (Sommer, 2017). The neocortical network identified for schema memories includes the vmPFC, the angular gyrus, and posterior cortical regions, e.g. retrosplenial cortex (Gilboa & Marlatte, 2017).

Special role of the mPFC in remote memories

Within the neocortex the mPFC (including the anterior cingulate cortex) has frequently been identified in rodent studies as the candidate structure in representing remote memories (e.g. Frankland et al., 2004; Frankland & Bontempi, 2005; Takehara-Nishiuchi, Nakao, Kawahara, Matsuki, & Kirino, 2006; Takehara-Nishiuchi & McNaughton, 2008; Restivo et al., 2009; Richards et al., 2014; Einarsson et al., 2014; Cullen, Gilman, Winiecki, Riccio, & Jasnow, 2015)[3]. While hippocampal activity during retrieval of memories decreases over time, mPFC activity increases (Bontempi et al., 1999) and lesions to the mPFC result in a deficit of retrieving remote memories while leaving recent memories intact (Frankland et al., 2004; Maviel et al., 2004; Takehara et al., 2003). Following these results, it was suggested that for remote memories the mPFC might replace the hippocampus in its linking role (Frankland & Bontempi, 2005; Frankland & Bontempi, 2006). In human studies, the vmPFC has also been repeatedly suggested as a key region in representing and retrieving remote memories (Takashima et al., 2006; Nieuwenhuis & Takashima, 2011; Bonnici, Chadwick, Lutti, et al., 2012; Bonnici & Maguire, 2018). In the context of the TTT, combining both an animal and a human experiment, Sekeres, Winocur, Moscovitch, Anderson, et al. (2018) also identified the mPFC as the key region for the retrieval of remote memories. Importantly, they thereby suggest that the memory represented in the mPFC is not identical to the originally detailed hippocampus-dependent memory but is a transformed, less detailed version.

[3] This dominance of the mPFC in rodent studies may also be attributed to the fact that animal studies often do not cover the whole brain in their investigations. Wheeler et al. (2013) identified a whole network of areas involved in long-term memory recall in rodents, nevertheless still suggesting that the prefrontal cortex and thalamus are hub-like regions that most probably have special roles within this larger network.

1.3.4 MTL memory system

As outlined in section 1.3.2 the debate about systems consolidation is centered on the role of the hippocampus in remote detailed memories. However, the hippocampus is just one structure within the MTL memory system. Other than the hippocampus, the human MTL also includes the entorhinal cortex, anteriorly the perirhinal cortex, posteriorly the parahippocampal cortex (called postrhinal cortex in rodents) and the amygdala. Although the hippocampus plays a central and important role within the MTL memory system, it has nevertheless been shown that damage restricted to the entorhinal, perirhinal, and parahippocampal cortices still leads to severe memory impairments (Squire & Zola-Morgan, 1991). Thus, these cortices also play decisive roles in memory.

Early accounts investigating functional differences between the MTL regions, suggested that the involvement of the different regions may be related to the distinction of recognition memory processes into recollection and familiarity. It was proposed that recollection is dependent on the hippocampus while the anterior perirhinal cortex mediates familiarity processes (Aggleton & Brown, 1999; Brown & Aggleton, 2001; Eichenbaum, Otto, & Cohen, 1994). Extensions of this model (Eichenbaum et al., 2007; Diana et al., 2007; Diana, Yonelinas, & Ranganath, 2010) added that the posterior parahippocampal cortex is also related to recollection in the sense that it contains representations of context information, while the perirhinal cortex contains representations of items. These item representations are sufficient to support familiarity processes but not recollection. According to this model, the hippocampus then binds together the context information from the parahippocampal cortex and the items from the perirhinal cortex, thereby supporting recollection processes (Eichenbaum et al., 2007; Diana et al., 2007; but see Squire et al., 2007 for a different view). These models focussed on recent memories but can also be related to the debate about systems consolidation. Winocur et al. (2010) sees support for the TTT in these accounts, as hippocampus-dependent recollection processes are necessary for the retrieval of remote truly episodic memories, while the retrieval of remote semantic memories may rely on the familiarity processes of the perirhinal cortex (Winocur et al., 2010).

However, related ideas suggest that the perirhinal cortex represents visual objects in general and can therefore even support recollection of object features, while the parahippocampal cortex represents scenes or spatial layouts and the hippocampus is responsible for the relational encoding of objects and scenes (Alvarado & Bachevalier, 2005; Davachi, 2006; Staresina & Davachi, 2006; Bird & Burgess, 2008; Staresina, Duncan, & Davachi, 2011).

Thus, in these latter accounts the main difference between anterior perirhinal and posterior parahippocampal cortex is the representation of objects versus scenes, but not familiarity versus recollection processes.

These descriptions of the proposed functions of different MTL regions are not exhaustive, yet it is beyond the scope of this thesis to review all available research. The important conclusion to keep in mind from these notions is that the hippocampus does not function in isolation but is strongly connected to surrounding MTL areas, and that these areas in themselves serve important memory functions.

1.3.5 Hippocampal subfields

Another important consideration is that the hippocampus itself can be subdivided into different subfields. In fact, when using the term *hippocampus*, one should always define which subfields one includes under this name, as this has not always been coherent in the literature[4]. In this thesis the term hippocampus is used for the combination of the following subfields: the cornu ammonis (CA; that again is divided into three layers: CA1, CA2 and CA3), the dentate gyrus (DG) and the subiculum.

Functions of the hippocampal subfields

While the involvement of the hippocampus in retrieval of remote detailed memories is debated, all researchers agree on the decisive role of the hippocampus during memory encoding and retrieval of recent memories. The anatomy of the hippocampus with the CA, DG, subiculum and their interconnections is quite unique and has been identified as the reason why the hippocampus is especially equipped to rapidly encode memories. Theoretical accounts (Marr, 1971; McNaughton & Morris, 1987; Treves & Rolls, 1994; O'Reilly & McClelland, 1994) and the complementary learning systems computational model (McClelland et al., 1995; O'Reilly & Norman, 2002) proposed that this unique anatomy of the hippocampus allows for two special mechanisms to occur during encoding and early retrieval: *pattern separation* and *pattern completion*.

[4] Some researchers use the term *hippocampal formation* to refer to the CA fields, DG and subiculum (Moscovitch et al., 2005), while others use the term *hippocampal region* for this combination of subfields and in turn include the entorhinal cortex as part of the term *hippocampal formation* (Squire & Zola, 1996). To make it even more confusing, the term *hippocampal complex* is, at times, used to refer to the hippocampus plus the adjacent entorhinal, perirhinal and parahippocampal cortices (Moscovitch et al., 2005).

Pattern separation is the orthogonalization of similar memories, i.e. similar neuronal patterns are assigned distinct representations to avoid interference and overwriting of previously learned similar material (McClelland et al., 1995). The models proposed that the DG granule cells and CA3 play a decisive role in this *pattern separation* process. Thus, while the information that arrives in the DG from the entorhinal cortex has overlapping similar representations, the signal that the DG then projects further onto the CA3 subfield through special mossy fibres is sparser and orthogonalized and CA3 additionally supports this distinct storage of similar memory representations (Marr, 1971; O'Reilly & McClelland, 1994; Yassa & Stark, 2011).

On the other hand, CA3 also receives direct input from the entorhinal cortex and has many recurrent connections (axons that circle back to the cells in the same subfield) that allow for the completion of previously stored patterns even with only partial or degraded input cues from those patterns during retrieval, thus enabling *pattern completion* (Marr, 1971; McNaughton & Morris, 1987; O'Reilly & McClelland, 1994).

Initially, the ideas of pattern completion and pattern separation in DG and CA3 were only based on these theoretical and computational models, but later electrophysiological studies in rodents (S. Leutgeb, Leutgeb, Treves, Moser, & Moser, 2004; Vazdarjanova & Guzowski, 2004; J. K. Leutgeb, Leutgeb, Moser, & Moser, 2007) and high-resolution neuroimaging studies in humans (Bakker, Kirwan, Miller, & Stark, 2008; Lacy, Yassa, Stark, Muftuler, & Stark, 2011; Bonnici, Chadwick, Kumaran, et al., 2012; Deuker, Doeller, Fell, & Axmacher, 2014) underlined the proposed mechanisms (Yassa & Stark, 2011).

While DG and CA3 have prominent roles in pattern separation and pattern completion, it has been suggested that CA1, on the other hand, acts as a match/mismatch detector. Thus, it receives perceptual input from the entorhinal cortex and "decides" if this matches the information that was retrieved in CA3 through pattern completion (Chen, Olsen, Preston, Glover, & Wagner, 2011; K. Duncan, Ketz, Inati, & Davachi, 2012).

1.3.6 Differentiation along the hippocampal long axis

It was proposed quite some time ago (Hughes, 1965; Nadel, 1968; Grant & Jarrard, 1968) not only to differentiate between the above described hippocampal subfields of CA1, CA2, CA3, DG and subiculum but to additionally examine structural, functional and connectivity differences along the hippocampal long axis (for reviews see Poppenk, Evensmoen, Moscovitch, & Nadel, 2013; Strange, Witter, Lein, & Moser, 2014). An important research goal of this thesis is to relate the hippocampal long axis literature to the above

described debate about systems consolidation theories. The following sections will therefore first briefly review hippocampal long axis research in general before connecting it to the systems consolidation debate in particular.

Dividing the hippocampal long axis

Anatomically the hippocampus has a long, curved form stretching from anterior MTL to posterior MTL, with a head, body and tail. Most often functional distinctions are made between the anterior hippocampus (or ventral hippocampus in rodents), corresponding to the head and the posterior hippocampus (or dorsal hippocampus in rodents), corresponding to the tail and body (Poppenk et al., 2013; Zeidman & Maguire, 2016). Other researchers have divided the hippocampus into three functional parts along the long axis, introducing an intermediate, mid-portion area corresponding to the body (Fanselow & Dong, 2010; Collin, Milivojevic, & Doeller, 2015). Additionally, it has been proposed that there are continuous gradients of function along the long axis, superimposed on the distinctive functional fields of anterior, mid-portion and posterior hippocampus (Kjelstrup et al., 2008; Strange et al., 2014).

Differences in connectivity along the hippocampal long axis

As the anatomy would suggest, the anterior hippocampus receives more input from the perirhinal cortex, lying adjacent in the anterior part of the MTL, while the posterior hippocampus receives more input from the more posterior parahippocampal cortex (Libby, Ekstrom, Ragland, & Ranganath, 2012). The connectivity beyond the MTL also differs: the anterior hippocampus shows stronger connectivity to regions in the lateral temporal cortex extending into the temporal pole (Kahn, Andrews-Hanna, Vincent, Snyder, & Buckner, 2008), the amygdala (Duvernoy, 2005), hypothalamus, insula and nucleus accumbens (Poppenk et al., 2013). The posterior hippocampus, on the other hand, has stronger connections to the lateral parietal cortex, posterior cingulate cortex, retrosplenial cortex (Kahn et al., 2008), precuneus, anterior cingulate cortex and parts of the thalamus (Poppenk et al., 2013).

Early proposals of functional distinctions

These differences in connectivity originally contributed to a prominent theory suggesting that based on the preferential input from visual, somatosensory, and spatial regions the posterior two-thirds of the hippocampus are involved in cognitive (spatial) memory processes

(Moser & Moser, 1998, Greicius et al., 2003, Bannerman et al., 2004, Fanselow & Dong, 2010) while the anterior hippocampus, based on its prominent connectivity with the amygdala, is involved in arousal, emotional responses, reward (Moser & Moser, 1998; Bannerman et al., 2004, Fanselow & Dong, 2010) or decision-making (Viard, Doeller, Hartley, Bird, & Burgess, 2011). However, there have been many reports of the anterior hippocampus also contributing to memory, and over the years many alternative theories and suggestions have emerged regarding the functional segregation along the axis (Poppenk et al., 2013; Strange et al., 2014). For example, one model suggested that the anterior hippocampus is responsible for memory encoding while the posterior hippocampus is activated by memory retrieval (Lepage, Habib, & Tulving, 1998). Another line of research has suggested that the anterior hippocampus responds to stimulus novelty (Strange, Fletcher, Henson, Friston, & Dolan, 1999; Kohler, Crane, & Milner, 2002; Daselaar, Fleck, & Cabeza, 2006), with the posterior hippocampus responding to previously encountered stimuli (Strange et al., 1999; Daselaar et al., 2006).

Local to global spatial gradients

Another prominent distinction, concerning global to local spatial representations, has been based on influential rodent studies on place cells: it was shown that place cells exist throughout the hippocampus, but that the spatial radius of the place fields of individual neurons increases along the long axis (Jung, Wiener, & McNaughton, 1994; Maurer, Vanrhoads, Sutherland, Lipa, & McNaughton, 2005), from less than one meter in the dorsal (posterior) pole to more than 10 meters in the ventral (anterior) pole (Kjelstrup et al., 2008). This gradient in place cell scaling along the long axis was also found by Keinath et al. (2014), but they additionally looked at neuronal populations within each subregion and found that, although single cells in the ventral hippocampus could not match the spatial precision of cells in the dorsal hippocampus, precise locations could still be represented in the anterior hippocampus through a population activity code. Furthermore, they used neural network modelling to suggest that the differences in place field size along the longitudinal axis may allow for memory details to be guarded against interference in the dorsal (posterior) hippocampus while commonalities among memories can be successfully generalized in the ventral (anterior) hippocampus (Keinath et al., 2014).

In humans a similar global-local gradient was shown in a virtual environment fMRI study that suggested that anterior hippocampus activity was related to the usage of coarse global environmental representations for navigation, while using fine-grained, local environmental

representations during the navigation task was related to posterior hippocampal activation (Evensmoen et al., 2013). The same gradient was found in an fMRI study using an object room geometry association task, with results suggesting that the most fine-grained positional representations were represented in posterior hippocampus with more coarse-grained representations in anterior hippocampus (Evensmoen et al., 2015). And Javadi et al. (2017) used graph measures of real-world topology and a navigation task in the scanner to show that anterior hippocampus activity is associated with global and posterior hippocampus activity with local metrics of graph-theoretic centrality. Thus, when performing spatial tasks there seems to be a local to global gradient in the activity along the hippocampal long axis in humans equivalent to that found in animal place cells. Interestingly, similar results were also seen for spatial aspects of autobiographical memories: the posterior hippocampus was preferentially activated when retrieving detailed spatial relations within personal event memories, while the anterior hippocampus was preferentially activated when retrieving general information about places (Nadel, Hoscheidt, & Ryan, 2013). But does this global-local gradient only hold for spatial aspects of memories? Collin et al. (2015) tested whether a similar gradient could be found for representations of multi-event narratives and indeed found that memory representations vary in scale along the anterior-posterior axis: individual event-pair associations (e.g. event A- event B) were represented in the posterior hippocampus, multiple event-pair associations were represented in the mid-portion hippocampus (both event pair A-B and event pair B-C) and finally complete narratives including indirectly related events (event A and event C, that were never learned together but only bridged by event B) were represented in the anterior hippocampus.

Anatomical distinction and complementary learning systems

Moreover, it has been shown that the relative percentage of DG and CA1-CA3 differs along the hippocampal axis. In the anterior hippocampus the CA1-CA3 subfields are over-represented, while the DG, on the other hand, is relatively more present in the posterior hippocampus (Malykhin, Lebel, Coupland, Wilman, & Carter, 2010). Schapiro, Turk-Browne, Botvinick, and Norman (2017) have recently suggested that the *complementary learning system* model can be applied to two different learning systems within the hippocampus (in comparison to the original *complementary learning system* model that was based on one system in the hippocampus and the other in the neocortex; McClelland et al., 1995): the system responsible for storing individual episodes through pattern separation relies on the trisynaptic pathway (entorhinal cortex-DG-CA3-CA1), while a complimen-

tary system is based on a monosynaptic pathway (entorhinal-CA1) and enables the rapid statistical learning that had previously been shown to take place within the hippocampus (Schapiro, Kustner, & Turk-Browne, 2012; Schapiro, Gregory, Landau, McCloskey, & Turk-Browne, 2014; Schapiro, Turk-Browne, Norman, & Botvinick, 2016). Consequently, Schapiro et al. (2017) suggest that the anterior hippocampus, having a higher relative percentage of CA1 (and lower proportion of DG) is better suited for the monosynaptic pathway and therefore rapid statistical learning, instead of storing individual episodes through pattern separation.

Integrating the long axis into the TTT

Thus, there has been a wealth of proposals of functional segregations along the long axis and as suggested by Strange et al. (2014) it could well be that different gradients exist simultaneously and might also be superimposed on distinct functional fields.

Most importantly for this thesis, the distinction along the hippocampal axis has also recently been introduced to the systems consolidation debate by the proponents of the TTT (Robin & Moscovitch, 2017; Sekeres, Winocur, & Moscovitch, 2018; Sekeres, Winocur, Moscovitch, Anderson, et al., 2018). In this thesis, I will term this extension of the TTT the *long axis TTT proposal*. In this proposal the authors suggest, that specific detailed memories are represented in the posterior hippocampus while gist-like memories of events are found in the anterior hippocampus, and only schematic and semantic memories can be supported by exclusively neocortical structures (Sekeres, Winocur, & Moscovitch, 2018). Importantly, initial formulations of the TTT (Winocur et al., 2010; Winocur & Moscovitch, 2011) did not differentiate between gist-like memories, semantic memories and schemas and thus proposed that all these types of memories were represented in the neocortex. This, therefore, changed in the long axis TTT proposal, as the authors now elaborate on the difference between these three types of memory (see section 1.2.2). This distinction now allowed them to suggest that gist-like memories are hippocampus-dependent, in comparison to semantic memories and schemas (Robin & Moscovitch, 2017; Sekeres, Winocur, & Moscovitch, 2018).

Sekeres, Winocur, and Moscovitch (2018) argue that their proposal is supported by the connectivity of the anterior hippocampus to the mPFC, representing schemas and the anterior temporal lobe, representing semantic memories and the connectivity of the posterior hippocampus to posterior neocortical structures implicated in perception (e.g. precuneus, retrosplenioal cortex, ventral temporal cortex). In other words, the gist area of the hip-

pocampus is connected to schema and semantic areas, while the detail area of the hippocampus is connected to the perceptual input. With time, there is normally a greater reliance on gist-memory, which would translate into a reduction of activity in posterior hippocampus and an increased reliance on the anterior hippocampus. However, the pattern of activation at memory retrieval will also depend on the task demands.

In fact, the distinction into gist (anterior hippocampus) and details (posterior hippocampus) had already previously been included in the review by Poppenk et al. (2013) as one possible distinction. There they suggested that evidence for this distinction could be found in a reduction of posterior hippocampal volume size in posttraumatic stress disorder patients compared to controls while anterior hippocampus volume was not affected (Bonne et al., 2008), a fact that might be related to the over-reliance on gist memory in posttraumatic stress disorder patients (Poppenk et al., 2013). Additionally, Poppenk and Moscovitch (2011) showed that in healthy participants, better recollection memory was positively correlated with the volume size of the posterior hippocampus and negatively with the size of the anterior hippocampus.

Viard, Desgranges, Eustache, and Piolino (2012) also focussed on the hippocampal long axis in a meta-analysis of 58 fMRI studies examining either the retrieval of autobiographical memories or the imagination of future episodic events and reported, in line with the long axis TTT proposal, that the posterior hippocampus was more active for "strictly episodic events" (spatio-temporally unique personal episodes accompanied by contextual and phenomenological details) compared to, what they term "episodic events" (spatio-temporally unique personal episodes without details), thus gist-like memories in our definition (see section 1.2.2).

One can also relate the long axis TTT proposal to results from a fMRI connectivity study by McCormick, St-Laurent, Ty, Valiante, and McAndrews (2015): they showed, that during an initial construction phase of memory retrieval the anterior hippocampus interacted with prefrontal areas and during the following elaboration phase the posterior hippocampus interacted with visual perceptual areas. Critically, they thereby suggested that the construction phase includes retrieval of gist-like information, while the elaboration phase includes retrieval of details of events.

Support for the long axis TTT proposal can additionally be found in a recent rodent study, showing that retrieval of remote context-generalized fear memories (comparable to gist-like memories in humans) require an intact ventral (anterior) hippocampus, while the dorsal (posterior) hippocampus played a role in the representation of context-specific pre-

cise memories (Cullen et al., 2015).

Of course, in a wider sense the long axis TTT proposal could also be related to the local (posterior) to global (anterior) spatial gradient discussed above, in the sense that details of events are more like the local features of spatial contexts and the gist-like memory of events could be defined as the global features of an event.

I will return to a discussion about the hippocampal long axis and the long axis TTT proposal in section 3.3 after introducing our own study about the time-dependent memory transformation along the hippocampal axis in section 2.1.

1.3.7 Laterality of hippocampal and MTL activation

One last issue to briefly consider about systems consolidation of initially hippocampus-dependent declarative memories, is the laterality of hippocampal/ MTL activation. Often memory studies report MTL or hippocampal activity lateralized to one of the hemispheres. However, results are not coherent and quite often contradictory. Using autobiographical memory studies as an example, there have been reports of a left-sided hippocampal dominance (e.g. Maguire & Mummery, 1999; Piefke et al., 2003; Svoboda & Levine, 2009), yet again other results showed predominantly right-sided hippocampus activations (e.g. Fink et al., 1996; Steinvorth, Corkin, & Halgren, 2006), while still others have found bilateral hippocampal activity (e.g. Ryan et al., 2001; Piolino et al., 2004; Gilboa et al., 2004; Rekkas & Constable, 2005; Viard et al., 2010). Maguire and Frith (2003) reported a temporal gradient of memories in the right hippocampus (more activity for recent than remote memories), while the left hippocampus activity did not decrease across time. Results from a meta-analysis suggested that the differences in hippocampal laterality found in autobiographical studies might depend on methodology issues and that truly episodic events elicited more activity in the left hippocampus (Viard et al., 2012).

All in all, the issue of laterality seems not yet fully conclusive and definitely requires further research. However, this topic will not be further addressed in this thesis, as it was neither a research goal to examine laterality, nor could we derive new clear conclusions about laterality from our own data.

1.4 Systems consolidation of nondeclarative memories

As described in detail in the last sections, systems consolidation of initially hippocampus-dependent declarative memories has been the topic of many research papers and has produced vigorous debates. But what happens to nondeclarative memories which are not initially dependent on the hippocampus, such as perceptual and motor skills? Do they also undergo a memory consolidation process over time? Synaptic consolidation, as introduced in section 1.1.2, is considered universal, thus occurring in all types of memory systems (Dudai et al., 2015). But can we also see changes at brain systems level in nondeclarative memory representation?

Indeed, since the 1990s researchers have investigated *early* systems consolidation processes for nondeclarative perceptual skills (e.g. Karni & Sagi, 1993; Karni, Tanne, Rubenstein, Askenasy, & Sagi, 1994) or motor skills (e.g. Karni et al., 1995; Karni et al., 1998; Brashers-Krug, Shadmehr, & Bizzi, 1996; Shadmehr & Holcomb, 1997). Nevertheless, the amount of research on systems consolidation of nondeclarative memories is far less than for declarative memories, and, to my knowledge, *prolonged* systems consolidation of nondeclarative memories has not been explicitly tested so far. One of the research goals of this thesis was therefore to test prolonged systems consolidation of nondeclarative memories and to compare it to systems consolidation of declarative memories.

As described in section 1.2, the term nondeclarative memory is an umbrella term for many different types of memory systems. The type of nondeclarative memory most frequently tested in early systems consolidation studies is *motor sequence memory*. In our second fMRI study (see section 2.2), we, therefore, also used a motor sequence learning task. Thus, the next section will focus explicitly on systems consolidation of motor sequence memories.

1.4.1 Systems consolidation of motor sequence memories

Motor sequence memories are memories of a specific sequence of movements, most often tested by a serial reaction time task (SRTT; Nissen & Bullemer, 1987), prompting a target sequence of button presses with specific fingers in response to respective visual cues. Thereby memory is tested by an decrease in reaction times in response to the target sequence compared to other random sequences. Indications for a consolidation process of motor sequence memories can be observed in behavioural data: (1) Performing a second motor task within the first few hours after learning a first motor skill, disrupts memory for

the newly established motor skill, equivalent to declarative memories and suggesting that motor memories are also initially unstable and vulnerable to interference (Brashers-Krug et al., 1996; Korman et al., 2007). (2) During a motor sequence training session one can observe immediate large performance gains (measured by reaction times and/or error rates) after only a few practice trials (online gains). Interestingly, however, performance gains are also observed between two practice sessions (offline gains), suggesting that consolidation processes proceeding within the first hours after training not only stabilize memories but even enhance performance offline (Karni et al., 1998; Korman, Raz, Flash, & Karni, 2003; Krakauer & Shadmehr, 2006; Dayan & Cohen, 2011).

Are these observed behavioural indicators of consolidation accompanied by reorganizations of memory representations at brain systems level? In order to test this, it first needs to be established which brain networks are involved in representing motor sequence memories initially. The learning of motor sequences has been shown to recruit a cortico-striatal system, including the motor and somatosensory neocortical areas and the putamen and caudate nucleus (Doyon et al., 2009; Doyon, Penhune, & Ungerleider, 2003; Hardwick, Rottschy, Miall, & Eickhoff, 2013). However, it has additionally been reported that the hippocampus is involved in performing motor sequence tasks (Albouy et al., 2008; Albouy, King, Maquet, & Doyon, 2013; Albouy, Sterpenich, et al., 2013; Albouy et al., 2015). This seems at odds with the fact that the whole notion of nondeclarative memories was based on MTL-damaged patients performing normally in e.g. motor learning tasks, thus leading to the notion that these memories are hippocampus-independent. Yet, just because these tasks do not necessarily depend on the hippocampus, does not automatically mean they are naturally performed without involvement of this important brain area in humans with normally functioning MTLs (Dudai et al., 2015). Most probably, both the hippocampus-dependent system and the cortico-striatal system interact (Coynel et al., 2010; Debas et al., 2014), reflected by the fact that healthy humans, in comparison to amnesic patients, can at least remember performing a motor sequence task, even if they may not be explicitly aware of the trained sequence (Dudai et al., 2015).

Despite this interaction, the question remains whether there are brain system changes and reorganizations of memory representations within the cortico-striatal system over time, parallel to the changes seen in the hippocampus-dependent system. Indeed, researchers focussing on early systems consolidation processes occurring within the first night after learning a motor sequence task, have shown sleep-dependent changes of activity in the striatum (Debas et al., 2010), greater between-regions interactions within the cortico-striatal sys-

tem (Debas et al., 2014), and an execution induced repetition enhancement in the primary motor cortex (Gabitov, Manor, & Karni, 2014). Most recently researchers have employed simultaneous electroencephalography (EEG) - fMRI recordings to capture reactivations of memory traces while participants were sleeping, shortly after performing a motor sequence learning session (Vahdat, Fogel, Benali, & Doyon, 2017; Fogel et al., 2017). Vahdat et al. (2017) reported a downscaling of functional connectivity within a cortical memory representation and a gradual reorganization of it into a more subcortically-dominant trace. Interestingly, Fogel et al. (2017) added that particularly the striatum was reactivated during post-learning sleep and reported a reorganization of the memory trace within subregions of the striatum, from the rostrodorsal to the caudoventral subregion. Importantly, all of the above studies reported correlations of the respective neuronal changes with the offline performance gains found in the behavioural data (Debas et al., 2010; Debas et al., 2014; Gabitov et al., 2014; Vahdat et al., 2017; Fogel et al., 2017).

Thus, it has been shown that early systems consolidation processes take place within the cortico-striatal system. Several other studies have investigated the effects of extensive training of a motor sequence task over several weeks and reported further changes within the brain systems (for reviews see Krakauer & Shadmehr, 2006; Ungerleider, Doyon, & Karni, 2002). In particular, a transfer of motor memories from an associative, premotor circuit to a sensorimotor circuit after extensive motor sequence training has been suggested (Coynel et al., 2010; Dayan & Cohen, 2011, Lehericy et al., 2005; but see Kupferschmidt, Juczewski, Cui, Johnson, & Lovinger, 2017). However, it remains unclear whether and how representations of motor memories learned within a single training session evolve over weeks without further training. In other words, is there evidence for a prolonged systems consolidation process of motor sequence memories unfolding within the weeks after the above reported early systems consolidation processes even without further training? We addressed this question in our second study (see section 2.2).

1.5 Memory dynamics and the influence of external factors on systems consolidation

In all the preceding discussions about declarative and nondeclarative systems consolidation, consolidation was considered to be an uninterrupted continuous process that takes place between memory encoding and the later memory retrieval without any external influences. However, in everyday life many external factors, such as emotions and stress may considerably influence all memory processes, including consolidation (Holland & Kensinger, 2010; Schwabe, Joels, Roozendaal, Wolf, & Oitzl, 2012). We therefore took this aspect into consideration in our studies by examining the influence of emotional stimulus material on memory consolidation in both study 1 (section 2.1) and study 4 (section 2.4) and by additionally testing the influence of stress on memory consolidation and transformation in study 4.

Furthermore, as described in the introduction to this thesis, memories are considered to be highly dynamic entities that can be changed by each reactivation, e.g. through external cues (Nadel et al., 2012; Kroes & Fernandez, 2012). Although this may, on the one hand, lead to false memories (St Jacques, Olm, & Schacter, 2013), it is, on the other hand, immensely important that we are able to dynamically update our memories when new information becomes available to maintain memory relevance (Exton-McGuinness, Lee, & Reichelt, 2015). In study 3 (section 2.3) we, therefore, tested which brain regions enable us to perform memory updating in the light of new information.

The later sections of this chapter will contain brief and concise summaries of the theoretical backgrounds of both memory updating (section 1.5.3) as well as the influence of emotions (section 1.5.4) and stress (section 1.5.5). Yet beforehand, the general concepts of memory replay and reactivation, both during sleep (section 1.5.1) and actively during wakefulness (section 1.5.2) will be introduced, as these processes are essential for memory dynamics.

1.5.1 The role of sleep

Sleep plays a special role in memory consolidation (Diekelmann, Wilhelm, & Born, 2009; Diekelmann & Born, 2010; Born & Wilhelm, 2012). First of all, sleep provides an optimal opportunity of reprocessing memories without the interference of new incoming sensory information (Born & Wilhelm, 2012; Kroes & Fernandez, 2012). Furthermore, sleep seems to play an active role in systems consolidation (Diekelmann & Born, 2010). It has been repeatedly shown that during sleep memories are replayed. This was first demonstrated

in rodent studies: during sleep, place cells in the hippocampus reactivated in the same sequence as they had done in prior spatial learning tasks (Pavlides & Winson, 1989; Wilson & McNaughton, 1994). Later, equivalent results were found in humans: Maquet et al. (2000) showed that the brain areas activated during a SRTT were active again during subsequent sleep and Peigneux et al. (2004) reported that the hippocampal areas active during learning of routes in a virtual environment were also active during sleep. Subsequently, Rasch, Buchel, Gais, and Born (2007) demonstrated that replay of odor-card associations in the hippocampus can be enhanced by administering the respective odors during post-learning sleep thereby increasing subsequent memories of the cards, a result later replicated with auditory cues (Rudoy, Voss, Westerberg, & Paller, 2009). Thus, these cueing studies demonstrated a causal contribution of reactivations during sleep to memory consolidation. Consequently, it was proposed and demonstrated that the reorganization of newly encoded memory representations across brain systems takes place mainly during sleep, especially during the slow-wave sleep phase (Gais et al., 2007; Diekelmann & Born, 2010; Born & Wilhelm, 2012). Additionally, it was demonstrated that this active systems consolidation during sleep is selective, i.e. not all memories are replayed and stored as long-term memory representations, instead sleep-dependent memory consolidation is motivationally driven, specifically strengthening memories relevant for future goals (Fischer & Born, 2009; Wilhelm et al., 2011; Born & Wilhelm, 2012). Furthermore, sleep can promote insight into hidden rules and structures within tasks (Wagner, Gais, Haider, Verleger, & Born, 2004) and can help form explicit knowledge of implicitly learned rules (Fischer, Drosopoulos, Tsen, & Born, 2006).

In summary, sleep selectively enhances specific memories and can dynamically change the quality of memories (Born & Wilhelm, 2012). Nevertheless, reactivation and replay of memories does not exclusively take place during sleep. Karlsson and Frank (2009) demonstrated that memory replay can also occur during wake rest periods, even without externally cueing the memory, suggesting that sleep-like consolidation processes can advance whenever the hippocampus is not occupied by new encoding processes (Karlsson & Frank, 2009; Squire et al., 2015). Reactivation of memories by external cues, on the other hand, can lead to destabilization-reconsolidation processes and memory updating (Lee, Nader, & Schiller, 2017), processes discussed in the following two sections 1.5.2 and 1.5.3.

1.5.2 Memory reconsolidation

For a long time, the textbook view of memory consolidation was that after an initial memory consolidation process is complete, memories are fixed and stable and cannot be disrupted by e.g. interference (Nadel et al., 2012; Nadel et al., 2000). Around the turn of the millennium this view was seriously challenged by emerging animal studies (Przybyslawski & Sara, 1997; Nader, Schafe, & Le Doux, 2000; Sara, 2000)[5]. These studies demonstrated that reactivating a memory (e.g. through the presentation of a conditioned stimuli in a conditioning paradigm) would transiently (for some hours) return the memory into a labile state, corresponding to the vulnerable time after initial encoding. Injecting a protein synthesis inhibitor during this crucial time disrupts the memory, while injecting such an inhibitor did not disturb memories that had not been previously reactivated (Nader et al., 2000). Thus, destabilising a memory through reactivation, changes an inactive, stable memory into an active, unstable memory for a certain time window (Exton-McGuinness et al., 2015). The existence of such a reconsolidation window has been repeatedly confirmed, both by using different tasks and in many different species (including humans). Furthermore, molecular and cellular correlates of reconsolidation have also been reported (for reviews see e.g. Nader & Einarsson, 2010; Nader, 2015; Schwabe, Nader, & Pruessner, 2014; Agren, 2014; Haubrich & Nader, 2016 but see Hardwicke, Taqi, & Shanks, 2016 for failed replications of the reconsolidation phenomenon). Importantly, researchers have also identified boundary conditions for reconsolidation, such as the strength or age of a memory (Suzuki et al., 2004; Exton-McGuinness et al., 2015; Walker & Stickgold, 2016).

1.5.3 Memory updating through new information

From a practical and evolutionary perspective, it is sensible that long-term memories can be updated. Environments are constantly changing and memories have to somehow be updated to allow the integration of this new information and thereby maintain the relevance of the memories in day-to-day life (Lee, 2009; Exton-McGuinness et al., 2015). Although memory updating via memory reactivation can, at times, also introduce unwanted memory distortions leading to false memories of older events (St Jacques et al., 2013; Loftus & Pickrell, 1995; Kroes & Fernandez, 2012), it is nevertheless important that we can update

[5] In fact, similar results had already been demonstrated in rats about 30 years prior to the emergence of the reconsolidation theories by Misanin, Miller, and Lewis (1968), Schneider and Sherman (1968), Miller and Springer (1972), and Lewis (1979), however these early accounts were initially ignored by most researchers (Nadel et al., 2000).

memories and abstract knowledge from similar experiences in order to maintain and support the predictive function of memories for future behaviour (Kroes & Fernandez, 2012). Memory updating is triggered through new competitive information. One question to be considered is, in which cases new information actually leads to the updating of old memories and when it just simply leads to the encoding of a new parallel memory trace. First and foremost, the original memory is only updated if the new learning material is similar enough to the original memory. Thus, the degree of similarity may act as a switch between memory updating and new learning (Besnard, 2012). In a related idea, it has been suggested that gradual changes in environments lead to updating, while abrupt changes lead to new learning (Gershman, Radulescu, Norman, & Niv, 2014).

Connection between memory updating and memory reconsolidation

As presented above, the reconsolidation-window was originally discovered by injecting amnesic agents after memory reactivation. It was therefore initially thought of as a window of vulnerability. But, indeed, in real life amnesic agents are rarely encountered but memories are constantly reactivated and therefore destabilized. Thus, the question arises what functional role this reconsolidation process serves (Lee et al., 2017). Recently, it has been suggested that the whole existence of a reconsolidation process serves the function of memory updating (Lee, 2009, Nader & Hardt, 2009, Lee et al., 2017, Rodriguez-Ortiz & Bermudez-Rattoni, 2017). In fact, it has been suggested that reconsolidation only occurs when memory has to be updated (Exton-McGuinness et al., 2015), thus only when information is present during reactivation that leads to a prediction error signal (Lee, 2009; Sevenster, Beckers, & Kindt, 2013; Exton-McGuinness et al., 2015). If performance is already asymptotic (no further gains possible) in memory tasks, then a further training trial will not trigger a reconsolidation process as the behaviour is already at its optimum and nothing needs to be updated. According to this theory, reconsolidation will only occur if a memory is either not yet well-established, so that it still requires strengthening, or if new information is present that needs to be integrated. Thus, well-established memories will only be destabilized if new information is available (Sevenster et al., 2013; Exton-McGuinness, Patton, Sacco, & Lee, 2014; Exton-McGuinness et al., 2015).

Memory updating at brain systems level

Therefore, memory updating has been linked to memory reconsolidation processes and has indeed been identified as the functional reason for reconsolidation to occur in the first

place (Lee, 2009; Exton-McGuinness et al., 2015; Lee et al., 2017). Thus, it seems as if reconsolidation processes are involved and responsible for memory updating at the synaptic level. However, recent animal research has shown that updating processes are not only traceable at molecular and cellular levels but that memory updating after exposure to new information can indeed reorganize the neural circuit representing established memories on a systems level (Kwapis, Jarome, Ferrara, & Helmstetter, 2017).

In humans, memory updating has been tested using associative pairing paradigms: an originally thoroughly learned A-B pairing is later explicitly and consciously updated to an A-C pairing (Nyberg et al., 2009; Kuhl, Bainbridge, & Chun, 2012). Using such a paradigm, Nyberg et al. (2009) showed that the process of updating long-term memories leads to an increase of left prefrontal cortex activity. Later, Kuhl et al. (2012) showed that during retrieval of visual A-C pairings both old and new (B and C) visual memories were reactivated in the ventral temporal cortex, but stronger inferior frontal gyrus activation during the encoding of the A-C pairs predicted less competition in the ventral temporal cortex during the following retrieval phase. On the other hand, increased activity in the anterior cingulate cortex was associated with stronger reactivation of the older memories. Importantly, however, in both these studies all phases of the respective paradigms were conducted in one experimental session, without delays that would allow systems consolidation processes to take place before memory updating. It therefore remains unclear which brain areas are needed to update memories that have already undergone a systems consolidation process. We investigated this question in our third study (see section 2.3).

1.5.4 Influence of emotions

When thinking of one's own autobiographical memories of the past, it is easily noticeable that emotional experiences are often remembered longer and more clearly than neutral events (Holland & Kensinger, 2010). This benefit of emotional arousal on memories has also been demonstrated in experimental tasks using emotional stimuli (for reviews see Hamann, 2001; LaBar & Cabeza, 2006; Talmi, 2013; Yonelinas & Ritchey, 2015). Emotion can thereby have a beneficial effect on the initial process of memory encoding, but importantly the effect additionally increases after a period of memory consolidation (Sharot & Phelps, 2004; Yonelinas & Ritchey, 2015).

The main brain area associated with emotions is the amygdala and it is therefore not surprising that the beneficial effect of emotional arousal on memory is orchestrated by the amygdala (Cahill, Babinsky, Markowitsch, & McGaugh, 1995). The activity within

the amygdala during encoding correlates with subsequent memory for emotional material (Cahill et al., 1996; Dolcos, LaBar, & Cabeza, 2004; Murty, Ritchey, Adcock, & LaBar, 2010). Furthermore, the normal advantage of emotional material over neutral material is reduced or even eliminated after amygdala damage (Adolphs, Cahill, Schul, & Babinsky, 1997; Markowitsch et al., 1994). The amygdala is also crucially involved in mediating the effects of stress hormones on memory (McGaugh, 2000), which will be discussed in the following section 1.5.5.

We included the factor emotion in our studies 1 and 4 by using negative emotional pictures as stimuli and comparing them to neutral pictures. We could therefore analyse the influence of emotion on systems consolidation and the transformation of memories from detailed to gist-like versions.

1.5.5 Influence of stress

One factor having a considerable effect on all learning and memory processes in our everyday life is stress (Roozendaal, 2002; Wolf, 2009; Schwabe et al., 2012). Stressful experiences lead to the release of stress hormones such as gluccocorticoids and noradrenalin (Joels & Baram, 2009) and brain areas important for memory, such as the hippocampus or prefrontal cortex are especially susceptible to effects of these stress hormones (Oei et al., 2007; Schwabe et al., 2012). The timing of the stressful experience in relation to the different memory phases influences whether stress enhances or impairs memory (Roozendaal, 2002; Schwabe, Wolf, & Oitzl, 2010; Schwabe et al., 2012; Cadle & Zoladz, 2015). Impairing effects on memory performance are most often seen when stress occurs shortly before memory retrieval (e.g. Kuhlmann, Piel, & Wolf, 2005; Buchanan, Tranel, & Adolphs, 2006; Smeets, Otgaar, Candel, & Wolf, 2008). On the other hand, studies targeting memory encoding and/or memory consolidation by administering the psychosocial stressor shortly before learning (affecting both encoding and consolidation) or shortly after memory encoding (affecting only memory consolidation) have mostly shown enhancing effects of stress on subsequent memory performances (e.g. Buchanan & Lovallo, 2001; Cahill, Gorski, & Le, 2003; Andreano & Cahill, 2006; Beckner, Tucker, Delville, & Mohr, 2006; Preuss & Wolf, 2009; McCullough & Yonelinas, 2013; but see Trammell and Clore (2014) for an opposite effect). Additionally, it was shown that pre-learning stress influences the learning system used to perform memory tasks, promoting a shift from a cognitve hippocampus-dependent system to a habit striatum-dependent system (Schwabe, Wolf, & Oitzl, 2010; Schwabe, Schachinger, de Kloet, & Oitzl, 2010; Schwabe, 2013; Schwabe & Wolf, 2013).

In our fourth study (see section 2.4), we tested whether the time course of memory transformation from detailed to gist-like memories is influenced by the amount of stress present during the very early consolidation process. This question was recently addressed in a rodent study by Pedraza et al. (2016). There they demonstrated that stress, indeed, had an influence on the memory transformation time course, yet in an inverted U-shaped fashion: moderate stress led to long-lasting detailed representation of the memory, while high levels of stress during learning, on the other hand, led to a transformation to generalized, gist-like versions of the memories earlier than normally expected (Pedraza et al., 2016). We tested, whether these results could also be found in humans.

1.6 Research goals

The overall aim of this thesis is to contribute to the ongoing systems consolidation debate between the SCT and the MTT/TTT and to connect these systems consolidation theories to the notions of memory dynamics. The four studies presented in this thesis therefore investigate the dynamics of the systems consolidation process by focussing on memory transformation (studies 1 and 4), memory updating (study 3), as well as parallels and differences between systems consolidation processes in distinct memory systems (study 2). Out of the four studies, study 1 most directly targets the ongoing systems consolidation debate by investigating the role of the hippocampus for detailed versus transformed remote memories. By using a memory paradigm that compares the memory retrieval of 1 day old memories to 28 days old memories of picture stimuli, we examined the prolonged systems consolidation process of declarative memories over several weeks. Furthermore, by differentiating the hippocampus along its long axis in the data analysis, study 1 focussed on including this important anatomical distinction into the systems consolidation debate. In study 2, we investigated whether prolonged systems consolidation processes can also be seen in nondeclarative memories that are not initially hippocampus-dependent. As the systems consolidation debate so far focussed primarily on declarative memories, and at the same time researchers examining the consolidation of nondeclarative memories only focussed on early systems consolidation processes after one night of sleep, the process of prolonged systems consolidation of nondeclarative memories over several weeks, has, to my knowledge, not been investigated so far. Finding parallels and differences between prolonged systems consolidation of declarative and nondeclarative memories can contribute to an overall better understanding of systems consolidation processes and may also indirectly inform the ongoing debate about declarative systems consolidation.

Another important issue to consider when investigating systems consolidation processes and theories is the fact that memories are highly dynamic and systems consolidation processes are unlikely to unfold in a straightforward manner without any influences from external factors or cues. In study 3, we therefore focussed on memory dynamics by investigating the neural correlates of actively updating memories through new information. Researchers testing the process of memory updating so far, have used experimental set-ups that did not test how memories which have already undergone a systems consolidation process are updated. In our study we therefore allowed memories to undergo an early systems consolidation process before updating them. Understanding how memories after

systems consolidation are updated may also allow indirect conclusions about the systems consolidation process itself.

Finally, in study 4, we tested the influence of post-encoding stress on memory consolidation and transformation processes. Research on the influence of stress on memory consolidation in general has shown beneficiary effects of stress. However, how post-encoding stress influences the process of memory transformation from detailed to gist-like memories has not been investigated in humans so far. Understanding how stress affects systems consolidation and memory transformation processes may additionally provide useful information on systems consolidation in general and may also contribute to the ongoing systems consolidation debate.

In the following chapter, the four studies[6] will be introduced in more detail.

[6] Note that I use the term study here to refer to four different memory tasks and questions investigated and presented in four different papers or manuscripts. Indeed, studies 1 and 2 were thereby conducted within one combined *experimental study* utilizing the same sample, with participants performing both tasks one after another. Nevertheless, I describe the two tasks here as two distinct studies, as each had its own hypotheses, each data set was analysed separately and the results were presented in two different papers/manuscripts. Studies 3 and 4 were conducted in two other unique samples, respectively.

2 Experimental studies

2.1 Study I: Time-dependent memory transformation along the hippocampal anterior–posterior axis

Published in Nature Communications (Dandolo & Schwabe, 2018). The full publication including a detailed methods and results section can be found in Appendix A.

2.1.1 Background

Systems consolidation of initially hippocampus dependent episodic memories has been controversially discussed for decades (Nadel et al., 2007; Squire et al., 2015; Sekeres et al., 2017). While the SCT expects all memories to become hippocampus independent with time (Squire & Bayley, 2007), the TTT suggests that detailed episodic memories remain hippocampus dependent and only the transformed gist-like versions of the memories can be retrieved without the involvement of the hippocampus (Winocur & Moscovitch, 2011). However, within this debate the hippocampus has most often been regarded as one unitary structure. Yet, a wealth of recent research has shown that the hippocampus can be subdivided along its long axis into anterior and posterior parts and that these parts differ in function, structure and connectivity to other brain areas (Fanselow & Dong, 2010; Poppenk et al., 2013; Strange et al., 2014). In this fMRI study we therefore examined memory consolidation and transformation of episodic memories over time, focussing on the role of anterior and posterior hippocampal areas.

2.1.2 Methods

Procedure

We tested 48 healthy young adults (24 men, 24 women) in two experimental sessions. Day 1 consisted of a memory picture encoding task performed outside the MRI scanner and day 2 of a recognition memory task in the scanner. Critically, we varied the time between

encoding and recognition testing in a between-subject design with half of the participants returning to the laboratory for experimental day 2 after one night of systems consolidation (1 day group), and the other half after four weeks (28 days group).

Memory task

During memory encoding participants memorized 60 pictures that were presented in three encoding runs, each run followed by a free recall task. Half of the pictures contained emotionally negative scenes or objects, while the other half contained neutral contents.
In the memory recognition task participants were presented with a total of 180 pictures: the 60 old encoded pictures, 60 completely novel pictures and, critically, 60 related pictures that carried the same gist as one of the old pictures respectively but depicted different details.

2.1.3 Results

Behavioural results

We found a sharp increase in the false alarm (FA) rate for related pictures in the 28 days group that was four times higher than the FA rate for novel pictures, suggesting a transformation towards more gist-like memories after four weeks, as participants had particular problems in differentiating between the details of the two related pictures. Our behavioural data further suggest that stimulus-related emotional arousal influenced the transformation to gist-like memories: after 28 days negative pictures were more often transformed to gist-like versions than neutral ones, while neutral pictures, on the other hand, were more often completely forgotten.

Univariate fMRI results

Overall, task-related hippocampus activity was lower in the 28 days group than the 1 day group. Dividing the hippocampus along its long axis, however, revealed that the time-dependent decrease in activity was restricted to the anterior and mid-portion hippocampus, with activity in the posterior hippocampus not differing between the 1 day group and 28 days group (see Figure 1A). To further examine whether the decrease in anterior hippocampus activity could be directly linked to the change in the nature of remembering, we correlated the activity in the hippocampal subregions with a behavioural index of memory specificity, i.e. the FA rate for related lures. These analyses showed that

specifically the anterior hippocampus activity was positively associated with the specificity of memory (see Figure 1B).

Multivariate fMRI results

Using representational similarity analysis (RSA) we constructed representational dissimilarity matrices (RDMs) in each hippocampal subregion. Visualizations of average RDMs per group suggested that in the 1 day group multivariate activity patterns for old pictures were more similar to each other than patterns of related and novel pictures in all hippocampal subregions. However, in the 28 days group the multivariate activity patterns seemed to be quite random in the anterior hippocampus, while the posterior hippocampus showed similarities between the patterns of old and related pictures (see Figure 1C).

2.1.4 Conclusion

In this study, we could show that memories of pictures are transformed to more gist-like versions after a month of systems consolidation. Importantly, we could link this memory transformation to changes along the hippocampal anterior-posterior axis. Activity in the anterior hippocampus decreased over time and was higher for participants with more detailed memories. On the other hand, overall task activity in the posterior hippocampus was stable across time with multivariate activity patterns for all old and related pictures becoming more similar to each other in this area after 4 weeks.

Overall, we can conclude from this study that the debate about systems consolidation has to go beyond considering the hippocampus as one unitary structure. In other words, the functional differentiation along the hippocampal long axis has to be considered within systems consolidation theories.

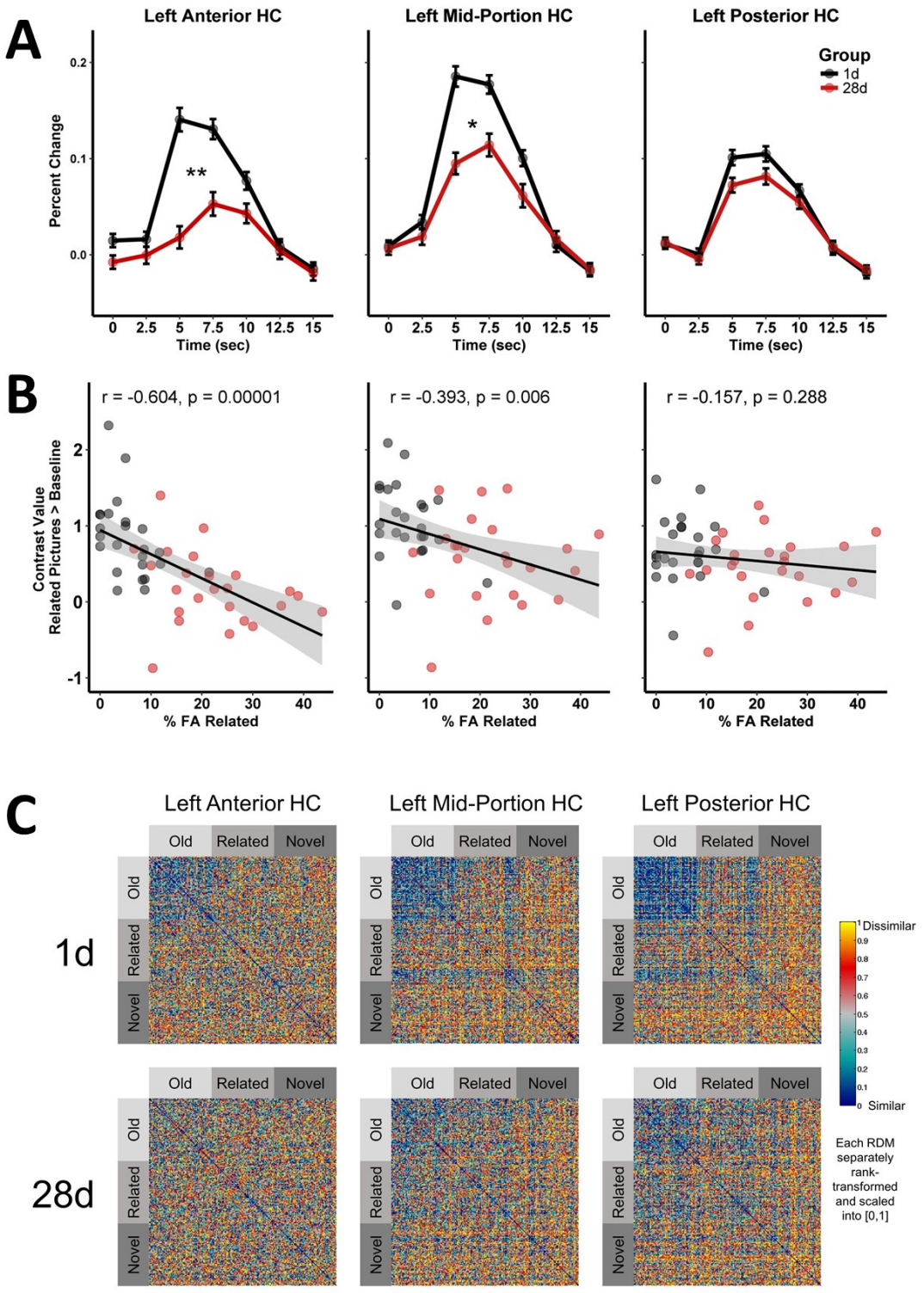


Figure 1: Main results of study 1: **(A)** Finite impulse response time courses over the first 15 seconds for all picture types combined. The activity in the 28 days group compared to the 1 day group decreased in the anterior hippocampus and mid-portion hippocampus but not in the posterior hippocampus. **(B)** Correlations of the percentage of FAs to related pictures with the contrast value for related pictures vs baseline: across groups, there was a negative correlation in the anterior hippocampus and mid-portion hippocampus, but not in the posterior hippocampus. **(C)** The group average RDMs of the long axis hippocampal regions of interest (ROIs) in the two groups. Blue colors= most similar, bright colors= most dissimilar; HC = hippocampus.

2.2 Study II: Time-dependent motor memory representations in prefrontal cortex

Published in NeuroImage (Dandolo & Schwabe, 2019). The full publication including a detailed methods and results section can be found in Appendix B.

2.2.1 Background

The process of systems consolidation has most often been investigated for initially hippocampus dependent declarative memories (Squire et al., 2015). However, there are multiple memory systems (Squire & Dede, 2015) and the question whether memories which are not initially hippocampus dependent undergo a prolonged systems consolidation process has only received little attention so far (Dudai et al., 2015). In this study we investigated motor sequence memory, a form of nondeclarative memory which initially relies on cortico-striatal and cortico-cerebellar systems (Doyon et al., 2003). Studies testing consolidation of these motor sequence memories have focussed either on early systems consolidation processes after one night of sleep (Vahdat et al., 2017; Fogel et al., 2017) or on reorganization processes associated with extensive training (Coynel et al., 2010; Wymbs & Grafton, 2015). However, it remains unclear whether several weeks old remote motor sequence memories, which were not extensively trained, rely on the same neural circuits as more recent (e.g. one day old) motor memories. In other words, is there a prolonged systems consolidation-like process in motor sequence memory?

2.2.2 Methods

Procedure

We tested 48 healthy, right-handed young adults on two experimental days. Due to various reasons (see Appendix B), we had to exclude 9 participants, leaving a final sample of 39 participants for analysis. Day 1 consisted of the learning phase of a SRTT outside the scanner and day 2 of the corresponding test phase in the fMRI scanner. Importantly, the time interval between the learning phase and the test phase depended on the experimental group: participants in the 1 day group performed the test phase one day after learning, while participants assigned to the 28 days group performed the test phase four weeks after learning.

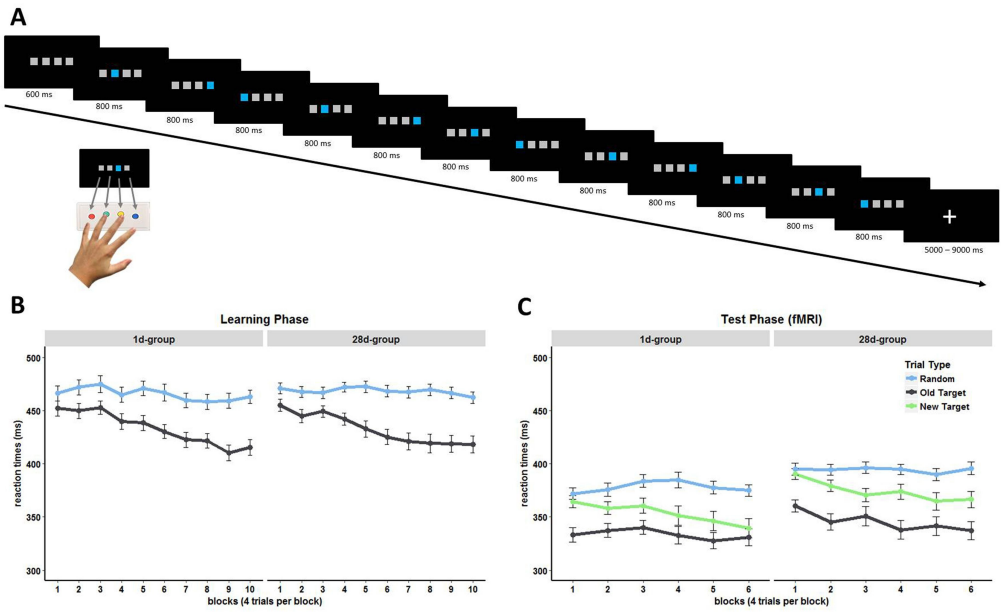


Figure 2: Memory task and behavioural results of study 2: **(A)** Schematic overview of one trial of the SRTT. The bottom left picture shows the visual cue-button-finger correspondence. **(B)** Reaction times in the learning phase were faster for old target trials than for random trials and the reaction times for the target trials decreased throughout the blocks in both groups. **(C)** Reaction times in the test phase for old target trials were, both in the 1 day group and in the 28 days group, faster than reaction times for both random and new target trials, in particular at the beginning of the test phase.

SRTT

In the SRTT participants saw changing visual cues and were instructed to respond to each cue as quickly as possible with a specific corresponding button press (Figure 2A). Unbeknownst to the participants, a specific target sequence of twelve cues was presented 40 times throughout a learning phase of the experiment, pseudo-randomly intermixed with 40 trials containing random sequences. In the test phase on experimental day 2 this first target sequence was presented 24 times, pseudo-randomly intermixed with 24 trials of a second new target sequence and 24 trials of random sequences.

2.2.3 Results

Behavioural results

Reaction time data from the learning phase showed that participants from both groups learned the target sequence very well on day 1 (Figure 2B). At the beginning of the test phase on day 2 the reaction times were faster for old target trials compared to both random and new target trials in both groups (Figure 2C). This finding clearly shows intact motor sequence memory in the 1 day group and, importantly, also in the 28 days group.

Univariate fMRI results

Using the classical general linear model estimation method, we did not find any significant group differences across the whole brain or predefined ROIs. Using Bayesian second level analysis, we found that the probability that the 1 day group showed more activation than the 28 days group across all trial types was higher in the left post central gyrus and, conversely, the probability that the 28 days group showed more activation than the 1 day group during the task was higher in the bilateral middle frontal gyrus, the frontal pole and the occipital pole.

Multivariate fMRI results

We tested six models in a searchlight RSA approach and looked for areas where model fits were different between the two groups. We thereby found two clusters, one in the right superior frontal gyrus and one in the right frontal pole extending into the middle frontal gyrus which showed better model fits in the 28 days group in comparison to the 1 day group for those 3 RDM models which predicted similar activity patterns for all regressors of the old target sequence and similar patterns for all regressors of the new target sequence, but dissimilar patterns between the two sequences (Figure 3). We suggest that these frontal cortex areas are more involved in separating remote (4 week old) motor sequence memory representations from newly learned sequence representations than separating recent (1 day old) sequence memory representation from newly learned sequence representation. There were no differences between the two groups in three control RDM models.

2.2.4 Conclusion

In this study we showed time-dependent changes in the neural system underlying motor sequence memory. More specifically, our findings indicated an increased involvement of lateral prefrontal cortex areas in the representation of remote compared to recent motor sequence memories. However, in contrast to the proposed systems consolidation of episodic memories (Dudai et al., 2015; Squire et al., 2015), this time-dependent increase in the involvement of neocortical areas after four weeks was not paralleled by a decrease in those cortico-striatal areas that supported initial motor learning. Instead, the additional recruitment of lateral prefrontal areas might contribute to a more distributed representation of remote motor memories. These findings provide first insights into how memories beyond the hippocampus evolve over a longer period of time.

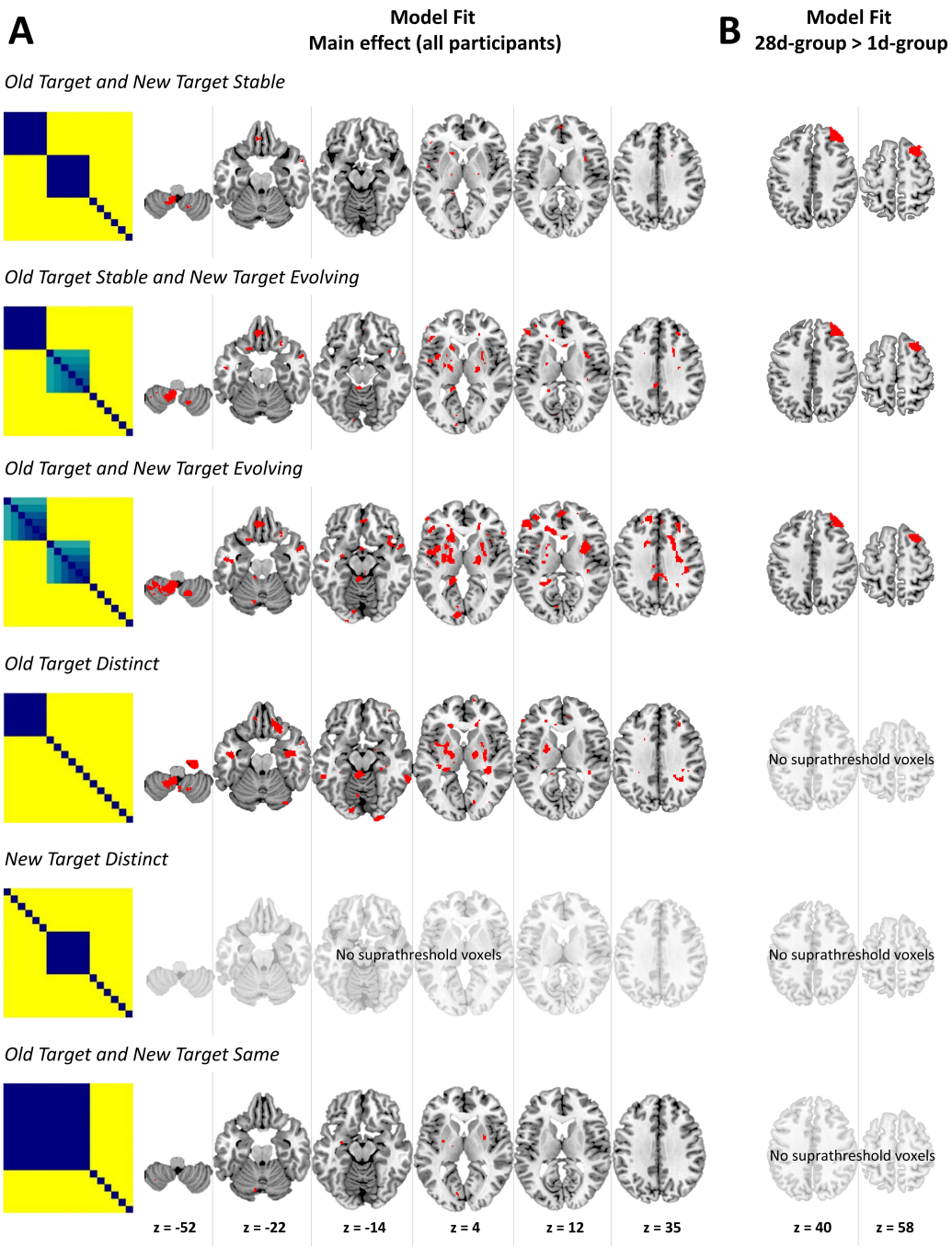


Figure 3: Results from the RSA searchlight analysis in study 2: **(A)** Main effects of the model fits (for all participants) for each of the six models separately. **(B)** Group differences for each model. In right prefrontal areas, we obtained better model fits in the 28 days group compared to the 1 day group for those three models that predicted similar activity patterns for all regressors of the old target sequence and similar activity patterns for all regressors of the new target sequence respectively, but dissimilar patterns between the two sequences. There were no group differences in the other three control models.

2.3 Study III: Dorsolateral prefrontal cortex enables updating of established memories

Published in Cerebral Cortex (Kluen, Dandolo, Jocham, & Schwabe, 2018). The full publication including a detailed methods and results section can be found in Appendix C.

2.3.1 Background

Our environments are constantly changing, requiring our memory systems to be adaptive. In light of new information, established memories need to be updated. Memory updating is also fundamental to accomplish prospective functions of memories, such as imagining and simulating upcoming events (Schacter et al., 2007; Addis et al., 2007), or decision-making (Wimmer & Shohamy, 2012).

The neural correlates of memory updating processes have recently been investigated for immediate episodic memories, thus for memories that have not yet undergone early systems consolidation processes (Kuhl et al., 2012; Nyberg et al., 2009), indicating an important role of prefrontal areas. There is also evidence for an important role of the dorsolateral prefrontal cortex (dlPFC) in working memory updating (D'Ardenne et al., 2012). However, it remains unclear to what extent updating processes for memories that have already undergone an early phase of systems consolidation resemble those for immediate and working memories.

2.3.2 Methods

Procedure

We tested 49 healthy adults in a 3-day study design, including an MRI scanning session on the second day. One participant was excluded because of minimal performance in the task, thus leaving 48 participants for analyses (25 women, 23 men).

Memory updating task

On day 1, participants learned 75 face–city name associations (a person and their last holiday location) in an encoding session comprised of four runs. On day 2, an updating phase was conducted during a fMRI measurement. Participants were again presented with 75 face-city pairs: 25 pairs were identical as on day 1 (old identical trials), 25 pairs involved a face that had been presented on day 1 now paired with a new city (old updated trials),

and 25 pairs were completely new (new trials). On day 3, memory updating success was assessed in a recognition test: each face was presented with four cities and participants were instructed to choose the correct city; in the case of updated trials, they were instructed to choose the new updated city.

2.3.3 Results

Behavioural results

Results from the recognition test on day 3 showed that, for the face–city pairs that were updated on day 2, the updated city name was correctly chosen in about 35% of trials, while in about 57% of trials participants incorrectly persevered with the old wrong city name. Still, city names that were not presented on either day 1 or day 2 were endorsed only in about 4% of trials each. The finding that participants correctly chose the updated city name at average in more than a third of the trials demonstrates participants' capability for updating some of the established memory traces.

Univariate fMRI results

The left dlPFC showed significantly increased activity when old updated trials were shown compared to both old identical trials (Figure 4A), and new trials. In contrast the vmPFC, the angular gyrus, and the anterior cingulate cortex were significantly less activated when old updated trials were presented compared to old identical trials (Figure 4B). Moreover, activity in the left dlPFC during old updated trials was positively correlated with subsequent updating performance (Figure 4C). We additionally found a significant increase in connectivity between the left dlPFC and the left hippocampus for updated trials vs. old identical trials and this crosstalk was also positively correlated with subsequent updating performance.

Multivariate fMRI results

We compared brain RDMs of the ROIs to different model RDMs and then compared model fits between successful and poor updaters. The pattern of results suggests that successful updaters may be characterized by showing an especially consistent activity pattern during the updating process in the left dlPFC. While, in poor updaters, there was no evidence for such a specifically consistent activity pattern during the updating process in the left dlPFC (Figure 4D).

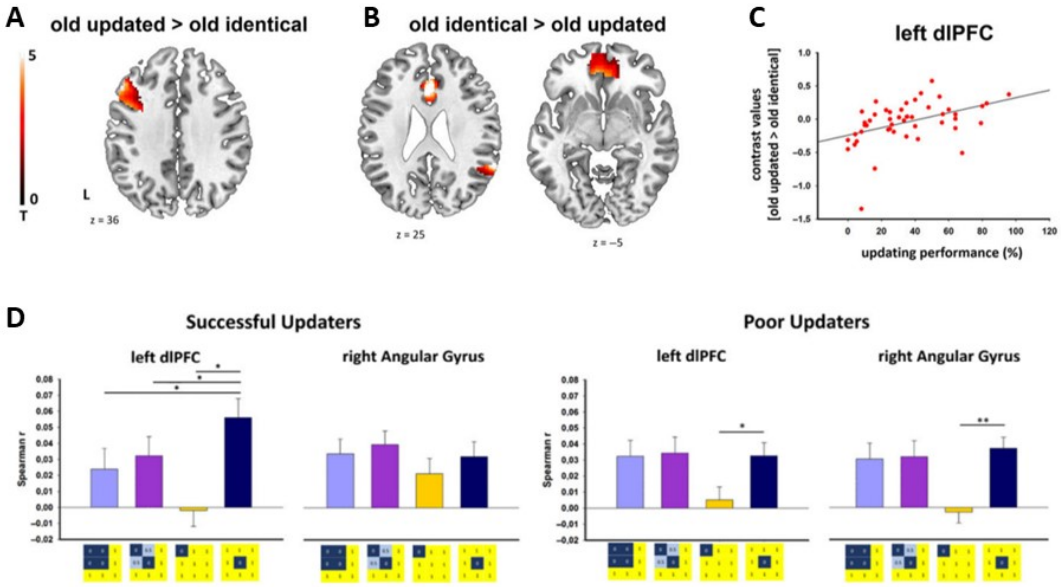


Figure 4: Main results of study 3: **(A)** The comparison of old updated and old identical trials revealed significant activation in the left dlPFC. **(B)** Activity in the anterior cingulate gyrus, bilateral vmPFC, and right angular gyrus was higher for old identical than old updated trials. **(C)** Activity in the left dlPFC was positively correlated with updating success as assessed on day 3. **(D)** Model representation distinguished between successful and poor updaters in the left dlPFC: for successful updaters, a model with high similarity values for old updated trials had the highest model fit value; for poor updaters this model did not have higher model fit values than other control models.

2.3.4 Conclusion

This study indicates a key role of the left dlPFC in the updating of memories after early systems consolidation. The dlPFC has previously been assigned a critical role in working memory updating (D'Ardenne et al., 2012). However, in working memory, there is no need to interact with long-term storage sites. Critically, our results also showed task-related connectivity between the dlPFC and hippocampus that was connected to the updating process. Thus, we demonstrated that the dlPFC, most likely through its interaction with the hippocampus, is essential for keeping memories up to date.

2.4 Study IV: Moderate post-encoding stress shows no influence on memory transformation rate after four weeks of systems consolidation

Unpublished Manuscript. The full manuscript including a detailed methods and results section can be found in Appendix D.

2.4.1 Background

Studies testing the influence of stress on memory consolidation in general have mostly shown enhancing effects of post-encoding stress on subsequent memory performances (e.g. Buchanan & Lovallo, 2001; Cahill et al., 2003; Andreano & Cahill, 2006; McCullough & Yonelinas, 2013). However, these studies have not tested the influence of stress on memory quality. Over time, episodic memories transform from detailed memories to more gist-like versions (Winocur & Moscovitch, 2011). Is this memory transformation over time influenced by the amount of stress present during initial encoding and early consolidation processes? A recent rodent study suggests that the rate of memory transformation indeed depends on the strength of the stress response during learning: moderate stress led to a detailed representation of the memory which remained detailed for a long period of time; a high level of stress, on the other hand, led to a faster rate of memory transformation, resulting in generalized, gist-like versions of the memory after only 2 weeks (Pedraza et al., 2016). We tested whether similar results can also be found in humans.

2.4.2 Methods

Procedure

In total, we tested 162 healthy young adults. After excluding 4 participants for various reasons (see Appendix D) we analysed a final sample of 158 participants. All participants encoded emotionally negative and neutral pictures on experimental day one, shortly before undergoing the standardized socially evaluated cold-pressor test (SECPT; Schwabe, Haddad, & Schachinger, 2008; Schwabe & Schachinger, 2018; see also Appendix D) or a warm water control condition. In a between-subject design, participants performed a free recall task and a recognition task either 1 day or 28 days after encoding. Thus, participants were pseudo-randomly assigned to four different experimental groups: 1d-stress group, 28d-stress group, 1d-control group, 28d-control group.

Memory task

We used the same memory task as in study 1 (see section 2.1.2). However, we added some new pictures, increasing the total number of pictures from 180 to 240.

2.4.3 Results

Overall effective stress manipulation with a moderate responder rate

Overall the stress manipulation was effective. However, when we classified participants into cortisol responders and non-responders only 49% of stressed participants were classified as responders. Therefore, we must acknowledge that we have a rather moderate responder rate. Nevertheless, this moderate rate can benefit correlational analyses of the cortisol increase in the stress group with subsequent memory scores (Schwabe & Schachinger, 2018).

Stress benefits the free recall of negative pictures in females

For female participants stress after encoding when compared to a control task after encoding led to a better subsequent recall performance of negative picture stimuli. This, importantly, was the case for both the 1d-group and the 28d-group.
We did not find effects of stress or cortisol increase on the overall percentage of hits in the recognition task, nor did we find effects when dissociating the recognition memory into recollection and familiarity components based on confidence ratings.

No effect of post-encoding stress on the degree of memory transformation captured after four weeks

We captured the transformation process in three different ways. First, we asked participants to recall as many details as possible during the free recall task on experimental day 2. Secondly, we compared the FA rate of related pictures to the FA rate of novel pictures. Third, we classified picture pairs into detailed, transformed and forgotten pairs. All three ways of capturing a possible transformation process, suggest that, across time, memories are transformed from detailed versions to gist-like versions. Importantly, however, we did not find an effect of the treatment group (stress vs. control) on this transformation process.

2.4.4 Conclusion

We found an enhancing effect of post-encoding stress on the subsequent recall of negative pictures in females. This result fits to the common opinion that stress after encoding benefits subsequent memory retrieval (Roozendaal, 2002, Schwabe et al., 2012; Cadle & Zoladz, 2015). However, our results also showed that this stress effect is not robustly found in all circumstances but depends on many factors (e.g. sex of participant).

Most importantly, we did not find any effects of post-encoding stress or cortisol increase on the degree of memory transformation tested after four weeks of systems consolidation. This is at odds with a recent rodent study (Pedraza et al., 2016) and does not align with human studies that, although not explicitly testing the time course of memory transformation, nevertheless showed an over-reliance on gist-like memories after pre -or post-encoding stress (Payne, Nadel, Allen, Thomas, & Jacobs, 2002; Pardilla-Delgado, Alger, Cunningham, Kinealy, & Payne, 2016). It is important to note here, that the stress response to the SECPT in the present study was in general rather moderate in comparison to other stress studies. Therefore, future studies might profit from stressors eliciting higher stress responses or from a pharmacological administration of cortisol.

3 General discussion

Understanding how memories consolidate and transform across time is one of the fundamental tasks of memory research. Deciphering the neuronal representations of memory and bringing light into the reorganization of these memory representations across brain systems and time are essential building blocks for clinical research e.g. on degenerative memory diseases. Additionally, it is important to understand how external factors such as stress or new conflicting information influence the dynamics of memory processes across time. Knowledge about the neural correlates of these influencing factors may have important implications in clinical research of conditions such as post-traumatic stress disorder, dealing with memory intrusions and distortions.

In this dissertation, I have introduced four studies contributing to the understanding of systems consolidation processes and memory dynamics. In particular, in the first two studies we tracked the neural correlates of systems consolidation processes across time of two kinds of memories, namely declarative episodic memories in study 1 and nondeclarative motor sequence memories in study 2. In study 1, we found that the transformation of episodic memories from detailed to gist-like memories across time is accompanied by changes along the hippocampal anterior-posterior axis. Study 2 is, to our knowledge, the first study to investigate explicitly the prolonged systems consolidation processes of nondeclarative motor memories. We found that after four weeks of systems consolidation without further training, areas in the lateral prefrontal cortex were more involved in performing the motor task and representing the remote compared to recent motor sequence memories. In study 3 and 4, we focussed on memory dynamics by examining external factors which may influence and change systems consolidation processes. In particular, in study 3 we investigated the neural correlates of updating memories through new conflicting information after these memories have already undergone early systems consolidation processes and found that the dlPFC is decisively involved in the actual updating process. In study 4 we stressed participants after memory encoding and tested whether this stressor would influence the transformation of memories across time. However, we did not find any clear effects of

moderate stress or cortisol increase on memory transformation after 4 weeks.

In the following sections, I will discuss the implications of these findings within the framework of existing theories. I will thereby focus on the ongoing debate about systems consolidation of declarative memories between proponents of the two most important theories to date, the SCT and MTT/TTT. After summarizing the problems researchers have so far faced during the examination of these two theories, I will elaborate on how asking more precise research questions may contribute to solving these problems in future research. I will discuss the role of time in systems consolidation, proposing a differentiation of this process into two sub-phases: early and prolonged systems consolidation. Furthermore, I will compare our hippocampal long axis data from study 1 to the recent long axis TTT proposal as introduced in section 1.3.6. As our conclusion and this proposal are contradictory, I will present possible reasons for this contradiction. Additionally, I will integrate the distinction into different systems consolidation phases (early vs. prolonged) in this new hippocampal long axis debate. In the last three sections, I will discuss our findings from study 2 (non-declarative systems consolidation), study 3 (memory updating) and study 4 (influence of post-encoding stress) within the new proposed framework of early and prolonged systems consolidation, respectively.

3.1 Contributions to the systems consolidation debate

The debate about systems consolidation has gone on for decades. Yet, there is still no consensus between memory researchers. Proponents of both the SCT and the MTT/TTT continue to find support for their theories (see e.g. Squire et al., 2015 for the SCT and Sekeres et al., 2017 for the TTT). The main topic of the debate has been whether the hippocampus is needed for the retrieval of remote detailed episodic memories or not. At first glance, one would think this seems to be a rather simple question and amazing that despite all the available methods in neuroscience, this issue is still unresolved, even after 20 years of research. Hence, the question arises: what factors may contribute to this continuing debate?

Methodological limitations

Although the question at the heart of the debate seems rather simple, a review of the research over the last 20 years shows that it is not easy to find an experimental approach and memory paradigm which accounts for all the dimensions this question comprises.
Originally, the debate centred around studies of patients with hippocampal damage leading to retrograde amnesia (e.g. Scoville & Milner, 1957; Moscovitch et al., 1999; Cipolotti et al., 2001; Rosenbaum et al., 2004; Steinvorth et al., 2005; Bayley et al., 2005; Kirwan et al., 2008). However, it is challenging to find patients with complete bilateral hippocampal damage but otherwise intact brains, leading to a few single case studies that lack generalizability (e.g. Rosenbaum et al., 2000; Maguire et al., 2006). In all other cases, if remote memories do not survive hippocampal damage, proponents of the SCT point to damages in other brain areas, that may alternatively be responsible for this memory loss (Squire & Bayley, 2007). By contrast, when amnesia patients show remaining remote detailed memories, proponents of the MTT/TTT point to remaining hippocampal tissue that may support those memories (Nadel et al., 2007). Furthermore, to assess retrograde amnesia, researchers often rely on autobiographical memories, and although new scoring techniques were developed to try and assess explicitly the vividness of remote autobiographical memories (Moscovitch et al., 1999), the problem of verifying the correctness of these memories remains. In addition, researchers cannot control the timing between encoding, lesion occurrence and retrieval in patient studies.
In animal studies, on the other hand, one can control lesion site, size and the timing between encoding and retrieval more easily (e.g. Winocur et al., 2007; Wiltgen & Silva, 2007;

Wang et al., 2009). Yet, in animal studies the fundamental question, whether the tested memories are truly episodic, remains. Can we really compare context-dependent episodic like memories in animals to truly detailed episodic human memories? According to Tulving (2002) animal memories can never be truly episodic. Moreover, we cannot always directly relate findings from non-primate animals at brain systems level to the human brain, as the anatomy is not equivalent, especially in prefrontal brain areas (Seamans, Lapish, & Durstewitz, 2008; Laubach, Amarante, Swanson, & White, 2018).

The emergence of neuroimaging techniques enabled researchers to measure hippocampal activity in healthy humans' brains (e.g. Gilboa et al., 2004; Viard et al., 2007; Bonnici, Chadwick, Lutti, et al., 2012). This allowed for a variety of experimental set-ups, including prospective studies in humans, in which researchers can control the timing between encoding and retrieval (e.g. Takashima et al., 2009; Furman et al., 2012; Harand et al., 2012). However, a common criticism of these studies is that hippocampal activity tested during memory retrieval tasks can be explained by new learning and encoding during the task, instead of actual memory retrieval (Squire et al., 2015), although researchers have tried to account for this through adequate control conditions (e.g. Gilboa et al., 2004).

Moreover, the question of hippocampus dependency of memories is a challenge in all studies with healthy participants (both animals or humans). In healthy participants, the hippocampus may be involved in the retrieval of certain memories, although the retrieval of these memories is not necessarily *dependent* on an intact hippocampus. In other words, it may well be that other brain areas could be utilized to retrieve the respective memories in the case of hippocampus damage.

Overall, each of the experimental approaches has certain disadvantages and methodological limitations. Thus, whatever results have been found in patient, animal or neuroimaging studies, there always remains a reason for criticism, doubt and possible alternative explanations.

Influence of confirmation bias on data interpretation

As a consequence of these possible alternative explanations, researchers have at times reviewed identical experimental research studies and have still come to opposite conclusions. An illustrative example of this phenomenon are two systems consolidation reviews published within days of each other, one by Squire and Bayley (2007), in favour of the SCT, and the other by Nadel et al. (2007), in favour of the MTT. Both reviewed, more or less, the same experimental research papers but nevertheless came to opposite conclusions. This

showcases the power of confirmation bias, defined as the tendency to gather data, knowledge and information that supports one's prior beliefs, assumptions, expectations and hypotheses (Nickerson, 1998). This is a natural human behaviour that can be observed in many actions of everyday life (Glick, 2017). When it comes to research, confirmation bias can, in general, influence all phases of the research process: e.g. the choice of experimental design, the behaviour of experimenters during data collection, the choice of data analysis techniques, and most importantly, in combination with the phenomenon of apophenia (a human tendency of seeing patterns in random data) the interpretation of results (Glick, 2017; Munafò et al., 2017). This last point not only includes the interpretation of one's own experimental data but also the way one perceives, interprets and critiques other people's research. However, seen from a positive perspective, it can be quite beneficial for memory consolidation research that there are two (or more, see section 1.3.1) opposing theories, as the proponents of the different theories challenge each other, eventually leading to more thoroughly designed experimental studies and more precisely explained analysis choices and interpretations. In other research areas, where only one theory dominates, it may well be that this theory is unjustly confirmed due to confirmation bias, without being challenged by alternative ideas. Thus, two opposing theories can, in a way, protect against the pitfalls of confirmation bias. As suggested by Glick (2017) we should therefore seek and encourage disagreement and embrace uncertainty, yet at the same time be humble enough to consider whether some of one's own beliefs are actually true or whether one simply wants them to be true.

Formulating more precise research questions

After reviewing the available research on systems consolidation and considering both sides of the ongoing debate, another problem crystallized: the main research question - whether the hippocampus is necessary for retrieving detailed remote memories - is formulated too broadly and generally. In order to adequately explain the complex process of systems consolidation, it may well be necessary to formulate more precise research questions.

One issue with this broad, general question and the original descriptions of the different systems consolidation theories is that they consider the hippocampus to be one homogeneous brain structure. Only recently has it been suggested that one should consider that the hippocampus varies in structure and function along the hippocampal long axis (Poppenk et al., 2013; Fanselow & Dong, 2010; Sekeres, Winocur, & Moscovitch, 2018). We therefore included the hippocampal long axis in the data analysis of our first study. I

will discuss this issue in more detail in section 3.3.

Another source of confusion is that the terms gist-like, semantic and schema memories have often been used as synonyms in the systems consolidation literature. In fact, in our first study, we used both the terms *gist-like* and *semantic* memory interchangeably. Only recently, Sekeres, Winocur, and Moscovitch (2018) have suggested the need to clearly distinguish between these three terms (see section 1.2.2). In particular, clarification that gist-like memories, although being coarse-grained memories without details, are nevertheless memories of one single episode can be, in my opinion, very useful for more precise data interpretations. Especially when it comes to reviewing and comparing results from studies using different memory paradigms, one should clarify in which cases gist-like memories, semantic or schema memories are being tested, as each kind of memory may rely on different brain areas (Sekeres, Winocur, & Moscovitch, 2018). Applying the proposed definitions by Sekeres, Winocur, and Moscovitch (2018), it should be clarified here that the memories we test with our picture stimuli paradigm are gist-like memories, and not semantic memories. More precisely, when participants can remember that one of the pictures they learned during the memory encoding session depicted e.g. a tractor, but cannot remember the details of this tractor, they have a gist-like memory of the event of seeing a tractor on one of the pictures during encoding. We can therefore not directly compare our results to other recent studies, which tested either schema memories (e.g. Sommer, 2017) or semantic memories (e.g. Chadwick et al., 2016).

Finally, another important issue, which emerged while writing this thesis, is that the general question, whether remote detailed memories are dependent on an intact hippocampus or not, is not precise in its definition of the term *remote*. As a consequence, researchers have used a vast amount of different time intervals in their experimental studies to test *remote* memories and some differences in results may be related to this issue. I will discuss this inconsistent usage of time intervals in more detail in the following section.

3.2 Early versus prolonged systems consolidation

Researchers investigating systems consolidation are well aware of the problem, that the time intervals used in different experimental studies are inconsistent and that the time course of memory consolidation is poorly defined (Takashima et al., 2006; Sekeres et al., 2017). Especially the time frame of patient and autobiographical studies (several years or decades) diverges substantially from the time frame of prospective experimental studies (mostly weeks to months) and it has been suggested before that the term *systems consolidation* may thereby be used for two separated kinds of operation (Takashima et al., 2006). Additionally, Ritchey, Montchal, Yonelinas, and Ranganath (2015) suggested that future research should focus on how very early stages of systems consolidation after only one night of sleep may relate to longer-lasting changes after weeks and whether these different time scales reflect different kinds of changes in respect to memory quality.
Nevertheless, although these researchers mentioned the issue of time intervals in their respective discussions, studies utilizing different time intervals are still often directly compared to each other in reviews and the choice of time intervals in memory designs is not always sufficiently explained.

Importance of first night of sleep

As introduced in section 1.5.1, sleep plays an active role in consolidation of memories, as memories are replayed and strengthened during sleep (Diekelmann & Born, 2010; Born & Wilhelm, 2012). Moreover, sleep selectively enhances specific memories (Fischer & Born, 2009) or changes the quality of memories (Fischer et al., 2006). Notably, there have been several studies showing that not only changes at the cellular level occur in these early stages of consolidation, but also changes at brain systems level are already evident after only one night of sleep (e.g. Bosshardt, Schmidt, et al., 2005; Takashima et al., 2006; Ritchey et al., 2015; Albouy et al., 2008; Debas et al., 2010). In other words, not only the fast, synaptic consolidation takes place within the first night of sleep, but also systems consolidation. Importantly, the first night of sleep after memory encoding seems to be of special importance to the systems consolidation process (Orban et al., 2006; Gais et al., 2007; Sterpenich et al., 2009; Alger, Chambers, Cunningham, & Payne, 2015). Studies using sleep deprivation showed that changes made to the memory trace within the first night of sleep influences the brain networks used for later memory retrieval (Orban et al., 2006), even 6 months after encoding (Gais et al., 2007; Sterpenich et al.,

2009). Thus, cortical activation during retrieval 6 months after encoding still reflected whether the participants had slept during the first night after encoding or not (Gais et al., 2007). Therefore, sleep in the second and following nights after encoding cannot compensate for a deprivation of sleep in the first night after encoding. This insight from sleep research suggests that the systems consolidation process within the first night of sleep is of special importance and may therefore also be qualitatively different from further systems consolidation processes during the following weeks and months.

Introducing two distinct systems consolidation phases

In line with both these insights from sleep research and the above introduced suggestions from Ritchey et al. (2015), I propose here to differentiate between two phases of systems consolidation: early and prolonged systems consolidation.

Early systems consolidation comprises all changes at brain systems level occurring in the first night of sleep, while prolonged systems consolidation describes the changes in memory representations across brain systems seen over several weeks and months.

The debate about systems consolidation might benefit substantially from testing, describing and discussing the effects of early and prolonged systems consolidation separately. When reading and comparing experimental studies, one should be aware that in some cases early systems consolidation processes are being tested while in other cases prolonged systems consolidation processes are. Thus, differences in results may be due to the different phase tested.

Terminology: recent versus remote memories

One difficulty in comparing the systems consolidation phases tested in different studies is that the usage of the terms related to time is not consistent. In particular, the terms *recent* and *remote* memories are used variably. In our studies 1, 2 and 4 we used a between subject design with two groups: a 1 day group, in which participants returned for memory testing one day after encoding and a 28 day group, in which participants returned for memory testing four weeks after encoding. When comparing the memories of these two groups we refer to the memories of the 1 day group as *recent* memories and to the memories of the 28 day group as *remote* memories. Although this usage of recent and remote is comparable to a few studies (e.g. Harand et al., 2012), it quite often differs. For example, the term *recent* is frequently used for memories tested immediately or within a few hours after encoding (e.g. Takashima et al., 2006; Takashima et al., 2009; Yamashita et al., 2009;

Tompary & Davachi, 2017) yet, in contrast, *recent* has also been used for several years' old autobiographical memories (e.g. up to 5 years old in Gilboa et al., 2004). Similarly, the term *remote* is sometimes used for memories only 24 hours old (e.g. Takashima et al., 2006, Takashima et al., 2009), while in other cases for several decades' old autobiographical memories (Gilboa et al., 2004; Bonnici & Maguire, 2018), and, indeed, for any time interval lying in-between these two extremes. It is important to note how the terms overlap. For instance, one day old memories have been referred to as either recent (e.g. in our studies) or remote (e.g. Takashima et al., 2009). This may definitely lead to confusion, especially if studies are directly compared to each other, without carefully describing the different time intervals utilized to test the so-called recent and remote memories in the respective paradigms.

In order to avoid this confusion in terminology, I propose here to use the following three terms for different types of memories depending on which phase of systems consolidation they have already undergone: (1) *immediate* memories are memories directly after encoding up to a few hours after encoding, without including a night of sleep, thus memories that have not yet undergone either early or prolonged systems consolidation; (2) *recent* memories are memories tested 1 day after encoding, including at least one night of sleep, thus memories which have undergone an early systems consolidation phase but not yet the prolonged phase; and (3) *remote* memories, tested after approximately a week of consolidation or at later time intervals (several weeks, months or years), thus memories that have undergone both early and prolonged systems consolidation. To be precise, the transition from immediate to recent memories takes place within the early systems consolidation phase, while the transition from recent to remote memories takes place within the prolonged systems consolidation phase.

Insights about early versus prolonged systems consolidation from prospective declarative memory studies

In our studies testing declarative memories using the picture stimuli paradigm (study 1, see section 2.1 and study 4, see section 2.4), we compare the retrieval of recent memories to remote memories, yet we do not test the retrieval of immediate memories. Therefore, we did not capture early systems consolidation processes and effects in our studies, but focussed exclusively on prolonged systems consolidation. This is important to remember when interpreting our results and especially when comparing them to results of other prospective memory studies.

Unfortunately, in quite a number of the prospective studies testing declarative systems consolidation one cannot distinguish between early and prolonged consolidation as only immediate and remote, but no recent memories were tested (e.g. Yamashita et al., 2009; Viskontas et al., 2009; Furman et al., 2012; Tompary & Davachi, 2017). Studies that tested exclusively early systems consolidation processes showed varying results with either increasing (Bosshardt, Schmidt, et al., 2005) or decreasing (Takashima et al., 2009) hippocampal activity after one night of consolidation. I will return to this controversy when discussing the importance of differentiating along the hippocampal long axis in this early systems consolidation phase (see section 3.3). Prospective studies testing exclusively prolonged systems consolidation processes are rare. In one such study, Harand et al. (2012) showed that a hippocampal activity decrease was observed after prolonged systems consolidation for memories that transformed to more gist-like versions, while the hippocampal activity remained stable for episodic memories. This might suggest, that changes in hippocampal activity in this prolonged phase may be related to changes in memory quality. Indeed, in our first study we also showed that changes in the activity along the hippocampal long axis during prolonged systems consolidation is related to changes in memory quality.

However, comparing the two phases indirectly in different studies is not ideal, as many other factors, such as the used paradigm, analysis technique and sample may have confounding influences. To my knowledge, there has been only one study which assessed immediate, recent and remote memories, enabling a direct comparison between the early and prolonged phase. In this fMRI study by Takashima et al. (2006) participants performed four recognition tests in the scanner on day 1, day 2, day 30 and day 90. They found a decrease in hippocampus activity for confident hits across time and a parallel increase in activity in the vmPFC, supporting the SCT. Interestingly, the largest hippocampal activity decrease was found between day 1 and day 2, while recognition performance did not decrease between these two time points. The authors of the study emphasize that this speaks against the TTT as the decrease in hippocampal activity cannot be related to memory vividness, as this vividness remains the same between the first two time points. This study suggests that the brain systems changes seen after the early systems consolidation phase cannot be directly related to changes in memory quality. Unfortunately, they did not explicitly test whether the further hippocampal decrease seen between day 2 and the later delays was related to memory vividness.

Considering these studies, and in line with the idea of Ritchey et al. (2015), it might well be that there are two different types of mechanisms underlying early and prolonged systems

consolidation. While early brain systems changes may not have an influence on memory vividness (Takashima et al., 2006), later changes in brain systems might be in correspondence to the transformation of memories from episodic to gist-like memories (Dandolo & Schwabe, 2018; Harand et al., 2012).

Timing within prolonged systems consolidation

As seen above my definition for the time interval of early systems consolidation is clearly defined as the first night of sleep after encoding. However, less clear is how to define the prolonged systems consolidation phase, demonstrated by my failure to decide from which point onwards one should use the word *remote* for memories (my suggestion above is approximately one week, but that is a rather vague suggestion). In my proposed definition, I consider the starting point of the prolonged systems consolidation phase to be after early systems consolidation is complete, thus after the first night of sleep after encoding. However, the question is from which point onwards one should say that memories have undergone a prolonged systems consolidation phase. Have two days old memories undergone a meaningful part of this prolonged phase already or is this the case only after a week, or maybe a month?

Related to the above proposal that changes in brain systems in the prolonged systems consolidation phase might be in correspondence to the transformation of memories from episodic to gist-like memories, one idea might be to define prolonged systems consolidation as the time interval it takes for memories to transform from a detailed to a gist-like version. However, as TTT suggests, both a gist-like and an episodic version of the memory may exist in parallel and it may depend on the memory task which memory is retrieved (Furman et al., 2012; Winocur et al., 2010). This makes it difficult to base the time interval on the change in memory quality. It will therefore require further studies and theoretical debates to define this memory phase more clearly and it may well be that this prolonged phase has to be further separated into sub-phases. Nevertheless, subdividing systems consolidation into an early and prolonged phase is a first step in trying to take time intervals into account.

3.3 The hippocampal long axis debate

Other than time, another important point is to acknowledge that the hippocampus is not a homogeneous structure and that we have to consider a functional differentiation along its long axis when discussing systems consolidation of declarative memories. In our first study we examined the hippocampal long axis and concluded that there is a time-dependent reorganization along the anterior–posterior axis, which is related to the transformation of detailed memories into gist-like representations. In particular, we suggest that anterior hippocampal activity is associated with detailed memory and decreases across time, while posterior hippocampal representations are more gist-like at a later retention interval. Our conclusion is therefore contrary to the idea of the long axis TTT proposal that I introduced in section 1.3.6. In this proposal Robin and Moscovitch (2017) and Sekeres, Winocur, and Moscovitch (2018) suggest that specific detailed memories are represented in the posterior hippocampus while gist-like memories of events are found in the anterior hippocampus. This discrepancy is intriguing and may well start a new debate. In the following sections, I will compare our results to the data which led to the long axis TTT proposal and will discuss how confirmation bias, or a lack thereof, may have contributed to the discrepancy. I will also relate the hippocampal long axis debate to the distinction between early and prolonged systems consolidation as introduced above.

Confirmation bias in the hippocampal long axis debate

At this point, it is important to acknowledge that we were not aware of the long axis TTT proposal as stated in Robin and Moscovitch (2017) and Sekeres, Winocur, and Moscovitch (2018) during the analysis of our data and the writing of our paper. We included the hippocampal long axis into our analyses due to other reports of structural and functional differentiation along this axis (e.g. Poppenk et al., 2013; Strange et al., 2014; Fanselow & Dong, 2010), and although the idea of "gist memory = anterior hippocampus and detailed memory = posterior hippocampus" is included in the review by Poppenk et al. (2013), this is just one of many possible distinctions discussed in that review and it was not yet clearly related to the systems consolidation debate and memory transformation idea. We therefore performed exploratory analyses without clear hypotheses regarding the long axis and memory transformation. Our conclusions were therefore purely based on the pattern of data we found in our analysis, without being influenced by a confirmation bias towards the long axis TTT proposal. But what if we would have been aware of this proposal?

Might we have been influenced in interpreting our data differently? In retrospect, this seems unlikely as there are different aspects in our data that cannot easily be aligned with the long axis TTT proposal. Our univariate fMRI data clearly shows that the anterior hippocampus is less active during the memory recognition task after 28 days than after 1 day. At the same time the behavioural data shows that participants rely more on their gist memory to solve the recognition task after 28 days, reflected in the especially high FA rate for related pictures. If the anterior hippocampus is involved in representing the gist, as proposed by the long axis TTT proposal, then why would this area be less involved in performing the task after 28 days? Indeed, we find a negative correlation between the FA rate for related pictures and the anterior hippocampal activity, showing that the more participants rely on gist-like memories the less active the anterior hippocampus is. Moreover, the univariate results additionally show that activity in the posterior hippocampus does not change after 28 days compared to 1 day. If the posterior hippocampus represents the detailed memories as suggested by the long axis TTT proposal, then one would expect a decrease in activity after 28 days in this area, corresponding to the decrease of detailed memories across time seen in the behavioural data. Thus, on the basis of our univariate data, it is hard to imagine how we could have interpreted our findings in line with the long axis TTT proposal.

Therefore, the question arises on which experimental studies the long axis TTT proposal is based, and why our memory paradigm leads to a different conclusion compared to the respective paradigms used in those studies. The long axis TTT theory was introduced in review articles by Robin and Moscovitch (2017) and Sekeres, Winocur, and Moscovitch (2018) and is based on many different experimental studies, as introduced in section 1.3.6. At first sight it seems as if all the available research data is in line with the long axis TTT proposal. However, during the comparison of our data to some other recent multivariate fMRI studies cited in support of the long axis TTT proposal (e.g. Ritchey et al., 2015, Tompary & Davachi, 2017, Bonnici & Maguire, 2018), a closer inspection of the original research papers revealed that some of those multivariate studies were not interpreted in line with the long axis TTT proposal by the original authors of the respective studies.

For example, in a group of studies Bonnici and colleagues (Bonnici, Chadwick, Lutti, et al., 2012; Bonnici, Chadwick, & Maguire, 2013; Bonnici & Maguire, 2018) used multivariate pattern analysis to distinguish between different "recent" (2 week old) and different "remote" (10 year old) vivid autobiographical memories in the anterior and posterior hippocampus. In Bonnici, Chadwick, Lutti, et al. (2012) they found that both recent and

remote memories were represented in the anterior and posterior hippocampus. However, in the posterior hippocampus the multivariate patterns of the different remote memories were more distinguishable from each other than the patterns of the different recent memories. In Bonnici and Maguire (2018) they then retested the same participants two years later with the same memories and showed that again the anterior hippocampus contained equally decodable information about all memories, while the memories which were now 2 years old were more easily distinguished in the posterior hippocampus than they used to be when they were 2 weeks old. The same laboratory had recently proposed that, within the scope of the *Scene Construction Theory* (see section 1.3.1), the anterior hippocampus (especially the medial part of it) is critically involved in the scene construction process of retrieving consolidated memories or imagining future scenes, while the posterior hippocampus is activated during the perception and recollection of externally presented visual scenes but not internally constructed imagination (Zeidman, Mullally, & Maguire, 2015; Zeidman, Lutti, & Maguire, 2015; Zeidman & Maguire, 2016). In line with this work, Bonnici and Maguire (2018) interpreted their multivariate data as follows: the fact that the anterior hippocampus contained distinguishable representations of all the vivid memories accords with the idea that the anterior hippocampus is particularly relevant for scene-based cognition and therefore decisive for the reconstruction of detailed scenes in perpetuity. They speculated that the results they found in the posterior hippocampus might reflect the processing of spatial contexts which could require more re-instantiation for remote memories than recent ones. Thus, Bonnici and Maguire (2018) see the anterior hippocampus as the area representing detailed memories and not gist-like memories, in line with the results from our first study. Interestingly, their data was interpreted contrarily by Robin and Moscovitch (2017) and Sekeres, Winocur, and Moscovitch (2018), underlining the argument that confirmation bias can change the interpretations of results.

In another recent multivariate fMRI analysis, Tompary and Davachi (2017) used a paradigm which paired different trial-unique objects to one of four repeating scenes. They then tested half of the object-scene associations immediately after learning and the other half a week later. They focussed their analysis on the neural similarity between objects which were paired with the identical scene during learning (overlapping similarity) and compared this to the neural similarity between objects which were paired with different scenes during learning (non-overlapping similarity). Their results concerning the hippocampal axis showed that in the posterior hippocampus the neural patterns of overlapping remote memories had become more similar after one week of consolidation when comparing it to the

overlapping similarity of recent memories. The neural similarity of non-overlapping memories stayed the same for recent and remote memories. In the anterior hippocampus, on the other hand, the neural similarity for overlapping memories was comparable for recent and remote memories, while the similarity for non-overlapping memories decreased across time. The authors suggested that these results can be related to the concept portrayed in Davachi (2006) about the perirhinal and parahippocampal cortex (see section 1.3.4), in the sense that the anterior hippocampus, preferentially connected to the perirhinal cortex, is sensitive to specific items and may be responsible for separating distinct memories, while the posterior hippocampus and parahippocampal cortex are more sensitive to spatial contexts and may have a bias in representing overlapping scenes, leading to the integration of related memories (Tompary & Davachi, 2017). Therefore, the authors of this experimental study suggest an interpretation which can, to some extent, be compared to our proposal of detailed memory representation in the anterior hippocampus and gist-like representations in the posterior hippocampus. Nevertheless, once again their data was interpreted to be in line with the contrary long axis TTT proposal by Sekeres, Winocur, and Moscovitch (2018).

Hence, although there is much data in support of the long axis TTT proposal as introduced in section 1.3.6, on closer inspection of the available research not all the data is as clearly in line with the proposal as seems at first sight. On the contrary, some studies can be interpreted in line with our suggestion concerning the hippocampal long axis (e.g. Bonnici, Chadwick, Lutti, et al., 2012, Bonnici & Maguire, 2018, Tompary & Davachi, 2017). An additional observation drawn from the above considerations is that, while multivariate analysis techniques allow exciting new possibilities in data analysis, they may also contribute to the flexibility in data interpretation discerned in the above examples.

Multivariate analysis techniques and the flexibility of interpretation

In general, multivariate analysis techniques bring a new, exciting and meaningful dimension to fMRI studies. Considering the complexity of our brains and our brain functions, it seems reasonable and appropriate to assume that information in the brain is represented by multivariate patterns. Yet, multivariate analysis techniques for fMRI data are still rather new, complex and indeed still under development. Different techniques and toolboxes exist, but currently there are no standardized recommendations on how best to employ these techniques. In many cases the experimental paradigms used in fMRI studies are not yet optimally designed for multivariate techniques leading, at times, to problems during the

interpretation of the results. Another concern is that multivariate techniques are hard to fully understand for readers who themselves may not yet have used these techniques. This can lead to misconceptions and even wrong descriptions of other researchers' data[1], which then might be cited incorrectly in reviews and related studies. As a consequence, multivariate techniques, not only bring new possibilities, exciting innovations and new insights into brain functions, but at this moment in time also lead to a higher flexibility in data interpretation and thus more possibility for confirmation bias[2].

Our own multivariate RSA data in study 1 show highly distinct activity patterns in the anterior hippocampus, compared to the mid-portion and posterior hippocampus for both recent and remote memories. We interpreted this in the sense that the activity patterns of different pictures in the anterior hippocampus are highly distinct as they represent memory details. One might find ways to interpret this data differently, e.g. one could suggest that the distinct representational patterns represent the gist of the pictures, as the gist also differs between e.g. two different old pictures, therefore making the activity patterns between two old pictures dissimilar. In the posterior hippocampus we found higher similarities between all old pictures for recent memories (compared to the anterior hippocampus), and higher similarities between all old and related pictures after 28 days, leading us to suggest that posterior hippocampal representations are more gist-like at a later retention interval,

[1] In the reviews of Robin and Moscovitch (2017) and Sekeres, Winocur, and Moscovitch (2018) it was claimed that the results by Bonnici, Chadwick, Lutti, et al. (2012) and Bonnici and Maguire (2018) show that representations of detailed memories "increased" in the posterior hippocampus, apparently underlining the importance of the posterior hippocampus for detailed memories in comparison to the anterior hippocampus. Importantly, however, the multivariate data actually showed that the classifier accuracy was higher in the posterior hippocampus for different "remote" memories than for different "recent" memories. This does not mean that the posterior hippocampus involvement "increased", but that multivariate patterns for different remote memories became more distinguishable. It is important to be precise in the description of data, as a higher classifier accuracy cannot be interpreted in the same way as an activity increase in an area.

[2] During the interpretation of the multivariate data of Tompary and Davachi (2017) in the review by Sekeres, Winocur, and Moscovitch (2018), it was first admitted that the increases in representational similarity among related events in the posterior hippocampus seem puzzling. Nevertheless, the authors of the review find a way of changing the interpretation of those results to align them with their long axis TTT proposal. They thereby claim that because the posterior hippocampus codes for details it also codes for the relations among them, and as overlapping elements of stimuli will decay less with time, those "details" will lead to greater similarity among related remote memories in the posterior hippocampus. This again shows the powerful influence of confirmation bias and it seems as though the multivariate data here, once more, allows for an especially high flexibility in interpretation.

as the similarity between old and related pictures increases. We thereby infer, that the activity patterns between all old and all related pictures become more similar with time, and not only the patterns between the actual matching old and related picture pairs. Thus, we assume that the activity pattern does not directly code the actual gist information, but instead the process of using the gist memory to solve the memory task. Again, there may be alternative explanations for and interpretations of the representational patterns in the posterior hippocampus. For example, it may be that the similar activity patterns for old and related pictures in the posterior hippocampus after 28 days represent a pattern separation process, which is more needed after 28 days than after 1 day. Indeed, it has been suggested elsewhere that the posterior hippocampus is active whenever a fine discrimination between old stimuli and related lures is required (Moscovitch et al., 2016). Therefore, it could be that our multivariate results may be interpreted differently by other researchers, especially when trying to align them with other prior beliefs and theories. Having said that, when interpreting our multivariate data in accordance with our univariate data, I still consider our initial interpretation to be the most coherent and conclusive. Nevertheless, future research on this topic using RSA should focus on developing a memory paradigm optimally designed for RSA analysis thereby limiting these alternative explanations.

Overall, although multivariate techniques seem to allow a high flexibility in data interpretation at this moment in time, the benefits of these techniques still prevail. Surely the knowledge of researchers about the correct usage and interpretation of multivariate techniques will grow in coming years and experimental designs will be optimized, leading to clearer conclusions from multivariate data.

Early versus prolonged systems consolidation in the hippocampal long axis debate

Just as with the systems consolidation studies in general, experimental studies focussing on differences along the hippocampal long axis in systems consolidation use a variety of time intervals, from one night of sleep (Ritchey et al., 2015), to a week of consolidation (Tompary & Davachi, 2017), up to differences between 2 weeks and 10 year old memories (Bonnici, Chadwick, Lutti, et al., 2012).

Ritchey et al. (2015) focussed on differences between the anterior and posterior hippocampus using a memory paradigm testing immediate and recent memories, thus capturing early systems consolidation processes. Using a similar paradigm as in the above described study by Tompary and Davachi (2017), Ritchey et al. (2015) also linked different objects to one of eight rooms, and then compared neural similarity patterns of objects linked to the same

room to similarities of objects linked to different rooms. After a delay of only one night, they did not find an increase over time in same-room similarity, neither in the posterior nor the anterior hippocampus[3]. They did report univariate differences in recollection-related activity along the long axis: activity in the posterior hippocampus decreased after one night of sleep, while activity stayed stable over 1 day in the anterior hippocampus. This pattern of results indeed fits with other early systems consolidation studies. Although not focussing on the hippocampal long axis, a closer look at the results of Takashima et al. (2009) shows that the reported decrease in hippocampal activity after one night of sleep was indeed located in the posterior hippocampus. Furthermore, a closer look at the Bosshardt, Schmidt, et al. (2005) study shows that, although they found an increase in left hippocampal activity after 1 day, they also found reduced right posterior hippocampal activity after one night of sleep. However, the right posterior hippocampal activity was neither correlated with retrieval quantity nor retrieval quality, while they found that the relation between left anterior hippocampal activity and retrieval success became closer after a night of sleep.

Harand et al. (2012), testing prolonged systems consolidation, also reported differences between the anterior and posterior hippocampus, although this was not their main focus of attention. They found that in the posterior hippocampus the activation remained stable for memories which were rated as remembered memories at both 3 days and 3 months, while the anterior hippocampus was more activated for recent than remote memories. This corresponds well with our univariate data which shows the same pattern of results.

Taken together, the picture emerging is that activity in the posterior hippocampus decreases within the first night of sleep but then remains rather stable during prolonged consolidation. On the other hand, no activity decrease has been reported in the anterior hippocampus after early systems consolidation (in fact Bosshardt, Schmidt, et al. (2005) reported an increase), but instead a following decrease during prolonged systems consolidation can be observed. Combining this with the previous considerations, I suggest that

[3] They did find a significantly higher same-room than different-room similarity exclusively in the anterior hippocampus in the immediate condition and a correlation between this difference and a behavioural parameter of context memory in the 1 day delayed condition. Overall, Ritchey et al. (2015) suggest that their neural similarity effects in the anterior hippocampus might reflect the representation of generalized context information, while the posterior hippocampus might code for more specific representations of context, thus this interpretation of the data is in line with the long axis TTT proposal. However, one could also interpret the representation of the specific context in the anterior hippocampus as a reflection of a detailed context-specific memory.

there might be two different mechanisms active during early and prolonged systems consolidation respectively: (1) an early decrease in activity in the posterior hippocampus which is not directly related to changes in observed memory quality and (2) a later decrease in anterior hippocampus activity which is related to the transformation of memories from detailed to gist-like versions over time.

To be precise, the activity decreases after one night of sleep do not seem to be related to those changes in memory quality which can be observed using recognition or recall memory tasks, or remember/know procedures. Thus, memories after one night of sleep are still detailed enough to differ between a learned picture and a related lure picture, and subjectively participants feel as if their memories from the day before are still in the "remember" category. However, it could be that there are changes in memory quality over night that are subtler, not affecting memory task behaviour or remember/know responses. These changes might include early forgetting processes of less important event details that do not affect subjective feeling of memory vividness.

Of course, all the above suggestions remain to be tested in future hippocampal long axis studies, designed to capture both early and prolonged systems consolidation, as well as memory quality changes at different levels.

3.4 Interim summary

The main research goal of this thesis is to contribute to the ongoing systems consolidation debate between SCT and MTT/TTT and relate systems consolidation to the concept of memory dynamics. A thorough review of research papers regarding the systems consolidation debate has underlined the importance of formulating more precise research questions. As can be seen in our first study, dividing the hippocampus along its long axis provides valuable new information, suggesting that the transformation of memories is related to a change along its anterior-posterior axis. A recent extension of the TTT (Sekeres, Winocur, & Moscovitch, 2018; Robin & Moscovitch, 2017) has also included the hippocampal long axis, although the proposed differentiation along the long axis is contrary to our suggestion. A comparison of the data leading to their proposal and our data has shown that the interpretation of multivariate fMRI results is quite flexible. Future research should therefore concentrate on finding memory paradigms optimally designed for multivariate analysis which allow a clearer interpretation of the data, eliminating alternative explanations. While comparing our data to former research, it additionally crystallized that the time interval used in systems consolidation studies is quite variable. Informed by insights from sleep research, I propose a differentiation between an early (first night of sleep after encoding) and a prolonged systems consolidation phase. As described in section 3.3, taking both the hippocampal long-axis and the distinction between early and prolonged systems consolidation into account simultaneously has led to new suggestions about the dynamics of memory representations across both the hippocampal long axis and time in relation to changes in memory quality, which should be tested in future studies.

While our first experimental study is directly related to the main research goal of this thesis and has therefore been extensively considered in the above discussion, the other experimental studies focus on important related aspects: systems consolidation of non-declarative motor memories (study 2), the dynamic process of updating memories with new information (study 3) and the influence of post-encoding stress on systems consolidation and memory transformation (study 4). The remaining sections will discuss the results of these three studies with a focus on how they may also contribute to the ongoing systems consolidation debate by interpreting them with respect to the newly proposed distinction of early versus prolonged systems consolidation.

3.5 Early and prolonged systems consolidation of motor sequence memories

Most of the systems consolidation debate has focussed on declarative memories. However, as described in section 1.2, there are several other memory systems comprised under the umbrella term nondeclarative memories. In our second fMRI study we investigated one type of nondeclarative memory, i.e. motor sequence memories. Up to now, studies investigating consolidation of motor sequence memories have exclusively tested either early systems consolidation processes after one night of sleep (Debas et al., 2010; Debas et al., 2014; Gabitov et al., 2014; Vahdat et al., 2017; Fogel et al., 2017), or a type of learning that is characterized by extensive repetitive motor sequence training over several weeks (Krakauer & Shadmehr, 2006; Ungerleider et al., 2002). This type of repetitive training can be related to everyday life situations such as the learning of a specific piece of music on an instrument through many rehearsals. However, there are also several situations in life, where one learns a specific movement one day and then does not repeat it for a long time, similar to an episodic memory of an event which is not frequently retrieved. Indeed, our behavioural results and other behavioural studies (Julius & Adi-Japha, 2015; Savion-Lemieux & Penhune, 2005) have shown, that one is capable of expressing several weeks old motor memories, even if these were not extensively trained. Thus, investigating the neural correlates of a prolonged systems consolidation phase of motor sequence memories which were not extensively trained is of general interest. Surprisingly, we are not aware of a previous study that explicitly tested this prolonged systems consolidation process of motor sequence memories.

Studies investigating the early systems consolidation phase of motor sequence memories have shown that, after only one night of sleep, there are already changes at brain systems level in the representation of motor memories. As described in more detail in section 1.4.1, these include changes of activity in the striatum (Debas et al., 2010) and greater between-regions interactions within the cortico-striatal system (Debas et al., 2014). Interestingly, a recent study using simultaneous EEG-fMRI recordings to directly capture reactivations of motor sequence memory traces during post-learning sleep reported a reorganization of the memory trace within subregions of the striatum, from the rostrodorsal to the caudoventral subregion (Fogel et al., 2017). This suggests that there may be differences within the striatum along its axis, just as there are differences along the hippocampal long axis. An interesting observation that should be addressed in future studies and again underlines the

importance of investigating anatomical subregions more precisely, and not merely averaging across regions.

Another interesting point taken from the early systems consolidation studies of motor sequence memories is, that all of the observed changes at brain systems level can be directly related to offline gains observed in the behavioural data (Debas et al., 2010; Debas et al., 2014; Gabitov et al., 2014; Vahdat et al., 2017; Fogel et al., 2017), suggesting that early consolidation processes not only stabilize memories but even enhance performance (Karni et al., 1998; Korman et al., 2003; Krakauer & Shadmehr, 2006; Dayan & Cohen, 2011). Therefore, these early brain systems changes are related to the quality of the motor memories in a positive direction, thus increasing performance. As discussed earlier, studies investigating early systems consolidation of declarative memories found changes at brain systems level which were not directly related to memory quality (Takashima et al., 2006). As I suggested above, it might nevertheless be that there are changes of memory quality related to the brain systems changes, e.g. early forgetting processes, which are just not captured by the respective memory tasks. In light of the offline gains seen in connection with early systems consolidation of motor sequences, one could speculate that changes at brain systems level of declarative memories may also be related to a strengthening or even enhancement of memories after one night of sleep. One possibility might be that there is a parallel decrease of less important event details and increase of the important event details, strengthening the relevant memories. Motor sequence memories are tested by comparing the reaction times of target and random sequences, thereby capturing even small differences in performance, enabling us to detect offline gains. Thus, more precise measures of memory quality in declarative memory paradigms may also reveal changes of memory quality not captured by the simple yes/no answers in recognition tests. Thus, these insights from early systems consolidation of motor memories underline the importance of establishing memory paradigms which can test memory quality at different levels.

In our study of prolonged systems consolidation of motor sequence memories, we showed time-dependent changes in the neural representation of these memories, in particular a stronger involvement of (lateral) prefrontal areas, mainly the middle frontal gyrus and the superior frontal gyrus, in differentiating remote motor sequence memories from new learning, compared to differentiating recent motor sequence memories from new learning. Interestingly, we did not find any areas in which the multivariate activity pattern of the motor sequence memory was more distinctly represented in the 1 day group than the 28 days group. In other words, the cortical, striatal, and cerebellar regions relevant for motor

memory (and active in our analyses of overall task-related activity) after 1 day appeared to be equally relevant for motor memory after 28 days. Thus, we did not find any brain area showing a decrease in activity for remote motor memories that could be compared to the decrease in activity found in the hippocampus for declarative memories. One explanation for this could be that the temporal profile of systems consolidation processes for motor memories is different than what we know from episodic memory. Therefore, it could be that at a time interval longer than 4 weeks we would find a decrease in e.g. striatal involvement (as we do in the hippocampus for declarative memories at the four weeks interval). Consequently, testing parallels and differences between the reorganization of motor and episodic memories at different time intervals would be highly interesting.

Another interesting aspect for future studies would be to test whether motor sequence memories also transform to less precise versions over time, parallel to the transformation of episodic memories to gist-like memories during prolonged systems consolidation of declarative memories. In the discussion of our paper we therefore suggest that future studies could address this question by using new sequences in the test session which resemble the initially learned target sequence to a certain extent.

Although our study only gives first insights into prolonged systems consolidation of motor sequence memories, it identifies one area, namely the (lateral) prefrontal cortex, which seems to be important for the representation of remote motor memories. However, as prolonged systems consolidation of motor sequence memories has been neglected so far by memory researchers, it is hard to relate our findings to previous research. Moreover, as we identified this area in an exploratory searchlight analysis, the next step is to explicitly test the importance of this prefrontal area for remote motor memories using confirmatory study designs. As outlined above, many other questions concerning prolonged systems consolidation of motor memories remain to be investigated in future studies, e.g. testing different time intervals, as well as examining the quality of the remote motor memories. Thus, prolonged systems consolidation of motor sequence memories should be given more attention in future studies.

3.6 Updating of memories after early systems consolidation

In our first two studies we focussed on prolonged systems consolidation processes under normal circumstances. In other words, we did not systematically interfere with or actively change the systems consolidation process. In our third study we focussed instead on the process of actively updating information which had previously been learned. The ability to update information is essential in everyday life as the environment, information and facts about the world around us are constantly changing (Lee, 2009; Exton-McGuinness et al., 2015). In order to act and make decisions, we need to be able to use the newest information available to us, without interference from old outdated information. Therefore, the ability to successfully update memories without persevering on old information is an important feature of our memory system. However, the actual process of actively updating memories with new information has hardly been investigated so far. Nyberg et al. (2009) and Kuhl et al. (2012) tested the updating of immediate memories, thus memories that have not yet undergone any systems consolidation processes. In our study we focussed instead on the important question, how updating of recent memories after early systems consolidation is accomplished. We identified the dlPFC as the main area in orchestrating the memory updating process, most likely through its interaction with the hippocampus. Interestingly, similar areas were identified to be involved in the updating of immediate memories: Nyberg et al. (2009) also identified an area in left prefrontal cortex and Kuhl et al. (2012) reported an area in the inferior frontal gyrus, not too far from the area we found in the dlPFC. So, it seems that the control executed by the dlPFC during memory updating is similar no matter whether the updated memories have undergone early systems consolidation processes or not.

When considering everyday life scenarios, it often happens that one receives new information about facts learned quite some time ago. An interesting idea for future studies would therefore be to test the updating of remote memories after a prolonged systems consolidation phase. Several questions could be of interest in this regard: Is the dlPFC still mainly involved in the updating process, even for remote memories? Can gist-like memories also be successfully updated? If yes, is the updating success higher or lower when updating gist-like memories compared to detailed memories? Does the dlPFC interact more with other neocortical structures when updating remote memories, than it does during the updating of recent memories?

We also showed that the interaction of the dlPFC with the hippocampus is related to the

memory updating success. Another interesting idea for future studies could be to investigate with which part of the hippocampal long axis the dlPFC interacts. Consequently, it would be interesting to see if this changes over time. In other words, if the dlPFC interacts with different hippocampal long axis regions, depending on whether the memories having to be updated are recent or remote, assuming that the dlPFC is still involved in the updating of remote memories.

In summary, memory updating is an essential process within our memory system, but its neural correlates have not yet been investigated sufficiently and adequately enough. Our study gives first important insights into the updating process of consolidated memories after early systems consolidation, but several other questions have been identified and outlined above, especially concerning the updating of remote memories after prolonged systems consolidation. Thus, several aspects concerning memory updating remain to be investigated in future studies, only then can conclusions be drawn which may also inform the ongoing systems consolidation debate.

3.7 Influence of emotions and stress on memory transformation during prolonged systems consolidation

Our personal experiences from everyday life suggest that not all memories are retained at the same level of detail over time. When remembering our past experiences, one common observation is that emotional and stressful situations seem to be remembered for a longer period of time than neutral events. Memory quality can thereby be either influenced by internal emotional features of the events themselves or by external factors such as stress occurring in close proximity to otherwise neutral events. In this thesis we therefore tested the influence of emotional arousing stimuli (study 1 and study 4) and external post-encoding stress (study 4) on the time course of memory transformation. Former research has shown an enhancing effect of both emotional arousing stimuli (Sharot & Phelps, 2004; Yonelinas & Ritchey, 2015) and pre- and post-encoding stress (e.g. Buchanan & Lovallo, 2001; Cahill et al., 2003; Andreano & Cahill, 2006) on memory consolidation in general. In both cases it has been suggested that the beneficial effects on overall memory strength are orchestrated by the amygdala (Cahill et al., 1995; Dolcos et al., 2004; Murty et al., 2010; McGaugh, 2000). But do emotional arousing stimuli and/or post-encoding stress also influence the transformation of memories from episodic to gist-like memories?

Influence of emotional arousing stimuli on memory transformation

In our first study, we did find an influence of emotional arousing stimuli on memory transformation: negative pictures were significantly more often transformed to gist-like versions than neutral pictures. On the other hand, neutral pictures were more often forgotten than negative ones. This is in line with former accounts suggesting a trade-off between an overall superior memory for emotional material and reduced memory for contextual details of these emotional memories (Kensinger, Garoff-Eaton, & Schacter, 2007; Waring & Kensinger, 2011). Importantly, this emotion induced memory trade-off can already be observed before memory consolidation, directly after encoding (Kensinger, Piguet, Krendl, & Corkin, 2005). Nevertheless, a same kind of trade-off seemed to dictate memory transformation in our first study: on the one hand emotional stimuli are superior to neutral ones, as they are less often completely forgotten, but on the other hand, the emotional memories are more often transformed to gist-like versions than neutral ones, reducing memory details.

However, we could only replicate parts of this trade-off in our fourth study using the same

memory paradigm. What we could replicate is the effect, that neutral pictures were more often forgotten than negative pictures after 28 days, thus the overall memory strength was again higher for emotional material. However, in this study we did not find a higher number of transformed memories for negative pictures than neutral pictures after 28 days. Thus, the effect that negative pictures are more often transformed to gist-like versions could not be replicated. It will therefore require further studies to draw clear conclusions about the influence of emotional arousing stimuli on memory transformation during prolonged systems consolidation.

Influence of post-encoding stress on memory transformation

We did not find any clear effects of post-encoding stress or cortisol increase on the degree of memory transformation tested after four weeks of systems consolidation in our fourth study. Contrary to our findings, a recent rodent study had suggested an inverted-U-shape relationship between stress and memory transformation with moderate stress during encoding leading to subsequent detailed memories for a long period of time, while high amounts of stress during encoding, on the other hand, led to an earlier transformation to gist-like memories (Pedraza et al., 2016). Additionally, human studies also showed an over-reliance on gist-like memories after pre -or post-encoding stress (Payne et al., 2002; Pardilla-Delgado et al., 2016), although these studies did not explicitly test the time course of memory transformation.

It is important to note here, that the stress response to the stress test in the present study was in general rather moderate in comparison to other stress studies. Thus, when testing for quadratic relationships between the cortisol increase and the memory transformation scores, the small number of participants with high stress responses might have limited the power to find an inverted-U-shape relationship as suggested by Pedraza et al. (2016). Therefore, future studies with stressors eliciting higher stress responses or with pharmacological administration of cortisol, might find effects of stress or cortisol on memory transformation which could not be uncovered by our moderate stress response.

Additionally, future studies might also benefit from testing the influence of post-encoding stress on the time course of memory transformation at different time points. It could be that our interval of 28 days does not capture potential variations in the time course of memory transformation induced by stress. It may be that stress effects are observable at later time intervals when memories encoded under normal circumstances, i.e. without stress, start fading more rapidly.

It might also be that only stressors more directly related to the learned material have an effect on memory transformation. In our experiment the participants learned picture stimuli and were then stressed with a subsequent socially evaluated cold water stress test performed in a different room. Thus, the learned material and stressor were only connected by a close proximity in time, but not in context. Testing the effects of a stressor more closely related to the learned material might also lead to different results in respect to memory transformation.

Overall, although we did not find stress effects in our post-encoding stress study, it may well be that there are circumstances under which stress could influence memory transformation, e.g. higher stress responses or testing the transformation at later time points. It will therefore require further studies, e.g. using other stressors and time intervals, to draw clear conclusions about the influence of post-encoding stress on memory transformation during prolonged systems consolidation.

3.8 Conclusion

Although memory functions are essential in our everyday lives and memories form the base of many other cognitive functions, such as decision making (Wimmer & Shohamy, 2012), we do not yet know how exactly memories are consolidated and stored in our brains. Different theories exist about the concept of systems consolidation and as recent reviews show (e.g. Sekeres, Winocur, & Moscovitch, 2018) available theories are still in the process of being updated and modified by new information. Thus, just as memories themselves evolve over time, memory consolidation theories are also still evolving. The studies discussed in this thesis contribute to this ongoing process. Comparing them to other available research generated new ideas, e.g. the division of systems consolidation into an early and prolonged phase. I suggest, that the early phase comprises all changes at systems level after one night of sleep, while the prolonged phase tracks brain systems changes over the following weeks. Furthermore, I propose that these different phases are related to distinct kinds of functional processes in relation to memory quality: while memories are transformed from detailed to gist-like versions in the prolonged systems consolidation phase, early changes in brain systems during the first night of sleep are not directly related to memory vividness, at least not when tested using recognition or remember/know tasks.

Our second and third study both examined processes which have, to my knowledge, not been investigated so far, i.e. prolonged systems consolidation of motor sequence memories (study 2) or the updating of memories with new information after an early systems consolidation phase (study 3). They therefore present first interesting insights into the respective processes and lay the foundations for future research.

However, the biggest proportion of this thesis concentrated on the discussion of our results from our first fMRI study about the transformation of memories along the hippocampal long axis, suggesting that the anterior hippocampus represents detailed memories while the posterior hippocampus is involved in representing gist-like memories. By including the differentiation along the hippocampal long axis in our analysis, we investigated a highly important, exciting and up-to-date research question. Intriguingly, our results are in direct contradiction to a recent long axis proposal by Robin and Moscovitch (2017) and Sekeres, Winocur, and Moscovitch (2018). Instead of being discouraged by the fact that our data does not support the current theory, it is important to remember that researchers should in fact seek and embrace disagreement (Glick, 2017), as this encourages fruitful debates, more thoroughly designed experimental studies and analysis techniques and can weaken the

influence of confirmation bias. Therefore, the main contribution of this thesis to systems consolidation research is to challenge the current theory introduced by Sekeres, Winocur, and Moscovitch (2018) about the role of the hippocampal long axis in memory representation over time. In fact, a closer inspection of the experimental studies cited in Sekeres, Winocur, and Moscovitch (2018) demonstrated that some of those studies (e.g. Bonnici & Maguire, 2018; Tompary & Davachi, 2017) present results and conclusions not necessarily in line with the long axis theory proposed by Sekeres, Winocur, and Moscovitch (2018), but instead can be interpreted in line with our results.

Moreover, our fMRI studies demonstrate the benefits of using multivariate analysis techniques, in particular RSA, to analyse data. Multivariate activity patterns can represent valuable information not captured by the mass univariate approach. Nevertheless, in this thesis I also described an increased flexibility in interpreting multivariate data which can, at times, lead to misinterpretations. It is therefore essential that the research community increases the collective knowledge about these multivariate techniques and the correct usage and interpretation of the data. Most importantly, experimental tasks should be developed in ways optimally designed for the respective analysis technique. If this can be achieved, future research will benefit substantially from multivariate techniques.

In general, there are still many questions remaining about the process of systems consolidation. For example, the time line of systems consolation is still poorly understood. I proposed here a differentiation in an early and prolonged systems consolidation phase. However, the prolonged phase may well need to be subdivided into further sub-phases. In addition to the integration of the hippocampal long axis into systems consolidation theories, it may also be important to clarify the roles of hippocampal subfields, such as CA, DG and subiculum, as well as the surrounding MTL cortices, such as the perirhinal and parahippocampal cortex, in systems consolidation. As outlined in sections 3.5 - 3.7, several questions regarding the systems consolidation of nondeclarative memories, the updating of memories after prolonged systems consolidation and the influence of emotional stimuli and stress on memory transformation also remain to be answered. Thus, the coming years and decades should bring interesting, challenging and exciting new insights and updates of memory theories. Only the future will show how our results will have contributed to the ever-evolving theories of systems consolidation, memory transformation and memory updating.

References

Addis, D. R., Moscovitch, M., Crawley, A. P., & McAndrews, M. P. (2004). Recollective qualities modulate hippocampal activation during autobiographical memory retrieval. *Hippocampus*, *14*, 752-762.

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, *45*, 1363-1377.

Adolphs, R., Cahill, L., Schul, R., & Babinsky, R. (1997). Impaired declarative memory for emotional material following bilateral amygdala damage in humans. *Learning and Memory*, *4*, 291-300.

Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behavioral and Brain Sciences*, *22*, 425-489.

Agren, T. (2014). Human reconsolidation: A reactivation and update. *Brain Research Bulletin*, *105*, 70-82.

Akers, K. G., & Frankland, P. W. (2009). Grading gradients: Evaluating evidence for time-dependent memory reorganization in experimental animals. *Journal of Experimental Neuroscience*, *2*, JEN.S2391.

Albouy, G., Fogel, S., King, B. R., Laventure, S., Benali, H., Karni, A., ... Doyon, J. (2015). Maintaining vs. enhancing motor sequence memories: respective roles of striatal and hippocampal systems. *NeuroImage*, *108*, 423-434.

Albouy, G., King, B. R., Maquet, P., & Doyon, J. (2013). Hippocampus and striatum: dynamics and interaction during acquisition and sleep-related motor sequence memory consolidation. *Hippocampus*, *23*, 985-1004.

Albouy, G., Sterpenich, V., Balteau, E., Vandewalle, G., Desseilles, M., Dang-Vu, T., ... Maquet, P. (2008). Both the hippocampus and striatum are involved in consolidation of motor sequence memory. *Neuron*, *58*, 261-272.

Albouy, G., Sterpenich, V., Vandewalle, G., Darsaud, A., Gais, S., Rauchs, G., ... Maquet, P. (2013). Interaction between hippocampal and striatal systems predicts subsequent

consolidation of motor sequence memory. *PLoS One*, *8*, e59490.

Alger, S. E., Chambers, A. M., Cunningham, T., & Payne, J. D. (2015). The role of sleep in human declarative memory consolidation. *Current Topics in Behavioral Neurosciences*, *25*, 269-306.

Alvarado, M. C., & Bachevalier, J. (2005). Comparison of the effects of damage to the perirhinal and parahippocampal cortex on transverse patterning and location memory in rhesus macaques. *The Journal of Neuroscience*, *25*, 1599-1609.

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: a simple network model. *Proceedings of the National Academy of Sciences of the United States of America*, *91*, 7041-7045.

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

Andreano, J. M., & Cahill, L. (2006). Glucocorticoid release and memory consolidation in men and women. *Psychological Science*, *17*, 466-470.

Baddeley, A. (2003). Double dissociations: Not magic, but still useful. *Cortex*, *39*, 129-131.

Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. (2008). Pattern separation in the human hippocampal ca3 and dentate gyrus. *Science*, *319*, 1640-1642.

Bannerman, D. M., Rawlins, J. N. P., McHugh, S. B., Deacon, R. M. J., Yee, B. K., Bast, T., ... Feldon, J. (2004). Regional dissociations within the hippocampus-memory and anxiety. *Neuroscience and Biobehavioral Reviews*, *28*, 273-283.

Bayley, P. J., Gold, J. J., Hopkins, R. O., & Squire, L. R. (2005). The neuroanatomy of remote memory. *Neuron*, *46*, 799-810.

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*, *37*, 135-144.

Bayley, P. J., Hopkins, R. O., & Squire, L. R. (2006). The fate of old memories after medial temporal lobe damage. *The Journal of Neuroscience*, *26*, 13311-13317.

Beckner, V. E., Tucker, D. M., Delville, Y., & Mohr, D. C. (2006). Stress facilitates consolidation of verbal memory for a film but does not affect retrieval. *Behavioral Neuroscience*, *120*, 518-527.

Besnard, A. (2012). A model of hippocampal competition between new learning and memory updating. *The Journal of Neuroscience*, *32*, 3281-3283.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? a critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*, 2767-2796.

Bird, C. M., & Burgess, N. (2008). The hippocampus and memory: insights from spatial processing. *Nature Reviews Neuroscience*, *9*, 182-194.

Bolhuis, J. J., Stewart, C. A., & Forrest, E. M. (1994). Retrograde amnesia and memory reactivation in rats with ibotenate lesions to the hippocampus or subiculum. *The Quarterly Journal of Experimental Psychology*, *47*, 129-150.

Bonne, O., Vythilingam, M., Inagaki, M., Wood, S., Neumeister, A., Nugent, A. C., ... Charney, D. S. (2008). Reduced posterior hippocampal volume in posttraumatic stress disorder. *The Journal of Clinical Psychiatry*, *69*, 1087-1091.

Bonnici, H. M., Chadwick, M. J., Kumaran, D., Hassabis, D., Weiskopf, N., & Maguire, E. A. (2012). Multi-voxel pattern analysis in human hippocampal subfields. *Frontiers in Human Neuroscience*, *6*, 290.

Bonnici, H. M., Chadwick, M. J., Lutti, A., Hassabis, D., Weiskopf, N., & Maguire, E. A. (2012). Detecting representations of recent and remote autobiographical memories in vmpfc and hippocampus. *The Journal of Neuroscience*, *32*, 16982-16991.

Bonnici, H. M., Chadwick, M. J., & Maguire, E. A. (2013). Representations of recent and remote autobiographical memories in hippocampal subfields. *Hippocampus*, *23*, 849-854.

Bonnici, H. M., & Maguire, E. A. (2018). Two years later - revisiting autobiographical memory representations in vmpfc and hippocampus. *Neuropsychologia*, *110*, 159-169.

Bontempi, B., Laurent-Demir, C., Destrade, C., & Jaffard, R. (1999). Time-dependent reorganization of brain circuitry underlying long-term memory storage. *Nature*, *400*, 671-675.

Born, J., & Wilhelm, I. (2012). System consolidation of memory during sleep. *Psychological Research*, *76*, 192-203.

Bosshardt, S., Degonda, N., Schmidt, C. F., Boesiger, P., Nitsch, R. M., Hock, C., & Henke, K. (2005). One month of human memory consolidation enhances retrieval-related hippocampal activity. *Hippocampus*, *15*, 1026-1040.

Bosshardt, S., Schmidt, C. F., Jaermann, T., Degonda, N., Boesiger, P., Nitsch, R. M., ... Henke, K. (2005). Effects of memory consolidation on human hippocampal activity during retrieval. *Cortex*, *41*, 486-498.

Brashers-Krug, T., Shadmehr, R., & Bizzi, E. (1996). Consolidation in human motor memory. *Nature*, *382*, 252-255.

Brown, M. W., & Aggleton, J. P. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience*, *2*, 51-61.

Buchanan, T. W., & Lovallo, W. R. (2001). Enhanced memory for emotional material following stress-level cortisol treatment in humans. *Psychoneuroendocrinology*, *26*, 307-317.

Buchanan, T. W., Tranel, D., & Adolphs, R. (2006). Impaired memory retrieval correlates with individual differences in cortisol response but not autonomic response. *Learning and Memory*, *13*, 382-387.

Cabeza, R., & St Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in Cognitive Sciences*, *11*, 219-227.

Cadle, C. E., & Zoladz, P. R. (2015). Stress time-dependently influences the acquisition and retrieval of unrelated information by producing a memory of its own. *Frontiers in Psychology*, *6*, 910.

Cahill, L., Babinsky, R., Markowitsch, H. J., & McGaugh, J. L. (1995). The amygdala and emotional memory. *Nature*, *377*, 295-296.

Cahill, L., Gorski, L., & Le, K. (2003). Enhanced human memory consolidation with post-learning stress: interaction with the degree of arousal at encoding. *Learning and Memory*, *10*, 270-274.

Cahill, L., Haier, R. J., Fallon, J., Alkire, M. T., Tang, C., Keator, D., ... McGaugh, J. L. (1996). Amygdala activity at encoding correlated with long-term, free recall of emotional information. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 8016-8021.

Cermak, L. S., & O'Connor, M. (1983). The anterograde and retrograde retrieval ability of a patient with amnesia due to encephalitis. *Neuropsychologia*, *21*, 213-234.

Chadwick, M. J., Anjum, R. S., Kumaran, D., Schacter, D. L., Spiers, H. J., & Hassabis, D. (2016). Semantic representations in the temporal pole predict false memories. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 10180-10185.

Chen, J., Olsen, R. K., Preston, A. R., Glover, G. H., & Wagner, A. D. (2011). Associative retrieval processes in the human medial temporal lobe: hippocampal retrieval success and ca1 mismatch detection. *Learning and Memory*, *18*, 523-528.

Cipolotti, L., Shallice, T., Chan, D., Fox, N., Scahill, R., Harrison, G., ... Rudge, P.

(2001). Long-term retrograde amnesia: The crucial role of the hippocampus. *Neuropsychologia*, *39*, 151-172.

Clark, R. E., Broadbent, N. J., & Squire, L. R. (2005a). Hippocampus and remote spatial memory in rats. *Hippocampus*, *15*, 260-272.

Clark, R. E., Broadbent, N. J., & Squire, L. R. (2005b). Impaired remote spatial memory after hippocampal lesions despite extensive training beginning early in life. *Hippocampus*, *15*, 340-346.

Clark, R. E., & Squire, L. R. (1998). Classical conditioning and brain systems: A key role for awareness. *Science*, *280*, 77-81.

Cohen, N., & Squire, L. R. (1980). Preserved learning and retention of pattern analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, *210*, 207-209.

Collin, S. H., Milivojevic, B., & Doeller, C. F. (2015). Memory hierarchies map onto the hippocampal long axis in humans. *Nature Neuroscience*, *18*, 1562-1564.

Coynel, D., Marrelec, G., Perlbarg, V., Pelegrini-Issac, M., Van de Moortele, P. F., Ugurbil, K., ... Lehericy, S. (2010). Dynamics of motor-related functional integration during motor sequence learning. *NeuroImage*, *49*, 759-766.

Cullen, P. K., Gilman, T. L., Winiecki, P., Riccio, D. C., & Jasnow, A. M. (2015). Activity of the anterior cingulate cortex and ventral hippocampus underlie increases in contextual fear generalization. *Neurobiology of Learning and Memory*, *124*, 19-27.

Damasio, A. R., Eslinger, P. J., Damasio, H., & Van Hoesen, G. (1985). Multimodal amnesic syndrome following bilateral temporal and basal forebrain damage. *Archives of neurology*, *42*, 252-259.

Dandolo, L. C., & Schwabe, L. (2018). Time-dependent memory transformation along the hippocampal anterior-posterior axis. *Nature Communications*, *9*, 1205.

Dandolo, L. C., & Schwabe, L. (2019). Time-dependent motor memory representations in prefrontal cortex. *NeuroImage*, *197*, 143-155.

Daselaar, S. M., Fleck, M. S., & Cabeza, R. (2006). Triple dissociation in the medial temporal lobes: recollection, familiarity, and novelty. *Journal of Neurophysiology*, *96*, 1902-1911.

Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, *16*, 693-700.

Dayan, E., & Cohen, L. G. (2011). Neuroplasticity subserving motor skill learning. *Neuron*, *72*, 443-454.

Debas, K., Carrier, J., Barakat, M., Marrelec, G., Bellec, P., Hadj Tahar, A., ... Doyon, J. (2014). Off-line consolidation of motor sequence learning results in greater integration within a cortico-striatal functional network. *NeuroImage*, *99*, 50-58.

Debas, K., Carrier, J., Orban, P., Barakat, M., Lungu, O., Vandewalle, G., ... Doyon, J. (2010). Brain plasticity related to the consolidation of motor sequence learning and motor adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 17839-17844.

Denny, C. A., Kheirbek, M. A., Alba, E. L., Tanaka, K. F., Brachman, R. A., Laughman, K. B., ... Hen, R. (2014). Hippocampal memory traces are differentially modulated by experience, time, and adult neurogenesis. *Neuron*, *83*, 189-201.

Deuker, L., Doeller, C. F., Fell, J., & Axmacher, N. (2014). Human neuroimaging studies on the hippocampal ca3 region - integrating evidence for pattern separation and completion. *Frontiers in Cellular Neuroscience*, *8*, 64.

Diana, R., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends in Cognitive Sciences*, *11*, 379-386.

Diana, R., Yonelinas, A. P., & Ranganath, C. (2010). Medial temporal lobe activity during source retrieval reflects information type, not memory strength. *Journal of Cognitive Neuroscience*, *22*, 1808-1818.

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, *11*, 114-126.

Diekelmann, S., Wilhelm, I., & Born, J. (2009). The whats and whens of sleep-dependent memory consolidation. *Sleep Medicine Reviews*, *13*, 309-321.

Dolcos, F., LaBar, K. S., & Cabeza, R. (2004). Interaction between the amygdala and the medial temporal lobe memory system predicts better memory for emotional events. *Neuron*, *42*, 855-863.

Doyon, J., Bellec, P., Amsel, R., Penhune, V., Monchi, O., Carrier, J., ... Benali, H. (2009). Contributions of the basal ganglia and functionally related brain structures to motor learning. *Behavioural Brain Research*, *199*, 61-75.

Doyon, J., Penhune, V., & Ungerleider, L. G. (2003). Distinct contribution of the cortico-striatal and cortico-cerebellar systems to motor skill learning. *Neuropsychologia*, *41*, 252-262.

Dudai, Y. (2002). *Memory from a to z. keywords, concepts and beyond.* Oxford, UK: Oxford University Press.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram. *Annual Review of Psychology*, *55*, 51-86.

Dudai, Y. (2012). The restless engram: consolidations never end. *Annual Review of Neuroscience*, *35*, 227-247.

Dudai, Y., Karni, A., & Born, J. (2015). The consolidation and transformation of memory. *Neuron*, *88*, 20-32.

Duncan, C. (1949). The retroactive effect of electroshock on learning. *Journal of Comparative and Physiological Psychology*, *42*, 32-44.

Duncan, K., Ketz, N., Inati, S. J., & Davachi, L. (2012). Evidence for area ca1 as a match/mismatch detector: a high-resolution fmri study of the human hippocampus. *Hippocampus*, *22*, 389-398.

Duvernoy, H. (2005). *The human hippocampus: Functional anatomy, vascularization, and serial sections with mri.* New York, NY, USA: Springer.

D'Ardenne, K., Eshel, N., Luka, J., Lenartowicz, A., Nystrom, L. E., & Cohen, J. D. (2012). Role of prefrontal cortex and the midbrain dopamine system in working memory updating. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 19900-19909.

Eichenbaum, H., Otto, T., & Cohen, N. J. (1994). Two functional components of the hippocampal memory system. *Behavioral and Brain Sciences*, *17*, 449-517.

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, *30*, 123-152.

Einarsson, E. O., Pors, J., & Nader, K. (2014). Systems reconsolidation reveals a selective role for the anterior cingulate cortex in generalized contextual fear memory expression. *Neuropsychopharmacology*, *40*, 480-487.

Evensmoen, H. R., Ladstein, J., Hansen, T. I., Moller, J. A., Witter, M. P., Nadel, L., & Haberg, A. K. (2015). From details to large scale: the representation of environmental positions follows a granularity gradient along the human hippocampal and entorhinal anterior-posterior axis. *Hippocampus*, *25*, 119-135.

Evensmoen, H. R., Lehn, H., Xu, J., Witter, M. P., Nadel, L., & Haberg, A. K. (2013). The anterior hippocampus supports a coarse, global environmental representation and the posterior hippocampus supports fine-grained, local environmental representations. *Journal of Cognitive Neuroscience*, *25*, 1908-1925.

Exton-McGuinness, M. T. J., Lee, J. L. C., & Reichelt, A. C. (2015). Updating memories–the role of prediction errors in memory reconsolidation. *Behavioural Brain Research*,

278, 375-384.

Exton-McGuinness, M. T. J., Patton, R. C., Sacco, L. B., & Lee, J. L. C. (2014). Reconsolidation of a well-learned instrumental memory. *Learning and Memory*, *21*, 468-477.

Fanselow, M. S., & Dong, H. W. (2010). Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron*, *65*, 7-19.

Fellows, L. (2016). The neuroscience of human decision-making through the lens of learning and memory. In R. Clark & S. Martin (Eds.), *Behavioral neuroscience of learning and memory. current topics in behavioral neurosciences, vol 37.* Cham: Springer.

Fink, G. R., Markowitsch, H. J., Reinkemeier, M., Bruckbauer, T., Kessler, J., & Heiss, W. D. (1996). Cerebral representation of one's own past: neural networks involved in autobiographical memory. *The Journal of Neuroscience*, *16*, 4275-4282.

Fischer, S., & Born, J. (2009). Anticipated reward enhances offline learning during sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1586-1593.

Fischer, S., Drosopoulos, S., Tsen, J., & Born, J. (2006). Implicit learning-explicit knowing: a role for sleep in memory system interaction. *Journal of Cognitive Neuroscience*, *18*, 311-319.

Foerde, K., Knowlton, B., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 11778-11783.

Fogel, S., Albouy, G., King, B. R., Lungu, O., Vien, C., Bore, A., ... Doyon, J. (2017). Reactivation or transformation? motor memory consolidation associated with cerebral activation time-locked to sleep spindles. *PLoS One*, *12*, e0174755.

Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience*, *6*, 119-130.

Frankland, P. W., & Bontempi, B. (2006). Fast track to the medial prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 509-510.

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

Furman, O., Mendelsohn, A., & Dudai, Y. (2012). The episodic engram transformed: Time reduces retrieval-related brain activity but correlates it with memory accuracy.

Learning and Memory, *19*, 575-587.

Gabitov, E., Manor, D., & Karni, A. (2014). Done that: short-term repetition related modulations of motor cortex activity as a stable signature for overnight motor memory consolidation. *Journal of Cognitive Neuroscience*, *26*, 2716-2734.

Gais, S., Albouy, G., Boly, M., Dang-Vu, T. T., Darsaud, A., Desseilles, M., ... Peigneux, P. (2007). Sleep transforms the cerebral trace of declarative memories. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 18778-18783.

Gerard, R. (1949). Physiology and psychiatry. *The American Journal of Psychiatry*, *106*, 161-173.

Gershman, S. J., Radulescu, A., Norman, K. A., & Niv, Y. (2014). Statistical computations underlying the dynamics of memory updating. *PLOS Computational Biology*, *10*, e1003939.

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? a historical perspective on current neuroscience literature. *Neuropsychologia*, *53*, 104-114.

Gilboa, A., & Marlatte, H. (2017). Neurobiology of schemas and schema-mediated memory. *Trends in Cognitive Sciences*, *21*, 618-631.

Gilboa, A., Winocur, G., Grady, C. L., Hevenor, S. J., & Moscovitch, M. (2004). Remembering our past: Functional neuroanatomy of recollection of recent and very remote personal events. *Cerebral Cortex*, *14*, 1214-1225.

Glick, M. (2017). Believing is seeing: Confirmation bias. *Journal of the American Dental Association*, *148*, 131-132.

Goshen, I., Brodsky, M., Prakash, R., Wallace, J., Gradinaru, V., Ramakrishnan, C., & K., D. (2011). Dynamics of retrieval strategies for remote memories. *Cell*, *147*, 678-689.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *13*, 45-53.

Grant, L. D., & Jarrard, L. E. (1968). Functional dissociation within hippocampus. *Brain Research*, *10*, 392-401.

Greicius, M., Krasnow, B., Boyett-Anderson, J. M., Eliez, S., Schatzberg, A. F., Reiss, A. L., & Menon, V. (2003). Regional analysis of hippocampal activation during memory encoding and retrieval: fmri study. *Hippocampus*, *13*, 164-174.

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences*, *5*, 394-400.

Harand, C., Bertran, F., La Joie, R., Landeau, B., Mezenge, F., Desgranges, B., ... Rauchs, G. (2012). The hippocampus remains activated over the long term for the retrieval of truly episodic memories. *PLoS One*, *7*, e43495.

Hardt, O., & Nadel, L. (2018). Systems consolidation revisited, but not revised: The promise and limits of optogenetics in the study of memory. *Neuroscience Letters*, *680*, 54-59.

Hardwick, R. M., Rottschy, C., Miall, R. C., & Eickhoff, S. B. (2013). A quantitative meta-analysis and review of motor learning in the human brain. *NeuroImage*, *67*, 283-297.

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

Hassabis, D., Kumaran, D., & Maguire, E. A. (2007). Using imagination to understand the neural basis of episodic memory. *The Journal of Neuroscience*, *27*, 14365-14374.

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 1726-1731.

Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences*, *11*, 299-306.

Hassabis, D., & Maguire, E. A. (2009). The construction system of the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 1263-1271.

Haubrich, J., & Nader, K. (2016). Memory reconsolidation. *Current Topics in Behavioral Neurosciences*.

Hebb, D. (1949). *The organization of behaviour*. New York, NY, USA: Wiley-Interscience.

Holland, A. C., & Kensinger, E. A. (2010). Emotion and autobiographical memory. *Physics of Life Reviews*, *7*, 88-131.

Hughes, K. R. (1965). Dorsal and ventral hippocampal lesions and maze learning: Influence of preoperative environment. *Canadian Journal of Psychology*, *19*, 325-332.

James, W. (1890). *Principles of psychology*. New York, NY, USA: Holt.

Javadi, A.-H., Emo, B., Howard, L. R., Zisch, F. E., Yu, Y., Knight, R., ... Spiers, H. J. (2017). Hippocampal and prefrontal processing of network topology to simulate the future. *Nature Communications*, *8*, 14652.

Joels, M., & Baram, T. Z. (2009). The neuro-symphony of stress. *Nature Reviews Neuroscience*, *10*, 459-466.

Julius, M. S., & Adi-Japha, E. (2015). Learning of a simple grapho-motor task by young children and adults: similar acquisition but age-dependent retention. *Frontiers in Psychology*, *6*, 225.

Jung, M. W., Wiener, S. I., & McNaughton, B. L. (1994). Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *The Journal of Neuroscience*, *14*, 7347-7356.

Kahn, I., Andrews-Hanna, J. R., Vincent, J. L., Snyder, A. Z., & Buckner, R. L. (2008). Distinct cortical anatomy linked to subregions of the medial temporal lobe revealed by intrinsic functional connectivity. *Journal of Neurophysiology*, *100*, 129-139.

Kandel, E. (2001). The molecular biology of memory storage: a dialogue between genes and synapses. *Science*, *294*, 1030-1038.

Kandel, E., Dudai, Y., & Mayford, M. (2014). The molecular and systems biology of memory. *Cell*, *157*, 163-186.

Karlsson, M. P., & Frank, L. M. (2009). Awake replay of remote experiences in the hippocampus. *Nature Neuroscience*, *12*, 913-918.

Karni, A., Meyer, G., Jezzard, P., Adams, M. M., Turner, R., & Ungerleider, L. G. (1995). Functional mri evidence for adult motor cortex plasticity during motor skill learning. *Nature*, *377*, 155-158.

Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M. M., Turner, R., & Ungerleider, L. G. (1998). The acquisition of skilled motor performance: fast and slow experience-driven changes in primary motor cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 861-868.

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature*, *365*, 250-252.

Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J., & Sagi, D. (1994). Dependence on rem sleep of overnight improvement of a perceptual skill. *Science*, *265*, 679-682.

Keinath, A. T., Wang, M. E., Wann, E. G., Yuan, R. K., Dudman, J. T., & Muzzio, I. A. (2014). Precise spatial coding is preserved along the longitudinal hippocampal axis. *Hippocampus*, *24*, 1533-1548.

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2007). Effects of emotion on memory specificity: Memory trade-offs elicited by negative visually arousing stimuli. *Journal of Memory and Language*, *56*, 575-591.

Kensinger, E. A., Piguet, O., Krendl, A. C., & Corkin, S. (2005). Memory for contextual details: Effects of emotion and aging. *Psychology and Aging*, *20*, 241–250.

Kim, J. J., Clark, R. E., & Thompson, R. F. (1995). Hippocampectomy impairs the memory of recently, but not remotely, acquired trace eyeblink conditioned responses. *Behavioral Neuroscience*, *109*, 195-203.

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

Kirwan, C. B., Bayley, P. J., Galvan, V. V., & Squire, L. R. (2008). Detailed recollection of remote autobiographical memory after damage to the medial temporal lobe. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 2676-2680.

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

Kjelstrup, K. B., Solstad, T., Brun, V. H., Hafting, T., Leutgeb, S., Witter, M. P., ... Moser, M. B. (2008). Finite scale of spatial representation in the hippocampus. *Science*, *321*, 140-143.

Kluen, L. M., Dandolo, L. C., Jocham, G., & Schwabe, L. (2018). Dorsolateral prefrontal cortex enables updating of established memories. *Cerebral Cortex*, bhy298.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*, 1399-1402.

Kohler, S., Crane, J., & Milner, B. (2002). Differential contributions of the parahippocampal place area and the anterior hippocampus to human memory for scenes. *Hippocampus*, *12*, 718-723.

Korman, M., Doyon, J., Doljansky, J., Carrier, J., Dagan, Y., & Karni, A. (2007). Daytime sleep condenses the time course of motor memory consolidation. *Nature Neuroscience*, *10*, 1206-1213.

Korman, M., Raz, N., Flash, T., & Karni, A. (2003). Multiple shifts in the representation of a motor sequence during the acquisition of skilled performance. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 12492-12497.

Krakauer, J. W., & Shadmehr, R. (2006). Consolidation of motor memory. *Trends in Neurosciences*, *29*, 58-64.

Kroes, M. C., & Fernandez, G. (2012). Dynamic neural systems enable adaptive, flexible memories. *Neuroscience and Biobehavioral Reviews*, *36*, 1646-1666.

Kuhl, B. A., Bainbridge, W. A., & Chun, M. M. (2012). Neural reactivation reveals mechanisms for updating memory. *The Journal of Neuroscience*, *32*, 3453-3461.

Kuhlmann, S., Piel, M., & Wolf, O. T. (2005). Impaired memory retrieval after psychosocial stress in healthy young men. *The Journal of Neuroscience*, *25*, 2977-2982.

Kupferschmidt, D. A., Juczewski, K., Cui, G., Johnson, K. A., & Lovinger, D. M. (2017). Parallel, but dissociable, processing in discrete corticostriatal inputs encodes skill learning. *Neuron*, *96*, 476-489.e5.

Kwapis, J. L., Jarome, T. J., Ferrara, N. C., & Helmstetter, F. J. (2017). Updating procedures can reorganize the neural circuit supporting a fear memory. *Neuropsychopharmacology*, *42*, 1688-1697.

LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, *7*, 54-64.

Lacy, J. W., Yassa, M. A., Stark, S. M., Muftuler, L. T., & Stark, C. E. (2011). Distinct pattern separation related transfer functions in human ca3/dentate and ca1 revealed using high-resolution fmri and variable mnemonic similarity. *Learning and Memory*, *18*, 15-18.

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*, 42-55.

Lambon Ralph, M. A., & Patterson, K. (2008). Generalisation and differentiation in semantic memory: insights from semantic dementia. *Annals of the New York Academy of Sciences*, *1124*, 61-76.

Lashley, K. (1929). *Brain mechanisms and intelligence: A quantitative study of injuries to the brain.* Chicago, IL, USA: Chicago Univ. Press.

Laubach, M., Amarante, L. M., Swanson, K., & White, S. R. (2018). What, if anything, is rodent prefrontal cortex? *eNeuro*, *5*, e0315-18.

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

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

Lehericy, S., Benali, H., Van de Moortele, P. F., Pelegrini-Issac, M., Waechter, T., Ugurbil, K., & Doyon, J. (2005). Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 12566-12571.

Lehmann, H., Sparks, F. T., Spanswick, S. C., Hadikin, C., McDonald, R. J., & Sutherland, R. J. (2009). Making context memories independent of the hippocampus. *Learning*

and Memory, *16*, 417-420.

Lepage, M., Habib, R., & Tulving, E. (1998). Hippocampal pet activations of memory encoding and retrieval: the hiper model. *Hippocampus*, *8*, 313-322.

Leutgeb, J. K., Leutgeb, S., Moser, M. B., & Moser, E. I. (2007). Pattern separation in the dentate gyrus and ca3 of the hippocampus. *Learning and Memory*, *14*, 745-757.

Leutgeb, S., Leutgeb, J. K., Treves, A., Moser, M. B., & Moser, E. I. (2004). Distinct ensemble codes in hippocampal areas ca3 and ca1. *Science*, *305*, 1295-1298.

Lewis, D. J. (1979). Psychobiology of active and inactive memory. *Psychological Bulletin*, *86*, 1054-1083.

Libby, L. A., Ekstrom, A. D., Ragland, J. D., & Ranganath, C. (2012). Differential connectivity of perirhinal and parahippocampal cortices within human hippocampal subregions revealed by high-resolution functional imaging. *The Journal of Neuroscience*, *32*, 6550-6560.

Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, *25*, 720-725.

Maguire, E. A., & Frith, C. D. (2003). Lateral asymmetry in the hippocampal response to the remoteness of autobiographical memories. *The Journal of Neuroscience*, *23*, 5302-5307.

Maguire, E. A., Henson, R. N., Nummerg, C. J., & Frith, C. D. (2001). Activity in prefrontal cortex, not hippocampus, varies parametrically with the increasing remoteness of memory. *Neuroreport*, *12*, 441-444.

Maguire, E. A., & Mullally, S. L. (2013). The hippocampus: a manifesto for change. *Journal of Experimental Psychology: General*, *142*, 1180-1189.

Maguire, E. A., & Mummery, C. J. (1999). Differential modulation of a common memory retrieval network revealed by positron emission tomography. *Hippocampus*, *9*, 54-61.

Maguire, E. A., Nannery, R., & Spiers, H. (2006). Navigation around london by a taxi driver with bilateral hippocampal lesions. *Brain*, *129*, 2894-2907.

Malykhin, N. V., Lebel, R. M., Coupland, N. J., Wilman, A. H., & Carter, R. (2010). In vivo quantification of hippocampal subfields using 4.7 t fast spin echo imaging. *NeuroImage*, *49*, 1224-1230.

Maquet, P., Laureys, S., Peigneux, P., Fuchs, S., Petiau, C., Phillips, C., ... Cleeremans, A. (2000). Experience-dependent changes in cerebral activation during human rem sleep. *Nature Neuroscience*, *3*, 831-836.

Markowitsch, H. J., Calabrese, P., Würker, M., Durwen, H. F., Kessler, J., Babinsky, R.,

... Gehlen, W. (1994). The amygdala's contribution to memory–a study on two patients with urbach-wiethe disease. *Neuroreport*, *5*, 1349-1352.

Marr, D. (1971). Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *262*, 23-81.

Martin, S. J., de Hoz, L., & Morris, R. G. (2005). Retrograde amnesia: Neither partial nor complete hippocampal lesions in rats result in preferential sparing of remote spatial memory, even after reminding. *Neuropsychologia*, *43*, 609-624.

Maurer, A. P., Vanrhoads, S. R., Sutherland, G. R., Lipa, P., & McNaughton, B. L. (2005). Self-motion and the origin of differential spatial scaling along the septo-temporal axis of the hippocampus. *Hippocampus*, *15*, 841-852.

Maviel, T., Durkin, T. P., Menzaghi, F., & Bontempi, B. (2004). Sites of neocortical reorganization critical for remote spatial memory. *Science*, *305*, 96-99.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419-457.

McCormick, C., St-Laurent, M., Ty, A., Valiante, T. A., & McAndrews, M. P. (2015). Functional and effective hippocampal-neocortical connectivity during construction and elaboration of autobiographical memory retrieval. *Cerebral Cortex*, *25*, 1297-305.

McCullough, A. M., & Yonelinas, A. P. (2013). Cold-pressor stress after learning enhances familiarity-based recognition memory in men. *Neurobiology of Learning and Memory*, *106*, 11-17.

McDermott, K. B., & Roediger, H. L. (2019). Memory(encoding, storage, retrieval). In R. Biswas-Diener & E. Diener (Eds.), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers.

McDermott, K. B., Szpunar, K. K., & Christ, S. E. (2009). Laboratory-based and autobiographical retrieval tasks differ substantially in their neural substrates. *Neuropsychologia*, *47*, 2290-2298.

McDonald, R. J., & White, N. (1994). Parallel information processing in the water maze: Evidence for independent memory systems involving dorsal striatum and hippocampus. *Behavioral and Neural Biology*, *61*, 260-270.

McDougall, W. (1923). *Outline of psychology.* New York, NY, USA: Scribners.

McGaugh, J. L. (2000). Memory–a century of consolidation. *Science*, *287*, 248-251.

McIntyre, C. K., Gold, P. E., & Marriott, L. K. (2003). Cooperation between memory systems: Acetylcholine release in the amygdala correlates positively with performance on a hippocampus-dependent task. *Behavioral Neuroscience*, *117*, 320-326.

McNaughton, B., & Morris, R. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends in Neurosciences*, *10*, 408-415.

Miller, R. R., & Springer, A. D. (1972). Induced recovery of memory in rats following electroconvulsive shock. *Physiology and Behavior*, *8*, 645-651.

Milner, B. (1962). Les troubles de la memoire accompagnant des lésions hippocampiques bilaterales. In *Physiologie de l'hippocampe* (p. 257-272). Paris, France: Centre National de la Recherche Scientifique.

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

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

Moscovitch, M., Nadel, L., Winocur, G., Gilboa, A., & Rosenbaum, R. S. (2006). The cognitive neuroscience of remote episodic, semantic and spatial memory. *Current Opinion in Neurobiology*, *16*, 179-190.

Moscovitch, M., Rosenbaum, R. S., Gilboa, A., Addis, D. R., Westmacott, R., Grady, C., ... Nadel, L. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. *Journal of Anatomy*, *207*, 35-66.

Moscovitch, M., Yaschyshyn, T., Ziegler, M., & Nadel, L. (1999). Remote episodic memory and retrograde amnesia: was endel tulving right all along? In E. Tulving (Ed.), *Memory, consciousness and the brain: the tallinn conference* (p. 331-345). New York, NY, USA: Psychology Press.

Moser, M. B., & Moser, E. I. (1998). Functional differentiation in the hippocampus. *Hippocampus*, *8*, 608-619.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., du Sert, N. P., ... Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*, 0021.

Murty, V. P., Ritchey, M., Adcock, R. A., & LaBar, K. S. (2010). fmri studies of successful

emotional memory encoding: A quantitative meta-analysis. *Neuropsychologia*, *48*, 3459-3469.

Müller, G. E., & Pilzecker, A. (1900). Experimentelle beiträge zur lehre vom gedächtnis [experimental contributions to the science of memory]. *Zeitschrift fuer Psychologie. Ergaenzungsband*, *1*, 1-300.

Nadel, L. (1968). Dorsal and ventral hippocampal lesions and behavior. *Physiology and Behavior*, *3*, 891-900.

Nadel, L. (1991). The hippocampus and space revisited. *Hippocampus*, *1*, 221-229.

Nadel, L. (1995). The role of the hippocampus in declarative memory: A comment on zola-morgan, squire, and ramus (1994). *Hippocampus*, *5*, 232-234.

Nadel, L., Hoscheidt, S., & Ryan, L. R. (2013). Spatial cognition and the hippocampus: the anterior-posterior axis. *Journal of Cognitive Neuroscience*, *25*, 22-8.

Nadel, L., Hupbach, A., Gomez, R., & Newman-Smith, K. (2012). Memory formation, consolidation and transformation. *Neuroscience and Biobehavioral Reviews*, *36*, 1640-1645.

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

Nadel, L., & Moscovitch, M. (2001). The hippocampal complex and long-term memory revisited. *Trends in Cognitive Sciences*, *5*, 228-230.

Nadel, L., Samsonovich, A., Ryan, L., & Moscovitch, M. (2000). Multiple trace theory of human memory: computational, neuroimaging, and neuropsychological results. *Hippocampus*, *10*, 352-368.

Nadel, L., Winocur, G., Ryan, L., & Moscovitch, M. (2007). Systems consolidation and hippocampus: two views. *Debates in Neuroscience*, *1*, 55-66.

Nader, K. (2015). Reconsolidation and the dynamic nature of memory. *Cold Spring Harbor Perspectives in Biology*, *7*, a021782.

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

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

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

Nickerson, R. S. (1998). Confirmation bias: a ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175-220.

Nieuwenhuis, I. L., & Takashima, A. (2011). The role of the ventromedial prefrontal cortex in memory consolidation. *Behavioural Brain Research*, *218*, 325-334.

Nieuwenhuis, I. L., Takashima, A., Oostenveld, R., McNaughton, B. L., Fernandez, G., & Jensen, O. (2012). The neocortical network representing associative memory reorganizes with time in a process engaging the anterior temporal lobe. *Cerebral Cortex*, *22*, 2622-2633.

Niki, K., & Luo, J. (2002). An fmri study on the time-limited role of the medial temporal lobe in long-term topographical autobiographic memory. *Journal of Cognitive Neuroscience*, *14*, 500-507.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1-32.

Norman, K. A. (2010). How hippocampus and cortex contribute to recognition memory: Revisiting the complementary learning systems model. *Hippocampus*, *20*, 1217-1227.

Nyberg, L., Andersson, M., Forsgren, L., Jakobsson-Mo, S., Larsson, A., Marklund, P., ... Backman, L. (2009). Striatal dopamine d2 binding is related to frontal bold response during updating of long-term memory representations. *NeuroImage*, *46*, 1194-1199.

Oei, N. Y., Elzinga, B. M., Wolf, O. T., de Ruiter, M. B., Damoiseaux, J. S., Kuijer, J. P., ... Rombouts, S. A. (2007). Glucocorticoids decrease hippocampal and prefrontal activation during declarative memory retrieval in young men. *Brain Imaging and Behavior*, *1*, 31-41.

Orban, P., Rauchs, G., Balteau, E., Degueldre, C., Luxen, A., Maquet, P., & Peigneux, P. (2006). Sleep after spatial learning promotes covert reorganization of brain activity. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 7124–7129.

O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off. *Hippocampus*, *4*, 661-682.

O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. *Trends in Cognitive Sciences*, *6*, 505-510.

O'Keefe, J. (1991). The hippocampal map and navigational strategies. In J. Paillard (Ed.), *Brain and space* (p. 173-295). Oxford, UK: Oxford University Press.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, *34*, 171-175.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford, UK: Oxford

University Press.

Packard, M., & McGaugh, J. L. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially effects expression of place and response learning. *Neurobiology of Learning and Memory*, *65*, 65-72.

Pardilla-Delgado, E., Alger, S. E., Cunningham, T. J., Kinealy, B., & Payne, J. D. (2016). Effects of post-encoding stress on performance in the drm false memory paradigm. *Learning and Memory*, *23*, 46-50.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? the representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*, 976-987.

Pavlides, C., & Winson, J. (1989). Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. *The Journal of Neuroscience*, *9*, 2907-2918.

Payne, J. D., Nadel, L., Allen, J. J., Thomas, K. G., & Jacobs, W. J. (2002). The effects of experimentally induced stress on false recognition. *Memory*, *10*, 1-6.

Pedraza, L. K., Sierra, R. O., Boos, F. Z., Haubrich, J., Quillfeldt, J. A., & Alvares, L. D. (2016). The dynamic nature of systems consolidation: Stress during learning as a switch guiding the rate of the hippocampal dependency and memory quality. *Hippocampus*, *26*, 362-371.

Peigneux, P., Laureys, S., Fuchs, S., Collette, F., Perrin, F., Reggers, J., ... Maquet, P. (2004). Are spatial memories strengthened in the human hippocampus during slow wave sleep? *Neuron*, *44*, 535-545.

Piefke, M., Weiss, P. H., Zilles, K., Markowitsch, H. J., & Fink, G. R. (2003). Differential remoteness and emotional tone modulate the neural correlates of autobiographical memory. *Brain*, *126*, 650-668.

Piolino, P., Giffard-Quillon, G., Desgranges, B., Chetelat, G., Baron, J.-C., & Eustache, F. (2004). Re-experiencing old memories via hippocampus: a pet study of autobiographical memory. *NeuroImage*, *22*, 1371-1383.

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, *414*, 546-550.

Poppenk, J., Evensmoen, H. R., Moscovitch, M., & Nadel, L. (2013). Long-axis specialization of the human hippocampus. *Trends in Cognitive Sciences*, *17*, 230-240.

Poppenk, J., & Moscovitch, M. (2011). A hippocampal marker of recollection memory

ability among healthy young adults: contributions of posterior and anterior segments. *Neuron*, *72*, 931-937.

Preuss, D., & Wolf, O. T. (2009). Post-learning psychosocial stress enhances consolidation of neutral stimuli. *Neurobiology of Learning and Memory*, *92*, 318-326.

Price, A. R., Bonner, M. F., Peelle, J. E., & Grossman, M. (2015). Converging evidence for the neuroanatomic basis of combinatorial semantics in the angular gyrus. *The Journal of Neuroscience*, *35*, 3276-3284.

Przybyslawski, J., & Sara, S. J. (1997). Reconsolidation of memory after its reactivation. *Behavioural Brain Research*, *84*, 24-26.

Ranganath, C. (2010). A unified framework for the functional organization of the medial temporal lobes and the phenomenology of episodic memory. *Hippocampus*, *20*, 1263-1290.

Rasch, B., Buchel, C., Gais, S., & Born, J. (2007). Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science*, *315*, 1426-1429.

Rekkas, P. V., & Constable, R. T. (2005). Evidence that autobiographic memory retrieval does not become independent of the hippocampus: an fmri study contrasting very recent with remote events. *Journal of Cognitive Neuroscience*, *17*, 1950-1961.

Restivo, L., Vetere, G., Bontempi, B., & Ammassari-Teule, M. (2009). The formation of recent and remote memory is associated with time-dependent formation of dendritic spines in the hippocampus and anterior cingulate cortex. *The Journal of Neuroscience*, *29*, 8206-8214.

Ribot, T. (1882). *Diseases of memory.* New York, NY, USA: Appleton-Century-Crofts.

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*, *17*, 981-986.

Ritchey, M., Montchal, M. E., Yonelinas, A. P., & Ranganath, C. (2015). Delay-dependent contributions of medial temporal lobe regions to episodic memory retrieval. *Elife*, *4*.

Robin, J., & Moscovitch, M. (2017). Details, gist and schema: hippocampal-neocortical interactions underlying recent and remote episodic and spatial memory. *Current Opinion in Behavioral Sciences*, *17*, 114-123.

Rodriguez-Ortiz, C. J., & Bermudez-Rattoni, F. (2017). Determinants to trigger memory reconsolidation: The role of retrieval and updating information. *Neurobiology of Learning and Memory*, *142*, 4-12.

Roozendaal, B. (2002). Stress and memory: opposing effects of glucocorticoids on memory consolidation and memory retrieval. *Neurobiology of Learning and Memory*, *78*, 578-595.

Rosenbaum, R. S., McKinnon, M. C., Levine, B., & Moscovitch, M. (2004). Visual imagery deficits, impaired strategic retrieval, or memory loss: Disentangling the nature of an amnesic patient's autobiographical memory deficit. *Neuropsychologia*, *42*, 1619-1635.

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*, 1044-1048.

Rosenbaum, R. S., Winocur, G., & Moscovitch, M. (2001). New views on old memories: Re-evaluating the role of the hippocampal complex. *Behavioural Brain Research*, *127*, 183-197.

Ross, R. S., & Eichenbaum, H. (2006). Dynamics of hippocampal and cortical activation during consolidation of a nonspatial memory. *The Journal of Neuroscience*, *26*, 4852-4859.

Rozin, P. (1976). The psychobiological approach to human memory. In R. Rozenzweig & E. L. Bennett (Eds.), *Neural mechanisms of learning and memory.* Cambridge, MA, USA: MIT Press.

Rudoy, J. D., Voss, J. L., Westerberg, C. E., & Paller, K. A. (2009). Strengthening individual memories by reactivating them during sleep. *Science*, *326*, 1079.

Russell, N. P., W.R.W. (1946). Traumatic amnesia. *Brain*, *69*, 280-300.

Ryan, L., Nadel, L., Keil, K., Putnam, K., Schnyer, D., Trouard, T., & Moscovitch, M. (2001). Hippocampal complex and retrieval of recent and very remote autobiographical memories: Evidence from functional magnetic resonance imaging in neurologically intact people. *Hippocampus*, *11*, 707-714.

Ryle, G. (1949). *The concept of mind.* San Francisco, CA, USA: Hutchinson.

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

Savion-Lemieux, T., & Penhune, V. B. (2005). The effects of practice and delay on motor skill learning and retention. *Experimental Brain Research*, *161*, 423-431.

Schacter, D. L. (1999). The seven sins of memory. insights from psychology and cognitive neuroscience. *American Psychologist*, *54*, 182-203.

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine

the future: the prospective brain. *Nature Reviews Neuroscience*, *8*, 657-661.

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. *Journal of Cognitive Neuroscience*, *26*, 1736-1747.

Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology*, *22*, 1622-1627.

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*.

Schapiro, A. C., Turk-Browne, N. B., Norman, K. A., & Botvinick, M. M. (2016). Statistical learning of temporal community structure in the hippocampus. *Hippocampus*, *26*, 3-8.

Schneider, A. M., & Sherman, W. (1968). Amnesia: a function of the temporal relation of footshock to electroconvulsive shock. *Science*, *159*, 219-221.

Schwabe, L. (2013). Stress and the engagement of multiple memory systems: integration of animal and human studies. *Hippocampus*, *23*, 1035-1043.

Schwabe, L., Haddad, L., & Schachinger, H. (2008). Hpa axis activation by a socially evaluated cold-pressor test. *Psychoneuroendocrinology*, *33*, 890-895.

Schwabe, L., Joels, M., Roozendaal, B., Wolf, O. T., & Oitzl, M. S. (2012). Stress effects on memory: an update and integration. *Neuroscience and Biobehavioral Reviews*, *36*, 1740-1749.

Schwabe, L., Nader, K., & Pruessner, J. C. (2014). Reconsolidation of human memory: brain mechanisms and clinical relevance. *Biological Psychiatry*, *76*, 274-280.

Schwabe, L., & Schachinger, H. (2018). Ten years of research with the socially evaluated cold pressor test: Data from the past and guidelines for the future. *Psychoneuroendocrinology*, *92*, 155-161.

Schwabe, L., Schachinger, H., de Kloet, E. R., & Oitzl, M. S. (2010). Corticosteroids operate as a switch between memory systems. *Journal of Cognitive Neuroscience*, *22*, 1362-1372.

Schwabe, L., & Wolf, O. T. (2013). Stress and multiple memory systems: from 'thinking' to 'doing'. *Trends in Cognitive Sciences*, *17*, 60-68.

Schwabe, L., Wolf, O. T., & Oitzl, M. S. (2010). Memory formation under stress: quantity

and quality. *Neuroscience and Biobehavioral Reviews*, *34*, 584-591.

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

Seamans, J. K., Lapish, C. C., & Durstewitz, D. (2008). Comparing the prefrontal cortex of rats and primates: Insights from electrophysiology. *Neurotoxicity Research*, *14*, 249-262.

Sekeres, M. J., Moscovitch, M., & Winocur, G. (2017). Mechanisms of memory consolidation and transformation. In N. Axmacher & B. Rasch (Eds.), *Cognitive neuroscience of memory consolidation* (p. 17-44). Cham: Springer International Publishing.

Sekeres, M. J., Winocur, G., & Moscovitch, M. (2018). The hippocampus and related neocortical structures in memory transformation. *Neuroscience Letters*, *680*, 39-53.

Sekeres, M. J., Winocur, G., Moscovitch, M., Anderson, J. A. E., Pishadian, S., Wojtowicz, J. M., ... Grady, C. L. (2018). Changes in patterns of neural activity underlie a time-dependent transformation of memory in rats and humans. *Hippocampus*, *28*, 745-764.

Sevenster, D., Beckers, T., & Kindt, M. (2013). Prediction error governs pharmacologically induced amnesia for learned fear. *Science*, *339*, 830-833.

Shadmehr, R., & Holcomb, H. H. (1997). Neural correlates of motor memory consolidation. *Science*, *277*, 821-825.

Sharot, T., & Phelps, E. A. (2004). How arousal modulates memory: Disentangling the effects of attention and retention. *Cognitive, Affective, and Behavioral Neuroscience*, *4*, 294-306.

Shrager, Y., Levy, D. A., Hopkins, R. O., & Squire, L. R. (2008). Working memory and the organization of brain systems. *The Journal of Neuroscience*, *28*, 4818-4822.

Smeets, T., Otgaar, H., Candel, I., & Wolf, O. T. (2008). True or false? memory is differentially affected by stress-induced cortisol elevations and sympathetic activity at consolidation and retrieval. *Psychoneuroendocrinology*, *33*, 1378-1386.

Sommer, T. (2017). The emergence of knowledge and how it supports the memory for novel related information. *Cerebral Cortex*, *27*, 1906-1921.

Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*, 195-231.

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of Learning and Memory*, *82*, 171-177.

Squire, L. R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a

neurobiological perspective. *Current Opinion in Neurobiology*, *5*, 169-177.

Squire, L. R., & Bayley, P. J. (2007). The neuroscience of remote memory. *Current Opinion in Neurobiology*, *17*, 185-196.

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

Squire, L. R., & Dede, A. J. (2015). Conscious and unconscious memory systems. *Cold Spring Harbor Perspectives in Biology*, *7*, a021667.

Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory consolidation. *Cold Spring Harbor Perspectives in Biology*, *7*, a021766.

Squire, L. R., & Knowlton, B. J. (1995). Memory, hippocampus, and brain systems. In M. S. Gazzinga (Ed.), *The cognitive neurosciences* (p. 825-837). Cambridge, MA, USA: MIT Press.

Squire, L. R., & Wixted, J. T. (2011). The cognitive neuroscience of human memory since h.m. *Annual Review of Neuroscience*, *34*, 259-288.

Squire, L. R., Wixted, J. T., & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: A new perspective. *Nature Reviews Neuroscience*, *8*, 872-883.

Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 13515-13522.

Squire, L. R., & Zola-Morgan, S. (1988). Memory: Brain systems and behavior. *Trends in Neurosciences*, *11*, 170-175.

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*, 1380-1386.

Staresina, B. P., & Davachi, L. (2006). Differential encoding mechanisms for subsequent associative recognition and free recall. *The Journal of Neuroscience*, *26*, 9162-9172.

Staresina, B. P., Duncan, K. D., & Davachi, L. (2011). Perirhinal and parahippocampal cortices differentially contribute to later recollection of object- and scene-related event details. *The Journal of Neuroscience*, *31*, 8739-8747.

Steinvorth, S., Corkin, S., & Halgren, E. (2006). Ecphory of autobiographical memories: an fmri study of recent and remote memory retrieval. *NeuroImage*, *30*, 285-298.

Steinvorth, S., Levine, B., & Corkin, S. (2005). Medial temporal lobe structures are needed to re-experience remote autobiographical memories: Evidence from h.m. & w.r. *Neuropsychologia*, *43*, 479-496.

Sterpenich, V., Albouy, G., Darsaud, A., Schmidt, C., Vandewalle, G., Dang Vu, T. T., ... Maquet, P. (2009). Sleep promotes the neural reorganization of remote emotional memory. *Journal of Neuroscience*, *29*, 5143-5152.

St Jacques, P. L., Olm, C., & Schacter, D. L. (2013). Neural mechanisms of reactivation-induced updating that enhance and distort memory. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 19671-19678.

Strange, B. A., Fletcher, P. C., Henson, R. N., Friston, K. J., & Dolan, R. J. (1999). Segregating the functions of human hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 4034-4039.

Strange, B. A., Witter, M. P., Lein, E. S., & Moser, E. I. (2014). Functional organization of the hippocampal longitudinal axis. *Nature Reviews Neuroscience*, *15*, 655-669.

Sutherland, R. J., & Lehmann, H. (2011). Alternative conceptions of memory consolidation and the role of the hippocampus at the systems level in rodents. *Current Opinion in Neurobiology*, *21*, 446-451.

Sutherland, R. J., Sparks, F. T., & Lehmann, H. (2010). Hippocampus and retrograde amnesia in the rat model: a modest proposal for the situation of systems consolidation. *Neuropsychologia*, *48*, 2357-2369.

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

Svoboda, E., & Levine, B. (2009). The effects of rehearsal on the functional neuroanatomy of episodic autobiographical and semantic remembering: a functional magnetic resonance imaging study. *The Journal of Neuroscience*, *29*, 3073-3082.

Svoboda, E., McKinnon, M. C., & Levine, B. (2006). The functional neuroanatomy of autobiographical memory: a meta-analysis. *Neuropsychologia*, *44*, 2189-2208.

Takashima, A., Nieuwenhuis, I. L., Jensen, O., Talamini, L. M., Rijpkema, M., & Fernandez, G. (2009). Shift from hippocampal to neocortical centered retrieval network with consolidation. *The Journal of Neuroscience*, *29*, 10087-10093.

Takashima, A., Petersson, K. M., Rutters, F., Tendolkar, I., Jensen, O., Zwarts, M. J., ... Fernandez, G. (2006). Declarative memory consolidation in humans: a prospective functional magnetic resonance imaging study. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 756-761.

Takehara, K., Kawahara, S., & Kirino, Y. (2003). Time-dependent reorganization of the brain components underlying memory retention in trace eyeblink conditioning. *The*

Journal of Neuroscience, *23*, 9897-9905.

Takehara-Nishiuchi, K., & McNaughton, B. L. (2008). Spontaneous changes of neocortical code for associative memory during consolidation. *Science*, *322*, 960-963.

Takehara-Nishiuchi, K., Nakao, K., Kawahara, S., Matsuki, N., & Kirino, Y. (2006). Systems consolidation requires postlearning activation of nmda receptors in the medial prefrontal cortex in trace eyeblink conditioning. *The Journal of Neuroscience*, *26*, 5049-5058.

Talmi, D. (2013). Enhanced emotional memory: Cognitive and neural mechanisms. *Current Directions in Psychological Science*, *22*, 430-436.

Teyler, T. J., & DiScenna, P. (1985). The role of hippocampus in memory: a hypothesis. *Neuroscience and Biobehavioral Reviews*, *9*, 377-389.

Teyler, T. J., & DiScenna, P. (1986). The hippocampal memory indexing theory. *Behavioral Neuroscience*, *100*, 147-154.

Tompary, A., & Davachi, L. (2017). Consolidation promotes the emergence of representational overlap in the hippocampus and medial prefrontal cortex. *Neuron*, *96*, 228-241 e5.

Trammell, J. P., & Clore, G. L. (2014). Does stress enhance or impair memory consolidation? *Cognition and Emotion*, *28*, 361-374.

Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, *4*, 374-391.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., ... Morris, R. G. (2007). Schemas and memory consolidation. *Science*, *316*, 76-82.

Tse, D., Takeuchi, T., Kakeyama, T., Kajii, Y., Okuno, H., Tohyama, C., ... Morris, R. G. (2011). Schema-dependent gene activation and memory encoding in neocortex. *Science*, *333*, 891-895.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (p. 381-403). New York, NY, USA: Academic Press.

Tulving, E. (1983). *Elements of episodic memory.* Cambridge, MA: Oxford University Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*, 1-12.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*, 1-25.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, *247*, 301-306.

Tulving, E., Schacter, D. L., McLachland, D., & Moscovitch, M. (1988). Priming of semantic autobiographical knowledge: a case study of retrograde amnesia. *Brain and Cognition*, *8*, 3-20.

Ungerleider, L. G., Doyon, J., & Karni, A. (2002). Imaging brain plasticity during motor skill learning. *Neurobiology of Learning and Memory*, *78*, 553-564.

Vahdat, S., Fogel, S., Benali, H., & Doyon, J. (2017). Network-wide reorganization of procedural memory during nrem sleep revealed by fmri. *Elife*, *6*.

van Kesteren, M. T. R., Fernandez, G., Norris, D. G., & Hermans, E. J. (2010). Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 7550-7555.

van Kesteren, M. T. R., Ruiter, D. J., Fernandez, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*, 211-219.

Vazdarjanova, A., & Guzowski, J. F. (2004). Differences in hippocampal neuronal population responses to modifications of an environmental context: Evidence for distinct, yet complementary, functions of ca3 and ca1 ensembles. *The Journal of Neuroscience*, *24*, 6489-6496.

Viard, A., Desgranges, B., Eustache, F., & Piolino, P. (2012). Factors affecting medial temporal lobe engagement for past and future episodic events: an ale meta-analysis of neuroimaging studies. *Brain and Cognition*, *80*, 111-125.

Viard, A., Doeller, C. F., Hartley, T., Bird, C. M., & Burgess, N. (2011). Anterior hippocampus and goal-directed spatial decision making. *The Journal of Neuroscience*, *31*, 4613-4621.

Viard, A., Lebreton, K., Chételat, G., Desgranges, B., Landeau, B., Young, A., ... Piolino, P. (2010). Patterns of hippocampal- neocortical interactions in the retrieval of episodic autobiographical memories across the entire life-span of aged adults. *Hippocampus*, *20*, 153-165.

Viard, A., Piolino, P., Desgranges, B., Chételat, G., Lebreton, K., Landeau, B., ... Eustache, F. (2007). Hippocampal activation for autobiographical memories over the entire lifetime in healthy aged subjects: An fmri study. *Cerebral Cortex*, *17*, 2453-2467.

Viskontas, I. V., Carr, V. A., Engel, S. A., & Knowlton, B. J. (2009). The neural correlates of recollection: hippocampal activation declines as episodic memory fades. *Hippocampus*, *19*, 265-272.

Wagner, U., Gais, S., Haider, H., Verleger, R., & Born, J. (2004). Sleep inspires insight. *Nature*, *427*, 352-355.

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

Wang, S. H., Teixeira, C. M., Wheeler, A. L., & Frankland, P. W. (2009). The precision of remote context memories does not require the hippocampus. *Nature Neuroscience*, *12*, 253-255.

Waring, J. D., & Kensinger, E. A. (2011). How emotion leads to selective memory: Neuroimaging evidence. *Neuropsychologia*, *49*, 1831-1842.

Wheeler, A. L., Teixeira, C. M., Wang, A. H., Xiong, X., Kovacevic, N., Lerch, J. P., ... Frankland, P. W. (2013). Identification of a functional connectome for long-term fear memory in mice. *PLOS Computational Biology*, *9*, e1002853.

White, N., & McDonald, R. J. (2002). Multiple parallel memory systems in the brain of the rat. *Neurobiology of Learning and Memory*, *77*, 125-184.

Wilhelm, I., Diekelmann, S., Molzow, I., Ayoub, A., Mölle, M., & Born, J. (2011). Sleep selectively enhances memory expected to be of future relevance. *The Journal of Neuroscience*, *31*, 1563-1569.

Wilson, M., & McNaughton, B. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, *265*, 676-679.

Wiltgen, B. J., & Silva, A. J. (2007). Memory for context becomes less specific with time. *Learning and Memory*, *14*, 313-317.

Wimmer, G. E., & Shohamy, D. (2012). Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science*, *338*, 270-273.

Winocur, G. (1990). Anterograde and retrograde amnesia in rats with dorsal hippocampal or dorsomedial thalamic lesions. *Behavioural Brain Research*, *38*, 145-154.

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

Winocur, G., & Moscovitch, M. (2011). Memory transformation and systems consolidation. *Journal of the International Neuropsychological Society*, *17*, 766-780.

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

Winocur, G., Moscovitch, M., Caruana, D. A., & Binns, M. A. (2005). Retrograde amnesia in rats with lesions to the hippocampus on a test of spatial memory. *Neuropsychologia*, *43*, 1580-1590.

Winocur, G., Moscovitch, M., & Sekeres, M. J. (2007). Memory consolidation or transformation: Context manipulation and hippocampal representations of memory. *Nature Neuroscience*, *10*, 555-557.

Winocur, G., Moscovitch, M., & Sekeres, M. J. (2013). Factors affecting graded and ungraded memory loss following hippocampal lesions. *Neurobiology of Learning and Memory*, *106*, 351-364.

Wolf, O. T. (2009). Stress and memory in humans: twelve years of progress? *Brain Research*, *1293*, 142-154.

Wymbs, N. F., & Grafton, S. T. (2015). The human motor system supports sequence-specific representations over multiple training-dependent timescales. *Cerebral Cortex*, *25*, 4213-4225.

Yamashita, K., Hirose, S., Kunimatsu, A., Aoki, S., Chikazoe, J., Jimura, K., ... Konishi, S. (2009). Formation of long-term memory representation in human temporal cortex related to pictorial paired associates. *The Journal of Neuroscience*, *29*, 10335-10340.

Yassa, M. A., & Stark, C. E. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*, 515-525.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1341-1354.

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1415-1434.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441-517.

Yonelinas, A. P., Aly, M., Wang, W. C., & Koen, J. D. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus*, *20*, 1178-1194.

Yonelinas, A. P., & Ritchey, M. (2015). The slow forgetting of emotional episodic memories: an emotional binding account. *Trends in Cognitive Sciences*, *19*, 259-267.

Zeidman, P., Lutti, A., & Maguire, E. A. (2015). Investigating the functions of subregions

within anterior hippocampus. *Cortex*, *73*, 240-256.

Zeidman, P., & Maguire, E. A. (2016). Anterior hippocampus: the anatomy of perception, imagination and episodic memory. *Nature Reviews Neuroscience*, *17*, 173-182.

Zeidman, P., Mullally, S. L., & Maguire, E. A. (2015). Constructing, perceiving, and maintaining scenes: hippocampal activity and connectivity. *Cerebral Cortex*, *25*, 3836-3855.

Zola-Morgan, S., Squire, L. R., & Ramus, S. J. (1994). Severity of memory impairment in monkeys as a hnction of locus and extent of damage within the medial temporal lobe memory system. *Hippocampus*, *4*, 483-495.

Appendices

Appendix A

Study I:
Time-dependent memory transformation along the hippocampal anterior–posterior axis

Published in Nature Communications

nature COMMUNICATIONS

ARTICLE

Corrected: Author correction

DOI: 10.1038/s41467-018-03661-7 OPEN

Time-dependent memory transformation along the hippocampal anterior–posterior axis

Lisa C. Dandolo[1] & Lars Schwabe[1]

With time, memories undergo a neural reorganization that is linked to a transformation of detailed, episodic into more semantic, gist-like memory. Traditionally, this reorganization is thought to involve a redistribution of memory from the hippocampus to neocortical areas. Here we report a time-dependent reorganization within the hippocampus, along its anterior–posterior axis, that is related to the transformation of detailed memories into gist-like representations. We show that mnemonic representations in the anterior hippocampus are highly distinct and that anterior hippocampal activity is associated with detailed memory but decreases over time. Posterior hippocampal representations, however, are more gist-like at a later retention interval, and do not decline over time. These findings indicate that, in addition to the well-known systems consolidation from hippocampus to neocortex, there are changes within the hippocampus that are crucial for the temporal dynamics of memory.

[1] Department of Cognitive Psychology, University of Hamburg, 20146 Hamburg, Germany. Correspondence and requests for materials should be addressed to L.S. (email: Lars.Schwabe@uni-hamburg.de)

Memories evolve over time. After initial encoding, new information becomes fixed at a cellular level and integrated within networks of existing memories[1,2]. This integration involves a reorganization of memory during which, with time, detailed, episodic memories are transformed into more semantic, gist-like representations[1,3]. Although, the neural underpinnings of this time-dependent memory reorganization are at the heart of the neuroscience of memory, the neural evolution of memories over time remains a topic of much controversy. In particular, whether the hippocampus, a critical hub for initial memory formation[4–8], is involved in remote memories or not has been controversial for decades[3,9–12].

The hippocampus can be subdivided into anterior and posterior parts—corresponding to the ventral and dorsal hippocampus, respectively, in rodents—and these parts differ in function, structure and their connections to cortical and subcortical areas[13–15]. A prominent proposal that was largely based on rodent data linked the ventral (anterior) hippocampus to emotion, stress, and affect, whereas the dorsal (posterior) hippocampus was implicated in cognitive functions such as learning, memory, and spatial navigation[16]. Electrophysiological and lesion studies in rodents, as well as human neuroimaging studies, however, suggest that this view may need to be revised and that both anterior and posterior hippocampal areas (aHC and pHC, respectively) may contribute to learning and memory processes, although the exact functional specialization is still unclear[14,15,17]. Further studies suggest that the aHC and pHC might be differentially involved in recent and remote memories[18–20], yet whether the transformation of memory over time may be linked to time-dependent changes in aHC and pHC involvement in memory is completely unknown.

Here we determine whether there are time-dependent changes in aHC and pHC contributions to memory and if so, whether they are associated with the transformation from detailed to gist-like memory. To do so, we combined functional magnetic resonance imaging (fMRI) and multivariate representational similarity analysis (RSA) with a task probing memory transformation. Participants learned 60 pictures of scenes and objects (30 neutral,

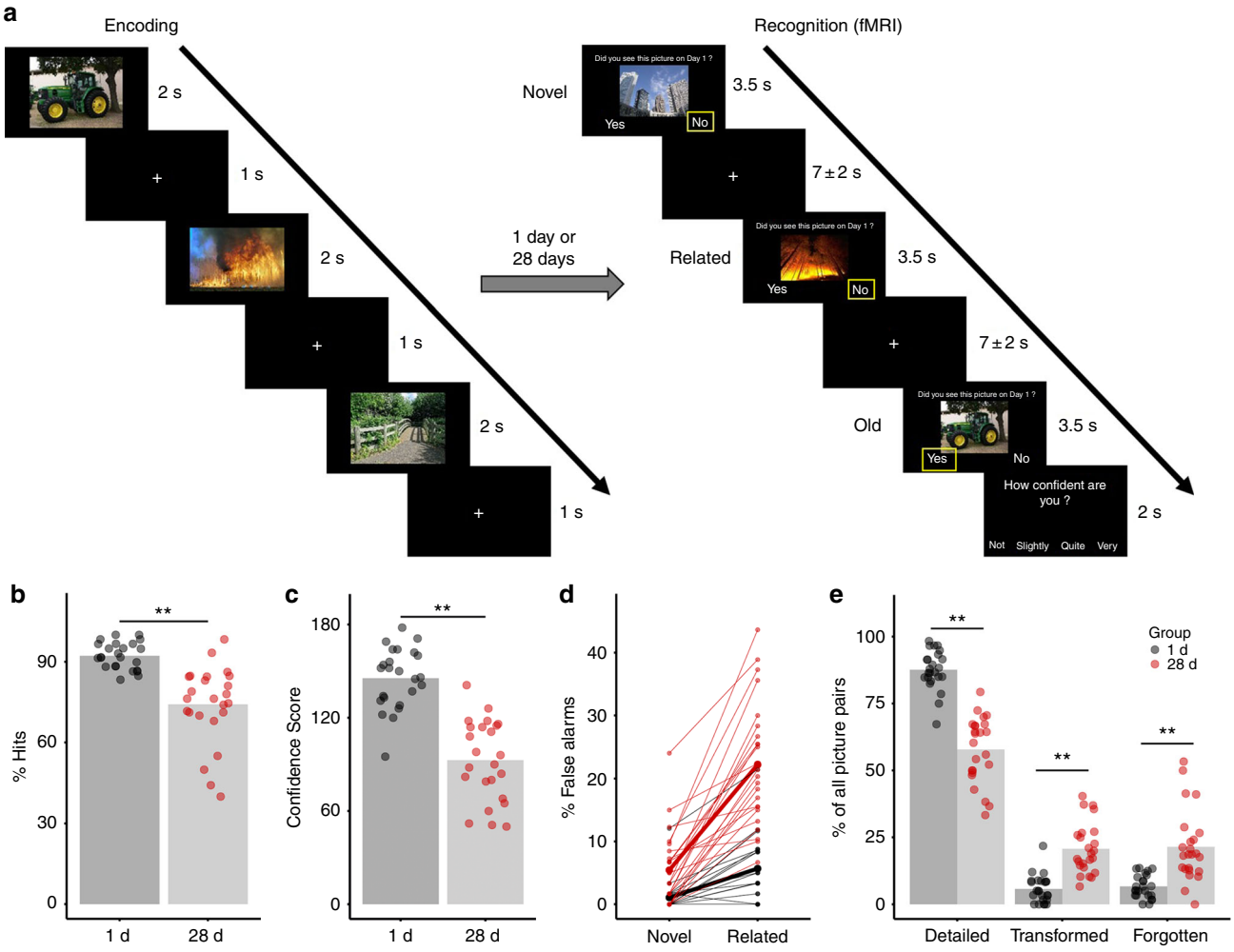
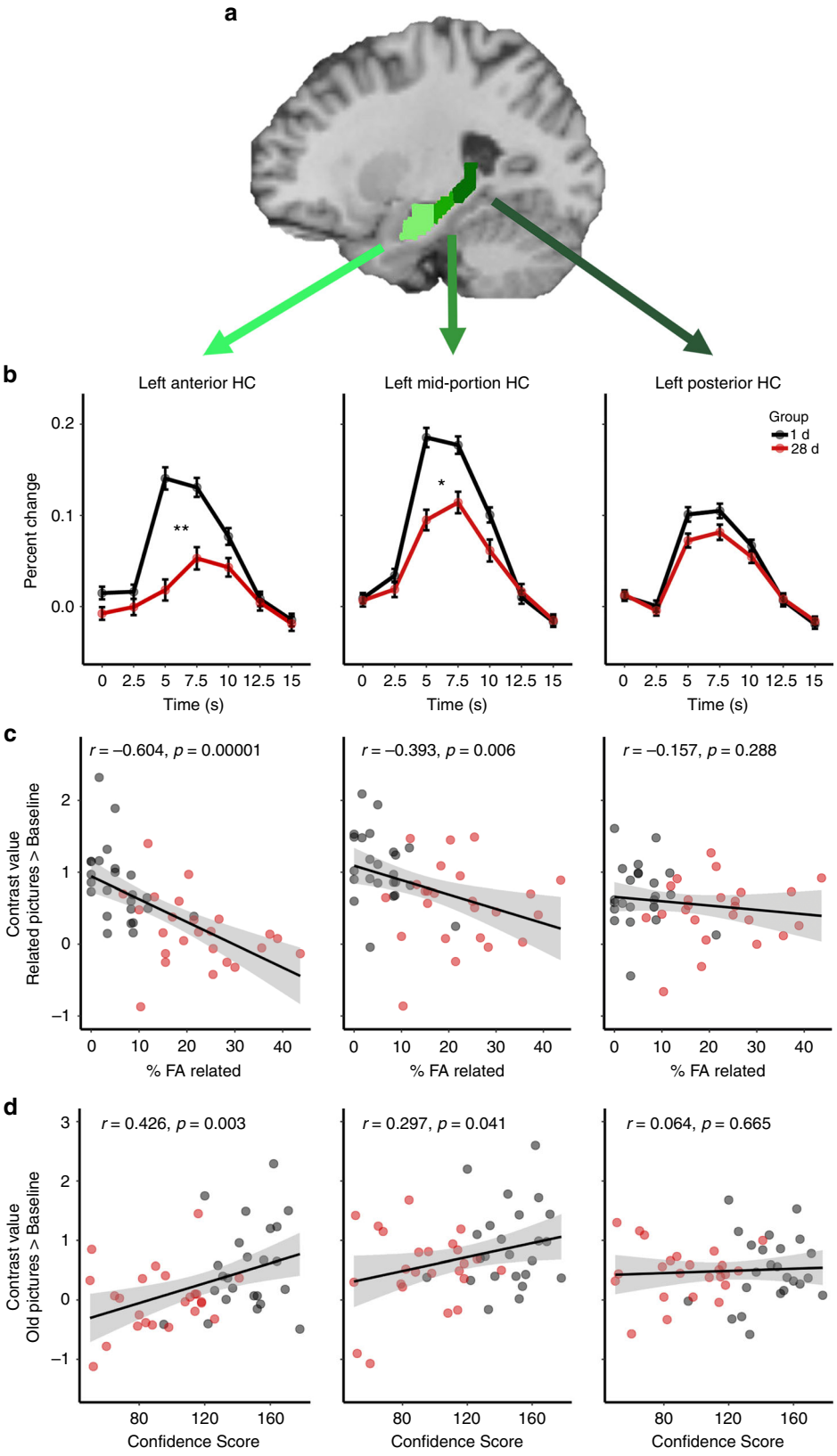


Fig. 1 Task and behavioral results. **a** Schematic overview of the picture encoding task (experimental day 1) and recognition task (experimental day 2). Images courtesy of Andreas Praefcke (tractor), Morio (skyscrapers), USDA (wildfires), and Acabashi (footbridge). **b** Percentage of Hits: participants in the 28 d group showed significantly less hits (main effect Group: $F_{(1,46)} = 33.57$, $p = 5.89e{-}07$, generalized $\eta^2 = 0.377$, $n = 48$). **c** Confidence Score: participants in the 28 d group had a significantly lower Confidence Score (main effect Group: $F_{(1,46)} = 63.33$, $p = 3.41e{-}10$, generalized $\eta^2 = 0.550$, $n = 48$). **d** Percentage of FA for related pictures and novel pictures: the increase in FA from the 1 d group to the 28 d group was more pronounced for related pictures than novel pictures (Picture Type × Group interaction: $F_{(1,46)} = 36.31$, $p = 2.65e{-}07$, generalized $\eta^2 = 0.155$, $n = 48$). **e** Percentage of picture pairs: the 28 d group compared to the 1 d group showed fewer detailed pairs (main effect Group: $F_{(1,46)} = 102.96$, $p = 2.55e{-}13$, generalized $\eta^2 = 0.607$, $n = 48$), more forgotten pairs (main effect Group: $F_{(1,46)} = 26.88$, $p = 4.72e{-}06$, generalized $\eta^2 = 0.335$, $n = 48$) and critically also more transformed pairs (main effect Group: $F_{(1,46)} = 45.15$, $p = 2.41e{-}08$, generalized $\eta^2 = 0.425$, $n = 48$). $^{**}p < 0.001$

30 negative) and performed a recognition test for these pictures in the MRI scanner either one day after encoding (1 d group) or four weeks later (28 d group). Critically, the recognition test included, in addition to the old pictures learned during encoding and completely novel pictures, related lure pictures that carried the semantic gist of the old pictures but had different details (Fig. 1a). The endorsement of related pictures as "old" provided a behavioral index of the time-dependent transformation from detailed to more gist-like memory representations.

As predicted, we found a strong increase in the endorsement of related pictures as old in the 28 d group relative to the 1 d group. This finding indicates a time-dependent memory transformation.



At the neural level, this transformation was paralleled by a time-dependent decrease in the aHC, the hippocampal subregion that was directly linked to memory specificity. The pHC, in turn, was not related to memory specificity and did not decline over time. Further, RSA revealed that activity patterns in the aHC were highly specific and differed between old and new memories in the 1 d group but not after 28 d. Representations in the pHC became more gist-like after 28 days. Together, these findings show that the aHC that supports memory specificity declines over time, whereas pHC remains stable over time but carries more gist-like representations. Our data suggest that the time-dependent transformation from detailed to gist-like memory is linked to a reorganization within the hippocampus.

Results

Time-dependent memory transformation. During encoding on experimental day 1, participants of the 1 d- and 28 d groups learned the pictures equally well (Supplementary Fig. 1). In the recognition test, either 1 day or 28 days after encoding, the hit rate was expectedly lower in the 28 d group than in the 1 d group (main effect Group: $F_{(1,46)} = 33.57$, $p = 5.89e{-}07$, generalized $\eta^2 = 0.377$, $n = 48$; Fig. 1b), yet memory was still clearly intact after 28 d as reflected by a hit rate of ~75% and a false alarm (FA) rate for novel lures of only 5%. Most importantly, however, participants in the 28 d group showed a sharp increase in the FA rate specifically for related pictures and to a significantly lesser extent for novel pictures (Picture Type × Group interaction: $F_{(1,46)} = 36.31$, $p = 2.65e{-}07$, generalized $\eta^2 = 0.155$, $n = 48$; main effect Group: $F_{(1,46)} = 45.63$, $p = 2.13e{-}08$, generalized $\eta^2 = 0.343$, $n = 48$; main effect Picture Type: $F_{(1,46)} = 113.18$, $p = 5.48e{-}14$, generalized $\eta^2 = 0.363$, $n = 48$; Fig. 1d). For related pictures, the FA rate rose to almost 25% after 28 days and was thus more than four times higher than the FA rate for novel pictures. This indicates that participants in the 28 d group particularly had difficulties differentiating between old pictures and related pictures carrying the gist of the old pictures, suggesting a transformation towards more gist-like memory. Additionally, participants were asked to rate their confidence on a 4-point-scale, whenever they indicated that they had seen the picture before (Fig. 1a), allowing us to calculate a Confidence Score as another measure of memory specificity. Participants in the 28 d group were, as expected, significantly less confident in their memory than participants in the 1 d group (main effect Group: $F_{(1,46)} = 63.33$, $p = 3.41e{-}10$, generalized $\eta^2 = 0.550$, $n = 48$; Fig. 1c).

We further analyzed the 60 matching picture pairs (i.e., old pictures learned during encoding and their respective related lures) and categorized memories for them as being either detailed, transformed or forgotten depending on whether participants endorsed solely the old pictures, both the old and related pictures, or none of them as "old". Participants in the 28 d group showed, compared to those of the 1 d group, significantly fewer detailed (main effect Group: $F_{(1,46)} = 102.96$, $p = 2.55e{-}13$, generalized $\eta^2 = 0.607$, $n = 48$) and more forgotten memories (main effect Group: $F_{(1,46)} = 26.88$, $p = 4.72e{-}06$, generalized $\eta^2 = 0.335$, $n = 48$), but critically also more transformed memories (main effect Group: $F_{(1,46)} = 45.15$, $p = 2.41e{-}08$, generalized $\eta^2 = 0.425$, $n = 48$; Fig. 1e), again in line with the proposed time-dependent memory transformation. Our behavioral data further suggest that stimulus-related emotional arousal influenced the transformation to gist-like memories: after 28 days significantly fewer negative pictures were forgotten than neutral ones (paired t-test: $t(23) = -5.00$, $p = 4.64e{-}05$, Cohen's $d = -1.02$, $n = 24$) and, even more interestingly, negative pictures were significantly more often transformed than neutral ones (paired t-test: $t(23) = 2.67$, $p = 0.0138$, Cohen's $d = 0.54$, $n = 24$; Supplementary Fig. 2), in line with findings[21,22] suggesting that superior memory for emotional material, indicated here by the slower forgetting rate, comes at the cost of reduced memory for contextual details, reflected here in an increase in transformed memories.

aHC but not pHC activity decreases over time. To elucidate the neural underpinnings of the memory dynamics over time, we first analyzed time-dependent changes in the hippocampus as a whole and other cortical and subcortical areas that have been implicated in episodic memory before. We obtained overall reduced activity in the hippocampus, parahippocampus, and the amygdala in the 28 d group compared to 1 d group (see Supplementary Fig. 3 and Supplementary Tables 1 and 2). In addition, we performed a psychophysiological interaction (PPI) analysis to test whether the cross-talk of the hippocampus with other areas critical for memory formation changed as a function of time. Our analysis showed specifically reduced functional connectivity between the right hippocampus and the right amygdala in the 28 d group compared to the 1 d group (Supplementary Fig. 4). This decrease in hippocampal-amygdala connectivity was of particular interest as the interaction of these areas is commonly linked to vivid memory[23].

As hippocampal involvement in remote memories has been argued to depend critically on memory vividness[3,20], we further explicitly looked at activity for old items that were recognized with high confidence. Even for those High Confidence Hits, overall hippocampal activity was lower in the 28 d group than in the 1 d group (Supplementary Fig. 5a). For neocortical areas involved in more semantic or schema-related memory processes, there was, however, no reliable difference in activity in the 28 d group vs. the 1 d group (when correcting for the number of ROIs; Supplementary Table 2). As we tested memory with a recognition test in which participants directly viewed all pictures, it may not

Fig. 2 Univariate analysis of the left HC long axis ROIs. **a** Depiction of the three hippocampal ROIs: aHC ($Y = -4$ to -18) = light green, mHC ($Y = -19$ to -29) = green, pHC ($Y = -30$ to -40) = dark green. Visualizations of the anatomical masks are superimposed on a sagittal section of a template image. **b** FIR time courses over the first 15 s (7TRs) for all picture types combined. Statistical comparisons were calculated for the peak response (average of the 5 s and 7.5 s time points). The activity in the 28 d group compared to the 1 d group decreased in the aHC (main effect Group: $F_{(1,46)} = 15.46$, $p = 0.0003$, generalized $\eta^2 = 0.1546$, $n = 48$) and mHC ($F_{(1,46)} = 9.24$, $p = 0.0038$, generalized $\eta^2 = 0.0999$, $n = 48$). There was no difference between the groups in the pHC ($F(1,46) = 1.70$, $p = 0.1986$, generalized $\eta^2 = 0.0230$, $n = 48$). For effects of Picture Type and Emotion see Supplementary Fig. 6. **c** Correlations of the percentage of FA to related items with the contrast value for related pictures vs baseline: across groups, there was a negative correlation in the aHC and mHC, but not in the pHC. Calculating the correlations separately for each group showed a significant correlation in the aHC for the 1 d group ($t(22) = -2.37$, $p = 0.027$, Pearson's $r = -0.45$, $n = 24$) and a trend in the 28 d group ($t(22) = -1.62$, $p = 0.121$, Pearson's $r = -0.37$, $n = 24$), while correlations for each group separately were not significant in the mHC and pHC. **d** Correlations of the Confidence Score with the contrast value for old pictures vs baseline: across groups, there was a positive correlation in the aHC and mHC, but not in the pHC. Note, however, that none of the correlations with the Confidence Score were significant when calculating them separately for each group. For analysis of the right HC long axis ROIs see Supplementary Fig. 7. *$p < 0.05$, **$p < 0.001$, all error bars are SEM

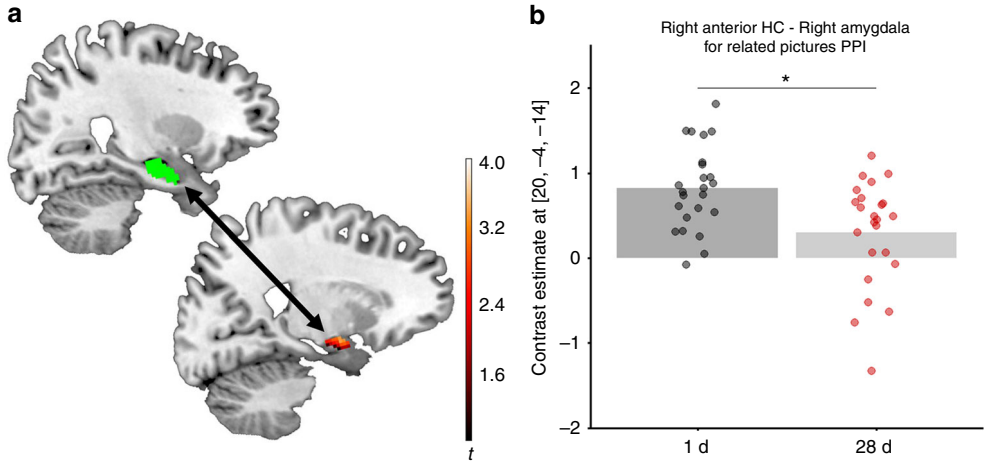


Fig. 3 Connectivity with amygdala. **a** Visualization of the connectivity between right aHC and right amygdala. Green represents the anatomical right aHC mask; red represents activation in the right amygdala for the contrast 1 d > 28 d of PPI interactions for related pictures with the right aHC as seed. Visualizations are superimposed on sagittal sections of a T1-weighted template image. **b** Parameter estimates in the peak voxel in the right amygdala. The right aHC showed a reduced connectivity to the right amygdala in the 28 d group compared to the 1 d group for related pictures (SVC peak level: $x = 20$, $y = -4$, $z = -14$, $t = 3.59$, p(FWE) = 0.0156, $k = 15$), but not for old or novel pictures. Note that neither using the left aHC nor the pHC (left or right) as seed regions, nor looking at other ROIs resulted in significant group differences in connectivity. See also Supplementary Table 3. *$p < 0.05$

be surprising that neocortical areas were similarly involved in the 1 d- and 28 d-group, as these areas might just reflect the processing of the currently viewed pictures. The fMRI findings so far are generally in line with the systems consolidation view, which would predict a decrease of hippocampal involvement in memory, irrespective of the specific picture type, over time[9,12,24].

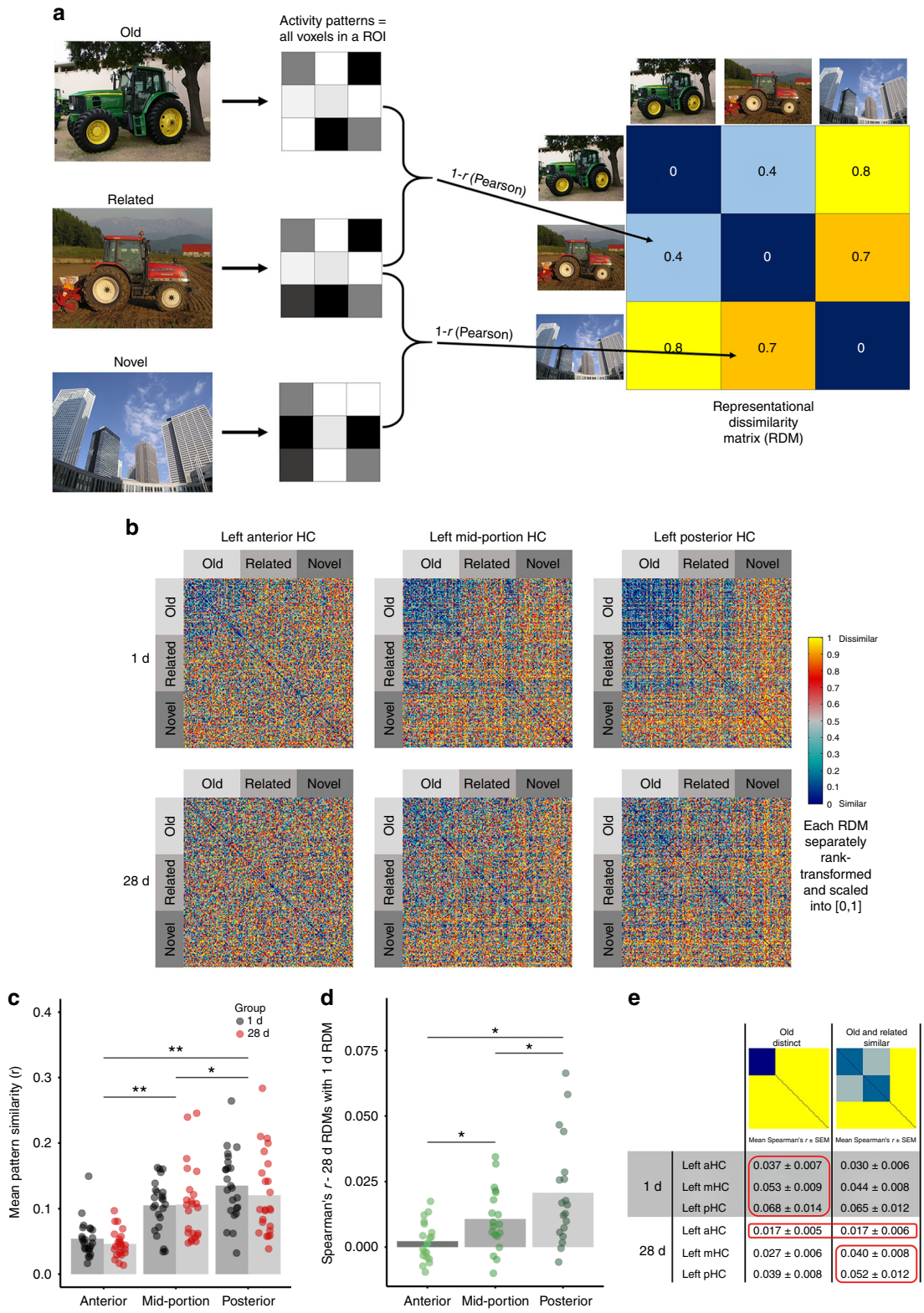
Looking at the hippocampal subregions along the long axis (Fig. 2a), however, revealed that the time-dependent decrease in activity was restricted to the aHC (left aHC main effect Group: $F_{(1,46)} = 15.46$, $p = 0.0003$, generalized $\eta^2 = 0.1546$, $n = 48$) and mid-portion hippocampus (left mHC: $F_{(1,46)} = 9.24$, $p = 0.0038$, generalized $\eta^2 = 0.0999$, $n = 48$). Activity in the pHC, however, did not significantly differ between the 1 d- and 28 d-group (left pHC: $F_{(1,46)} = 1.70$, $p = 0.1986$, generalized $\eta^2 = 0.0230$, $n = 48$; Fig. 2b). These differences between the ROIs were underlined by a significant Group × HC Long Axis interaction ($F(2,92) = 6.07$, $p = 0.0033$, generalized $\eta^2 = 0.036$, $n = 48$). The decrease in activity for high-confidence hits was also most pronounced in the aHC (Supplementary Fig. 5b). Moreover, connectivity analysis using the aHC and pHC as seed regions showed that the right aHC-right amygdala connectivity for related pictures (but not old or novel pictures) was significantly reduced in the 28 d- relative to the 1 d-group (SVC peak level: $x = 20$, $y = -4$, $z = -14$, $t = 3.59$, p(FWE) = 0.0156, $k = 15$; Fig. 3), while we found no significant differences between the groups in the connectivity to the amygdala when using the pHC as seed region, suggesting that it might be the connectivity between the aHC and the amygdala that is notably reduced in the 28 d group.

In order to further examine whether the decrease in aHC activity could be directly linked to the change in the nature of remembering, we correlated the activity in the hippocampal subregions with behavioral indices of memory specificity, i.e., the FA rate for related lures and the Confidence Score. These analyses showed that specifically the aHC was associated with the specificity of memory. In particular, aHC activity for related pictures was correlated negatively with the FA rate to related pictures (left aHC: $t_{(46)} = -5.15$, $p = 5.37e{-}06$, Pearson's $r = -0.60$, $n = 48$; Fig. 2c) and aHC activity for old pictures correlated positively with the Confidence Score (left aHC: $t(46) = 3.19$, $p = 0.0025$, Pearson's $r = 0.43$, $n = 48$; Fig. 2d). For the pHC, however, there were no such associations with memory specificity (left pHC, FA rate to related pictures: $t(46) = -1.08$, $p = 0.2879$, Pearson's $r = -0.16$, $n = 48$; Confidence Score: $t(46) = 0.44$, $p = 0.6645$, Pearson's $r = 0.06$, $n = 48$) and the correlations between activity and indicators of memory specificity were significantly distinct in the left aHC and pHC (FA related: Pearson and Filon's $z = -3.46$, $p = 0.0005$, $n = 48$; Confidence Score: Pearson and Filon's $z = 2.81$, $p = 0.0049$, $n = 48$). These correlations across the 1 d- and 28 d-groups indicate that the aHC and pHC are differentially linked to memory specificity. When we looked at the correlations separately in the 1 d- and 28 d-group, we obtained for the percentage of FA to related items, the key parameter of memory specificity, a significant correlation with aHC in the 1 d group only ($t(22) = -2.37$, $p = 0.027$, Pearson's $r = -0.45$, $n = 24$). For the 28 d group, this correlation did not reach significance ($t(22) = -1.62$, $p = 0.121$, Pearson's $r = -0.37$, $n = 24$), which might be related to the proposed reduced involvement of the aHC in memory in the 28 d group, although a lack of statistical power might also account for the non-significant correlation in the 28 d group. For the memory Confidence Score, the correlations with aHC activity did not reach significance in the separate groups (1 d group: $t(22) = 1.17$, $p = 0.255$, Pearson's $r = 0.24$, $n = 24$; 28 d group: $t(22) = 1.01$, $p = 0.326$, Pearson's $r = 0.21$, $n = 24$).

Although, we found a reduction of activity in the 28 d group in comparison to the 1 d group in the aHC and the mHC, it is important to note that there was no Group × Picture Type interaction (Supplementary Fig. 6a) in these ROIs. Thus, our univariate results show that the activity in the aHC, but not the pHC, is reduced in the 28 d group compared to the 1 d group for all picture types, suggesting that the contribution of the aHC in the task in general is reduced. Our brain-behavior correlations further show that the aHC, but not the pHC, is associated with memory specificity. It is not surprising that aHC activity was reduced irrespective of Picture Type after 28 d because a specific memory representation is required to both correctly identify an old item as old and to correctly reject novel or related items.

Specificity of mnemonic representations in aHC and pHC. The above univariate analyses showed that it was specifically the aHC that was associated with memory specificity and that specifically

activity in this area decreased at a longer retention interval of 28 days. While univariate analyses can show a general involvement of an area in a task, multivariate analysis allows the detection of specific patterns of activity across multiple voxels and may be more sensitive to the changing representations of the different picture types and more informative about the functional organization of memory at different time intervals. Therefore, we ran a RSA (Fig. 4a) to examine whether the mnemonic representations differed in the aHC, mHC and pHC, whether they changed depending on the retention interval, and to what extent such different representational patterns can be linked to the proposed memory transformation. We first created average representational dissimilarity matrices (RDMs) in each hippocampal subregion, separately for each group (Fig. 4b):



visualizations of these RDMs suggest the most similar activity patterns for combinations of old pictures in all hippocampal subregions in the 1 d group; whereas, in the 28 d group the representational pattern was less clear in the aHC and less-specific in the pHC. Note, however, that for the visualizations each RDM was separately rank transformed and scaled into [0, 1] preventing a direct descriptive comparison across hippocampal subregions. We therefore extracted the mean pattern similarity for each RDM: this showed that the mean similarity was highest in the pHC and lowest in the aHC (main effect HC Long Axis: $F_{(2,92)} = 84.37$, $p = 1.54e{-}21$, generalized $\eta^2 = 0.350$, $n = 48$; Fig. 4c), suggesting particularly distinct activity patterns in the aHC for different pictures which might allow for highly specific memories, whereas in the pHC the higher neural similarity across different pictures may reflect a larger degree of overlapping representations. We then looked at the similarity of the RDMs across groups and found that the overlap of the representational patterns of the 1 d- and 28 d-groups was significantly higher in the pHC than in the aHC (main effect HC Long Axis: $F_{(2,36)} = 12.72$, $p = 6.65e{-}05$, generalized $\eta^2 = 0.231$, $n = 19$; Fig. 4d). This suggests that pHC representational patterns changed less over time than aHC representational patterns did.

In order to directly test whether the time-dependent changes of the memory representations in anterior and posterior hippocampal areas were associated with the transformation from detailed to more gist-like memory, we finally compared brain and model-based RDMs. More specifically, we compared the brain RDMs of the hippocampal subregions with two model RDMs: (1) the model "Old Distinct" expects similar activity patterns for all old pictures that are distinct from patterns for related or novel pictures, a representation expected in areas that help detecting the old pictures as old and separate them from related pictures; (2) the model "Old and Related Similar" expects a more similar pattern for the old and related pictures, that is distinct from patterns for the novel pictures, a representation expected in areas that detect the gist as having been encoded but cannot separate between details (Fig. 4e). Based on the behavioral data, we reasoned that the "Old Distinct" model should, in general, fit better in the 1 d group as these participants still had detailed memories, while the "Old and Related Similar" model might fit better in the 28 d group, as for these participants part of the memories had been transformed to gist-like versions. Our analyses showed that, in the 1 d group, the "Old Distinct" model had, compared to the "Old and Related Similar" model, indeed a marginally better fit in the left aHC (one-tailed paired t-test: $t(23) = 1.58$, $p = 0.0635$, $n = 24$) and a better fit in left mHC (one-tailed paired t-test: $t(23) = 1.91$, $p = 0.0345$, $n = 24$), whereas in the left pHC both models were indistinguishable (one-tailed paired t-test: $t(23) = 0.61$, $p = 0.2737$, $n = 24$). In the 28 d group, on the other hand, the "Old and Related Similar" model had a better fit than the "Old Distinct" model in the left mHC (one-tailed paired t-test: $t(23) = -2.24$, $p = 0.0174$, $n = 24$) and tended to have a better fit in the left pHC (one-tailed paired t-test: $t(23) = -1.53$, $p = 0.0695$, $n = 24$), while in the left aHC both models were indistinguishable (one-tailed paired t-test: $t(23) = -0.07$, $p = 0.4740$, $n = 24$) and the respective model fits were generally rather low.

The brain data were largely comparable in both hemispheres (see Supplementary Figures 7 and 8 for results in the right HC). Group differences in the univariate analysis and the RSA were not modulated by stimulus emotionality (see Supplementary Figs. 6b and 9). We did, however, find time-dependent connectivity changes with hippocampal seed regions that were modulated by stimulus emotionality for some of the ROIs (Supplementary Table 3).

Discussion

How memories evolve over time is a fundamental issue of the neuroscience of memory. While previous research focused mainly on time-dependent changes in hippocampal and neocortical contributions to memory[3,9,10,12], here we show a time-dependent reorganization along the hippocampal long axis that is related to the transformation from detailed to gist-like memory. More specifically, our data indicate that the aHC is involved in memory specificity and represents actually encoded events distinctly from semantically related information at short retention intervals but shows a marked decrease in activity at longer retention intervals. Activity in the pHC, in turn, was largely unrelated to memory specificity and did not decrease over time, while pHC representational patterns seemed more gist-like at a longer retention interval.

The present data point to a possible involvement of the aHC in the specificity of memory. Previous data in rodents showed that firing fields of ventral hippocampal (corresponding to aHC in humans) place cells are larger than those in the dorsal hippocampus[25], which might translate into more abstract, large-scale aHC memory representations ([26], see also ref. [14]). However, the finding that an animals' exact location can be decoded from ventral hippocampal activity[27] is in line with the role of the aHC in memory specificity that we propose here. In addition, our results fit to a study showing stronger aHC activity for recent memories than for remote memories[20], to a study showing that aHC carries information about memory contexts in immediate and recent (1 day old) memories[28] and to studies showing a consistent implication of the aHC in memory of specific events[29]. Our results further dovetail with reports showing that the aHC specifically is associated with segregating events[30] and with

Fig. 4 Representational similarity analysis for the left HC long axis ROIs. **a** Schematic overview of the creation of a representational dissimilarity matrix (RDM; for illustration purposes only 3 pictures) modified from ref. [48]. Images courtesy of Andreas Praefcke (green tractor), Akiyoshi's Room (red tractor) and Morio (skyscrapers). **b** The group average RDMs of the left long axis hippocampal ROIs (1 d group in the first row, 28 d group in the second row). Blue colors = most similar, bright colors = most dissimilar; note that for the visualizations each RDM was separately rank transformed and scaled into [0, 1]. **c** Comparison of mean pattern similarities (Pearsons r) across ROIs: the mean similarity in the hippocampal subregions differed significantly. Note that all $n = 48$ participants are included here. **d** Comparison between groups: correlations (Spearman's r) of each single-subject RDM of the 28 d group to the respective average RDM of the 1 d group. The correlations between the two groups differed significantly in the three ROIs. Note that the data in **b** and **d** are from only 40 participants (1 d group = 21, 28 d group = 19) as the remaining eight participants had different sized RDMs and could therefore not be included in the average RDM for the group comparisons. **e** Comparison with two model RDMs: each cell in the table shows the mean of the correlations (Spearman's r + SEM) of the single-subject brain RDMs with the respective model RDM (first three rows = 1 d group, last three rows = 28 d group; $n = 48$). For each ROI the model with the higher correlation was marked by a red frame, in case of very similar correlations both values were marked. In the 1 d group, the "Old Distinct" model showed trends toward a better fit. In the 28 d group the "Old and Related Similar" model had a better fit in the mHC and a trend toward a better fit in the pHC, while in the aHC both models were indistinguishable and the model fits were generally rather low. For analysis of the right HC long-axis ROIs see Supplementary Fig. 8. $^*p < 0.05$, $^{**}p < 0.001$

novelty detection[17,31], both of which requires specific memory representations.

Whereas the activity of the aHC was reduced after 28 d, no such decline was observed for the pHC, suggesting that not all parts of the hippocampus decrease in activity over time. The RSA data, however, suggested a time-dependent change of the representational pattern in the pHC. In the 1 d group the pHC representation was already less-specific than the aHC representation, which corroborates the recent idea that there are complimentary learning systems within the hippocampus with one supporting gist-like representations[32,33]. In the 28 d group, the model RSA data even suggested that the representational patterns in the pHC resemble more gist-like patterns. This result suggests a time-dependent decrease in the specificity of the pHC memory representation. This idea is in line with a recent finding[34], showing that neural patterns of overlapping memories were more similar in the pHC after a week of consolidation. In this study, however, part of the memories were actually overlapping (e.g., same scene with different objects), whereas our study extends this finding by using two different pictures with only the semantic gist as overlap, thereby pointing to a memory transformation process.

Thus, there may be two time-dependent processes that contribute to more gist-like memory: a decrease in the aHC supporting memory specificity and an increase in the unspecificity of the mnemonic representation in the pHC, whose activity remains rather stable over time. Whether one process proceeds or follows the other or whether both occur independently remains to be shown.

Our findings suggest a functional specialization in which aHC representations support detailed memories and pHC representations are more gist-like after a longer time delay. Rodent data, however, suggest that the hippocampal long axis is organized along a gradient[15]. Most human studies did not address the mHC and rather little is known about the properties of this subregion. Our finding that the mHC was both with respect to its association with memory specificity and in terms of decreased activity in between the aHC and pHC is in line with the proposed functional gradient along the hippocampal long axis. Yet, how exactly the proposed different functions of the aHC and pHC are bridged is still unclear and remains a challenge for future research.

It is important to note that while we report this time-dependent reorganization within the hippocampus that was linked to memory transformation, we obtained also evidence for the proposed systems consolidation theory[9,12,24]. Hippocampal activity during recognition testing was significantly lower after 28 days than after 1 day, even for items remembered with high confidence. This latter point opposes the transformation hypothesis[2], which would not expect a reduction in hippocampal involvement for high confident, detailed memory. In addition, hippocampus-amygdala connectivity, known to be implicated in vivid memory[23,35], was reduced in the 28 d group compared to the 1 d group. However, this reduction in functional connectivity with the amygdala seemed to be specific to the aHC. This finding is in line with data suggesting that the aHC is connected to, among other regions, the amygdala, whereas the pHC is connected to areas involved in schematic memory such as the precuneus[14,16]. Thus, the aHC and pHC appear to be part of distinct neural networks that are involved in specific vs. gist-like memory and the observed reorganization along the hippocampal long axis is most likely concerted with the postulated large-scale redistribution (i.e., systems consolidation) of memory.

Finally, we would like to point out that the time-dependent changes reported here cannot be interpreted as a mere indication of a reduction in memory strength. In fact, we have designed this study explicitly to be able to differentiate between a general reduction in memory strength and memory transformation processes. In particular, we included related pictures that allowed us to probe memory specificity. If only memory strength was reduced after 28 d, this should be reflected in a comparable increase in the FA rates for related and novel pictures. We observed, however, a much stronger increase of FAs for related pictures than for novel pictures, which is in sharp contrast to the interpretation of a simple reduction in general memory strength but in line with the proposed transformation from detailed to gist-like memory. In addition, our model RSA data can also not be explained by a general reduction in memory strength. This view would imply that the memory for specific details and the gist memory decrease to a similar extent over time so that the relative representation of old and related items remains over time. Our data, however, show that the "Old Distinct" model best characterized activity in the 1 d group, whereas after 28 d the two models were indistinguishable in the aHC and the "Old and Related Similar" model seemed to fit better in the mHC and pHC. Together, these findings indicate that, in addition to the well-known decline in memory strength over time, there is also a change in the nature of memory, from detailed to more gist-like.

Our findings show that while the involvement of the hippocampus as a whole in memory decreases over time, this decrease is not present in all parts of the hippocampus. However, although there was a hippocampal memory representation even long after encoding, the nature (and origin) of the hippocampal contribution to remembering changed significantly with time. To conclude, we suggest here a time-dependent reorganization within the human hippocampus that is linked to a transformation from detailed to gist-like memory and might operate in tandem with the previously suggested large-scale reorganization of memory that occurs in the brain over time[9,12,24].

Methods

Participants. We tested 48 healthy, right-handed, young adults (24 men, 24 women; age: mean = 23.85 years, SD = 3.28 years) without a history of any psychiatric or neurological diseases, without medication intake or drug abuse and without circumstances preventing an MRI scan. All participants gave written informed consent and received monetary compensation for participation. The study protocol was approved by the ethics committee of the German Psychological Society (DGPs). Participants were pseudo-randomly assigned to the 1 d- or 28 d-group (12 women and 12 men per group). All experiments took place in the afternoon or early evening. The sample size corresponds to other studies on the neural underpinnings of memory processes and an a-priori power calculation with G* Power (http://www.gpower.hhu.de/; $f(U) = 0.5$, $\alpha = 0.05$, $1\text{-}\beta = 0.90$) for the decisive interaction effect Group × FA Picture Type (see Behavioral data analysis).

Study design and experimental paradigm. Testing took place on two experimental days: Day 1, encoding outside of the scanner and Day 2, recognition memory testing in the MRI scanner. Critically, the time interval between encoding and recognition testing was varied between the two experimental groups: for participants in the 1 d group recognition testing took place one day after encoding, while for participants in the 28 d group recognition memory was tested 28 days after encoding. The testing of the two groups was intermixed, so confounds related to changes in, for instance, the technical environment of the scanner over time cannot explain group differences.

Stimulus material. We used 180 pictures of natural scenes and objects as stimulus material. About one third of the pictures were taken from the International Affective Picture System (IAPS[36]), while the others were taken from open internet platforms. Half of the pictures contained emotionally negative scenes or objects while the other half contained neutral contents. Participants rated all pictures at the end of the experiment with respect to picture valence (scale from 0 = negative to 100 = positive, with 50 = neutral) and picture arousal (scale from 0 = not arousing to 100 = very arousing). In retrospect, these data confirmed that neutral pictures ($M = 57.38$, SEM = 0.79) were perceived as neutral and negative pictures ($M = 25.99$, SEM = 1.39) as more negative (paired t-test: $t_{(47)} = -18.38$, $p < 0.0001$, Cohen's $d = -2.65$, $n = 48$). Furthermore, negative pictures ($M = 47.50$, SEM = 3.01) had higher arousal ratings than neutral ones ($M = 11.15$, SEM = 1.90; paired t-test: $t_{(47)} = 13.42$, $p < 0.0001$, Cohen's $d = 1.94$, $n = 48$).

The 180 pictures were divided into three lists (each 30 negative and 30 neutral pictures): List A and List B contained semantically related pictures, i.e., for each

picture in List A there was a matching picture in List B that carried the same gist (e.g., mowing tractor) but different details (e.g., different brand, color, perspective, and background). List C, on the other hand, contained novel pictures that were not semantically related to either List A or List B pictures. Half of the participants learned List A during encoding and List B pictures were used as related lures and List C pictures as novel lures in the recognition test, while the other half of the participants learned List B during encoding and List A pictures were used as related lures and List C pictures as novel lures in the recognition test.

The semantic relatedness of the stimuli was rated by an independent sample ($n = 12$) on a scale from 1 ("not related") to 10 ("highly related"). Corresponding pictures of List A und List B were rated as highly related ($M = 8.25$, SEM = 0.062), in comparison to List A pictures compared to List C pictures ($M = 2.03$, SEM = 0.056, paired t-test: $t_{(19.84)} = 12.86$, $p < 0.0001$, Cohen's $d = 4.18$), or List B compared to List C pictures ($M = 1.98$, SEM = 0.055, paired t-test: $t_{(11)} = 14.85$, $p < 0.0001$, Cohen's $d = 4.49$).

Experimental day 1 (memory encoding). On the first experimental day, participants performed three encoding runs. In each run, the 60 pictures from the respective list (either A or B) were presented to the participant in random order on a computer screen, using MATLAB (www.mathworks.com) with the Psychophysics Toolbox extensions[37]. Each picture was presented for 2 s followed by a fixation cross of 1 s. Participants were instructed to memorize the pictures. Immediately after each encoding run a free recall task followed: participants verbally listed all the pictures they could remember while the investigator checked off the named pictures on a list and prompted the participant to a more detailed description in case the description of a picture was inconclusive. In total, the encoding session took about 20 min.

Experimental day 2 (memory testing). On the second experimental day, either 1 d or 28 d after encoding, participants first performed another free recall task outside the scanner and then a recognition task while fMRI measurements were taken. In the recognition task, participants saw the 60 old pictures, 60 related pictures, i.e., new pictures carrying the gist of the old pictures, and 60 novel pictures in random order. Each picture was shown for 3.5 s and participants were asked to indicate ("yes" vs. "no") by button press whether they had seen this picture during the encoding session or not. Critically, participants were informed before the task that some of the pictures may be similar to the original ones. Participants were further explicitly instructed to answer "Yes" only if they thought the picture was exactly the same as the one learned on experimental day 1. After participants' response, their choice was marked by a yellow box around the answer. If they answered "Yes" a confidence rating followed: they were asked to indicate on a 4-point scale how confident (not at all confident, slightly confident, quite confident, or very confident) they were that they had seen the picture on experimental day 1. This rating was shown for 2 s, and again their answer was marked by a yellow box. Each trial was followed by a fixation cross with a jittered presentation time of 7 ± 2 s.

Behavioral data analysis. To assess the performance in the recognition task in general we compared the percentages of hits for old pictures in a mixed-design ANOVA with Group (1 d vs 28 d) as between-subject factor and Emotionality (negative vs neutral) as within-subject factor. In order to assess the specificity of memory, we further analyzed the percentages of FA for related and novel lures in a mixed-design ANOVA with Group (1 d vs 28 d) as between-subject factor and FA Picture Type (related vs novel) and Emotionality (negative vs neutral) as within-subject factors. To additionally include information about the confidence of the participants when answering correctly, we calculated a Confidence Score by weighting each hit by the respective confidence (not at all confident = 0, slightly confident = 1, quite confident = 2 or very confident = 3), resulting in a score of 0–3 for each hit, before summing up overall hits. The maximum Confidence Score is therefore 180 (60 "very confident" hits). This score was again subjected to a mixed-design ANOVA.

We also analyzed the matching picture pairs (old picture and the corresponding related lure carrying the same gist) by assigning each pair to one of three categories: (1) detailed pairs, for which participants could reliably distinguish between old and related pictures and therefore correctly identified the old picture as old and correctly rejected the related picture as new. (2) transformed pairs, for which participants could not rely on detailed memories but still remembered the gist, as reflected by a FA to the related lure (irrespective of the response to the old picture), and (3) forgotten pairs, for which participants may have forgotten the whole picture (both gist and details), as reflected by a miss for the old picture and a correct rejection for the related picture.

Behavioral data analyses were performed with R version 3.3.2 (https://www.r-project.org/). All p-values are two-tailed and Welch's t-tests were used as default for between group comparisons[38]. The Shapiro–Wilk normality test was applied to the dependent variables: while the Confidence Score ($W = 0.96$, $p = 0.1246$) was normally distributed, this was not the case for the Hits ($W = 0.87$, $p = 0.0001$) and the FA ($W = 0.90$, $p = 0.0009$). Despite this violation of the normality assumption we applied the above described ANOVAs due to the robustness of these tests against the violation of this assumption[39].

MRI acquisition. MRI measurements were obtained with a 3T Skyra scanner (Siemens), equipped with a 32-channel head coil. For the functional images, a 3D echoplanar imaging (EPI) sequence (836 volumes) was used with the following parameters: 36 slices, slice thickness = 3 mm, distance factor 20%, repetition time (TR) = 2500 ms, echo time (TE) = 30 ms, voxel size 3.0 mm isotropic. We additionally acquired a high-resolution T1-weighted anatomical image (TR = 2.5 s, TE = 2.12 ms, 256 slices, voxel size = 0.8 × 0.8 × 0.9 mm) and a magnetic (B0) field map to unwarp the functional images.

Data preprocessing. The fMRI data were preprocessed using MATLAB and SPM12 (http://www.fil.ion.ucl.ac.uk/spm/). The first four functional images (10 s) were discarded from the rest of the analysis to allow for T1 equilibration. The remaining 832 functional images were first spatially realigned and unwarped using the field maps, then coregistered to the structural image, followed by a normalization to the MNI space. For the univariate analysis, the images were additionally spatially smoothed using an 8 mm full-width half-maximum Gaussian kernel.

General linear modeling and whole-brain analysis. For the univariate analysis, the data were analyzed using general linear modeling (GLM) as implemented in SPM12. Six separate regressors for each of the Picture Type × Emotionality combinations were modeled: old negative, old neutral, related negative, related neutral, novel negative, and novel neutral. The onsets of the confidence ratings were additionally included as a regressor of no interest and all regressors were convolved with the canonical hemodynamic response function. Note that we did not include movement regressors in the GLM, as we used the SPM unwarp function in the data preprocessing instead. A high-pass filter of 128 s was used to remove low-frequency drifts and serial correlations in the time series were accounted for using an autoregressive AR(1) model. To look at whole-brain activation differences between the 1 d- and 28 d-groups, we used a two-sample t-test design at second-level modeling.

ROI analysis. In addition to the whole-brain analysis, we performed regions of interest analyses that focused on brain areas that have previously been implicated in detailed and more semantic or schema-related memory processes[1,3,9,40]. To this end, we used the following anatomical masks from the Harvard-Oxford atlas using a probability threshold of 50%: hippocampus (left and right), anterior parahippocampal gyrus (left and right), posterior parahippocampal gyrus (left and right), precuneus, angular gyrus (left and right), anterior cingulate gyrus, inferior frontal gyrus pars opercularis (left and right), inferior frontal gyrus pars triangularis (left and right), temporal pole (left and right), and the amygdala (left and right). In addition, we used masks created with MARINA (http://www.bion.de/eng/MARINA.php) for the left and right ventromedial prefrontal cortex. The signal within the ROIs was deconvolved for each of the regressors from the GLM (old negative, old neutral, related negative, related neutral, novel negative, and novel neutral) using a finite impulse response function (FIR) on the time course averaged across all voxels of the ROI as implemented within MarsBar[41] for the first seven repetition times (TRs; 15 s). We choose FIR deconvolutions here to capture the shape of the HRF and allow for differences in this hemodynamic response across regions and participants. For statistical comparisons in R, we extracted the peak response from these FIR time courses: as described in refs. [42,43] the peak response was defined as the average signal over time points whose responses (collapsed across all conditions) did not significantly ($p > 0.05$) differ from the numerical peak tested across all participants. In most ROIs this procedure resulted in the peak response being the average across the 5 s and the 7.5 s time points (exceptions: left and right posterior parahippocampal gyrus and right angular gyrus = only the 5 s time point; left angular gyrus = 2.5 s and 5 s time points; precuneous cortex = 7.5 s and 10 s time points; left and right temporal pole = 5 s, 7.5 s and 10 s time points; in the left and right ventromedial prefrontal cortex the numerical peak did not significantly differ from any of the other time points, as no reliable peak response seems to be present in these two ROIs we did not perform further analysis on this data). Then the average signal over the respective time points were calculated for each condition separately (old negative, old neutral, related negative, related neutral, novel negative, and novel neutral) and used in a mixed ANOVA model, with Group (1 d vs 28 d) as between-subject factor and Picture Type (old vs related vs novel) and Emotion (negative vs neutral) as within-subject factors. In case of violation of the sphericity assumption, Greenhouse-Geisser corrections were applied. The p-value threshold was adjusted for multiple comparisons by the numbers of ROIs.

Differentiation along the hippocampal long axis. In order to look at differences across the long (anterior–posterior) axis of the hippocampus, we used the procedure described by ref. [26] to divide a hippocampal mask into three parts with approximately equal lengths along the long axis, using the WFU pick-atlas[44,45]: pHC from $Y = -40$ to -30, mHC from $Y = -29$ to -19, and aHC from $Y = -18$ to -4. For these new hippocampal ROIs, we then deconvolved the signal for each of the regressors using a finite impulse response function (FIR), and extracted the peak response from these FIR time courses for statistical comparisons in R as described in detail for the other ROIs above. In all of these hippocampal long axis ROIs this procedure resulted in the peak response being defined as the average across the 5 s and the 7.5 s time points. We then run a mixed ANOVA model on

these peak responses, with Group (1 d vs 28 d) as between-subject factor and Picture Type (old vs related vs novel) and Emotion (negative vs neutral) as within-subject factors in each of the ROIs separately. In addition, we performed another analysis to see if the delay manipulation (1 d vs 28 d) had a different effect on these three hippocampal ROIs, using an ANOVA with Group as between-subjects factor and HC Long Axis (aHC, mHC, and pHC) as within-subject factor.

Comparison of high- and low-confidence hits. In order to assess whether the time-dependent decrease in hippocampal activity was modulated by confidence, we created a second model that was based on the behavioral responses and confidence ratings of the participants. We modeled five regressors: High-confidence hits (hits with a confidence rating of 3), low-confidence hits (hits with a confidence rating of 0, 1 or 2), related CR, novel CR, and all incorrect answers (misses + related FA + novel FA + no presses). The onsets of the confidence ratings were again included as a regressor of no interest and all other procedures were the same as in the first GLM. We then deconvolved the signal for the high-confidence hits and low-confidence hits in the hippocampal ROIs using a finite impulse response function (FIR) with MarsBar, and extracted the peak response from these FIR time courses for statistical comparisons in R using the same time points as peak as in the analysis above. We then run a mixed ANOVA model on these peak responses, with Group (1 d vs 28 d) as between-subjects factor and Hit Type (high vs low) as within-subject factor in each of the ROIs separately. Group effects were followed up by post hoc Welchs *t*-tests comparing the groups (1 d vs 28 d) for each Hit Type separately.

Correlation with behavior. We extracted the contrast values from the main effects (condition vs baseline) of the first GLM for each Picture Type (old, related, novel) in each ROI again using the MarsBar toolbox, and computed correlations between behavioral memory scores (percentage of related FA; Confidence Score) and the contrast values for the respective picture types (related > baseline, old > baseline) in each of the hippocampal long axis ROIs outside SPM using R for all participants. We then also performed the correlations for each group (1 d vs 28 d) separately. In a next step, we compared the correlations in the hippocampal long axis ROIs with each other (e.g., correlation in aHC with the correlation in pHC), using the cocor package from R (http://comparingcorrelations.org/). As the related FA value (or Confidence Score) was used in all three correlations, we report the results of comparisons of two overlapping correlations based on dependent groups. The reported values are Pearson and Filon's *z*-scores.

Functional connectivity analysis. We used a generalized form of context-dependent psychophysiological interaction (gPPI)[46] to measure task-dependent connectivity with either the whole hippocampus (left and right) or the aHC or pHC (left and right) as seed regions. In contrast to the standard PPI implementation through SPM, the gPPI toolbox allows the inclusion of more than two task conditions in one PPI model and therefore allows a more flexible analysis: we entered the six task regressors from the first GLM model (old negative, old neutral, related negative, related neutral, novel negative, and novel neutral), plus a PPI Interaction term for each of these regressors, plus the time course from the respective seed region and the confidence ratings as regressor of no interest into our first-level PPI model. For second-level modeling, we entered the following contrast files from the first-level PPI analyses (main effects for PPI Old, PPI Related, and PPI Novel, and the differences contrast PPI old negative > PPI old neutral, PPI related negative > PPI related neutral, PPI novel negative > PPI novel neutral) into two-sample *t*-tests, comparing the 1 d group and 28 d group. We then applied a small volume correction (SVC) for all our other ROIs (see ROI analysis for a list of the ROIs), to find areas which show a significant difference in their connectivity to the seed region between the two groups. Voxels were regarded as significant when falling below a corrected voxel threshold of 0.05 (FWE) adjusted for the small volume. All areas with $k > 10$ significant voxels were reported.

Representational similarity analysis. Independent from the univariate analysis, we carried out a RSA[47] in the hippocampal long-axis ROIs using the rsatoolbox[48]. For each ROI and each subject, brain Representational Dissimilarity Matrices (RDMs) were computed based on a single trial univariate GLM estimated on unsmoothed, normalized functional images. The response-amplitude beta estimate maps associated with each trial were converted into *t*-maps and used to create vectors of activity patterns for each trial, separately for each ROI. These activity patterns were used to calculate the dissimilarity between two trials by correlation distances ($1-r$, Pearson linear correlation). Next, the dissimilarities based on each combination of trials were placed into the respective cells of the 180×180 RDMs (Fig. 4a). Due to technical failure, we did not have functional data for all trials in some of the participants. Thus 8 participants had RDMs of slightly different sizes (in the 1 d group three participants had 179×179 RDMs; in the 28 d group two participants had 179×179 RDMs, two participants had 178×178 RDMs and one participant had a 176×176 RDM). For visualization of the RDMS we created average RDMs for each group from the single-subject RDMs (Fig. 4b). As this required RDMs of the same size, the visualizations only include the data of the 40 participants with 180×180 RDMs.

Comparison of pattern similarities across ROIs. We extracted the mean pattern similarity (r) from each single-subject RDM in order to get the overall similarity of activity patterns when performing the recognition task in general, irrespective of picture types, per ROI. We then compared these mean pattern similarities across the hippocampal long axis by conducting an ANOVA with the factors HC Long Axis (aHC, mHC, pHC), and Group (1 d vs 28 d), and post hoc Bonferroni corrected paired *t*-tests to compare each region to each of the other regions. Note that for this analysis all 48 participants were included.

Comparison of RDMs between groups. We next compared the RDMs of the two groups in the respective ROIs: for this we extracted the Spearman correlation for each single-subject RDM of the 28 d group with the group average RDM of the 1 d group. This similarity of the RDMs between groups was then again compared between the hippocampal long axis ROIs with an ANOVA with the factor HC Long Axis (aHC, mHC, and pHC), and post hoc Bonferroni corrected paired t-tests to compare each region to each of the other regions. Note that for this analysis only the 40 participants with 180×180 RDMs could be included as we aimed to directly compare the RDMs based on single trials, irrespective of picture category. Therefore all RDMs had to be of the same size for this particular analysis.

Comparison with model RDMs. We also compared the brain RDMs to two model RDMs (Fig. 4e) that were based on the expected similarities of the different picture types: the model "Old Distinct" expects similar activity patterns for all old pictures that are distinct from patterns for related or novel pictures and the model "Old and Related Similar" expects a similar pattern for the old and related pictures, that is distinct from patterns for the novel pictures. We calculated Spearman's rank correlation coefficient for each single-subject brain RDM and these a-priori model RDMs. This rank coefficient is beneficial if it is not possible to assume a direct linear match between the RDMs that are compared[47], as is the case here. We then calculated the mean of these Spearman's r's for each group separately (1 d vs. 28 d) to find the model with the overall best fit in each group and ROI. We then statistically compared the two models in each ROI per group with one-tailed paired *t*-tests, expecting a better fit for the "Old Distinct" model in the 1 d group and, on the other hand, a better fit for the "Old and Related Similar" model in the 28 d group. Note that for this analysis all 48 participants could be included by creating model RDMs of the respective size matching each participants brain RDM.

Data availability. All data and codes are available from the corresponding authors upon request. The data are not publicly available yet because they contain information that could compromise research participant privacy and consent. In the near future, they will be de-identified at the level of contemporary best practices and made publicly available, together with relevant code, at the corresponding author's GitHub repository (https://github.com/LarsSchwabeHamburg/transformation).

Received: 22 September 2017 Accepted: 2 March 2018
Published online: 23 March 2018

References

1. Dudai, Y. The restless engram: consolidations never end. *Annu. Rev. Neurosci.* **35**, 227–247 (2012).
2. McKenzie, S. & Eichenbaum, H. Consolidation and reconsolidation: two lives of memories? *Neuron* **71**, 224–233 (2011).
3. Winocur, G., Moscovitch, M. & Bontempi, B. Memory formation and long-term retention in humans and animals: convergence towards a transformation account of hippocampal-neocortical interactions. *Neuropsychologia* **48**, 2339–2356 (2010).
4. Backus, A. R., Bosch, S. E., Ekman, M., Grabovetsky, A. V. & Doeller, C. F. Mnemonic convergence in the human hippocampus. *Nat. Commun.* **7**, 11991 (2016).
5. Chadwick, M. J., Hassabis, D., Weiskopf, N. & Maguire, E. A. Decoding individual episodic memory traces in the human hippocampus. *Curr. Biol.* **20**, 544–547 (2010).
6. Ekstrom, A. D. & Ranganath, C. Space, time, and episodic memory: the hippocampus is all over the cognitive map. *Hippocampus* **00**, 1–8 (2017).
7. Schacter, D. L. & Wagner, A. D. Medial temporal lobe activations in fMRI and PET studies of episodic encoding and retrieval. *Hippocampus* **9**, 7–24 (1999).
8. Scoville, W. B. & Milner, B. Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* **20**, 11–21 (1957).
9. Frankland, P. W. & Bontempi, B. The organization of recent and remote memories. *Nat. Rev. Neurosci.* **6**, 119–130 (2005).
10. Nadel, L. & Moscovitch, M. Memory consolidation, retrograde amnesia and the hippocampal complex. *Curr. Opin. Neurobiol.* **7**, 217–227 (1997).
11. Squire, L. R. & Alvarez, P. Retrograde amnesia and memory consolidation: a neurobiological perspective. *Curr. Opin. Neurobiol.* **5**, 169–177 (1995).

12. Squire, L. R. & Bayley, P. J. The neuroscience of remote memory. *Curr. Opin. Neurobiol.* **17**, 185–196 (2007).
13. Moser, M. B. & Moser, E. I. Functional differentiation in the hippocampus. *Hippocampus* **8**, 608–619 (1998).
14. Poppenk, J., Evensmoen, H. R., Moscovitch, M. & Nadel, L. Long-axis specialization of the human hippocampus. *Trends Cogn. Sci.* **17**, 230–240 (2013).
15. Strange, B. A., Witter, M. P., Lein, E. S. & Moser, E. I. Functional organization of the hippocampal longitudinal axis. *Nat. Rev. Neurosci.* **15**, 655–669 (2014).
16. Fanselow, M. S. & Dong, H. W. Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron* **65**, 7–19 (2010).
17. Zeidman, P. & Maguire, E. A. Anterior hippocampus: the anatomy of perception, imagination and episodic memory. *Nat. Rev. Neurosci.* **17**, 173–182 (2016).
18. Bonnici, H. M. et al. Detecting representations of recent and remote autobiographical memories in vmPFC and hippocampus. *J. Neurosci.* **32**, 16982–16991 (2012).
19. Bonnici, H. M., Chadwick, M. J. & Maguire, E. A. Representations of recent and remote autobiographical memories in hippocampal subfields. *Hippocampus* **23**, 849–854 (2013).
20. Gilboa, A., Winocur, G., Grady, C. L., Hevenor, S. J. & Moscovitch, M. Remembering our past: functional neuroanatomy of recollection of recent and very remote personal events. *Cereb. Cortex.* **14**, 1214–1225 (2004).
21. Kensinger, E. A., Garoff-Eaton, R. J. & Schacter, D. L. Effects of emotion on memory specificity: memory trade-offs elicited by negative visually arousing stimuli. *J. Mem. Lang.* **56**, 575–591 (2007).
22. Waring, J. D. & Kensinger, E. A. How emotion leads to selective memory: neuroimaging evidence. *Neuropsychologia* **49**, 1831–1842 (2011).
23. Richardson, M. P., Strange, B. A. & Dolan, R. J. Encoding of emotional memories depends on amygdala and hippocampus and their interactions. *Nat. Neurosci.* **7**, 278–285 (2004).
24. Squire, L. R., Genzel, L., Wixted, J. T. & Morris, R. G. Memory consolidation. *Cold Spring Harb. Perspect. Biol.* **7**, a021766 (2015).
25. Kjelstrup, K. B. et al. Finite scale of spatial representation in the hippocampus. *Science* **321**, 140–143 (2008).
26. Collin, S. H., Milivojevic, B. & Doeller, C. F. Memory hierarchies map onto the hippocampal long axis in humans. *Nat. Neurosci.* **18**, 1562–1564 (2015).
27. Keinath, A. T. et al. Precise spatial coding is preserved along the longitudinal hippocampal axis. *Hippocampus* **24**, 1533–1548 (2014).
28. Ritchey, M., Montchal, M. E., Yonelinas, A. P., Ranganath, C. Delay-dependent contributions of medial temporal lobe regions to episodic memor retrieval. *eLife* **4**, e05025 (2015).
29. Svoboda, E., McKinnon, M. C. & Levine, B. The functional neuroanatomy of autobiographical memory: a meta-analysis. *Neuropsychologia* **44**, 2189–2208 (2006).
30. Milivojevic, B., Vicente-Grabovetsky, A. & Doeller, C. F. Insight reconfigures hippocampal-prefrontal memories. *Curr. Biol.* **25**, 821–830 (2015).
31. Kumaran, D. & Maguire, E. A. Which computational mechanisms operate in the hippocampus during novelty detection? *Hippocampus* **17**, 735–748 (2007).
32. Schapiro, A. C., Kustner, L. V. & Turk-Browne, N. B. Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr. Biol.* **22**, 1622–1627 (2012).
33. Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., Norman, K. A. Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372**, 20160049 (2017).
34. Tompary, A. & Davachi, L. Consolidation promotes the emergence of representational overlap in the hippocampus and medial prefrontal cortex. *Neuron* **96**, 228–241 (2017). e225.
35. Kensinger, E. A., Addis, D. R. & Atapattu, R. K. Amygdala activity at encoding corresponds with memory vividness and with memory for select episodic details. *Neuropsychologia* **49**, 663–673 (2011).
36. Lang, P. J., Bradley, M. M. & Cuthbert, B. N. *International Affective Picture System (IAPS): Technical Manual and Affective Ratings* (National Institute of Mental Health Center for the Study of Emotion and Attention, Gainesville, 1997).
37. Brainard, D. H. The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
38. Delacre, M., Lakens, D. & Leys, C. Why psychologists should by default use Welch's t-test instead of Student's t-test. *Int. Rev. Social. Psychol.* **30**, 92 (2017).
39. Tabachnick B. G., Fidell L. S. *Using Multivariate Statistics* 6th edn (Pearson Education, Boston, 2013).
40. van Kesteren, M. T., Ruiter, D. J., Fernandez, G. & Henson, R. N. How schema and novelty augment memory formation. *Trends Neurosci.* **35**, 211–219 (2012).
41. Brett, M., Anton, J. L., Valabregue, R. & Poline, J. B. Region of interest analysis using an SPM toolbox. In *8th International Conference on Functional Mapping of the Human Brain*. Available in *NeuroImage* **16**, abstr. 497 (Sendai, Japan, 2002).
42. Turk-Browne, N. B., Simon, M. G. & Sederberg, P. B. Scene representations in parahippocampal cortex depend on temporal context. *J. Neurosci.* **32**, 7202–7207 (2012).
43. Turk-Browne, N. B., Yi, D. J. & Chun, M. M. Linking implicit and explicit memory: common encoding factors and shared representations. *Neuron* **49**, 917–927 (2006).
44. Lancaster, J. L. et al. Automated Talairach atlas labels for functional brain mapping. *Hum. Brain Mapp.* **10**, 120–131 (2000).
45. Maldjian, J. A., Laurienti, P. J., Kraft, R. A. & Burdette, J. H. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage* **19**, 1233–1239 (2003).
46. McLaren, D. G., Ries, M. L., Xu, G. & Johnson, S. C. A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. *Neuroimage* **61**, 1277–1286 (2012).
47. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).
48. Nili, H. et al. A toolbox for representational similarity analysis. *PLoS Comput. Biol.* **10**, e1003553 (2014).

Acknowledgements

The authors thank Daniel Kutzner, Wiebke Schmidt, Sandra Erbach, Kristin Medel, and Amina Shah for assistance during data collection and Dr. Lynn Nadel as well as Dr. Tobias Sommer for helpful comments on a previous version of this manuscript. This work was supported by the German Research Foundation (DFG; SCHW1357/12-1).

Author contributions

L.C.D. collected the data, analyzed the data, and wrote the manuscript. L.S. conceived and designed the experiment, supervised the project, and wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-03661-7.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix B

Study II: Time-dependent motor memory representations in prefrontal cortex

Published in NeuroImage

NeuroImage 197 (2019) 143–155

Contents lists available at ScienceDirect

NeuroImage

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

Time-dependent motor memory representations in prefrontal cortex

Lisa C. Dandolo, Lars Schwabe *

Department of Cognitive Psychology, University of Hamburg, 20146, Hamburg, Germany

ARTICLE INFO

Keywords:
Middle frontal gyrus
Motor memory
Motor sequence learning
Serial reaction time task
Superior frontal gyrus

ABSTRACT

How memories evolve over time is fundamental for understanding memory. Hippocampus-dependent episodic memories are generally assumed to undergo a time-dependent neural reorganization involving an increased reliance on neocortical areas. Yet, whether other forms of memory undergo a similar reorganization over time remains unclear. Here, we examined whether the neural underpinnings of motor sequence memories change over time. Participants were trained on a motor sequence learning task. Either 1d or 28d later, they performed a retention test for this task in the fMRI scanner. Sequence-specific motor memory was observed both 1d and 28d after initial training. Bayesian second-level fMRI analyses suggested a higher probability for task activity in the middle frontal gyrus and frontal pole 28d compared to 1d after initial motor learning. Searchlight representational similarity analysis indicated that areas in middle and superior frontal cortex were more involved in differentiating between multivariate activity patterns for old motor sequence memories and newly learned motor sequences in the 28d-group compared to the 1d-group. This increased involvement of lateral frontal areas during the task after 28 days was not paralleled by a decrease in those areas that were involved in performing the motor sequence retention task after 1d. These novel findings provide insights into how memories beyond the hippocampus evolve over time.

1. Introduction

Memory changes over time. For episodic and spatial memories, encoded by the hippocampus (Eichenbaum, 1999; O'Keefe and Nadel, 1978; Squire and Zola-Morgan, 1991), these time-dependent changes are known to be accompanied by a neural reorganization that is referred to as systems consolidation (Dudai et al., 2015; Frankland and Bontempi, 2005). It is assumed that during this consolidation process initially hippocampus-dependent memories are, as time proceeds, increasingly represented by neocortical areas (Squire and Alvarez, 1995; Squire et al., 2015) and it is debated whether the hippocampus is at all involved in remote episodic or spatial memories (Nadel et al., 2007; Squire and Bayley, 2007; Winocur and Moscovitch, 2011). The time-dependent reorganization of memories is fundamental for understanding memory. However, although it is well known that there are multiple memory systems beyond the hippocampus (Eichenbaum and Cohen, 2001; Squire, 2004; White et al., 2013), the long-term temporal dynamics of memory have been mostly investigated in (initially) hippocampus-dependent forms of memory (e.g. Bonnici et al., 2012; Furman et al., 2012; Gilboa et al., 2004). The time-dependent neural reorganization of other forms of memory, such as motor sequence memory, has only rather recently been discussed within the framework of the systems consolidation debate (Dudai et al., 2015) and studies so far focused on early systems consolidation processes or reorganization processes associated with repeated training (Albouy et al., 2008; Debas et al., 2010; Coynel et al., 2010; Dayan and Cohen, 2011).

Motor sequence learning relies mainly on cortico-striatal and cortico-cerebellar systems, including the putamen, caudate nucleus, cerebellum, thalamus as well as motor and somatosensory cortices (Doyon et al., 2003, 2009; Hardwick et al., 2013). It is well documented that motor memories undergo an initial synaptic consolidation process similar to the one known for hippocampal memories, with newly established memories being vulnerable to interference immediately after learning but not hours later (Brashers-Krug et al., 1996). This consolidation process is also paralleled by distinct neural changes at systems level within the first hours after learning (Shadmehr and Holcomb, 1997). Consolidation of motor sequence memory is further reflected in off-line gains in reaction times, i.e. improvements in motor task performance between two training sessions without further practice (Dayan and Cohen, 2011; Debas et al., 2014; Doyon and Benali, 2005; Robertson et al., 2004a). At least for explicit motor sequence memory, this off-line learning requires sleep (Debas et al., 2010; Robertson et al., 2004b). Elegant neuroimaging

* Corresponding author. University of Hamburg, Department of Cognitive Psychology, 20146, Hamburg, Germany.
E-mail address: Lars.Schwabe@uni-hamburg.de (L. Schwabe).

https://doi.org/10.1016/j.neuroimage.2019.04.051
Received 28 September 2018; Received in revised form 22 March 2019; Accepted 17 April 2019
Available online 20 April 2019
1053-8119/

studies revealed that the sleep-dependent early systems consolidation of motor memories is linked to changes in the striatum (Albouy et al., 2008; Debas et al., 2010; Fogel et al., 2017; Vahdat et al., 2017) and greater between-regions interaction within the cortico-striatal system (Debas et al., 2014), pointing to a neural reorganization of motor memory after sleep, similar as for hippocampus-dependent memory (Born and Wilhelm, 2012; Diekelmann et al., 2009). While these studies point to an early overnight consolidation process for motor memories, several studies have also shown that extensive training of motor sequences over weeks results in a transfer of motor memories from an associative, premotor circuit to an sensorimotor circuit (Coynel et al., 2010; Dayan and Cohen, 2011; Floyer-Lea and Matthews, 2005; Lehericy et al., 2005; but see Kupferschmidt et al., 2017), and that these extensively trained motor skills can be retained over long periods of time, even without further training (Park and Sternad, 2015; Penhune and Doyon, 2002; Romano et al., 2010). Studies on extensive training of motor skills over weeks, however, provide only little insight into the spontaneous, i.e. training-independent, reorganization of motor memory over several weeks that has been shown for episodic memory (Squire and Alvarez, 1995; Squire et al., 2015). Behavioral studies demonstrate that motor memories can be retained over long time periods, even with only minimal training (Julius and Adi-Japha, 2015; Savion-Lemieux and Penhune, 2005). However, it remains unclear whether several weeks old remote motor sequence memories, that were not extensively trained, rely on the same neural circuits as more recent (e.g. one day old) motor memories. In other words, is there a prolonged systems consolidation-like process in motor sequence memory that goes beyond the early system consolidation processes seen after the first nights of sleep (Albouy et al., 2008; Debas et al., 2010; Fogel et al., 2017; Vahdat et al., 2017)?

In the present experiment, we tested whether motor sequence memories undergo a neuronal reorganization over several weeks. To this end, participants were first trained in a modified version of the implicit Serial Reaction Time Task (SRTT) that has been used since decades to test motor sequence learning and memory (Nissen and Bullemer, 1987; Robertson, 2007). In this task, changing visual cues are presented to the participants with the instruction to respond to each cue as fast as possible with a respective button press (Fig. 1A). Not known to the participants, a specific sequence of cues is repeated several times to produce a recurring movement of sequential button presses. Faster reaction times in these *target* trials compared to those in trials with *random* sequences reflect successful motor sequence learning, which is typically observed after only a few trials (Doyon et al., 2009; Doyon and Benali, 2005). After the initial SRTT learning phase on a first experimental day, participants completed a SRTT testing phase in a functional magnetic resonance imaging (fMRI) scanner on a second experimental day. Critically, for half of the participants the interval between motor learning and retention testing was one day (1d-group), whereas for the other half of the participants this interval was four weeks (28d-group). This experimental set-up allowed us to determine whether remote motor sequence memories recruit the same or different brain areas as recent motor sequence memories, without any interference due to repeated testing. In the testing phase, we additionally introduced a second repeating sequence, allowing a distinction between new sequence learning on experimental day 2 and sequence-specific motor memory for the sequence learned on day 1. Importantly, successful motor skill learning can lead to increasing or decreasing activity in the related brain areas (Dayan and Cohen, 2011; Huang et al., 2013; Wiestler and Diedrichsen, 2013). Decreasing activity, however, does not necessarily imply that the respective area is less involved in motor sequence learning and memory but may point to more efficient encoding of the motor sequences (Poldrack et al., 2005; Ungerleider et al., 2002; Wiestler and Diedrichsen, 2013). Therefore, it has been suggested to complement traditional, univariate fMRI analyses by multivariate analyses that allow the detection of specific multivariate activity patterns irrespective of an overall increase or decrease in activity (Wiestler and Diedrichsen, 2013; Wymbs and Grafton, 2015). Thus, we combined here univariate fMRI analysis and multivariate search-light representational dissimilarity analysis (RSA) to investigate time-dependent changes in the neuronal representation of motor sequence memories. Based on the proposed systems consolidation process for hippocampus-dependent episodic memory, we predicted a similar, time-dependent reorganization of remote (several weeks old), compared to recent (one day old), motor sequence memories, with an increase in neocortical memory representation, possibly paralleled by a decrease in regions subserving initial motor sequence learning.

2. Materials and methods

2.1. Participants

We tested 48 healthy, right-handed young adults with normal or corrected-to-normal vision (24 men, 24 women; age: mean = 23.85 years, SD = 3.28 years). Exclusion criteria comprised a lifetime history of any psychiatric or neurological disorder, medication intake or drug abuse and circumstances preventing a MRI scan. One participant had to be excluded due to technical problems during data acquisition leading to an incomplete data set, four participants due to extensive movement during the fMRI measurements (more than 3mm/3°), one participant due to strong task-related movement (correlation of task onsets and the 6th rigid body motion parameter of r = - 0.51) and three participants due to poor overall performance in the task (more than 34% error rate), thus leaving a final sample of 39 participants for analysis (age: mean = 23.44 years, SD = 3.11 years; 1d-group: 9 men, 9 women; 28d-group: 11 men, 10 women). The study protocol was approved by the ethics committee of the German Psychological Association (LS 062013_012015). Participants gave written informed consent before participation and received monetary compensation. All testing took place in the afternoon or early evening.

2.2. Experimental design

The study consisted of two experimental days: on Day 1, participants performed the learning phase of the serial reaction time task (SRTT) outside the scanner. On Day 2, participants completed the test phase of the SRTT in the fMRI scanner. Importantly, the time interval between the learning phase and the test phase depended on the experimental group: participants assigned to the 1d-group performed the test phase one day after learning, while participants assigned to the 28d-group performed the test phase four weeks after the learning phase. This between-subject design allowed us to compare recent and remote motor sequence memories without any interference due to repeated testing. Participants were pseudo-randomly assigned to the 1d-group or 28d-group (12 women and 12 men per group). The testing of participants of the two groups in the fMRI scanner took place intermixed, so potential group differences cannot be explained by systematic changes in the technical environment of the scanner over time. In addition to the SRTT, participants performed an unrelated explicit picture encoding task or picture recognition task immediately before the SRTT learning phase or test phase on experimental Day 1 or Day 2, respectively, the results of which are reported elsewhere (Dandolo and Schwabe, 2018).

2.3. Serial reaction time task (SRTT)

We adapted a modified version of the original SRTT (Nissen and Bullemer, 1987) that was introduced by Tzvi et al. (2015). In this task, participants saw changing visual cues and were instructed to respond to each cue as quickly as possible with a specific button press. Unbeknownst to the participants, a specific sequence of twelve cues was repeated several times. To allow an event-related design we modified the block design used in Tzvi et al. (2015) by introducing breaks between each sequence of twelve cues. More specifically, at the beginning of each trial, consisting of a sequence of twelve cues, four grey squares appeared in a horizontal array in the middle of a computer screen (see Fig. 1A). Each

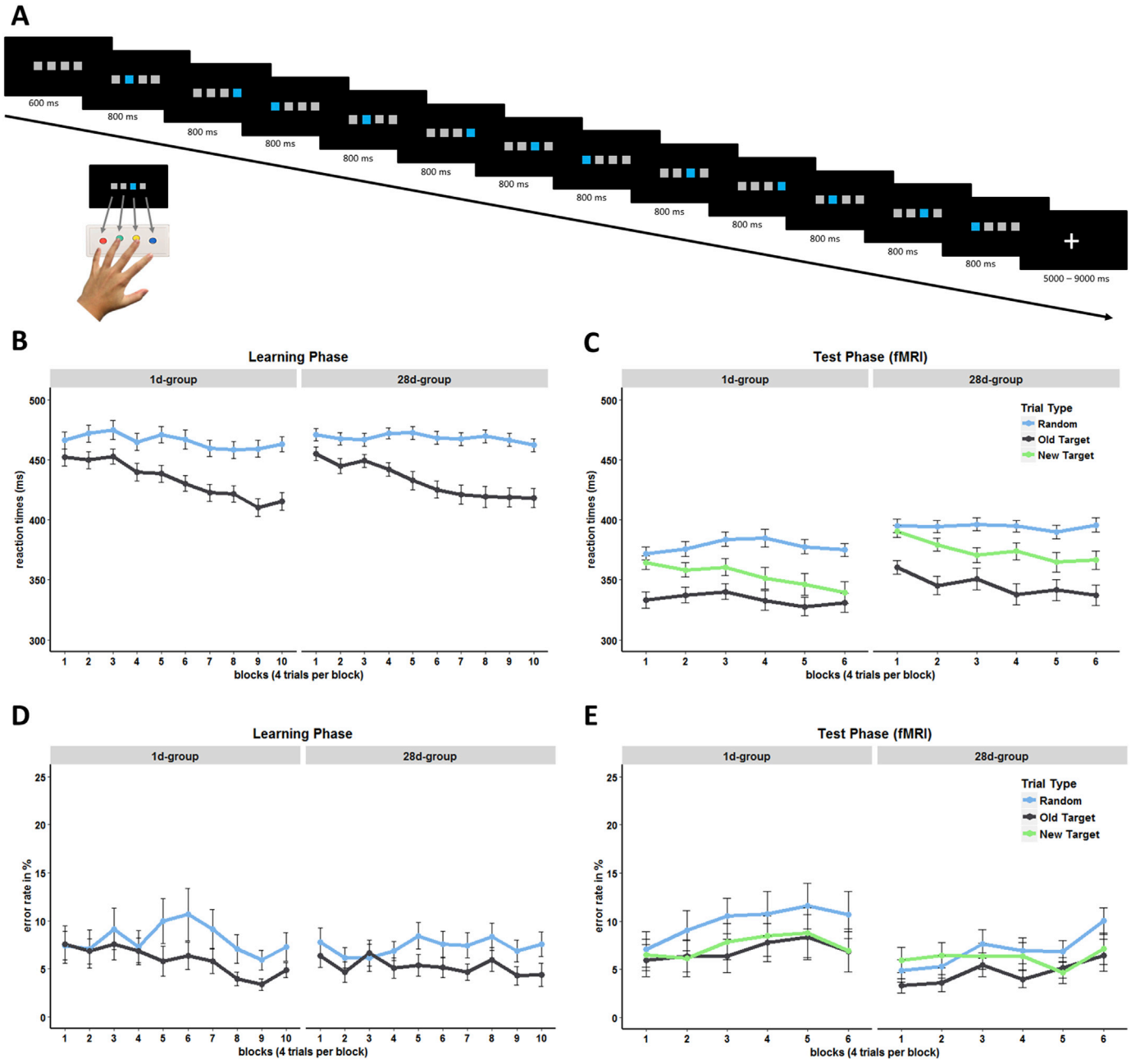


Fig. 1. (A) Schematic overview of one trial of the serial reaction time task (SRTT). First, four grey horizontal squares appeared on the screen for 600 ms to signal the start of a new trial. Then a sequence of 12 visual cues followed (800 ms each), with one of the squares turning blue each time to cue a corresponding finger press on a four-button response box using the left hand (see bottom left for visual cue-button-finger correspondence). Each finger was cued three times within one trial, but never twice in a row. A fixation cross was shown between trials. **(B) Reaction times in the learning phase.** Reaction times were faster for *old target* trials than for *random* trials and decreased throughout the blocks in both groups. Data from the 1d-group is presented in the left panel and data from the 28d-group in the right panel. The learning phase consisted of 10 blocks, with four trials per block for each trial type (*random* vs. *old target*), respectively. Error bars represent SEM. **(C) Reaction times in the test phase.** Reaction times for *old target* trials were, both for the 1d-group and for the 28d-group, faster than reaction times for both *random* and *new target* trials, in particular at the beginning of the test phase. Data from the 1d-group is presented in the left panel and data from the 28d-group in the right panel. The test phase consisted of 6 blocks, with four trials per block for each trial type (*random* vs. *old target* vs. *new target*) respectively. Error bars represent SEM. **(D) Error Rates in the learning phase**. We found a main effect of Trial Type with more errors in *random* trials than in *old target* trials, but no effects involving the factors Block or Group. Data from the 1d-group is presented in the left panel and data from the 28d-group in the right panel. The learning phase consisted of 10 blocks, with four trials per block for each trial type (*random* vs. *old target*) respectively. Error bars represent SEM. **(E) Error Rates in the test phase**. We found a main effect of Trial Type and a main effect of Block, but no interaction and no effects involving the factor Group. Data from the 1d-group is presented in the left panel and data from the 28d-group in the right panel. The test phase consisted of 6 blocks, with four trials per block for each trial type (*random* vs. *old target* vs. *new target*) respectively. Error bars represent SEM.

square was associated with a respective finger of the non-dominant (i.e., left) hand, placed on a response box (left square/button = small finger, second from left square/button = ring finger, second from right square/button = middle finger, right square/button = index finger). After 600 ms, one of the squares turned blue prompting participants to press the corresponding button on the response box with the corresponding finger. After 800 ms, a different square turned blue, cuing the next button press and so forth, until all twelve cues were presented. Within each trial, each of the four buttons was cued three times, with the same button never being cued twice in a row. We used a four-button fiber optic response box (Current Designs, Inc., Philadelphia, PA, USA) in both the learning phase outside the scanner and the test phase inside the scanner. The visual cues were presented and the button presses recorded using MATLAB and the Psychtoolbox (Brainard, 1997). Participants were instructed to press the correct buttons as quickly and as precisely as possible. They did not receive feedback about their performance during the task, neither during the learning nor during the test phase.

2.3.1. Learning phase

For each participant, an individual target sequence, consisting of 12 button presses, was generated at the beginning of the learning phase and unbeknownst to the participant this sequence was presented 40 times throughout the learning phase of the experiment, pseudo-randomly intermixed with 40 trials containing *random* sequences. The learning phase consisted of ten blocks in total, with four *target* trials and four *random* trials presented in random order in each block. This procedure assured that each trial type (*target* vs *random*) occurred at a similar rate throughout the learning phase. The blocks were presented directly one after another, without a break in between and without participants noticing a transition between blocks. Between trials a fixation cross was presented for 5 ± 2 s. Three random practice trials were presented before the learning phase to make sure participants understood the visual cue to finger correspondence.

2.3.2. Test phase

For the test phase we reused the *target* sequence generated for the respective participant on the first day, now referred to as *old target* sequence, and also generated a second *new target* sequence for each participant, again consisting of 12 button presses. The introduction of the *new target* sequence allowed us to differentiate between memory for the *old target* sequence and new learning processes during the test phase, which were likely to occur for both the *new target* and the *old target* sequence. Throughout the test phase the *old target* sequence was presented 24 times, pseudo-randomly intermixed with 24 *new target* sequences and 24 *random* sequences. The test phase consisted of six blocks in total, with four *old target* trials, four *new target* trials and four *random* trials presented in random order in each block. This procedure assured that each trial type (*old target* vs *new target* vs *random*) occurred at a similar rate throughout the test phase. The blocks were presented directly one after another, without a break in between and without participants noticing a transition between blocks. Between trials a fixation cross was presented for 7 ± 2 s.

2.4. Assessment of explicit awareness

After the test phase, we asked participants to reproduce the two visual cue sequences (*old target* and *new target*) that they had been presented with in the test phase by marking the respective presumed position of each visual cue on a paper sheet containing 12 rows of the horizontal grey squares array (one row for each cue in the sequence). We then compared these two reported sequences with both the *old target* and the *new target* sequence and noted the highest number of correctly recalled consecutive cues per sequence. Tzvi et al. (2015) used Monte Carlo simulations to find the number of consecutive correct hits that can be seen as "above chance level" and therefore interpreted as explicit awareness. In accordance with these simulations, we classified participants who correctly recalled six or more consecutive cues within a sequence as having gained explicit awareness of the respective sequence. In retrospect, ten participants were explicitly aware of at least one of the two target sequences, with three of these participants even recalling both target sequences correctly. We assessed if there were systematic differences in explicit awareness between the two groups (1d vs 28d) to make sure that potential group differences could not be explained by differences in explicit awareness. In retrospect, there were no significant differences in the number of participants with explicit awareness of either the *old target* sequence (4 out of 18 in the 1d group and 4 out of 21 in the 28d group, $p = 1$, Fisher's exact test) or the *new target* sequence (2 out of 18 in the 1d group and 3 out of 21 in the 28d group, $p = 1$, Fisher's exact test) between the two groups. Nevertheless, we further reanalyzed the data excluding the 10 participants with explicit awareness (leaving a sample of $n = 29$; 13 in the 1d-group and 16 in the 28d-group), to examine if the main results obtained in the analysis with all participants ($n = 39$) hold for the sample including only participants with implicit learning and memory ($n = 29$).

2.5. Assessment of patterns in target sequences

As we used different target sequences for each participant, we also assessed differences in target sequence difficulty, expressed by the potential occurrence of systematic patterns in the sequences. We checked for group differences and effects of the patterns on the behavioral results and could show that the behavioral effects were not driven by potential differences in target sequence difficulty (See Supplementary Results for a more detailed description of these analyses).

2.6. Behavioral data analysis

Behavioral data analyses were performed in R version 3.3.2 (https://www.r-project.org/). In case of a violation of the sphericity assumption in the ANOVAs, Greenhouse-Geisser corrections were applied.

2.6.1. Learning phase and test phase reaction times analyses

We analyzed the mean reaction time of all correct responses within a trial as our main behavioral parameter. For the learning phase, we subjected the reaction times to a mixed ANOVA with Block (10 blocks) and Trial Type (*target* vs *random*) as within-subjects factors and Group (1d vs 28d) as between-subjects factor. Likewise, for the test phase, we analyzed the reaction times by means of mixed ANOVAs with Block (6 blocks) and Trial Type (*old target* vs *new target* vs *random*) as within-subjects factors and again Group (1d vs 28d) as between-subjects factor.

2.6.2. Learning phase and test phase error rates analyses

In addition to reaction times, we analyzed the error rates defined as the percentage of incorrect and missing button presses within a trial. Error rates were subjected to the same ANOVAs as reaction times. Note that error rates overall were very low throughout the learning phase ($6.60\% \pm 0.22$ SEM) and the test phase ($6.92\% \pm 0.28$ SEM; i.e. in average less than one incorrect/missing button press per trial as 1 out of 12 would be 8.33%).

2.6.3. Differences between learning phase and test phase do not allow assessment of off-line learning

Previous studies have shown that, in addition to the fast online learning during the learning phase of a SRTT, there are also off-line learning effects, reflected in a decrease of reaction times for target sequences after learning even without further practice (Debas et al., 2014; Press et al., 2005). However, a direct assessment of off-line gains in our design was hardly possible due to differences in the experimental setup between our learning and test phases (different context and additional target sequences during the test session).

2.7. MRI acquisition

Whole-brain fMRI data were acquired on a 3T Skyra scanner (Siemens) equipped with a 32-channel head coil. For the functional images we used an echoplanar imaging (EPI) sequence, acquiring 513 vol with the following parameters: 36 slices, slice thickness = 3 mm, distance factor 20%, repetition time (TR) = 2500 ms, echo time (TE) = 30 ms, voxel size 3.0 mm isotropic. We also acquired a high-resolution T1-weighted anatomical image (TR = 2.5s, TE = 2.12 ms, 256 slices, voxel size = 0.8 × 0.8 × 0.9 mm) and a magnetic (B0) field map to unwarp the functional images.

2.8. fMRI data preprocessing

The fMRI data were preprocessed using MATLAB and SPM12 (http://www.fil.ion.ucl.ac.uk/spm/). The first four functional images (10 s) were discarded to allow for T1 equilibration. The remaining 509 functional images were first spatially realigned and unwarped using the field maps (utilizing the SPM *Calculate VDM* and SPM *Realign & Unwarp* modules), then coregistered to the structural image, followed by a normalization to the MNI space. For the univariate analysis, the images were additionally spatially smoothed using an 8 mm full-width half-maximum Gaussian kernel.

2.9. General linear modeling and whole brain analysis

The data was analyzed using general linear modeling (GLM), in the form of a random effects analysis, as implemented in SPM12.

2.9.1. First level modelling in SPM

For each of the six blocks in the test phase we created regressors for each Trial Type (*old target, new target, random*), resulting in 18 regressors (6 Blocks × 3 Trial Types). Thereby each regressor contained 4 trials and each trial was modeled as a short block, using the time the first square turned blue as the onset of the block and 9.6 s as duration, thus equal to the total trial length (12 cues × 800 ms presentation of each cue). We therefore modelled all 72 trials in the test phase (4 trials per regressor x 18 regressors = 72 trials). Because the error rate was overall very low (mean error rate 6.92% ± 0.28 SEM per trial) and trials were analyzed as 'mini-blocks' of 12 button presses each, we did not exclude trials containing a single erroneous button press. The division of the trials into the six blocks was implemented to test new learning across the test phase, utilizing the fact that the task was programmed in a pseudo-random way, ensuring that each trial type occurred at a similar rate throughout the test phase. Although including the blocks in the analysis enabled us to examine changes within the test phase due to new learning, it should nevertheless be noted that alternative analyses examining the modulation of brain activity at the trial level instead of blocks may have resulted in a different pattern of results.

Note that we did not include movement regressors in the GLM, as we used the SPM unwarp function in the data preprocessing instead. A high-pass filter of 128 s was used to remove low-frequency drifts and serial correlations in the time series were accounted for using an autoregressive AR(1) model.

2.9.2. Second level modelling in SPM

We first looked at the overall task activity across all subjects (using the first-level main contrast comparing all trial types to the implicit baseline) to verify that the expected motor areas are active when performing the SRTT in general. We then analyzed the following difference contrasts from the first level across all participants: *old target* > *random*, *new target* > *random* and *old target* > *new target*. To take new learning processes during the test phase into account, we additionally examined the main effect of Block (six blocks), and the interaction effect of Trial Type (*old target, new target, random*) x Block by setting up the relevant t-contrasts on the first-level (5 contrast for the main effect of block, 10 for the interaction effect Trial Type x Block; created using the spm_make_contrasts function). Using a two-sample *t*-test design at second level modelling we then compared differences between the 1d-group and 28d-group in all the above contrasts.

2.9.3. Classical and Bayesian estimation methods

To estimate the second level models we used both the classical (Restricted Maximum Likelihood) estimation and, in a second step, the Bayesian estimation of second level models as implemented in SPM12. For the results of the classical estimation method, we first used a voxel wise threshold of $p < 0.05$ (family-wise error correction (FWE)) at the whole brain-level and, if no voxels survived this threshold, used the less conservative FWE cluster-thresholding implemented in SPM12. Importantly, we used a high cluster forming threshold of $p = 0.001$ combined with a FWE corrected cluster extent threshold (and not an arbitrary extent threshold such as e.g. $k = 10$). These thresholds have been shown to be sufficient (Flandin and Friston, 2017) to avoid the otherwise inflated false-positive rates (Eklund et al., 2016).

The Bayesian second level estimation of the models as implemented in SPM12 uses an Empirical Bayes algorithm with global shrinkage priors, which embodies a prior belief that on average across all voxels there is no experimental effect (see SPM12 documentation). For this Bayesian second level analysis, we used an effect size threshold of 0.5 (medium effect size) and a log odds bayes factor threshold of logBF = 5 (very strong evidence (Han and Park, 2018)) for the posterior probability maps (PPMs) to test for group differences in the overall task activity and differences in the activity for the *old target* sequence, the *new target* sequence and the *random* sequence separately. We also computed evidence for the null hypothesis (no group differences) by first defining a F-contrast with the vector [1 -1] then using the Bayesian estimation method to create a posterior probability map using a logBF of 3 and then using the *spm_bms_test_null* function to create a map containing the log bayes factors for the null hypothesis.

2.10. ROI analysis

In addition to the whole-brain analysis, we specifically looked at regions that have previously been discussed as being involved in motor sequence learning (Hardwick et al., 2013). We used the following anatomical masks from the Harvard-Oxford atlas (Desikan et al., 2006) using a probability threshold of 50%: caudate nucleus, putamen, pallidum, thalamus, hippocampus and superior parietal lobule (left and right respectively), juxtapositional lobule (formerly supplementary motor cortex) and anterior cingulate gyrus. We also used the following anatomical masks from the Juelich atlas (Eickhoff et al., 2007), again using a probability threshold of 50%: Primary motor cortex BA4a, Primary motor cortex BA4p, Primary somatosensory cortex BA1, Primary somatosensory cortex BA2, Primary somatosensory cortex BA3a, Primary somatosensory cortex BA3b and Premotor cortex (left and right respectively). In addition, we used masks created with MARINA (http://www.bion.de/eng/MARINA.php) for the left and right dorsolateral prefrontal cortex (dlPFC) and a cerebellum mask retrieved from the Diedrichsen Lab (http://www.diedrichsenlab.org/imaging/propatlas.htm). We performed small volume correction for these ROIs to test whether there were differences between trial types (*old target* > *random*, *new target* > *random* and *old target* > *new target*) or differences between the groups (1d vs 28d) that could be detected when correcting the FWE-threshold for the number of voxels in the ROIs in comparison to the whole brain. Thus, voxels were regarded as significant when falling below a corrected voxel threshold of 0.05 (FWE) adjusted for the small volume. We report all areas with $k > 10$ significant voxels.

2.11. Searchlight representational similarity analysis

2.11.1. Searchlight approach

In addition to the univariate analysis, we performed a whole-brain

searchlight Representational Similarity Analysis (RSA; Kriegeskorte et al., 2008) using the *rsatoolbox* (Nili et al., 2014). For each participant, we extracted a representational dissimilarity matrix (RDM) for each spherical searchlight (15 mm radius), based on the t-maps of the 18 regressors (6 blocks × 3 trial types) from a univariate GLM estimated on unsmoothed, normalized functional images. In each sphere, we extracted the activity pattern across the sphere for each regressor and calculated the dissimilarity between two activity patterns by correlation distances (1−r, Pearson linear correlation). Next, these dissimilarities based on each combination of the 18 regressors were placed into the respective cells of the 18 × 18 RDMs. Then, these brain RDMs were compared to one of six model RDMs (see Section 2.11.2 for description of the models) in each sphere using Spearman's rank correlation coefficient r. The resulting r estimates were assigned to the center voxels of each sphere, thus creating a whole-brain map of model fits for each model and subject. For each model, we then used the r-maps of each subject as input to a Second-Level SPM estimation to find main effects across the whole group of subjects and differences between the model fits of the two groups (1d vs 28d).

2.11.2. Model RDMs

We performed the searchlight RSA for six different models: (1) the *Old Target and New Target Stable* model expects two different activity patterns, one for all regressors based on *old target* sequence and one for all regressors based on *new target* sequence, that are dissimilar from each other, but stable across the blocks; (2) the *Old Target Stable and New Target Evolving* model expects similar activity patterns for all regressors based on *old target* trials from the beginning on, and a separate pattern for *new target* trials that slowly emerges across the blocks, i.e. we modeled a linear increase in pattern similarity across the blocks for the *new target* sequences reflecting a learning process; (3) the *Old Target and New Target Evolving* model expects two different activity patterns for the *old target* and the *new target* respectively, both becoming more and more consistent across the blocks; (4) the *Old Target Distinct* model expects similar activity patterns for all regressors based on *old target* trials, but no consistent patterns for *new target* trials; (5) the *New Target Distinct* model expects similar activity patterns for all regressors based on *new target* trials, but no consistent patterns for *old target* trials; (6) the *Old Target and New Target Same* model expects similar activity patterns for all regressors based on either *old target* trials or *new target* trials. None of the models expect consistent patterns for *random* trials. We hypothesized that the first three models (the *Old Target and New Target Stable; Old Target Stable and New Target Evolving; Old Target and New Target Evolving)* will all capture general differences between the representations of previously learned old motor memories (after consolidation) and newly learned motor memories (before consolidation). Differences between the 1d-group and the 28d-group in these models can then be related to differences between representations of recent motor memories after early systems consolidation and representations of remote motor memories after prolonged systems consolidation in comparison to newly learned motor memories, respectively. We expected model 2 (*Old Target Stable and New Target Evolving)* to show better model fits in areas that additionally represented new learning of the new sequence across the test phase, while model 3 (*Old Target and New Target Evolving)* was expected to show a better model fit in areas representing new learning of both sequences across the test phase. Models 4 (*Old Target Distinct)*, 5 (*New Target Distinct)* and 6 (*Old Target and New Target Same)* were included as control models, to show areas that only represented the old memories (4), the newly learned memories (5), or had the same representation for both sequences (6).

3. Results

3.1. Successful motor sequence learning

On a first experimental day, participants completed ten blocks of a SRTT (Nissen and Bullemer, 1987; Tzvi et al., 2015) in which they were requested to respond to changing visual cues with a corresponding button press (Fig. 1A). Unbeknownst to the participants, a specific sequence of cues (*target* sequence) was presented repeatedly, enabling motor learning. As expected, participants showed significantly faster reaction times in *target* trials than in *random* trials in the learning phase (main effect Trial Type: $F_{(1,37)} = 47.81$, $p = 3.68e\text{-}08$, generalized $\eta^2 = 0.095$) and a significant decrease of reaction times across the ten learning blocks in *target* trials (main effect Block for *target* trials: $F_{(3.94,149.77)} = 15.06$, $p = 2.86e\text{-}10$, generalized $\eta^2 = 0.055$), but not in *random* trials (main effect Block for *random* trials: $F_{(9,342)} = 1.20$, $p = 0.2968$, generalized $\eta^2 = 0.004$; interaction effect Block × Trial Type: $F_{(4.53,167.65)} = 7.82$, $p = 3.00e\text{-}06$, generalized $\eta^2 = 0.011$, Fig. 1B). Post-hoc paired t-tests comparing *target* trials to *random* trials in each block were significant in all blocks (all $t > 3.93$, all $p < 0.0003$), showing that the reaction times were faster for target compared to *random* sequences already in the first block. Thus, participants show very early gains in reaction times for the target sequence (after only 4 repetitions of the sequence) comparable to steep learning curves shown in other studies using similar tasks (Tzvi et al., 2015; Wymbs and Grafton, 2015). However, Fig. 1B, additionally shows that the difference between the *random* and *target* sequence still increased further across the blocks, as the reaction times for the *target* sequence consistently decreased. In the last block of the learning phase, the mean difference between *random* and *target* trials was significantly larger than in the first block (47 ms vs 17 ms; paired *t*-test comparing block 1 to block 10: $t(38) = -4.36$, $p = 9.58e\text{-}05$, Cohen's $d = 0.70$). Importantly, both groups learned the *target* sequence equally well (main effect Group: $F_{(1,37)} = 0.005$, $p = 0.9444$, generalized $\eta^2 = 0.0001$, no significant interaction effects involving Group: all $F < 0.68$, all $p > 0.727$).

The analysis of the error rates in the learning phase also showed a main effect of Trial Type ($F_{(1,37)} = 24.49$, $p = 1.65e\text{-}05$, generalized $\eta 2 = 0.030$), with more errors for *random* trials than for *target* trials and no differences between the groups (main effect Group: $F_{(1,37)} = 0.27$, $p = 0.6093$, generalized $\eta^2 = 0.003$, Group × Trial Type: $F_{(1,37)} = 0.03$, $p = 0.8563$, generalized $\eta^2 = 0.00004$). There was only a trend for main effect of Block ($F_{(5.78,214.00)} = 1.83$, $p = 0.0969$, generalized $\eta^2 = 0.014$) and no interaction effects involving Block (all $F < 1.63$, all $p > 0.1370$), showing that the error rates stayed relatively constant throughout the learning phase (Fig. 1D). In sum, these data show that participants learned the target sequence very well on day 1, without any differences between the 1d- and 28d-groups.

3.2. Intact motor sequence-specific memory after 28 days

In the test phase, we introduced a *new target* sequence in addition to the *old target* sequence that was learned on day 1 in order to differentiate new sequence learning during the test phase from sequence-specific motor memory. The ANOVA showed a significant main effect of Trial Type ($F_{(2,74)} = 30.24$, $p = 2.52e\text{-}10$, generalized $\eta^2 = 0.105$), and a Block × Trial Type interaction ($F_{(5.83,215.69)} = 3.03$, $p = 0.0079$, generalized $\eta^2 = 0.005$; main effect Block: $F_{(2.40,88.94)} = 3.17$, $p = 0.0381$, generalized $\eta^2 = 0.006$). Most importantly, however, there were no interaction effects including the factor Group (all $F < 0.90$, all $p > 0.4247$) and no main effect of Group ($F_{(1,37)} = 1.09$, $p = 0.3017$, generalized $\eta^2 = 0.021$). Examining Fig. 1C shows that, for both groups, the reaction times for the *random* sequence remained rather constant across the test phase (main effect Block for *random* trials: $F_{(5,190)} = 1.05$, $p = 0.3899$, generalized $\eta^2 = 0.004$), while we found the expected learning effect for the *new target* sequence with reaction times decreasing across blocks (main effect Block for *new target* trials: $F_{(1.95,74.14)} = 4.41$, $p = 0.0162$, generalized $\eta^2 = 0.018$). Most importantly, however, the reaction times for the *old target* sequence that was trained on day 1 were consistently low throughout the testing phase. In particular in the first blocks the reaction times for the *old target* sequence were faster than not only the reaction times for the *random* sequence but also faster than the reaction times for the *new target* sequence, showing the expected

sequence-specific memory effect. Bonferroni-corrected post-hoc paired t-tests across both groups (0.05/18 tests = 0.0028 as significance threshold) showed that, in the first and second block, reaction times in the *old target* trials were significantly faster than in both the *random* trials (both $t > 5.98$, both $p < 0.0001$) and the *new target* trials (both $t > 4.21$, both $p < 0.0001$), while the *random* and *new target* trials did not differ significantly (both $t < 3.15$, both $p > 0.003$). In blocks four and five, the reaction times of *old target* trials were faster than the reaction times for *new target* trials and *random* trials, and the reaction times for *new target* trials were faster than the reaction times for *random* trials (all $t > 3.19$, all $p < 0.0028$). In blocks three and six, reaction times in *random* trials were significantly slower than both *old target* trials (both $t > 5.59$, both $p < 0.0001$) and *new target* trials (both $t > 3.58$, both $p < 0.0009$), while the *old target* and *new target* trials did not differ significantly (both $t < 3.03$, both $p > 0.004$). Although Fig. 1C suggests that across all trial types and blocks the reaction times in the 28d-group seemed to be somewhat slower than in the 1d-group, this difference was not significant (main effect of Group: $F_{(1,37)} = 1.09$, $p = 0.3017$, generalized $\eta^2 = 0.021$) and the pattern of results was clearly very similar in the 1d- and the 28d-groups. Thus, in both groups the reaction times were faster in the *old target* trials compared to both *random* and *new target* trials at the beginning of the test phase. This finding clearly shows intact sequence-specific motor memory in the 1d-group and the 28d-group.

For the error rates in the test phase, there was again a significant main effect of Trial Type ($F_{(2,74)} = 13.68$, $p = 8.79e\text{-}06$, generalized $\eta^2 = 0.023$), with more errors for *random* trials than for *old target* trials ($t(38) = -5.06$, $p = 1.09e\text{-}05$, Cohen's $d = -0.33$), and more errors for *random* trials than *new target* trials ($t(38) = -3.15$, $p = 0.0032$, Cohen's $d = -0.18$), and slightly less errors for *old target* trials than *new target* trials ($t(38) = 2.02$, $p = 0.0503$, Cohen's $d = 0.13$). There was also a main effect of Block ($F_{(3.84,142.04)} = 4.05$, $p = 4.37e\text{-}03$, generalized $\eta^2 = 0.014$), but no Trial Type × Block interaction ($F_{(6.77,250.56)} = 1.08$, $p = 0.3757$, generalized $\eta^2 = 0.005$). Importantly, there were again no differences between groups in the error rates (main effect Group: $F_{(1,37)} = 1.37$, $p = 0.2494$, generalized $\eta^2 = 0.023$, no interaction effects involving Group: all $F < 1.97$, all $p > 0.1051$, Fig. 1E).

3.3. Neural underpinnings of motor sequence learning and memory 1d and 28d after initial encoding

Our behavioral data from the test session on experimental day 2 suggest that sequence-specific motor memory was equally robust in participants tested one day after encoding and those tested 28 days after encoding. We next set out to determine whether the neural underpinnings of motor memory underwent time-dependent changes or not. In a first step, we analyzed overall task activity (for all trial types) against the implicit baseline across all participants to verify that overall the expected motor areas are involved during performance of the SRTT. This analysis showed, as expected, activity in areas implicated in motor sequence learning, including, for instance, the precentral gyrus, postcentral gyrus, caudate nucleus, putamen and cerebellum. Parts of the medial temporal lobe and frontal pole, in turn, showed a significant decrease in activation (compared to the implicit baseline) during the motor learning task ($p < 0.05$ (FWE); see Supplementary Fig. 1 for visualizations and Supplementary Table 1 for a complete list of increased and decreased activations). We then tested for differences between the trial types (*old target* > *random*, *new target* > *random* and *old target* > *new target*) across all participants at the whole-brain level and found higher activity for *old target* trials relative to *random* trials in the right caudate, the right putamen, the angular gyrus, the cuneal cortex and the frontal pole when applying a cluster forming threshold of $p < 0.001$ (unc.) and a cluster-extent of 254 voxels (see Supplementary Table 2 for exact coordinates and cluster extents). No areas survived this threshold for the *new target* > *random* and *old target* > *new target* contrasts. We then tested for effects related to the six different blocks within the testing phase and found a main effect of Block bilaterally in the lateral occipital cortex, inferior division (left: $k = 638$ voxels, t-value = 16.10, peak voxel: $x = -46$, $y = -70$ and $z = 2$; right: $k = 736$ voxels, t-value = 15.59, peak voxel: $x = 46$, $y = -68$ and $z = -2$) when applying a cluster forming threshold of $p < 0.001$ (unc.) and a cluster-extent of 165 voxels. However, no areas survived this threshold for the Trial Type × Block interaction. Next, we tested for group differences in all of the above described contrasts. These analyses, however, did not yield any significant voxels or clusters at the whole-brain level.

In a next step, we analyzed activity in a number of pre-defined ROIs that have been implicated in motor sequence memory before (see materials and method), using small volume correction for each ROI separately. Differences between *old target*, *new target*, and *random* sequences were obtained, across all participants, in the following regions (see Supplementary Table 3 for exact coordinates, p-values and number of voxels): we found higher activity for *old target* trials than *random* trials in the right caudate, left caudate, right putamen and left dlPFC, echoing the results of the whole-brain cluster-thresholded analysis for this contrast and suggesting that these areas are important for sequence-specific learning and memory. Further, we found higher activity for the *new target* trials than the *random* trials in the right caudate and the left hippocampus, suggesting that these areas are involved in new motor sequence learning. Finally, we found higher activity for *new target* trials than *old target* trials in the left primary somatosensory cortex BA2, suggesting that this area might be more involved in new motor sequence learning than in performing already trained motor sequences. Alternatively, this finding might also imply that the old sequence is already encoded more efficiently, leading to a decrease in activity, compared to the newly encoded sequence. Note, however, that after correcting for the number of ROIs used in these analyses, only the right caudate for the contrast *old target* > *random* remains significant. Most importantly, however, even before correcting for the number of ROIs in these small volume corrected analyses, we did not obtain evidence for any differences in brain activity between the 1d- and 28d-groups.

3.4. Bayesian analysis shows a higher probability of prefrontal involvement in motor sequence memory after 28 days

Our univariate analyses so far suggested that the neural underpinnings of motor memory are largely comparable in the 1d- and 28d-groups, implicating that the neural signature of motor memory remains, in contrast to episodic memory, stable over time (at least over 4 weeks). In order to explicitly test the evidence in favor of the null hypothesis (i.e. that there are no differences between the 1d- and 28d-groups) and the evidence in favor of the alternative hypothesis (i.e. that there are actually differences between the 1d- and 28d-groups), we run a Bayesian second level analysis. Bayesian second-level analysis has been shown to be more conservative than clusterwise FWE inference while being more sensitive than voxelwise FWE inference (Han and Park, 2018) and importantly also allows explicit testing of the null hypothesis. Using a logBF of 3 to create posterior probability maps for the null hypothesis we did, however, not find any evidence for the null hypothesis of comparable brain activity in the 1d- and 28d-groups, neither in the overall task activity nor in each trial type separately. On the contrary, when using an effect size threshold of 0.5 (medium effect size) and a log odds threshold of logBF = 5 (very strong evidence), we found that the probability that the 1d-group showed more activation than the 28d-group across all trial types was higher in the left post central gyrus (peak voxel: $x = -32$, $y = -38$, $z = 72$) and, conversely, the probability that the 28d-group showed more activation than the 1d-group across all trial types was higher in the bilateral middle frontal gyrus (peak voxel left: $x = -44$, $y = 26$, $z = 44$; peak voxel right: $x = 42$, $y = 34$, $z = 44$), the frontal pole (peak voxel: $x = -18$, $y = 68$, $z = 16$) and the occipital pole (peak voxel: $x = -14$, $y = -94$, $z = -2$; Fig. 2A). We then tested whether these group differences were related to new learning or memory processes and thus specific to certain trial types (i.e., specific to *old target* trials, *new target* trials and *random* trials). In *old target* trials, indicative of sequence-specific motor memory, there was a

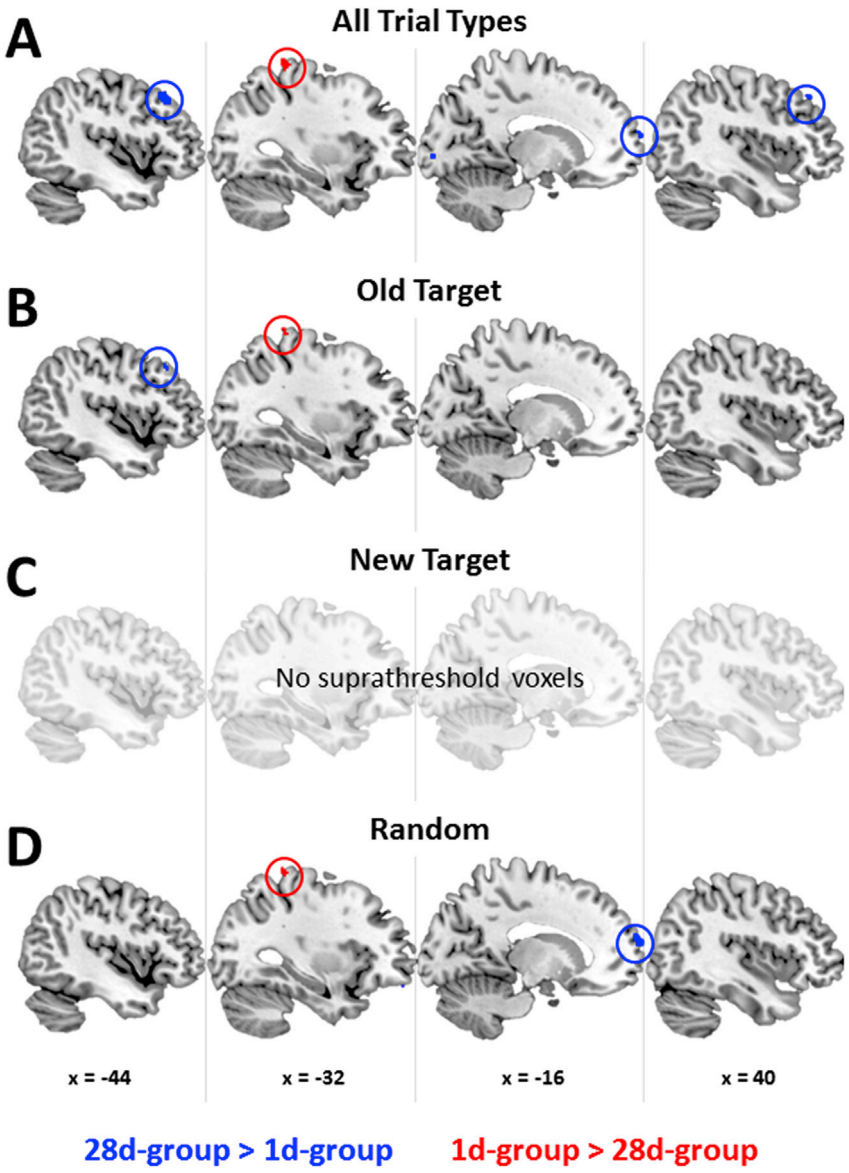


Fig. 2. Bayesian Second Level Group comparisons (n = 39). All voxels exceeding a threshold of logBF = 5 for PPMs created using an effect size threshold of 0.5 in the contrast 1d-group > 28d-group are coloured in red, and all voxels exceeding the same threshold in the contrast 28d-group > 1d-group are coloured in blue. Visualizations of these activations are superimposed on four sagittal slices (left hemisphere x = −44, −32, −16, right hemisphere x = 40) of a template image. Figures were created using MRIcron (https://www.nitrc.org/projects/mricron). **(A)** Group differences for the overall task activity (all trial types): the 1d-group showed a higher probability of activation compared to the 28d-group in the left postcentral gyrus (peak voxels: x = −32, y = −38, z = 72, k = 47; x = −30, y = −36, z = 50, k = 15), while there was a higher probability that the 28d-group showed more activation than the 1d-group in the left middle frontal gyrus (peak voxels: x = −44, y = 26, z = 44, k = 61; x = −32, y = 14, z = 66, k = 6), right middle frontal gyrus (peak voxel: x = 42, y = 34, z = 44, k = 13), occipital pole (peak voxel: x = - 14, y = −94, z = −2, k = 22) and frontal pole (peak voxel: x = −18, y = 68, z = 16, k = 12). Note that only areas with more than 5 voxels are explicitly listed here. **(B)** Group differences for *old target* trials: the 1d-group showed a higher probability of activation compared to the 28d-group in the left postcentral gyrus (peak voxel: x = −32, y = −38, z = 72, k = 11), while there was a higher probability that the 28d-group showed more activation than the 1d-group in the left middle frontal gyrus (peak voxel: x = −44, y = 26, z = 44, k = 7). **(C)** No voxels survived the threshold in the group comparisons of the *new target* trials. **(D)** Group differences for *random* trials: the 1d-group showed a higher probability of activation compared to the 28d-group in the left postcentral gyrus (peak voxel: x = −32, y = −38, z = 70, k = 16), while there was a higher probability that the 28d-group showed more activation than the 1d-group in the frontal pole (peak voxel: x = −16, y = 68, z = 16, k = 99) and the occipital pole (peak voxel: x = - 8, y = −92, z = −4, k = 8). Note that only areas with more than 5 voxels are explicitly listed here.

higher probability in the post central gyrus that the 1d-group showed more activation than the 28d-group (Fig. 2B). Yet, evidence for this difference (as well as a similar difference in the frontal pole and occipital pole) was also obtained for *random* trials (Fig. 2D), and thus seems to reflect differences between the groups in the general motor task rather than sequence-specific motor memory. Interestingly, however, there was selectively for *old target* trials a higher probability for activation in the left middle frontal gyrus in the 28d-group compared to the 1d-group, indicating that this area might be more strongly involved in sequence-specific memory after 28 days than after 1 day. In *new target* trials alone, there was no evidence for different activations in the two groups (Fig. 2C).

3.5. *Frontal cortex areas are more involved in separating between activity patterns for old sequence memories and patterns for newly learned sequences after 28 days than after 1 day*

The Bayesian Analysis showed a higher probability for the involvement of the middle frontal gyrus in motor memory after 28 days than 1 day after encoding. We also ran a searchlight RSA to find areas in the brain that show stable or evolving multivariate activity patterns for the *old* and/or *new target* sequence. In contrast to the mass univariate approach, the RSA allows the identification of information in the brain that is coded by patterns of activations in neighboring voxels. More specifically, we performed searchlight RSA comparing RDMs of each searchlight area to six different model RDMs reflecting different aspects of motor learning and memory: A first model reflected consistent multivariate representation of both the *old* learned sequence and the *new target* sequence throughout the test phase, reflecting differences between the two sequences without modeling any new learning processes across blocks (*Old Target and New Target Stable* model). The second model aimed to identify areas that show a consistent multivariate representation of the *old target* sequence throughout the test phase, reflecting an already established representation of the old sequence-specific motor memory, and a representation of the *new target* sequence that evolves over the course of the test phase, reflecting a new learning process (*Old Target Stable and New Target Evolving* model). Thus, this model reflects a representation of the motor memory of the target sequence learned on day 1, irrespective of any new learning processes for this *old target* sequence. The third model assumes also a re-learning process for the *old target* sequence, i.e. distinct representations for the *old target* and *new target* trials respectively that both become more consistent over the course of the test phase (*Old Target and New Target Evolving* model). The fourth

model assumed a specific and consistent multivariate representation of the *old target* sequence throughout the test phase that is clearly distinct from all other sequences, thereby not expecting any specific pattern for the *new target* sequence (*Old Target Distinct* model). The fifth model, on the other hand, assumed a specific and consistent multivariate representation of the *new target* sequence, without expecting any specific pattern for the *old target* sequence (*New Target Distinct* model). Finally, a sixth control model assumed consistent and indistinguishable multivariate representations for both target sequences (*Old Target and New Target Same* model), thereby not expecting the presence of sequence-specific information, but only a difference between re-occurring motor sequences and random sequences.

We first examined the main effects of model fits for each model, thereby searching for areas that showed, across all participants, a model fit that was significantly different from 0 (applying a $p < 0.05$ (FWE) threshold) and found that the first four models (*Old Target and New Target Stable* model, *Old Target Stable and New Target Evolving* model*, Old Target and New Target Evolving* model and *Old Target Distinct* model*)* all yielded positive results in a number of areas, including the putamen, thalamus, frontal pole, cerebellum, paracingulate gyrus, and temporal pole (Fig. 3A and Supplementary Tables 4–7 for complete lists of areas per model). However, for the *Old Target and New Target Same* model we found far less and smaller areas with significant model fits across all participants (Fig. 3A and Supplementary Table 8), while there were no areas with a significant model fit for the *New Target Distinct* model, demonstrating that a lot of areas involved in the task showed a specific multivariate activity pattern for the *old target* sequence that was distinct from the *new target* multivariate activity pattern. In a next step, we tested whether representations (i.e. model fits) were different between the 1d- and 28d-groups. When using a FWE cluster extent threshold (with a high cluster forming threshold of $p < 0.001$ unc.), we found two clusters, one in the right superior frontal gyrus (peak voxel: $x = 24$, $y = 20$, $z = 60$) and one in the right frontal pole extending into the middle frontal gyrus (peak voxel: $x = 38$, $y = 42$, $z = 42$) that showed significantly better model fits in the 28d-group in comparison to the 1d-group for the *Old Target and New Target Stable* model, *Old Target Stable and New Target Evolving* model and the *Old Target and New Target Evolving* model, i.e. those models that predict similar representational patterns for all regressors of the *old target* sequence and similar representational patterns for all regressors of the *new target* sequence respectively, that are, however, dissimilar from each other (see Fig. 3B and Supplementary Tables 9–11). Thus, these three models overlap to a certain extent and the reported clusters do not seem to differ between models that reflect an evolving learning process across the six blocks or a stable pattern across the six blocks. Importantly, however, we did not find group differences in these frontal cortex areas for the remaining three models (the *Old Target Distinct* model, the *New Target Distinct* model and the *Old Target and New Target Same* model), thus the decisive characteristic of a model that leads to differences between the 28d group and the 1d group in these frontal areas seems to be the separation between multivariate patterns for the old, consolidated motor sequence memory and patterns for a newly learned sequence. As this separation between the patterns of the two sequences seems to be higher after 28 days than after one day, one can assume that these frontal cortex areas are more involved in separating remote (4 week old) motor sequence memory representations from newly learned sequence representations than separating recent (1 day old) motor sequence memory representation from newly learned sequence representation. Notably, the cluster in the right frontal pole extending into the middle frontal gyrus included the right middle frontal area that showed a higher probability for an overall task-related involvement after 28d than after 1d in the univariate Bayesian analysis.

3.6. Results for subgroup with implicit awareness

In a last step, we examined whether the results obtained in the analysis with all participants ($n = 39$) held for a sample including only participants without explicit awareness of the sequences ($n = 29$; see methods section 2.4., Supplementary Figs. 1–3 and Supplementary Tables 12–19). Importantly, this analysis showed that the behavioral results remained largely comparable to those for the full sample (see Supplementary Fig. 2). In the Bayesian second level analysis, the results from the implicit sample showed a comparable pattern to that of the whole sample (Supplementary Fig. 3) for the overall task activity, although this time the higher probability of activation for the 28d-group in the frontal areas seemed to be especially reflected in *new target*trials and *random* trials. The search-light RSA results of the reduced sample, showed a comparable pattern to that of the whole sample (Supplementary Fig. 4), although the frontal pole clusters with a higher model fit in the 28d-group compared to the 1d-group (first three models) were located slightly inferior to those clusters found in the whole sample, yet all these clusters were located in the right frontal pole.

4. Discussion

Time-dependent neural reorganizations of hippocampus-dependent episodic or spatial memory have been a topic of intense scientific inquiry for decades (Dudai et al., 2015; Frankland and Bontempi, 2005; Moscovitch et al., 2016; Squire et al., 2015). Yet, if and how the neural representation of memories that are largely independent of the hippocampus changes over long time periods has received significantly less attention (Dayan and Cohen, 2011; Doyon et al., 2009). We tested here time-dependent changes in the neural representation of minimally trained motor sequence memories, known to rely mainly on cortico-striatal circuits (Dayan and Cohen, 2011; Doyon and Benali, 2005, but see Albouy et al., 2015; Albouy et al., 2013; Rose et al., 2011; Schendan et al., 2003 for evidence of an additional involvement of the hippocampus in implicit motor learning). Our behavioral findings show that even a single motor sequence learning session resulted in robust sequence-specific motor memory 28 days later. This motor memory was reflected in significantly faster reaction times for the target sequence learned during training compared to both a *random* sequence and a newly learned sequence during the test phase, which enabled us to separate actual motor memory from new learning during the test session. Our neuroimaging data are generally in line with previous findings implicating primarily cortico-striatal and cortico-cerebellar systems in motor learning tasks (Dayan and Cohen, 2011; Doyon et al., 2009; Doyon and Benali, 2005). Bayesian second-level fMRI analyses, which allowed us to directly test the evidence in favor and against the null hypothesis of similar motor memory-related activity after 1d and 28d, revealed a higher probability for overall task activity in the middle frontal gyrus and frontal pole 28d compared to 1d after initial motor learning. Further, searchlight RSA suggested that areas in the right middle frontal gyrus and frontal pole, including the area identified in the overall task-related Bayesian analysis, as well as areas in the right superior frontal gyrus are more involved in differentiating between multivariate activity patterns for old motor sequence memories and activity patterns of newly learned motor sequences in the 28d-group compared to the 1d-group. Together, these findings indicate time-dependent changes in the neural representation of motor sequence memory even without further training, in particular a stronger involvement of (lateral) prefrontal areas, mainly the middle frontal gyrus and the superior frontal gyrus, in differentiating remote motor sequence memories from new learning.

Although the middle frontal gyrus has been mainly linked to working memory (Leung et al., 2002; Ranganath et al., 2003) and inhibitory control processes (Garavan et al., 1999; Verbruggen and Logan, 2008), there is also some evidence for a role in motor learning and memory. Specifically, the middle frontal gyrus has been implicated in motor imagery (Decety et al., 1994) and in the initial acquisition of a motor skill (Shadmehr and Holcomb, 1997). Critically, however, the role of the middle frontal gyrus in the present study cannot be limited to the motor sequence learning process or the mere differentiation between two distinct motor sequences, as our findings show that the involvement of

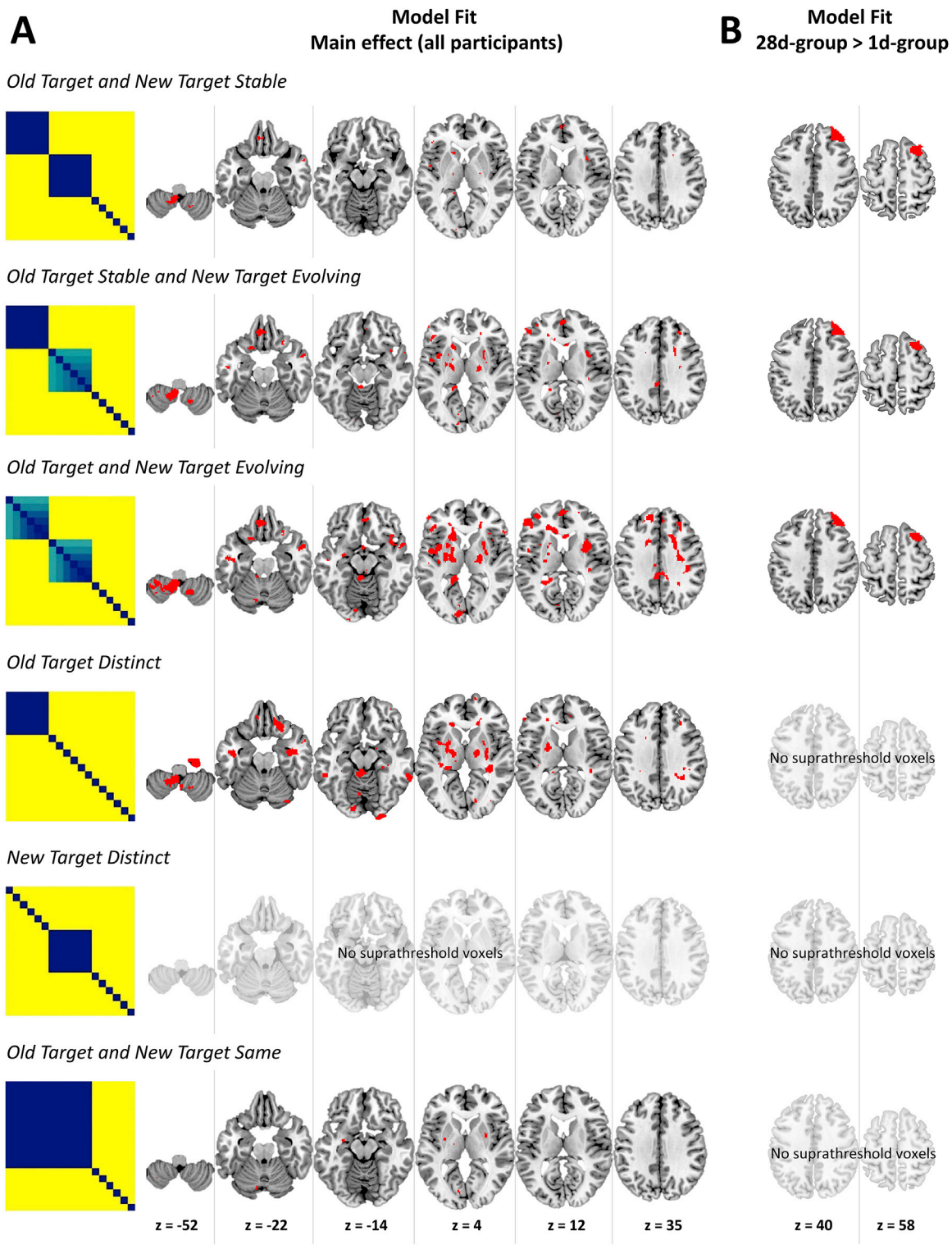


Fig. 3. Results from the RSA searchlight analysis (n = 39). **(A)** Main effects of the model fits (for all participants) for each of the six models separately: the *Old Target and New Target Stable* model in the first row, the *Old Target Stable and New Target Evolving* model in the second row, the *Old Target and New Target Evolving* model in the third row, the *Old Target Distinct* model in the fourth row, the *New Target Distinct* model in the fifth row and the *Old Target and New Target Same* model in the sixth row. All voxels surviving a threshold of $p < 0.05$ (FWE corrected) are coloured in red. For anatomical labels and coordinates see Supplementary Tables 4–8. **(B)** Group differences (28d-group > 1d-group) for each model. All voxels surviving a FWE cluster-threshold with a cluster forming threshold of $p = 0.001$ and a cluster-extent of 223 voxels (*Old Target and New Target Stable)*, 227 voxels (*Old Target Stable and New Target Evolving)* or 226 voxels (*Old Target and New Target Evolving*) are coloured in red. In right prefrontal areas, we obtained better model fits in the 28d-group compared to the 1d-group for the *Old Target and New Target Stable* model, the *Old Target Stable and New Target Evolving* model, and the *Old Target and New Target Evolving* model, but no suprathreshold voxels in the other three models. For anatomical labels and coordinates see Supplementary Tables 9–11. All visualizations are superimposed on axial slices of a template image. Figures were created using MRIcron (https://www.nitrc.org/projects/mricron).

the middle frontal gyrus (and superior frontal gyrus) in the differentiation of *old* and *new target* sequences was time-dependent. The separation of the old and new sequence representations was significantly more pronounced in the middle frontal gyrus (and superior frontal gyrus) after 28d than after 1d, indicating a time-dependent involvement of the middle frontal gyrus in the motor sequence memory representation.

The increased involvement of neocortical (prefrontal) areas across time is, in principle, in line with the proposed systems consolidation for hippocampus-dependent episodic memories (Dudai et al., 2015). However, for episodic memories, the time-dependent increase in the recruitment of the neocortex is thought to be accompanied by a decreased involvement and, ultimately, independence of the hippocampus (Frankland and Bontempi, 2005; Squire et al., 2015). Although we found decreased activity in the 28d-group compared to the 1d-group in the postcentral gyrus for the overall task activity, there were no areas in which the multivariate activity pattern for the *old target* sequence was more distinctly represented in the 1d-group than the 28d-group. In other words, the cortical, striatal, and cerebellar regions relevant for motor memory (and active in our analyses of overall task-related activity) after 1d appeared to be equally relevant for motor memory after 28d. Thus, our data suggest that there is no systems consolidation-like relocation of the motor sequence memories from some areas to others but instead an increased additional involvement of frontal areas in remote motor memories.

As noted above, several recent studies have suggested an additional involvement of the hippocampus during learning and early consolidation processes of a motor sequence task (Albouy et al., 2013, 2015; Rose et al., 2011; Schendan et al., 2003). We, however, found no increase in hippocampus activity during the overall performance of the task across both groups. On the contrary, some areas within the medial temporal lobe, including parts of the hippocampus, showed even reduced activation compared to the implicit baseline during performance of the task. Although, we did not focus on the involvement of the hippocampus in the present study and therefore cannot draw strong conclusions from these exploratory results, the reduced activation of the hippocampus during motor sequence learning may point to a competition between hippocampal areas, typically involved in episodic memory, and areas implicated in motor learning, such as the striatum (see also Poldrack et al., 2001; Poldrack and Packard, 2003). Future research is required to shed more light on the potential interaction of multiple memory systems during motor sequence learning. In addition, it is also important to note that our participants performed a declarative memory task before the motor sequence learning task and previous research suggested that declarative and procedural tasks performed one after another may influence each other (Keisler and Shadmehr, 2010; Robertson, 2012). Although group differences could hardly be explained by the prior performance of a declarative task because this task was performed by both groups in the same way, overall task-related activity (across groups) may have been influenced by the prior encoding task (Dandolo and Schwabe, 2018).

While we argue here that the representation of motor memories changes over time and that middle and superior frontal cortex areas represent, in addition to other cortico-striatal areas, remote motor memories, an alternative explanation might be that the increased involvement of these areas is due to an increased effort during the recall of the motor memories after 28d. Although lateral prefrontal areas have been associated with retrieval effort (Buckner and Wheeler, 2001; Henson et al., 1999), we consider this alternative rather unlikely. If participants had to show more effort to reproduce the learned sequence after 28d than after 1d, then this should be reflected in increased reaction times for the *old target* sequences in the 28 day group. Yet, we did not find a main effect of group or interaction effects including the factor group in the reaction time data. In addition, cognitive effort that is mediated by lateral prefrontal areas should be mainly relevant for the retrieval of explicit information. We used here, however, an implicit form of the SRTT and the vast majority of the participants were not explicitly aware of the *old target* sequence. Thus, retrieval effort may have been less relevant for most participants in the used task. Importantly, there were no differences between the 1d- and 28d-groups in terms of the number of participants that were explicitly aware of the target sequences and an additional analysis including only those participants without explicit awareness (see Supplementary Materials) showed mainly comparable results, although the area within the prefrontal cortex found in the RSA analyses was located more inferior in the sample of the participants without explicit awareness. However, we did not have enough participants to directly test whether the temporal dynamics are comparable for implicit vs. explicit motor sequence learning and future studies are needed to explicitly compare the areas involved in representing remote implicit or explicit motor memories. Overall, it should be noted that due to the necessary exclusion of nine participants (see Section 2.1.), the power of the study was lower than optimal, which makes it important for future research to run a replication study in a larger sample.

Furthermore, for episodic memories it has been suggested that the neuronal reorganization from hippocampus-dependent to more neocortical representations across time is accompanied by a transformation from rich, detailed memories to more semantic, gist-like representations (Moscovitch et al., 2016; Nadel et al., 2007; Winocur and Moscovitch, 2011). It would be very interesting to test whether the time-dependent changes in the neural representation of motor memories that we suggest here are also paralleled by a transformation in the nature or expression of motor sequence memories over time. Future studies might address this question by using new sequences in the test session that explicitly resemble the initially learned target sequence. Moreover, tracking the temporal dynamics of the reorganization of motor sequence memory representations is an important challenge for future research. We used here a 28d interval because earlier studies showed systems consolidation processes for episodic memories after several weeks (e.g. Takashima et al., 2006). However, the temporal profile of systems consolidation processes for motor memories may be different than what we know from episodic memory and testing parallels and differences between the reorganization of motor and episodic memories at different time intervals would be highly interesting. Finally, while our data showed more distinctive multivariate patterns for the target sequences in the 28d-group in the right, but not left, middle and superior frontal gyrus, it remains to be shown, for instance by employing brain stimulation techniques, whether the time-dependent involvement of lateral prefrontal cortex in motor memory is indeed lateralized.

In sum, we provide here evidence for time-dependent changes in the neural circuitry underlying motor sequence memory. More specifically, our findings show an increased involvement of lateral prefrontal cortex in the representation of remote compared to recent motor sequence memories. In contrast to the proposed systems consolidation of episodic or spatial memories (Dudai et al., 2015; Squire et al., 2015), however, the time-dependent increase in the involvement of neocortical areas after four weeks was not paralleled by a decrease in those cortico-striatal areas that supported initial motor learning. Instead, the additional recruitment of lateral prefrontal areas might contribute to a more distributed representation of remote motor memories. These findings provide insights into how memories beyond the hippocampus evolve over time.

Author contributions

L.C.D. collected the data, analyzed the data, and wrote the manuscript. L.S. conceived and designed the experiment, supervised the project, and wrote the manuscript.

Conflicts of interest

The authors declare no competing financial interests.

Acknowledgements

The authors thank Daniel Kutzner, Wiebke Schmidt, Sandra Erbach, Kristin Medel, and Amina Shah for assistance during data collection. This work was supported by the German Research Foundation (DFG; SCHW1357/12-1).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.04.051.

References

Albouy, G., Fogel, S., King, B.R., Laventure, S., Benali, H., Karni, A., Carrier, J., Robertson, E.M., Doyon, J., 2015. Maintaining vs. enhancing motor sequence memories: respective roles of striatal and hippocampal systems. Neuroimage 108, 423–434.

Albouy, G., King, B.R., Maquet, P., Doyon, J., 2013. Hippocampus and striatum: dynamics and interaction during acquisition and sleep-related motor sequence memory consolidation. Hippocampus 23, 985–1004.

Albouy, G., Sterpenich, V., Balteau, E., Vandewalle, G., Desseilles, M., Dang-Vu, T., Darsaud, A., Ruby, P., Luppi, P.H., Degueldre, C., Peigneux, P., Luxen, A., Maquet, P., 2008. Both the hippocampus and striatum are involved in consolidation of motor sequence memory. Neuron 58, 261–272.

Bonnici, H.M., Chadwick, M.J., Lutti, A., Hassabis, D., Weiskopf, N., Maguire, E.A., 2012. Detecting representations of recent and remote autobiographical memories in vmPFC and hippocampus. J. Neurosci. 32, 16982–16991.

Born, J., Wilhelm, I., 2012. System consolidation of memory during sleep. Psychol. Res. 76, 192–203.

Brainard, D.H., 1997. The psychophysics toolbox. Spatial Vis. 10, 433–436.

Brashers-Krug, T., Shadmehr, R., Bizzi, E., 1996. Consolidation in human motor memory. Nature 382, 252–255.

Buckner, R.L., Wheeler, M.E., 2001. The cognitive neuroscience of remembering. Nat. Rev. Neurosci. 2, 624–634.

Coynel, D., Marrelec, G., Perlbarg, V., Pelegrini-Issac, M., Van de Moortele, P.F., Ugurbil, K., Doyon, J., Benali, H., Lehericy, S., 2010. Dynamics of motor-related functional integration during motor sequence learning. Neuroimage 49, 759–766.

Dandolo, L.C., Schwabe, L., 2018. Time-dependent memory transformation along the hippocampal anterior-posterior axis. Nat. Commun. 9, 1205.

Dayan, E., Cohen, L.G., 2011. Neuroplasticity subserving motor skill learning. Neuron 72, 443–454.

Debas, K., Carrier, J., Barakat, M., Marrelec, G., Bellec, P., Hadj Tahar, A., Karni, A., Ungerleider, L.G., Benali, H., Doyon, J., 2014. Off-line consolidation of motor sequence learning results in greater integration within a cortico-striatal functional network. Neuroimage 99, 50–58.

Debas, K., Carrier, J., Orban, P., Barakat, M., Lungu, O., Vandewalle, G., Hadj Tahar, A., Bellec, P., Karni, A., Ungerleider, L.G., Benali, H., Doyon, J., 2010. Brain plasticity related to the consolidation of motor sequence learning and motor adaptation. Proc. Natl. Acad. Sci. U. S. A. 107, 17839–17844.

Decety, J., Perani, D., Jeannerod, M., Bettinardi, V., Tadary, B., Woods, R., Mazziotta, J.C., Fazio, F., 1994. Mapping motor representations with positron emission tomography. Nature 371, 600–602.

Desikan, R.S., Segonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31, 968–980.

Diekelmann, S., Wilhelm, I., Born, J., 2009. The whats and whens of sleep-dependent memory consolidation. Sleep Med. Rev. 13, 309–321.

Doyon, J., Bellec, P., Amsel, R., Penhune, V., Monchi, O., Carrier, J., Lehericy, S., Benali, H., 2009. Contributions of the basal ganglia and functionally related brain structures to motor learning. Behav. Brain Res. 199, 61–75.

Doyon, J., Benali, H., 2005. Reorganization and plasticity in the adult brain during learning of motor skills. Curr. Opin. Neurobiol. 15, 161–167.

Doyon, J., Penhune, V., Ungerleider, L.G., 2003. Distinct contribution of the cortico-striatal and cortico-cerebellar systems to motor skill learning. Neuropsychologia 41, 252–262.

Dudai, Y., Karni, A., Born, J., 2015. The consolidation and transformation of memory. Neuron 88, 20–32.

Eichenbaum, H., 1999. The hippocampus and mechanisms of declarative memory. Behav. Brain Res. 103, 123–133.

Eichenbaum, H., Cohen, N.J., 2001. From Conditioning to Conscious Recollection Memory Systems of the Brain. Oxford University Press, New York.

Eickhoff, S.B., Paus, T., Caspers, S., Grosbras, M.H., Evans, A.C., Zilles, K., Amunts, K., 2007. Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. Neuroimage 36, 511–521.

Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. Proc. Natl. Acad. Sci. U. S. A. 113, 7900–7905.

Flandin, G., Friston, K.J., 2017. Analysis of family-wise error rates in statistical parametric mapping using random field theory. Hum. Brain Mapp. 40, 2052–2054.

Floyer-Lea, A., Matthews, P.M., 2005. Distinguishable brain activation networks for short- and long-term motor skill learning. J. Neurophysiol. 94, 512–518.

Fogel, S., Albouy, G., King, B.R., Lungu, O., Vien, C., Bore, A., Pinsard, B., Benali, H., Carrier, J., Doyon, J., 2017. Reactivation or transformation? Motor memory consolidation associated with cerebral activation time-locked to sleep spindles. PLoS One 12, e0174755.

Frankland, P.W., Bontempi, B., 2005. The organization of recent and remote memories. Nat. Rev. Neurosci. 6, 119–130.

Furman, O., Mendelsohn, A., Dudai, Y., 2012. The episodic engram transformed: time reduces retrieval-related brain activity but correlates it with memory accuracy. Learn. Mem. 19, 575–587.

Garavan, H., Ross, T.J., Stein, E.A., 1999. Right hemispheric dominance of inhibitory control: an event-related functional MRI study. Proc. Natl. Acad. Sci. U. S. A. 96, 8301–8306.

Gilboa, A., Winocur, G., Grady, C.L., Hevenor, S.J., Moscovitch, M., 2004. Remembering our past: functional neuroanatomy of recollection of recent and very remote personal events. Cerebr. Cortex 14, 1214–1225.

Han, H., Park, J., 2018. Using SPM 12's second-level bayesian inference procedure for fMRI analysis: practical guidelines for end users. Front. Neuroinf. 12.

Hardwick, R.M., Rottschy, C., Miall, R.C., Eickhoff, S.B., 2013. A quantitative meta-analysis and review of motor learning in the human brain. Neuroimage 67, 283–297.

Henson, R.N., Shallice, T., Dolan, R.J., 1999. Right prefrontal cortex and episodic memory retrieval: a functional MRI test of the monitoring hypothesis. Brain 122 (Pt 7), 1367–1381.

Huang, Y., Zhen, Z., Song, Y., Zhu, Q., Wang, S., Liu, J., 2013. Motor training increases the stability of activation patterns in the primary motor cortex. PLoS One 8, e53555.

Julius, M.S., Adi-Japha, E., 2015. Learning of a simple grapho-motor task by young children and adults: similar acquisition but age-dependent retention. Front. Psychol. 6, 225.

Keisler, A., Shadmehr, R., 2010. A shared resource between declarative memory and motor memory. J. Neurosci. 30, 14817–14823.

Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2, 4.

Kupferschmidt, D.A., Juczewski, K., Cui, G., Johnson, K.A., Lovinger, D.M., 2017. Parallel, but dissociable, processing in discrete corticostriatal inputs encodes skill learning. Neuron 96, 476–489 e475.

Lehericy, S., Benali, H., Van de Moortele, P.F., Pelegrini-Issac, M., Waechter, T., Ugurbil, K., Doyon, J., 2005. Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. Proc. Natl. Acad. Sci. U. S. A. 102, 12566–12571.

Leung, H.C., Gore, J.C., Goldman-Rakic, P.S., 2002. Sustained mnemonic response in the human middle frontal gyrus during on-line storage of spatial memoranda. J. Cogn. Neurosci. 14, 659–671.

Moscovitch, M., Cabeza, R., Winocur, G., Nadel, L., 2016. Episodic memory and beyond: the Hippocampus and neocortex in transformation. Annu. Rev. Psychol. 67, 105–134.

Nadel, L., Winocur, G., Ryan, L., Moscovitch, M., 2007. Systems consolidation and hippocampus: two views. Debates in Neuroscience 1, 55–66.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., Kriegeskorte, N., 2014. A toolbox for representational similarity analysis. PLoS Comput. Biol. 10, e1003553.

Nissen, M.J., Bullemer, P., 1987. Attentional requirements of learning: evidence from performance measures. Cogn. Psychol. 19, 1–32.

O'Keefe, J., Nadel, L., 1978. The hippocampus as a Cognitive Map. Clarendon Press, Oxford.

Park, S.W., Sternad, D., 2015. Robust retention of individual sensorimotor skill after self-guided practice. J. Neurophysiol. 113, 2635–2645.

Penhune, V.B., Doyon, J., 2002. Dynamic cortical and subcortical networks in learning and delayed recall of timed motor sequences. J. Neurosci. 22, 1397–1406.

Poldrack, R.A., Clark, J., Paré-Blagoev, E.J., Shohamy, D., Creso Moyano, J., Myers, C., Gluck, M.A., 2001. Interactive memory systems in the human brain. Nature 414, 546–550.

Poldrack, R.A., Packard, M.G., 2003. Competition among multiple memory systems: converging evidence from animal and human brain studies. Neuropsychologia 41, 245–251.

Poldrack, R.A., Sabb, F.W., Foerde, K., Tom, S.M., Asarnow, R.F., Bookheimer, S.Y., Knowlton, B.J., 2005. The neural correlates of motor skill automaticity. J. Neurosci. 25, 5356–5364.

Press, D.Z., Casement, M.D., Pascual-Leone, A., Robertson, E.M., 2005. The time course of off-line motor sequence learning. Brain Res Cogn Brain Res 25, 375–378.

Ranganath, C., Johnson, M.K., D'Esposito, M., 2003. Prefrontal activity associated with working memory and episodic long-term memory. Neuropsychologia 41, 378–389.

Robertson, E.M., 2007. The serial reaction time task: implicit motor skill learning? J. Neurosci. 27, 10073–10075.

Robertson, E.M., 2012. New insights in human memory interference and consolidation. Curr. Biol. 22, R66–R71.

Robertson, E.M., Pascual-Leone, A., Miall, R.C., 2004a. Current concepts in procedural consolidation. Nat. Rev. Neurosci. 5, 576–582.

Robertson, E.M., Pascual-Leone, A., Press, D.Z., 2004b. Awareness modifies the skill-learning benefits of sleep. Curr. Biol. 14, 208–212.

Romano, J.C., Howard Jr., J.H., Howard, D.V., 2010. One-year retention of general and sequence-specific skills in a probabilistic, serial reaction time task. Memory 18, 427–441.

Rose, M., Haider, H., Salari, N., Buchel, C., 2011. Functional dissociation of hippocampal mechanism during implicit learning based on the domain of associations. J. Neurosci. 31, 13739–13745.

Savion-Lemieux, T., Penhune, V.B., 2005. The effects of practice and delay on motor skill learning and retention. Exp. Brain Res. 161, 423–431.

Schendan, H.E., Searl, M.M., Melrose, R.J., Stern, C.E., 2003. An FMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. Neuron 37, 1013–1025.
Shadmehr, R., Holcomb, H.H., 1997. Neural correlates of motor memory consolidation. Science 277, 821–825.
Squire, L.R., 2004. Memory systems of the brain: a brief history and current perspective. Neurobiol. Learn. Mem. 82, 171–177.
Squire, L.R., Alvarez, P., 1995. Retrograde amnesia and memory consolidation: a neurobiological perspective. Curr. Opin. Neurobiol. 5, 169–177.
Squire, L.R., Bayley, P.J., 2007. The neuroscience of remote memory. Curr. Opin. Neurobiol. 17, 185–196.
Squire, L.R., Genzel, L., Wixted, J.T., Morris, R.G., 2015. Memory consolidation. Cold Spring Harb Perspect Biol 7, a021766.
Squire, L.R., Zola-Morgan, S., 1991. The medial temporal lobe memory system. Science 253, 1380–1386.
Takashima, A., Petersson, K.M., Rutters, F., Tendolkar, I., Jensen, O., Zwarts, M.J., McNaughton, B.L., Fernández, G., 2006. Declarative memory consolidation in humans: a prospective functional magnetic resonance imaging study. Proc. Natl. Acad. Sci. U. S. A. 103, 756–761.
Tzvi, E., Stoldt, A., Witt, K., Kramer, U.M., 2015. Striatal-cerebellar networks mediate consolidation in a motor sequence learning task: an fMRI study using dynamic causal modelling. Neuroimage 122, 52–64.
Ungerleider, L.G., Doyon, J., Karni, A., 2002. Imaging brain plasticity during motor skill learning. Neurobiol. Learn. Mem. 78, 553–564.
Vahdat, S., Fogel, S., Benali, H., Doyon, J., 2017. Network-wide reorganization of procedural memory during NREM sleep revealed by fMRI. Elife 6.
Verbruggen, F., Logan, G.D., 2008. Response inhibition in the stop-signal paradigm. Trends Cognit. Sci. 12, 418–424.
White, N.M., Packard, M.G., McDonald, R.J., 2013. Dissociation of memory systems: the story unfolds. Behav. Neurosci. 127, 813–834.
Wiestler, T., Diedrichsen, J., 2013. Skill learning strengthens cortical representations of motor sequences. Elife 2, e00801.
Winocur, G., Moscovitch, M., 2011. Memory transformation and systems consolidation. J. Int. Neuropsychol. Soc. 17, 766–780.
Wymbs, N.F., Grafton, S.T., 2015. The human motor system supports sequence-specific representations over multiple training-dependent timescales. Cerebr. Cortex 25, 4213–4225.

Appendix C

Study III: Dorsolateral prefrontal cortex enables updating of established memories

Published in Cerebral Cortex

Cerebral Cortex, 2018; 1–15

doi: 10.1093/cercor/bhy298
Original Article

ORIGINAL ARTICLE

Dorsolateral Prefrontal Cortex Enables Updating of Established Memories

Lisa Marieke Kluen[1], Lisa Catherine Dandolo[1], Gerhard Jocham[2,3] and Lars Schwabe[1]

[1]Department of Cognitive Psychology, Institute of Psychology, University of Hamburg, 20146 Hamburg, Germany, [2]Institute of Experimental Psychology, Heinrich Heine University Düsseldorf, Germany and [3]Center for Behavioral Brain Sciences, Otto von Guericke University Magdeburg, Germany

Address correspondence to Lars Schwabe, Department of Cognitive Psychology, Institute of Psychology, University of Hamburg, Von-Melle-Park 5, 20146 Hamburg, Germany. Email: lars.schwabe@uni-hamburg.de

Abstract

Updating established memories in light of new information is fundamental for memory to guide future behavior. However, little is known about the brain mechanisms by which existing memories can be updated. Here, we combined functional magnetic resonance imaging and multivariate representational similarity analysis to elucidate the neural mechanisms underlying the updating of consolidated memories. To this end, participants first learned face–city name pairs. Twenty-four hours later, while lying in the MRI scanner, participants were required to update some of these associations, but not others, and to encode entirely new pairs. Updating success was tested again 24 h later. Our results showed increased activity of the dorsolateral prefrontal cortex (dlPFC) specifically during the updating of existing associations that was significantly stronger than when simple retrieval or new encoding was required. The updating-related activity of the dlPFC and its functional connectivity with the hippocampus were directly linked to updating success. Furthermore, neural similarity for updated items was markedly higher in the dlPFC and this increase in dlPFC neural similarity distinguished individuals with high updating performance from those with low updating performance. Together, these findings suggest a key role of the dlPFC, presumably in interaction with the hippocampus, in the updating of established memories.

Key words: hippocampus, memory, prefrontal cortex

Episodic memories allow us to mentally travel back in time and relive events from our past (Tulving 2002). Beyond remembering the past, these mnemonic representations support future survival. They enable us to imagine and simulate upcoming events (Schacter et al. 2007; Jing et al. 2017), to guide our attention and current decision-making (Chun and Turk-Browne 2007; Wimmer and Shohamy 2012). In order to accomplish these prospective functions, it is fundamental that memories are updated in light of new information. Indeed, it is increasingly acknowledged that memories are highly dynamic entities (Dudai 2012; Kroes and Fernandez 2012; Nadel et al. 2012; Schwabe et al. 2014) and there is considerable evidence that consolidated memories can be modified as a function of current experience (Baddeley and Dale 1966; Loftus 1975; Schiller et al. 2010; Zeithamova et al. 2012). However, although the updating of established memories is essential for our adaptation to changing environments, the neural mechanisms underlying the updating process of consolidated memories are not well understood.

A prime candidate that may contribute to the updating of established memories is the dorsolateral prefrontal cortex (dlPFC), a brain region that is thought to support relational encoding (Murray and Ranganath 2007; Blumenfeld et al. 2011) and strategic aspects of memory retrieval (Simons and Spiers

2003; Manenti et al. 2010). Decades of research have linked the dlPFC to cognitive control processes, such as monitoring and inhibition (Egner and Hirsch 2005; Cole and Schneider 2007), which are critical in the context of memory updating. Indeed, the dlPFC appears to play a causal role in the strengthening of memory after reactivation (Sandrini et al. 2013, 2014). Moreover, the dlPFC is directly connected to medial temporal lobe (MTL) areas, such as the hippocampus (Bilek et al. 2013; Preston and Eichenbaum 2013), that are crucial for memory formation and storage (Squire and Zola-Morgan 1991; Alvarez and Squire 1994). Through its interaction with the hippocampus, the dlPFC may orchestrate encoding and incorporation of updated information and suppress the reactivation of old memory representations (Anderson et al. 2004; Depue et al. 2007), thus representing a potential mechanism that enables updating of established memories.

Updating processes have recently been investigated in short-term memory (Kuhl et al. 2012; Schlichting and Preston 2016) and there is evidence for an important role of the dlPFC in working memory updating (D'Ardenne et al. 2012). However, the timescales of working memory processes and consolidated long-term memories are clearly distinct. While the dlPFC is known to be crucial for the maintenance of working memory representations (Fuster and Alexander 1971; Cohen et al. 1997), this is not the case for long-term memories, which are stored (at least transiently) in MTL areas (Squire and Zola-Morgan 1991; Burgess et al. 2002). It remains unclear to what extent updating processes in long-term memory resemble those in working memory and whether the dlPFC is implicated in the updating of established memories is completely unknown.

In the present study, we combined functional magnetic resonance imaging (fMRI) with multivariate representational similarity analysis (RSA) to elucidate the brain mechanisms supporting the updating of consolidated long-term memories. Participants were tested in a novel experimental paradigm on 3 consecutive days. On day 1, they learned a number of face–city name associations (Fig. 1A). Twenty-four hours later, some of the learned faces were paired with new cities, requiring participants to update the encoded associations. This updating phase was performed in the MRI scanner and included also face–city pairs that were not updated as well as entirely new face–city pairs, thus allowing us to control for simple retrieval as well as new learning, respectively. On day 3, memory updating success was assessed in a recognition test. We hypothesized that the dlPFC would be critically involved in successful memory updating, presumably in interaction with the hippocampus. In addition to the dlPFC and MTL areas, we focused on structures implicated in the representation of prior knowledge, such as the ventromedial prefrontal cortex (vmPFC) and angular gyrus, that have also been shown to be critically implicated in schema-based learning processes (van Kesteren et al. 2012; Wagner et al. 2015). Activity of these structures may therefore hamper memory updating.

Methods

Participants

We tested 49 healthy, right-handed adults in a 3-day study design, including an MRI scanning session and behavioral testing. One participant was excluded because of minimal performance in the task (more than 2 standard deviations below average performance in trial types where information did not change), thus leaving 48 participants (25 women; mean age 24.58 years, range: 19–32 years) for analyses. An a priori sample size calculation using G*POWER 3.19.2. showed that this sample size is sufficient to detect small to medium effects with a power of 0.95. Exclusion criteria were checked in a standardized telephone screening and comprised any current physical illnesses or medication intake, a lifetime history of any neurological or mental disorders, as well as any contraindications for MRI measurements, such as non-removable metal parts, pacemaker, pregnancy, or claustrophobia. All participants provided written informed consent before the beginning of testing and received a moderate monetary compensation. The study protocol was approved by the local ethics committee.

Stimulus Materials

In the memory-updating task, we used 100 pictures of faces and 400 well-known (non-German) city names. Pictures of faces were taken from the Radboud Faces Database (RaFD, Langner et al. 2010) and the Karolinska Directed Emotional Faces Database (KDEF, Lundqvist et al. 1998). Only Caucasian faces (50 men and 50 women) with a neutral expression were included. Pictures were resized (762 × 562 pixels) and formatted using Adobe Photoshop CS6 (64 bit) so that all faces had a white background. City names were checked for ease of spelling and possible similarity in their name with other city names. If a city was considered too difficult to spell or too similar to another city used, it was replaced by another city. The chosen city names were rated in an independent pilot study (n = 15) as to whether the city name was commonly known or not. The rating sale ranged from 0 ("never heard of") to 10 ("very well known"). Cities used in the task had an average familiarity rating of 6.1 and can thus be considered as well known.

Experimental Design and Procedure

To investigate the updating of already established memories, participants were tested in a novel memory-updating paradigm over the course of 3 consecutive days.

On experimental day 1, after arriving at the lab, participants were welcomed by the experimenter and informed about the general procedure, as well as the MRI procedure on the next day. Informed consent was obtained and any questions the participants had were answered. Participants then completed the "learning phase" (Fig. 1A) of the memory-updating paradigm, which took up to 100 min. They were informed that they would see a face of a person and next to it the name of this person's last vacation destination. The task of the participants was to memorize all face–city pairs they were presented with, 75 in total. Each pair was presented 4 times, each time for 3.5 s. In the first encoding run, participants were presented with 15 blocks of 5 face–city pairs and performed a cued recall test after each block for the 5 pairs shown. During the cued recall, they saw the face and had to type in the name of the city. Participants had to start typing in the word within the first 5 s of its presentation, after which they had enough time to finish typing the word. After their response, the correct city name was presented in green (until termination by the participant) above the city name typed in by the participant. In the second encoding run, participants saw 5 blocks of 15 face–city pairs and in the third run 3 blocks of 25 face–city pairs. In both runs, again each block was followed by a cued recall test for the pairs presented. Finally, in the fourth encoding run, all 75 pairs were presented one after the other, which was followed by a cued recall test for all face–city pairs. This encoding procedure

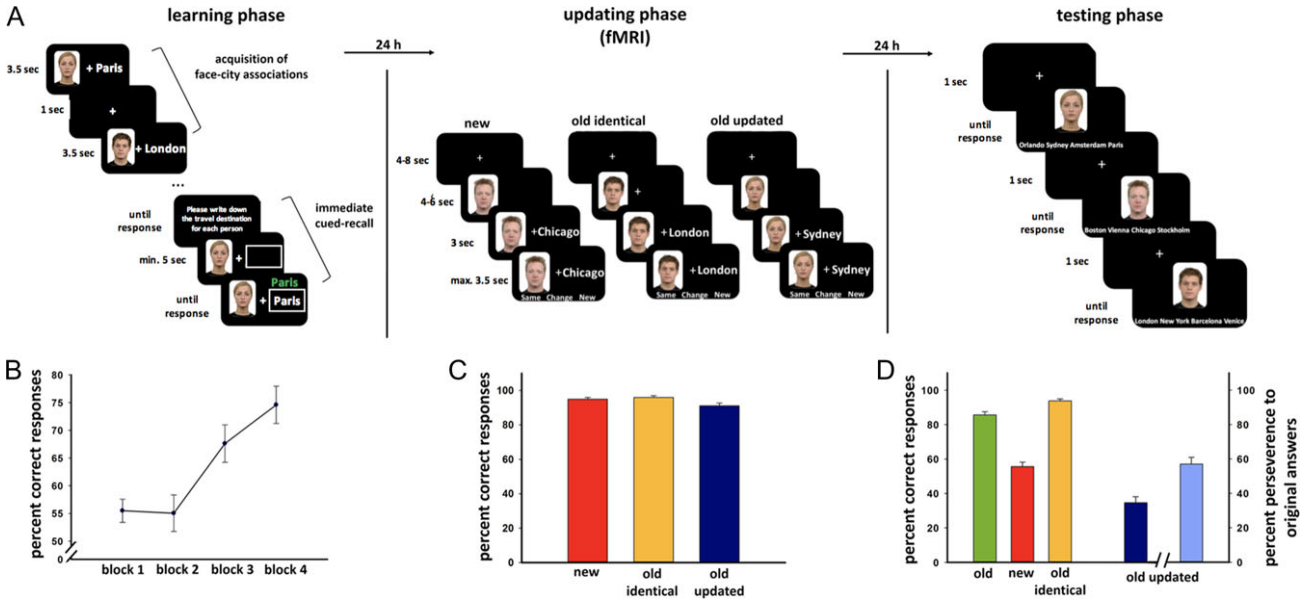


Figure 1. Memory updating paradigm and behavioral performance. (*A*) Memory updating paradigm. Participants performed the memory updating paradigm over the course of 3 consecutive days. During the learning phase on day 1, participants observed 75 face–city pairs, each presented 4 times and completed an immediate cued recall test followed by feedback showing the correct answer. On day 2, about 24 h after learning, participants completed the updating phase in the MRI scanner. In this phase, participants were presented with trials containing the same face–city pairs as on day 1 ("old identical" trials), entirely "new" trials (both face and city had not been shown before) and trials in which a known face was paired with a new city ("old updated" trials). Participants were asked to explicitly indicate whether the pair was the same as on day 1, whether the city was changed, or whether the pair was completely new. During the testing phase on day 3, participants completed a recognition task, comprising trials of face–city pairs that were only shown during the learning phase ("old" trials), "old identical" pairs, the "new" pairs learned on day 2 as well as trials for which the city was updated on day 2. Participants were presented with a face and had to select the correct city out of 4 alternatives. In the "old updated" trials, these 4 alternatives included both the city learned on day 1 and the updated city learned on day 2. Participants were explicitly instructed that if the face–city pair was updated on day 2, the city shown during the updating phase was the correct one. (*B*) Learning phase. Memory performance showed a significant increase across the learning blocks on day 1, reaching a final performance of 74.60 (SD ± 23.32) percent. (*C*) Updating phase. In the updating phase (day 2), performance differed significantly between trial types. More specifically, performance in "old updated" pairs was significantly lower (91.17 (SD ± 10.47) percent) than in "old identical" pairs (96.00 (SD ± 7.05) percent) as well as in "new" pairs (94.92 (SD ± 7.46) percent) while performance scores in "old identical" and "new" trials did not differ. However, despite the significant difference in performance, it is to note that participants were able to correctly identify more than 91% of trials as updated on day 2. (*D*) Testing phase. In the testing phase (day 3), performance between trial types differed significantly. Most importantly, however, participants named the updated city in about 34.7 (SD ± 24.21) percent of trials, whereas in 57.04 (SD ± 27.00) percent of the trials, the city associated with the face chosen on day 1. Error bars represent standard error of the mean. Error bars indicate standard error of the mean.

should help the participants to learn the 75 face–city pairs well. For each participant, the face–city associations as well as the cities that were updated on day 2 were predetermined in a face–city map. This predetermination ensured that cities with different familiarity scores from the pilot study were balanced across trial types. Trial order was also computed in advance for all phases for each participant and semi-randomized to avoid showing a trial type more than 3 times in a row in the later phases of the task. Participants were explicitly instructed to memorize the face–city pairs and they were informed that they could earn extra money if they perform well in the subsequent memory test

On experimental day 2, participants completed the "updating phase" of the paradigm in the fMRI scanner. Participants were instructed that they would see again face–city pairs. They were informed that some of the pairs would be the same as those learned on day 1. For other pairs, however, participants were told that some of the people they saw the faces of the day before made an error, and their last travel destination was actually a different city. Thus, in those pairs, a known face would be paired with an entirely novel city. Participants were asked to memorize this updated face–city association as this updated association would be the correct one (and the one learned on day 1 would be incorrect). Moreover, they were told, that they would also see new pairs in which both the face and the city were novel and they were also instructed to memorize these new pairs. Participants saw first a face (for a duration jittered 4–6 s), then a city appeared next to the face (for 3 s), after which 3 alternatives ("same", "changed", "new", for max 3.5 s) appeared below the face–city pair (see Fig. 1A). Participants were instructed to use a button box to indicate whether the pair was the same as the day before, whether the city changed and the face stayed the same or whether both face and city were new. Twenty-five of the pairs presented in this phase were identical as on day 1 ("old identical trials"), 25 pairs involved a face that had been presented on day 1 now paired with a new city ("old updated trials"), and 25 pairs were completely new ("new trials"). The face–city associations were taken from the face–city map that was computed at the beginning of the learning phase. Participants were explicitly told that they should again memorize the pairings. Once participants had pressed a button, the chosen alternative was highlighted for 0.1 s, which was then followed by a fixation cross shown for 6–8 s before a new trial started.

On experimental day 3, participants completed the "testing phase" (Fig. 1A). Participants saw all the faces they had seen on the previous 2 days, one after another in randomized order. Importantly, together with the face, they were shown 4 city names and were required to select the city that was correctly associated with the face. They were instructed that if the city

was changed for a face on day 2, they should select the city that was associated with the face on day 2. The face together with the alternatives were shown until participants made a choice (no time limit). Once participants made a selection, a fixation cross appeared (for 1 s) to signal the beginning of a new trial. Participants completed a total of 100 test trials, comprising 25 "old" trials that involved face–city pairs that were only presented during the learning phase on day 1, 25 "old identical" trials (neither face nor city changed) that were presented during the learning phase on day 1 and the updating phase on day 2, 25 "old updated" trials for which the face was associated with a different city name during the updating phase on day 2 than during learning on day 1, and 25 "new" trials that involved the faces and cities that appeared for the first time during the updating phase. Trial presentation was again semi-randomized as described above and the face–city map was used. Faces in "old updated" trials were presented with the city name from the learning phase, the 1 from the updating phase, and 2 other completely new city names. All other faces were presented with the city they had been associated with on days 1 and 2, respectively, together with 3 other cities that had never appeared before in the task.

Statistical Analysis

Performance (measured as percentage of correct trials) in the learning, updating, and testing phases was subjected to repeated measure ANOVAs, using block (learning phase) or trial type ("old identical," "old updated," "new"—in the updating phase; "old", "old identical", "old updated correctly", "old updated" perseverance, "new"—in the testing phase) as within subject factors.

Adjustment of recognition data (day 3). When an updated city was not correctly identified as updated on day 2, we assumed that participants would not be able to update their memory for these trials. We therefore removed these trials from the analysis of the day 3 recognition data and calculated the percent correctly updated over the reduced number of trials. This was also done for "old identical" and "new" trials. On average, we removed one trial from the "old identical" trial types, 2 trials from the "old updated" trial types and one trial from the "new" trial types.

Behavioral data were analyzed using SPSS (IBM 22). Significant main or interaction effects were pursued by appropriate post hoc tests if indicated. In the case of violations of sphericity, Greenhouse–Geisser correction was applied. All reported *P*-values are 2-tailed. In the case of multiple comparisons, we performed a Bonferroni correction where appropriate.

MRI Data Acquisition

MRI data were collected on a 3 T Siemens Skyra MRI Scanner, using a 32-channel head coil. A magnetic (B0) fieldmap for later unwarping of the functional images was recorded with a TR of 421 ms, a TE of 4.92 ms, and a voxel size of 3 × 3 × 3 mm. For the functional images, an echoplanar imaging sequence (476 volumes) was recorded with 3 × 3 × 3 mm voxel size (36 slices), a TR of 2.5 s and TE of 30 ms, and a flipangle of 90°. The slices were tilted 30° from the AC/PC line in order to minimize drop-out artifacts in medial temporal and orbitofrontal regions. A T1 structural image was acquired with a voxel size of 0.8 × 0.8 × 0.9 mm, a TR of 2.5 s, and a TE of 2.12 ms, with 256 slices.

MRI Data Preprocessing

The functional MRI data were preprocessed using SPM12 (http://www.fil.ion.ucl.ac.uk/spm/). The first 3 images were discarded to ensure T1 equilibration. We acquired a magnetic (B0) fieldmap to use the realign and unwarp function in SPM12. Using the FieldMap Toolbox in SPM12, a voxel displacement map was created as required for the unwarp function, which utilizes a combined static and dynamic distortion correction. The realignment function then realigns the acquired time series for each subject using a least squares approach and a 6-parameter (rigid body) spatial transformation. All scans are realigned to the first functional image that is used as a reference. Images were the coregistered to the structural image by using a rigid body model based on the work by Collignon et al. (1995). Data were then spatially normalized to fit the MNI space. To estimate the deformation, template data are deformed to match an individual scan. In the last step, data were spatially smoothed with an 8-mm full-width half-maximum Gaussian kernel.

Univariate fMRI Analyses

Data were analyzed using general linear modeling (GLM) as implemented in SPM12. Three separate regressors for each trial type ("old identical," "old updated," and "new") were modeled using the duration in seconds of the individual events. A high-pass filter of 128 s was used to eliminate low-frequency drifts and serial correlations in the time series were accounted for using an autoregressive AR(1) model. For the second-level models, contrast files for the contrasts of interest ("old updated" > "old identical"; "old identical" > "old updated"; "old updated" > "new") were tested using one-sample *t*-tests. Note that due to the large interindividual variation in updating success, an analysis with separate regressors for successfully versus not successfully updated items was not feasible as this would have resulted in a rather high number of participants for which one of the regressors had a low number of events and therefore had to be excluded from this analysis, resulting in a significant reduction of statistical power. We then performed region of interest (ROI) analyses that focused on brain areas that have been implicated in episodic long-term memory, the representation of prior knowledge, as this is critical during schema-based learning and memory, or cognitive control (Eichenbaum 1999; van Veen and Carter 2002; van Kesteren et al. 2012; Wagner et al. 2015; Gilboa and Marlatte 2017). From the Harvard–Oxford Atlas, we selected masks for the hippocampus, the anterior cingulate gyrus, medial frontal cortex as well as angular gyrus with a probability threshold of 50% as well as the left and right vmPFC and dlPFC masks, created using MARINA software (http://www.bion.de/eng/MARINA.php). Subsequently, we applied a small volume correction (svc) for the areas of interest. The svc was applied on voxel level. Voxels were regarded as significant, when falling below a corrected voxel threshold of 0.05 (family wise error (FWE) corrected) adjusted for the small volume. Only clusters within an ROI comprising $k \geq 5$ significant voxels are reported.

Furthermore, a second-level model was run that included the adjusted memory updating performance (i.e., percentage correct updated adjusted for false choices on day 2) that was measured on day 3 as a covariate. This way it was possible to account for the interindividual variability in updating performance. We then performed ROI analysis, with the above described ROIs and applied small volume correction, to obtain significantly activated clusters of voxels in our ROIs.

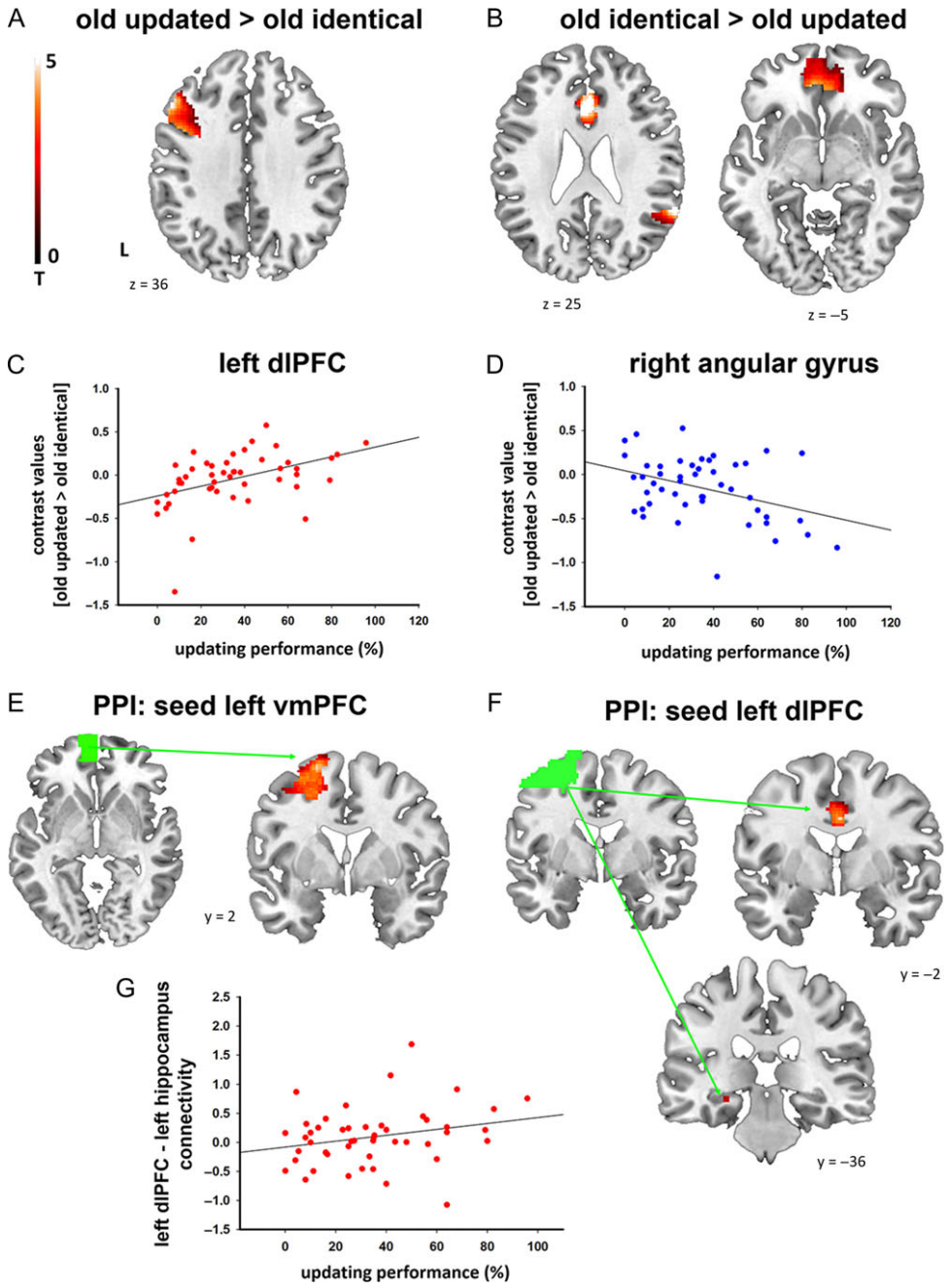


Figure 2. Neural underpinnings of memory updating. (*A*) The comparison of "old updated" and "old identical" face–city pairs during the updating phase revealed significant activation in the left dlPFC. (*B*) Activity in the anterior cingulate gyrus, bilateral ventromedial prefrontal cortex (vmPFC), and right angular gyrus was higher for "old identical" than "old updated" pairs. (*C*) Activity in the left dlPFC was positively correlated ($r = 0.434$) with updating success as assessed on day 3. For visualization purposes, we created activation masks using SPM Imcalc and the anatomical mask of the specific ROI. These masks were then converted to binary files compatible with the Marsbar toolbox (http://marsbar.sourceforge.net/). Marsbar was then used to extract the contrast values for each ROI that were above zero. These contrast values were then correlated with the adjusted memory updating scores (percent correctly updated in the testing phase adjusted for performance in the updating phase). (*D*) Updating performance was negatively correlated with updating success ($r = –0.386$) in the right angular gyrus. (*E*) A functional connectivity analysis revealed increased connectivity between the left vmPFC (seed region, indicated in green) and the left dlPFC for updated (vs. identical) face–city pairs. (*F*) When we used the left dlPFC as a seed region (green), we obtained increased functional connectivity for "old updated" versus "old identical" trials in the hippocampal dentate gyrus and the anterior cingulate gyrus. (*G*) Left dlPFC–left hippocampus connectivity was directly linked to memory updating success. For visualization purposes, we performed a correlation with the contrast values and the adjusted updating performance. Results indicate a positive correlation of the contrast values and updating performance ($r = 0.249$).

In addition, we performed a generalized form of context-dependent psychophysiological interaction (gPPI, https://www.nitrc.org/projects/gppi) to assess task-dependent connectivity with those ROIs that were most relevant in the previous analyses as seed regions. gPPI has the advantage over standard PPI, as implemented in SPM12, that it allows the inclusion of more than 2 task conditions (McLaren et al. 2012). We used the previously described regressors ("old identical," "old updated," and "new") plus a PPI

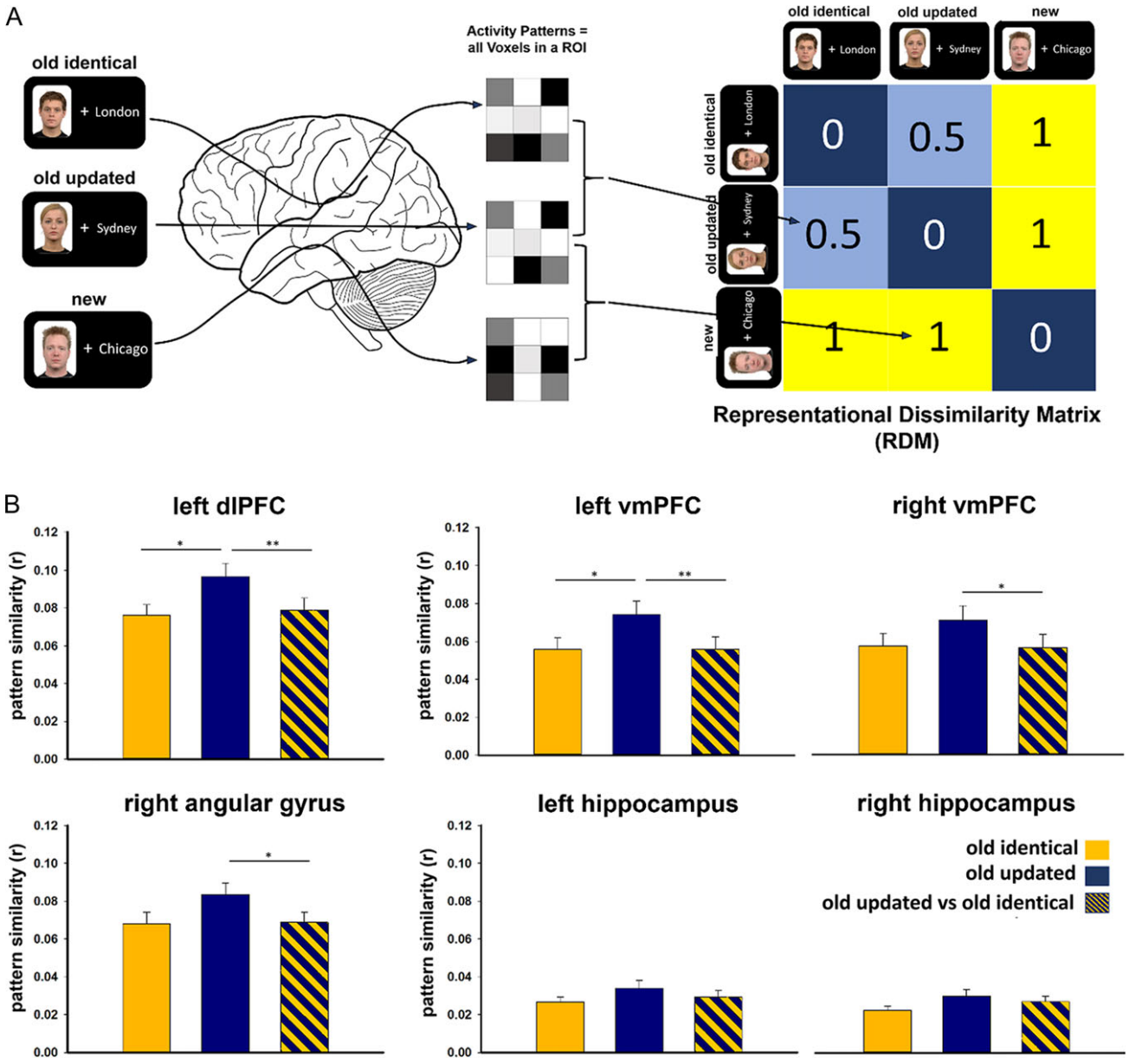


Figure 3. Neural pattern similarity within and between ROIs during memory updating. (A) Representational dissimilarity matrices (RDMs) used in the in the following analyses. The figure provides a schematic overview of the creation of an RDM for tree trial types (adapted from Nili et al.'s 2014). Quadrants representing the individual trial types and their hypothesized similarity with other trial types are indicated. (B) Pattern similarity in the left and right vmPFC indicated a significant difference between the quadrant "old updated" and the quadrant "old identical–old updated." In the left vmPFC the pattern similarity within the "old updated" quadrant was significantly higher compared with the "old identical" trials. In the right angular gyrus, pattern similarity in the "old updated" quadrant was again higher than in the "old updated–old identical" quadrant. Pattern similarity in the left dlPFC differed significantly between the "old updated" and "old identical" quadrants as well as the "old updated" and "old updated–old identical" quadrant. Figure shows results as Pearson correlations, to aid interpretability, but statistical tests show results performed after Fisher transformation was applied. Error bars represent standard error of the mean. $^{*}P < 0.05$, $^{**}P < 0.001$.

Interaction term for each of these regressors, plus the time course from the respective seed region in our first-level PPI model. For the second-level models, contrast files from the contrast PPI "old updated" > PPI "old identical," the contrast PPI "old identical" > PPI "old updated," the contrasts PPI "old identical" > PPI "old updated," PPI "old updated" > PPI "new," and PPI "new" > PPI "old identical" were tested, using a one sample *t*-test. Subsequently, we applied an svc for the other ROIs, to determine a difference in connectivity between respective regions depending on the trial types.

In order to analyze associations between the functional connectivity of ROIs and behavioral performance, we ran a second-level gPPI model with the contrasts PPI "old updated" > PPI "old identical" and PPI "old identical" > PPI "old updated," and the adjusted updating performance (i.e., percent correctly updated adjusted for performance on day 2) as covariate, and subsequently an svc that allowed us to obtain significant clusters of voxels within an ROI. For visualization purposes and only for these ROIs, we used SPM Imcalc and MarsBar to create binary activation maps including all voxels within an ROI with values above zero for the second-level contrast. For each of these activation maps, we then extracted an average time series for each participant using MarsBar and estimated the model on these,

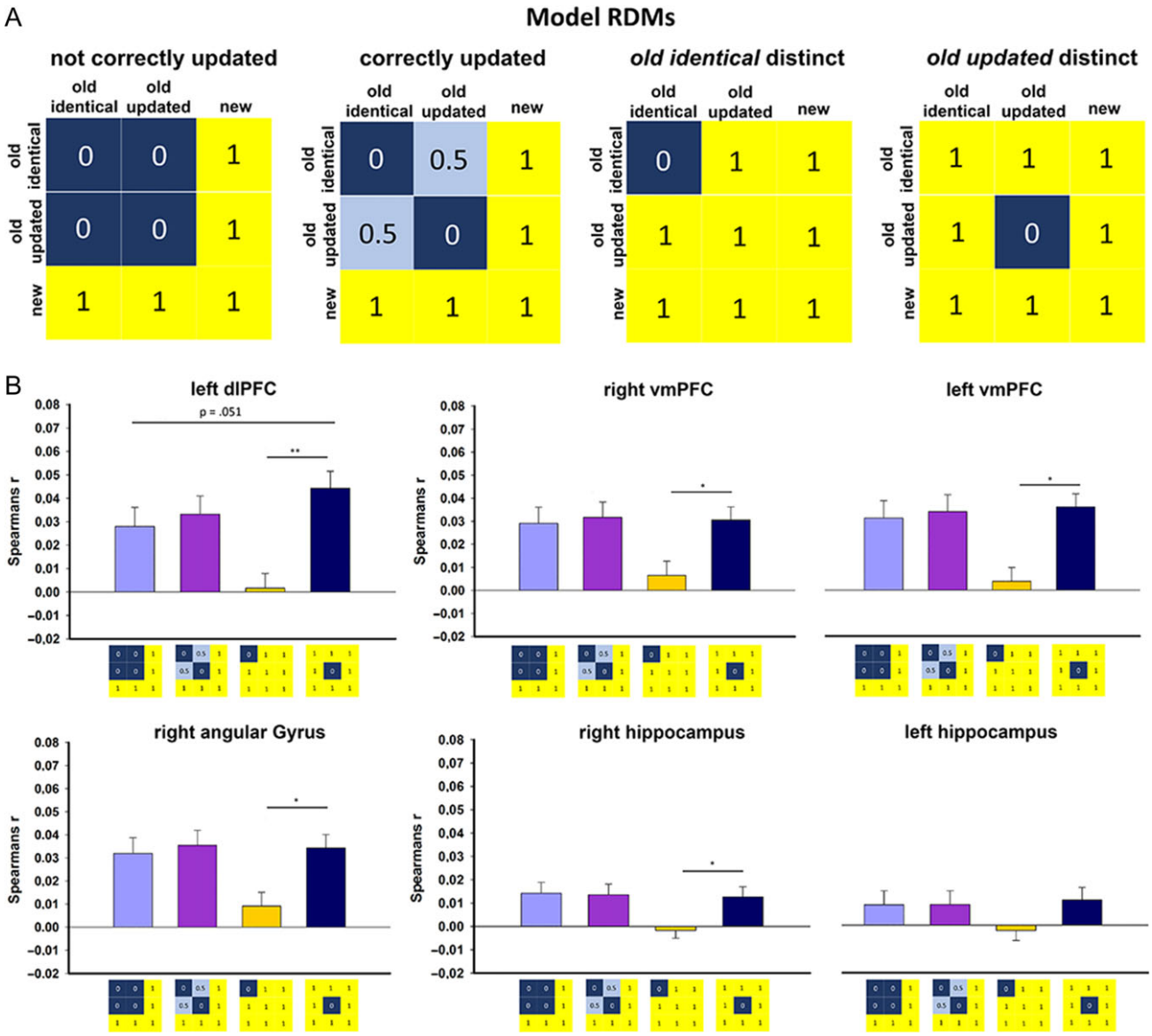


Figure 4. Distinct representation of updated items in the dlPFC. (*A*) Model RDMs. We performed a model fit analysis, comparing 4 model RDMs (not correctly updated, correctly updated, "old identical" distinct, "old updated" distinct) with the representational patterns observed in the individual subject RDMs. The model "correctly updated" represents the expected similarities for the "old identical" and "old updated" trials, suggesting a high similarity measure in the "old identical–old identical" and "old updated–old updated" quadrants (indicated by a blue color), and less similarity in the "old identical–old updated" quadrants (light blue color). The model "not correctly updated" does not distinguish between the "old identical" and "old updated" trials, that is, the highest similarity in the quadrants of "old identical," "old updated," and "old identical–old updated." The models "old identical distinct" and "old updated distinct" assume discrete activity patterns for the "old identical" and "old updated" trials. Model fit over all participants was assessed in 4 ROIs, the left and right vmPFC, left and right hippocampus, and the right angular gyrus and left dlPFC. The highest value of the Spearmans r indicates the best model fit. (*B*) The "old updated distinct" model shows a good fit in the left dlPFC, while this was less pronounced in the remaining ROIs. (*B*) The right angular gyrus and vmPFC showed similar fits for all models, while in the hippocampus the observed fits for all models were low (all Spearmans $r < 0.015$). Error bars indicate standard error of the mean. *$P < 0.05$, **$P < 0.001$, n.s. = nonsignificant.

resulting in one contrast value for each participant per ROI. The values from the first-level contrast "PPI old updated > PPI old identical" were correlated with the percentage correctly updated (i.e., the adjusted memory updating scores) from the testing phase.

Representational Similarity Analysis

In addition to the univariate analysis, we performed an RSA (Kriegeskorte et al. 2008; Nili et al. 2014) in those ROIs that turned out to be most relevant in the univariate analyses, that is, the left and right vmPFC, left and right hippocampus, right angular gyrus, and left dlPFC. For each of these ROIs and each subject, we computed representational dissimilarity matrices (RDMs) that were based on a single-trial univariate GLM (one regressor for each trial) that was estimated on native space functional images. We have utilized the data from the updating phase (day 2, in the MRI Scanner) when participants were presented with the cities, that were either new, known or updated. Due to technical difficulties, certain participants did not have all 25 trials for each trial type ("old identical", "old updated", and "new"). We therefore removed the last 2 trials of each trial

type to ensure that each subject RDM comprised 23 trials for each trial type, that is, 69 trials in total. We utilized the spm T-files for the single trial regressors, to create vectors of activity patterns for each trial, but separately for each ROI. The activity patterns were then utilized to compute the dissimilarity between 2 trials by correlation distances (1–*r*, Pearson linear correlation). After that, the dissimilarities based on each combination of trials were positioned into the corresponding cells of the 69 × 69 RDMs (see Fig. 3A).

Comparison of Pattern Similarities Between Trial Types

We extracted the mean pattern similarity (*r*, Pearson Correlation) out of specific RDM quadrants (within trial type: "old identical–old identical," "old updated–old updated," and between trial type: "old identical–old updated") from each single-subject RDM (Wolosin et al. 2013; Ritchey et al. 2015; Aly and Turk-Browne 2016). We then z-transformed the extracted mean pattern similarity values and compared 1) the within-trial-type similarity and 2) the within-trial-type similarity with the between-trial-type similarity; paired *t*-tests were performed using SPSS 22.

Comparison with Model RDMs

In addition, we compared the RDMs obtained to 4 model RDMs (Fig. 4A) that were constructed based on the expected similarities of the different trial types. The model "correctly updated" expects a high similarity between "old identical" and "old identical" trials and between "old updated" and "old updated" trials but less similarity between "old identical" and "old updated" trials, while all trials have the least similarity with "new" trials. The model "not correctly updated" does not distinguish between the "old identical" and "old updated" trials, hence it expects the highest similarity in the quadrants of "old identical," "old updated," and "old identical–old updated." The models "old identical distinct" and "old updated distinct" assume distinct activity patterns for the "old identical" and "old updated" trials, respectively, with a high similarity for trials within the respective quadrants. We used Spearman's rank correlation coefficient to obtain the correlation between the brain RDMs of the ROIs with the respective model RDMs, thus the pattern similarity values of the brain RDMs were rank transformed before calculating the correlation between the respective matrices (Nili et al. 2014). We then used 2 measures to test for the relatedness of each of the 4 model RDMs to the respective brain RDMs across subjects within each of the ROIs: 1) we first used the default one-tailed Wilcoxon signed-rank subject random-effects tests. Thus, first computing the Spearman's rank correlations of the respective brain RDM to the model RDM for each subject and then performing a Wilcoxon signed-rank test against the null hypothesis of a correlation of zero across all subjects (one-sided to test for a positive correlation only). 2) As an alternative, we further reanalyzed the data using the rather conservative stimulus-label randomization test. Here, the relatedness of the models and the brain RDMs are tested by randomizing the condition labels of the brain RDM and then calculating the Spearman's rank correlations of this randomized matrix to the model RDM. This randomization process is repeated 10 000 times to obtain a distribution of correlations simulating the null hypothesis that the brain RDMs and model RDMs are unrelated. Next one tests if the actual correlation of the (not randomized) brain RDM falls within the top 5% of the simulated distribution (Kriegeskorte et al. 2008).

In the next step, we compared the model fits of the 4 models to determine which of the models fits best in the respective ROIs, and if this differs across ROIs. For this, we extracted the Spearman's rank correlation coefficients for each model per subject in each ROI and then performed repeated measures ANOVA and post hoc *t*-tests using SPSS 22. The ANOVA contained the within subject factors ROI and model.

Moreover, we distinguished between successful and poor updaters based on whether the individual updating performance was above or below the median of all participants (median: 32% correctly updated). We again utilized a repeated measures ANOVA with the within subject factors ROI and model and the between subject factor Group (successful vs. poor updaters). Post hoc *t*-tests were performed when adequate using SPSS 22.

Results

Successful Learning and Memory Updating

During the learning phase on experimental day 1, participants (n = 48) learned a total of 75 face–city pairs. Each pair was presented 4 times and tested in 4 immediate cued recall tests (Fig. 1A). Memory performance increased significantly over the course of these 4 learning blocks ($F(1.718, 80.729) = 73.623$, $P < 0.001$, $\eta^2 = 0.610$), with a final cued recall performance of 74.60 (SD ± 23.32) percent (Fig. 1B), indicating that participants learned the face–city pairs well. There was no difference in performance between items that were subsequently updated or not (i.e., items used in later "old," "old identical," and "old updated" trials; $F(2,94) = 1.435$, $P = 0.243$, $\eta^2 = 0.030$).

On experimental day 2, about 24 h after initial learning, participants completed the memory updating phase in the MRI scanner (Fig. 1A). During this updating phase, participants saw 25 pairs that were identical to those presented during the learning phase, controlling for retrieval processes ("old identical"), 25 completely "new" pairs, that controlled for new learning (i.e., new face and new city name), and 25 pairs that included a known face now paired with a new city name ("old updated"), requiring participants to update the previously learned face–city associations. Participants were explicitly asked to memorize all pairs they were shown and, for "old updated" pairs, to retain the updated city as the correct one. Furthermore, participants were requested to indicate by button press whether the city paired with a particular face had changed from day 1 to day 2, whether it stayed the same, or whether the face–city pair was completely new (Fig. 1A). Participants were very well able to distinguish between "old identical," "old updated," and entirely "new" face–city pairs with an average performance of 94.03 (SD ± 6.03) percent correctly identified trials. Performance differed between the trial types though ($F(2, 94) = 5.838$, $P = 0.004$, $\eta^2 = 0.110$): performance for "old updated" pairs (91.17 (SD ± 10.47) percent correct, indicating that in over 91% of the trials participants were able to correctly identify a city as updated), was slightly lower than for "old identical" (96.00 (SD ± 7.05) percent correct, $t(47) = -3.705$, $P = 0.001$), and "new" pairs (94.92 (SD ± 7.46) percent correct, $t(47) = -2.186$, $P = 0.034$), while performance in "old identical" and "new" trials was comparable ($t(47) = 0.772$, $P = 0.444$; Fig. 1C).

Updating success was then tested on day 3 (testing phase), again about 24 h later, in a recognition test that included "old" pairs that were learned on day 1 but not shown on day 2, as well as "old identical," "old updated," and "new" pairs from day 2 (Fig. 1A). On each trial, participants were asked to choose the correct city name associated with a given face from 4

alternatives. For "old updated" trials, these alternatives included both the original and the updated city name. Performance differed significantly between trial types ($F(1.314, 61.735) = 68.422$, $P < 0.001$, $\eta^2 = 0.593$). Performance was highest in "old identical" pairs (94% correct) compared with both "old" (i.e., pairs that were only shown during the learning phase, 86% correct) and "new" pairs (about 56% correct, all $P < 0.001$). The lower performance in new pairs was expected given that those "new" pairs were presented only once on day 2. Most importantly, for the face–city pairs that were updated on day 2, the updated city name was correctly chosen in 34.7 (SD ± 24.21) percent of the trials, while in 57.04 (SD ± 27.00) percent of the "old updated" trials participants incorrectly persevered with the city name that was paired with the respective face during the learning phase ($t(47) = -3.055$, $P = 0.04$, Fig. 1*D*). Notably, updating performance or perseverance was not dependent on initial learning performance (comparison of day 1 memory for subsequently correctly updated vs. not correctly updated memories: $t(47) = -0.869$, $P = 0.389$). Perseverance with the old city names may well be explained by the fact that participants saw the updated items only once during the updating phase, while the old items were presented 4 times during the learning phase, each time followed by a cued recall test, which is known to boost subsequent memory (Karpicke and Roediger 2008). City names that were not presented on either day 1 or day 2 were endorsed only in about 4% of the trials each, that is, significantly less often than the updated city names from day 2. The finding that participants correctly chose, despite the differential strength of memory for the original and the updated information, the updated city name in more than a third of the trials clearly demonstrates participants' capability for updating established memory traces, which raises the question how the brain implements this memory updating capacity.

Neural Signature of Successful Memory Updating

To determine the neural mechanisms that facilitate successful memory updating, we assessed brain activity during the updating phase using fMRI. As a first step, we investigated the differences in brain activation during the presentation of a face–city pair that was updated ("old updated") compared with when face–city pairs were presented that did not require updating ("old identical"). The left dlPFC showed significantly increased activity when "old updated" face–city pairs were shown (peak coordinate: $xyz = -50, 26, 34$; $P_{svc} = 0.0026$; $t = 5.398$, $k = 32$; Fig. 2*A*), suggesting a crucial involvement of this region in the memory updating process. Moreover, the left and right dlPFC showed also significantly increased activity when comparing the activity during the presentation of "old updated" face–city pairs to the activity during the presentation of completely new face–city pairs (left peak coordinate: $xyz = -46, 22, 42$; $P_{svc} = 0.004365$; $k = 127$; right peak coordinate: $xyz = 46, 30, 36$; $P_{svc} = 0.01665$; $k = 38$), indicating that the activation of the (left) dlPFC in the contrast "old updated" versus "old identical" does not simply reflect the encoding of new items but is indeed linked to the updating of established associations. In contrast to the dlPFC, the vmPFC (left peak coordinate: $xyz = -8, 54, -2$; $P_{svc} = 0.00446$; $t = 4.6643$; $k = 16$; right peak coordinate: $xyz = 2, 44, -2$; $P_{svc} = 0.00113$; $t = 5.1508$, $k = 29$), the angular gyrus (peak coordinate: $x = 62$, $y = -48$, $z = 20$; $P_{svc} = 0.00002$; $t = 6.1211$, $k = 139$), and the anterior cingulate cortex (peak coordinate: $xyz = 0, 28, 22$; $P_{svc} = 0.00002$; $t = 6.5494$; $k = 930$), areas implicated in the representation of prior knowledge as well as selective attention and the recall of remote memories (Lenartowicz and McIntosh 2005; Weible 2013; Wagner et al. 2015; Gilboa and Marlatte 2017) were significantly less activated when "old updated" face–city pairs were presented, compared with when "old identical" items were shown (Fig. 2*B*).

In order to assess whether altered activity in those areas was directly related to the actual updating success, we correlated the neural activity during the presentation of "old updated" (vs. "old identical") face–city pairs with the percentage of correct choices of the updated city name on experimental day 3 (individual updating success) as a behavioral measure of successful memory updating. Activity in the left dlPFC was positively correlated with updating performance (peak coordinate: $xyz = -48, 12, 36$; $P_{svc} = 0.00813$; $t = 5.0338$, $k = 12$; Fig. 2*C*), indicating that this area was crucially involved in successful memory updating. In contrast, activity in the right angular gyrus was negatively correlated with updating success (peak coordinate: $xyz = 54, -48, 32$; $P_{svc} = 0.02026$; $t = 3.7943$, $k = 17$, Fig. 2*D*), suggesting that the recruitment of this area impedes the updating of established associations. Notably, parameter estimates for the contrast "old updated" versus "old identical" for the dlPFC and angular gyrus were not correlated ($r = 0.053$, $P = 0.719$), thus ruling out the possibility that the opposite correlation between updating success and activity in the left dlPFC and angular gyrus was simply due to a negative correlation between left dlPFC and angular gyrus activity. Additional activations observed in an exploratory whole-brain analyses are shown in Supplementary tables S1 and S2.

Crosstalk Between Left dlPFC and Hippocampus Supports Successful Memory Updating

In order to identify the networks that support memory updating, we next investigated which regions showed increased functional connectivity to those areas that were linked to successful memory updating. A gPPI analysis (McLaren et al. 2012) revealed that the attempt to update previously learned face–city pairs was associated with an increase in connectivity among specific prefrontal areas (i.e., the left dlFC, ACC, and left vmPFC; Fig. 2*E*,*F*). More precisely, taking the left vmPFC as a seed region, there was increased connectivity with the left dlPFC (peak coordinate: $xyz = -36, 36, 34$; $P_{svc} = 0.00587$; $t = 5.14074$, $k = 24$, Fig. 2*E*) for "old updated" (vs. "old identical") face–city pairs. When we used the left dlPFC as a seed region, we obtained increased functional connectivity for "old updated" vs. "old identical" trials with the anterior cingulate gyrus (peak coordinate: $xyz = 2, -2, 34$; $P_{svc} = 0.00354$; $t = 4.83124$, $k = 34$, Fig. 2*F*). Even more interestingly, however, when established associations were updated (vs. "old identical"), there was a significant increase in connectivity between the left dlPFC and the left hippocampus (peak coordinate in the dentate gyrus: $xyz = -26, -36, -2$; $P_{svc} = 0.02326$; $t = 3.19508$, $k = 6$; Fig. 2*F*), a critical hub in memory formation (Schacter et al. 1996; Eichenbaum 1999). When we correlated the connectivity between these areas with the individual updating success as assessed on experimental day 3, we found that specifically the crosstalk between the left dlPFC and the left hippocampus was associated with successful memory updating (peak coordinate: $xyz = -30, -26, -10$; $P_{svc} = 0.03967$; $t = 3.5807$, $k = 5$, Fig. 2*G*). To further investigate whether the dlPFC–hippocampus connectivity reflects rather new encoding or updating processes in particular, we ran an additional analysis investigating the contrast of PPI "old updated" versus PPI "new" face–city pairs. Results after a small volume correction did not reveal any difference for the "old updated" trials compared with the "new" trials in functional connectivity between the dlPFC and any of our other ROIs. This might suggest that the dlPFC–hippocampus connectivity reflects

mainly the encoding of new information, as a part of the updating process. However, this should then also be reflected in the contrast PPI "new" > PPI "old identical." However, our results showed no significant connectivity between the dlPFC and any other ROIs in this contrast. Thus, these results suggest a very specific process that is required when one is presented with the updated information (vs. "old identical"). More specifically, the familiar face in "old updated" trials may recruit a specific retrieval process, in combination with the detection of new information and, possibly, an attempt to integrate old and new information. For additional exploratory whole brain connectivity analyses, please see Supplementary Tables S3 and S4.

Neural Representations of Memory Updating

Our univariate fMRI data indicate a key role for the left dlPFC—and its functional connectivity with the hippocampus—for successful memory updating. If the left dlPFC indeed initiates specific processes that are crucial for the updating of established memories, then these processes should be paralleled by similar neural activity patterns in the left dlPFC specifically for updated information. To test this prediction, we examined in a next step the neural activity patterns of trials in which information was attempted to be updated and trials on which no updating was required. Specifically, we used multivariate RSA (Kriegeskorte et al. 2008; Nili et al. 2014) and analyzed neural activity pattern similarity for "old identical" and "old updated" (and "new") face–city pairs as well as the overlap between the activity patterns of "old updated" and "old identical" pairs in our ROIs (Wolosin et al. 2013; Ritchey et al. 2015; Aly and Turk-Browne 2016) (Fig. 3A; see Methods for details). In line with a proposed key role of the left dlPFC in memory updating, we observed a higher value for neural similarity for "old updated" face–city pairs in the left dlPFC, compared with both "old identical" pairs ($t(47) = -3.636$, $P = 0.001$) and the overlap between "old updated" and "old identical" pairs ($t(47) = 3.936$, $P < 0.001$, Fig. 3B). Comparable results were observed in the angular gyrus and bilateral vmPFC. More specifically, in the left vmPFC the pattern similarity within the "old updated" quadrant was significantly higher compared with both the "old identical" and "old updated–old identical" quadrants (both $P \leq 0.001$). In the right vmPFC, similarity within the old updated quadrant was significantly higher compared with the "old updated–old identical" similarity ($t(47) = 3.082$, $P = 0.003$), while the comparison between the "old updated" and "old identical" quadrant did not survive correction for multiple comparisons ($P = 0.018$). In the right angular gyrus, pattern similarity in the "old updated" quadrant was again higher than in the "old updated–old identical" quadrant ($t(47) = 2.87$, $P = 0.006$), yet the comparison between the "old identical" and "old updated" quadrants did not survive correction for multiple comparison ($P = 0.023$). This specific pattern in the vmPFC and angular gyrus may reflect the consistently reduced activity in those areas during "old updated" trials, as seen in the univariate analyses. In the hippocampus, neural similarity did not differ between trial types and was generally relatively low (Fig. 3B), in line with the idea that hippocampal activity patterns are fairly item specific, thus allowing pattern separation (Bakker et al. 2008; Yassa and Stark 2011).

Distinct Updating Representations in the Left dlPFC

In order to explicitly test which brain areas represent updated information distinctly from information that did not require updating, we next constructed 4 model RDMs: 1) a "not correctly updated" model, in which the activity patterns for "old updated" and "old identical" face–city pairs are comparable, 2) a "correctly updated" model, which assumes a certain overlap between activity patterns for "old updated" and "old identical" pairs (as a previously learned face is involved in both) but still expects the patterns to be more similar within each group of face–city pairs, 3) an "old identical distinct" model reflecting a clearly distinct activity pattern for the "old identical" face–city pairs but no specific pattern for "old updated" and "new" pairs, and 4) an "old updated distinct" model which assumes a highly distinct representation of "old updated" face–city pairs, as it purely reflects the processes active when information is updated without taking into account a specific overlap or similarity between "old identical" and "old updated" representations (Fig. 4A). To test whether the proposed models are significantly related to the brain RDMs in the respective ROIs we first considered results from the Wilcoxon signed rank test. These tests indicated a significant relatedness between the correctly updated, not correctly updated and "old updated" distinct models and the observed brain RDMs, respectively, within all ROIs (all $P < 0.01$), except for the left hippocampus (all $P > 0.119$). For the model old identical distinct, on the other hand, the results did not indicate any relatedness with the brain RDMs in none of the ROIs (all $P > 0.14$). We additionally performed the fixed effects condition label randomization test. Here, results indicated a non-significant trend (P (uncorrected) = 0.125) for the "old updated" distinct model in the dlPFC, while there were no other significant results in the other ROIs for any of the models. However, since the Wilcoxon signed rank test indicated a significant relatedness of the brain RDMs with most of the proposed models in most ROIs, we next compared the model fits of the 4 models in the ROIs statistically.

As the "old updated" distinct model reflects the unique representation of the updating process, we were primarily interested to specifically compare the fit of the "old updated" distinct model to the remaining models. Figure 4B suggests that the model fits were highest for the "old updated" distinct model in the left dlPFC, while all other regions showed a similar fit for the remaining models with specifically low fit values in the hippocampus (we obtained a strong trend for an ROI × model interaction: $F(5.856, 275.236) = 2.019$, $P = 0.065$, $\eta^2 = 0.041$). In the left dlPFC, post hoc t-tests indicated that the "old updated" distinct model may hold a superior fit compared with the "old identical" distinct model ($t(47) = -3.77$, $P < 0.001$) and a marginally significant better fit compared with the not correctly updated model ($t(47) = -2.003$, $P = 0.051$). The comparison between the "old updated" distinct and correctly updated model was, however, not significant ($t(47) = -1.513$, $P = 0.137$), which is not surprising since both models assume to a certain degree a differential representation of "old updated" and "old identical" trials, while the old updated distinct model purely reflects the processes active during updating of established memories with new information and does not take into account the similarity that may persist when an old face is shown. A different pattern than in the left dlPFC was observed in the angular gyrus, bilateral vmPFC, and right hippocampus. In these regions, the "old updated" distinct model was not significantly different from the "not correctly updated" model (all $P \geq 0.492$) and to a lesser extent than the left dlPFC from the "old identical" distinct model (all $P \leq 0.031$), indicating that there may be no clear distinction between the "old updated" and "not correctly updated" models in these regions (Fig. 5).

Successful Memory Updating Linked to Updating-Specific dlPFC Representation

Finally, we asked whether the above differences in the representational activity patterns between "old updated" versus "old identical" face–city pairs were predictive of subjects' behavioral memory updating performance. Our above findings suggest that the left dlPFC might show especially consistent activity patterns during the memory updating process. We therefore assumed that subjects highly proficient in updating memories were also the ones in which the neural pattern representation in the left dlPFC was most consistent during the updating process and therefore best characterized by the "old updated" distinct model. In contrast, in subjects performing poorly at updating their memories, we expected the angular gyrus to come to the fore, an area implicated, for example, in the representation of prior knowledge (van Kesteren et al. 2012; Wagner et al. 2015). To test this hypothesis, we used a median split to subdivide our participants into "successful updaters" (mean: 54% correct) and "poor updaters" (mean 15% correct), harnessing the large interindividual variance in updating performance (between 0% and 96 % correct). We then compared the RSA model fits for successful and poor updaters in our brain areas of interest, focusing on the left dlPFC and right angular gyrus (please see Fig. 5, for an overview of model fits in the remaining regions, see Supplemental Fig. S1) and again on the fit of the "old updated" distinct model compared with the remaining models. We again considered the Wilcoxon signed rank test in a first analysis and results showed a similar pattern for the poor updaters, similar to the analyses across all participants (with no significant relatedness between the "old identical distinct" model and the brain RDMs in any of the ROIs (all $P > 0.67$), and a significant relatedness between the remaining 3 models and the brain RDMs in all ROIs (all $P < 0.022$) except in the left hippocampus (all $P > 0.10$)), while in the successful updaters, the pattern was slightly different, indicating a significant relatedness between the brain RDMs and the "old identical" distinct model in the angular gyrus ($P = 0.025$), while this was not significant in any other ROI (all $P > 0.17$). The remaining models were significantly related to the brain RDMs in all ROIs ($P < 0.03$) except the left hippocampus in the correctly updated, not correctly updated and "old updated" distinct models ($P > 0.33$) and the right hippocampus in the "old updated" distinct model ($P = 0.13$). We again performed a fixed effects condition label randomization test. While results from the randomization test for each group separately indicated a trend in the dlPFC for the "old updated" distinct model (P (uncorrected) $= 0.0665$) for successful updaters, there was no other indication of a significant relatedness of any other model to the data in any other ROI for either group using the randomization test (all $P > 0.12$). However, since the Wilcoxon signed rank test indicated a significant relatedness between the brain RDMs and 3 of the proposed models we again compared the Spearman's rank correlation coefficient of the different models in the different ROIs using a repeated measures ANOVA, although results need to be interpreted with caution. Results indicated a significant interaction between ROI, model, and group (successful vs. poor updaters; $F(1.799, 82.737) = 6.663$, $P = 0.003$, $\eta^2 = 0.127$). In successful updaters, the "old updated" distinct model showed a significantly better fit in the left dlPFC compared with the "old identical" distinct model, indicated by post hoc t-tests ($t(23) = -3.131$, $P = 0.005$), the not correctly updated model ($t(23) = -2.537$, $P = 0.018$), and even the correctly updated model ($t(23) = -2.112$, $P = 0.046$; Figure 5). Thus, successful updaters appear to be characterized by showing an especially consistent activity pattern during the updating process in the left dlPFC. In the angular gyrus, however, there was no better fit of the "old updated" distinct model compared to any other model (all $P \geq 0.468$). Specifically, in successful updaters, the fit for the "old identical" distinct model was not significantly different from the fit for the "old updated" distinct model ($t(23) = -0.630$, $P = 0.535$). This may be due to a better representation of "old identical" information in the angular gyrus, allowing participants to better recognize information that did not change. In poor updaters, there was no evidence for such a specifically consistent activity pattern during the updating process in the left dlPFC. Specifically, whereas there was a better fit for the "old updated" distinct model compared with the "old identical" distinct model both in the left dlPFC and angular gyrus (both $P \leq 0.041$), none of the other model comparisons were significant (all $P \geq 0.447$). Together, these results suggest that successful updating of existing memories is driven by a consistent and distinct activity pattern during the updating process in the left dlPFC, and that a failure to show this distinct pattern may be linked to an increased inability to update already established memories with new information.

Discussion

The present findings provide novel insights into the neural mechanisms underlying the flexible updating of established memories. In particular, our data indicate a key role of the left dlPFC in this updating process. First, activity in left dlPFC and functional coupling of left dlPFC with the hippocampus were increased when learned associations had to be updated. Importantly, this increased (left) dlPFC activity during updating trials was seen both in comparison to "old identical" trials and "new" trials, showing that this activity was not due to simply retrieving old information or encoding new information but specifically to the updating of existing associations. Both, left dlPFC activity and prefrontal–hippocampal coupling were directly related to behavioral updating performance. Second, results of neural pattern similarity analysis in left dlPFC showed a distinct pattern during the updating process and the distinctiveness of these activity patterns separated individuals that were successful in memory updating from those that were not. Notably, these results hint to the fact that left dlPFC and hippocampus may act in concert to support memory updating. Unlike the left dlPFC, the hippocampus showed no increased activity during the presentation of trials that required updating in our univariate analyses. Similarly, hippocampal pattern similarity was relatively low in general and may lend support to the idea that hippocampal neural patterns are specific to particular associations (Leutgeb et al. 2007; Yassa and Stark 2011). Finally, activity in areas supporting prior knowledge, such as the angular gyrus, appeared to interfere with successful memory updating.

The dlPFC has been classically related to cognitive control and working memory (Curtis and D'Esposito 2003; Egner and Hirsch 2005) as well as attentional processes, novelty detection and cognitive load (Yamaguchi and Knight 1991; Johnson et al. 2007; Szczepanski and Knight 2014). All of these processes are highly relevant in the context of memory updating, which is a complex process that builds heavily on attention as well as cognitive control capacities and requires the individual to hold both the old and the new information in mind, thus posing a cognitive load. Without these basic cognitive processes memory updating would be unthinkable. Beyond these classical

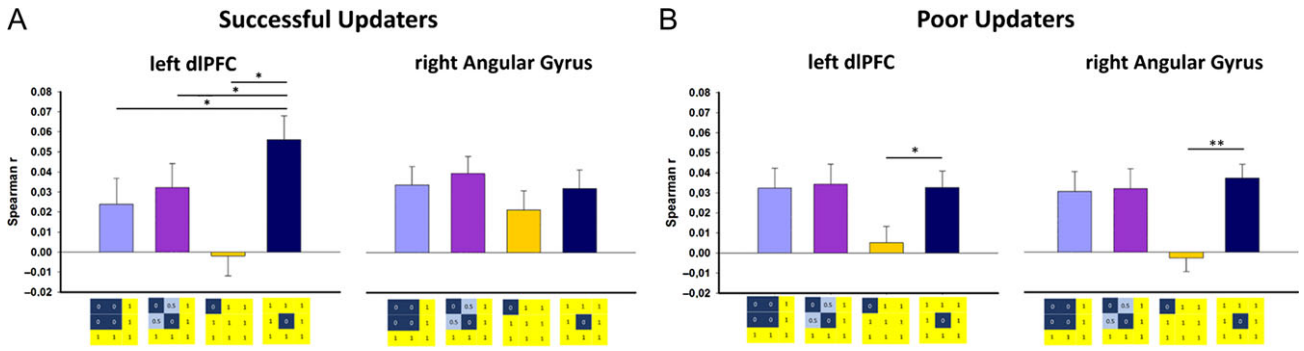


Figure 5. Model representation distinguishes between successful and poor updaters. We distinguished between (*A*) successful and (*B*) poor updaters based on a median split of updating success (successful updaters: mean performance of 54% correct, poor updaters: mean performance of 15% correct). (*A*) In the successful updaters, the "old updated distinct" model was represented with the highest fit value in the left dlPFC (Spearmans $r = 0.056$) and (*B*) while this was not the case in poor updaters (Spearmans $r = 0.033$). Error bars indicate standard error of the mean. $^{*}P < 0.05$, $^{**}P < 0.001$.

roles of the dlPFC, there is also evidence for its role in long-term memory processes (Rossi et al. 2001; Simons and Spiers 2003). Specifically, the dlPFC has been implicated in memory control processes, supporting strategic search, retrieval and evaluation of stored representations (Fletcher and Henson 2001). Moreover, the dlPFC has been linked to the verification and monitoring of recollected information (Burgess and Shallice 1996). These latter functions of the dlPFC dovetail with its key role in memory updating that we propose here, requiring the evaluation of and comparison between stored and currently presented associations in order to detect the new information. However, as participants identified the updated item (i.e., city) with high precision on experimental day 2, the mere recognition that associations have been modified does not appear to be sufficient for successful memory updating, nor does the mere attention to the updated information or novelty detection that are both reflected in participants judgments, and we presume that the role of the dlPFC goes beyond that of a "discrepancy detector." Instead, the dlPFC may orchestrate the retrieval of stored representations and the encoding of modified associations, facilitating actual memory updating, most likely in interaction with MTL areas. It may be this interaction and integration of stored representations and incoming information that distinguishes the updating of established memories from working memory updating.

The dlPFC has recently also been assigned a critical role in working memory updating (D'Ardenne et al. 2012). In working memory, however, both the original and the new information is maintained in the dlPFC (Curtis and D'Esposito 2003; Barbey et al. 2013) without the need to interact with long-term storage sites, whereas this interaction is essential for the updating of consolidated memories. We observed a significant increase in the crosstalk between the dlPFC and the hippocampus, when learned associations were updated and this functional connectivity was directly linked to updating success. Although PPI data do not allow conclusions about the direction of the interaction, the fact that we did not find a significant increase in hippocampal activity for updated items, nor an updating-specific activity pattern in the hippocampus, lets us assume that this crosstalk was mainly driven by the dlPFC.

Interactions between the dlPFC and hippocampus may facilitate memory updating in several ways. The dlPFC, when detecting modified associations, may suppress the activation of the original memory representations in the MTL, as shown in intentional forgetting paradigms (Anderson et al. 2004). Moreover, the dlPFC activation may foster the storage of the modified information in MTL areas or its incorporation into the existing trace. Our findings of a positive interaction between dlPFC and hippocampus, which was directly associated with updating success, speaks in favor of the latter alternative and renders a suppression effect, which should be reflected in a negative interaction, rather unlikely. It is to be noted, however, that the contrasts "old updated" versus "old identical" and "old updated" versus "new" may include a number of different processes, such as noticing that an item was updated, reactivating the existing memory trace, or the attempt to update memories. Although we cannot distinguish between these processes during the presentation of updated information in the present study, all of these processes may be relevant for successful memory updating. Interestingly, an earlier study that used an AB/AC interference paradigm, resembling our task design to test the relation between memory reactivation and integration of competing memories provided evidence for an involvement of the dlPFC in the processing of competing memories (Kuhl et al. 2011). Although there are important differences between this previous study and the present study, for instance, with respect to the degree to which participants were instructed to update their memories and to the age of the memories that were required to update, both studies agree in that they point to an important role of the dlPFC in memory updating.

While we consider a suppression-like mechanism as main source of successful updating rather unlikely, our data still suggest that representations of prior knowledge might hamper updating processes. Prior knowledge is thought to be represented in parietal areas, including the angular gyrus, and its relevance for the ongoing task is assumed to be detected by the vmPFC (van Kesteren et al. 2012). In line with these ideas, the angular gyrus and vmPFC were less active when modified information, compared with original information, was shown. Recruitment of the angular gyrus during the presentation of updated information was associated with impaired updating performance. Moreover, the specific increase in neural similarity in the angular gyrus for originally encoded item pairs distinguished successful from unsuccessful updaters. Together, these findings suggest that the recruitment of prior knowledge representations during the presentation of the original information is beneficial for updating performance, presumably as it allows a sharp discrimination between original and modified information, while the recruitment of those areas during the presentation of modified information impedes updating success. Albeit we did not find updating-related changes in dlPFC–angular gyrus connectivity, the dlPFC might also coordinate

angular gyrus activity through its connection to the vmPFC, which was increased for updated items in our study. An additional account describes the function of the medial prefrontal cortex (mPFC) in the process of integrating new memories into existing structures and storing these memories (Schlichting and Preston 2016) as seen in schema-based learning. This would also describe the current observations of heightened vmPFC activity when "old identical" information is presented. While an interaction between the mPFC and hippocampus is thought to facilitate this integration process (Kroes and Fernandez 2012; Schlichting and Preston 2016), our results did not show this pattern, which may however be due to differences in study design and task instructions.

It has been argued that the updating of consolidated memories is based on memory reconsolidation processes (for a recent review see Lee et al. 2017). Specifically, it is assumed that the reactivation of a consolidated memory renders the reactivated trace labile again so that it needs to be stabilized anew during a period of reconsolidation (Nadel and Land 2000; Hupbach et al. 2008; Nader and Einarsson 2010; Dudai 2012). During the reconsolidation window, memories can be weakened, strengthened, or updated (Dudai 2006; Alberini 2011). Reactivation-dependent memory modifications have by now been shown across tasks and species (Hupbach et al. 2007; Lee 2008; Stollhoff et al. 2008; Schwabe et al. 2012). Although we did not aim to probe reconsolidation processes, a reconsolidation-like mechanism might still have been active. In the updating phase, we presented the face first, which may have served as a reminder for the respective face–city pair, and the updated (or original) city name shortly thereafter, which may be considered a contextual reinstatement, possibly aiding an inability to successfully update memories (Gershman et al. 2013). Previous evidence from a study using TMS pointed also to a critical role of the dlPFC in the postreactivation strengthening of the episodic memories (Sandrini et al. 2013). In contrast to the present study, however, this study did not address the actual updating of memories by new information. In fact, we are not aware of any reconsolidation study that tested the neural underpinnings of the updating of long-term memories in humans. Closely related to the reconsolidation concept and to the current findings, however, is the fundamental and highly controversial issue of the fate of the original memory trace after memory updating (Hardt et al. 2009; Bermúdez-Rattoni and McGaugh 2017). Is the original trace overwritten by the representation of the updated information? Or are there 2 competing traces and successful updating reflects mainly the predominance of the trace representing the modified association? Experimental studies in humans, such as the present, can hardly solve this issue. Animal studies employing state-of-the art molecular techniques, however, could make a significant contribution to this long-standing debate. Moreover, as our data suggested updating-related activity mainly in the left dlPFC and there is some evidence for distinct roles of the right and left dlPFC in retrieval processes (Rossi et al. 2001; Javadi and Walsh 2012), future studies may use brain stimulation techniques to test whether memory updating processes are indeed lateralized.

In sum, our data provide novel insights into the neural mechanisms underlying a fundamental feature of memory, its ability to update in light of new information. In particular, we show that the dlPFC, most likely through its interaction with the hippocampus, is essential for keeping memories up to date, enabling them to effectively guide choice or simulate upcoming events (Schacter et al. 2007) and thus to prepare the organism for the future.

Supplementary Material

Supplementary material is available at *Cerebral Cortex* online.

Authors' Contributions

L.S. conceived and designed the experiment, L.M.K. performed research, L.M.K. and L.C.D. analyzed the data, L.S. and G.J. supervised research and analysis, L.M.K. and L.S. drafted the manuscript, all authors contributed to the manuscript.

Funding

This study was funded by the German Research Foundation (DFG; SCHW1357/16-1).

Notes

The authors thank Hendrik Heinbockel, Jan Patrick Kalweit, Svetlana Wiese, Vladimir Demchenko, Lea Seule, Simone Roik, and Felix Kalbe for their help during data collection and Lukas Mönch for programming the task. The authors further greatly acknowledge the help of Daniel Kutzner, Wiebke Schmidt, and Christoph Schaede during fMRI scanning. *Conflict of Interest*: All authors declare no conflict of interest.

References

Alberini CM. 2011. The role of reconsolidation and the dynamic process of long-term memory formation and storage. Front Behav Neurosci. 5:1–10.

Alvarez P, Squire LR. 1994. Memory consolidation and the medial temporal lobe: a simple network model. Proc Natl Acad Sci USA. 91:7041–7045.

Aly M, Turk-Browne NB. 2016. Attention stabilizes representations in the human hippocampus. Cereb Cortex. 26:783–796.

Anderson MC, Ochsner KN, Kuhl B, Cooper J, Robertson E, Gabrieli SW, Glover GH, Gabrieli JD. 2004. Neural systems underlying the suppression of unwanted memories. Science. 303:232–235.

Baddeley AD, Dale HCA. 1966. The effect of semantic similarty on retroactive inteference in long- and short-term memory. J Verbal Learning Verbal Behav. 5:417–420.

Bakker A, Kirwan CB, Miller M, Stark CE. 2008. Pattern separation in the human hippocampal CA3 and dentate gyrus. Science. 319:1640–1642.

Barbey AK, Koenigs M, Grafman J. 2013. Dorsolateral prefrontal contributions to human working memory. Cortex. 49: 1195–1205.

Bermúdez-Rattoni F, McGaugh JL. 2017. Memory reconsolidation and memory updating: two sides of the same coin? Neurobiol Learn Mem. 142:1–3.

Bilek E, Schafer A, Ochs E, Esslinger C, Zangl M, Plichta MM, Braun U, Kirsch P, Schulze TG, Rietschel M, et al. 2013. Application of high-frequency repetitive transcranial magnetic stimulation to the DLPFC alters human prefrontal-hippocampal functional interaction. J Neurosci. 33: 7050–7056.

Blumenfeld RS, Parks CM, Yonelinas AP, Ranganath C. 2011. Putting the pieces together: the role of dorsolateral prefrontal cortex in relational memory encoding. J Cogn Neurosci. 23:257–265.

Burgess N, Maguire EA, O'Keefe J. 2002. The human hippocampus and spatial and episodic memory. Neuron. 35:625–641.

Burgess PW, Shallice T. 1996. Confabulation and the control of recollection. Memory. 4:359–411.
Chun MM, Turk-Browne NB. 2007. Interactions between attention and memory. Curr Opin Neurobiol. 17:177–184.
Cohen JD, Perlstein WM, Braver TS, Nystrom LE, Noll DC, Jonides J, Smith EE. 1997. Temporal dynamics of brain activation during a working memory task. Nature. 386:604–608.
Cole MW, Schneider W. 2007. The cognitive control network: integrated cortical regions with dissociable functions. Neuroimage. 37:343–360.
Collignon A, Maes F, Delaere D, Vandermeulen D, Suetens P, Marchal G. 1995. Automated multi-modality image registration based on information theory. In: Bizais Y, Barillot C, Di Paola R, editors. Proc. information processing in medical imaging. Dordrecht, the Netherlands: Kluwer Academic Publishers. p. 263–274.
Curtis CE, D'Esposito M. 2003. Persistent activity in the prefrontal cortex during working memory. Trends Cogn Sci. 7:415–423.
Depue BE, Curran T, Banich MT. 2007. Prefrontal regions orchestrate suppression of emotional memories via a two-phase process. Science. 317:215–219.
Dudai Y. 2006. Reconsolidation: the advantage of being refocused. Curr Opin Neurobiol. 16:174–178.
Dudai Y. 2012. The restless engram: consolidations never end. Annu Rev Neurosci. 35:227–247.
D'Ardenne K, Eshel N, Luka J, Lenartowicz A, Nystrom LE, Cohen JD. 2012. Role of prefrontal cortex and the midbrain dopamine system in working memory updating. Proc Natl Acad Sci USA. 109:19900–19909.
Egner T, Hirsch J. 2005. Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. Nat Neurosci. 8:1784–1790.
Eichenbaum H. 1999. The hippocampus and mechanisms of declarative memory. Behav Brain Res. 103:123–133.
Fletcher PC, Henson R. 2001. Frontal lobes and human memory. Brain. 124:849–881.
Fuster JM, Alexander GE. 1971. Neuron activity related to short-term memory. Science. 173:652–654.
Gershman SJ, Schapiro AC, Hupbach A, Norman KA. 2013. Neural context reinstatement predicts memory misattribution. J Neurosci. 33:8590–8595.
Gilboa A, Marlatte H. 2017. Neurobiology of schemas and schema-mediated memory. Trends Cogn Sci. 21:618–631.
Hardt O, Wang SH, Nader K. 2009. Storage or retrieval deficit: the yin and yang of amnesia. Learn Mem. 16:224–230.
Hupbach A, Gomez R, Hardt O, Nadel L. 2007. Reconsolidation of episodic memories: a subtle reminder triggers integration of new information. Learn Mem. 14:47–53.
Hupbach A, Hardt O, Gomez R, Nadel L. 2008. The dynamics of memory: context-dependent updating. Learn Mem. 15:574–579.
Javadi A.H., Walsh, V. 2012. Transcranial direct current stimulation (tDCS) of the left dorsolateral prefrontal cortex modulates declarative memory. Brain Stimul. 5:231–241
Jing HG, Madore KP, Schacter DL. 2017. Preparing for what might happen: an episodic specificity induction impacts the generation of alternative future events. Cognition. 169:118–128.
Johnson JA, Strafella AP, Zatorre RJ. 2007. The role of the dorsolateral prefrontal cortex in bimodal divided attention: two transcranial magnetic stimulation. Studies. J Cogn Neurosci. 19:907–920.
Karpicke JD, Roediger HLI. 2008. The critical importance of retrieval for learning. Science. 319:966–968.
Kriegeskorte N, Mur M, Bandettini P. 2008. Representational similarity analysis—connecting the branches of systems neuroscience. Front Syst Neurosci. 2:1–28.
Kroes MC, Fernandez G. 2012. Dynamic neural systems enable adaptive, flexible memories. Neurosci Biobehav Rev. 36: 1646–1666.
Kuhl BA, Bainbridge WA, Chun MM. 2012. Neural reactivation reveals mechanisms for updating memory. J Neurosci. 32: 3453–3461.
Kuhl BA, Rissman J, Chung MM, Wagner AD. 2011. Fidelity of neural reactivation reveals competition between memories. PNAS. 108:5903–5908.
Langner O, Dotsch R, Bijlstra G, Wigboldus DHJ, Hawk ST, van Knippenberg A. 2010. Presentation and validation of the Radboud Faces Database. Cogn Emot. 24:1377–1388.
Lee JL. 2008. Memory reconsolidation mediates the strengthening of memories by additional learning. Nat Neurosci. 11: 1264–1266.
Lee JLC, Nader K, Schiller D. 2017. An update on memory reconsolidation updating. Trends Cogn Sci. 21:531–545.
Lenartowicz A, McIntosh AR. 2005. The role of anterior cingulate cortex in working memory is shaped by functional connectivity. J Cogn Neurosci. 17:1026–1042.
Leutgeb JK, Leutgeb S, Moser M-B, Moser EI. 2007. Pattern separation in the dentate gyrus and CA3 of the hippocampus. Science. 315:961–966.
Loftus EF. 1975. Leading questions and the eyewitness report. Cogn Psychol. 7:560–572.
Lundqvist D, Flykt A, Öhman A 1998. The Karolinska Directed Emotional Faces—KDEF, CD ROM from Department of Clinical Neuroscience.
Manenti R, Cotelli M, Calabria M, Maioli C, Miniussi C. 2010. The role of the dorsolateral prefrontal cortex in retrieval from long-term memory depends on strategies: a repetitive transcranial magnetic stimulation study. Neuroscience. 166: 501–507.
McLaren DG, Ries ML, Xu G, Johnson SC. 2012. A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. Neuroimage. 61:1277–1286.
Murray LJ, Ranganath C. 2007. The dorsolateral prefrontal cortex contributes to successful relational memory encoding. J Neurosci. 27:5515–5522.
Nadel L, Hupbach A, Gomez R, Newman-Smith K. 2012. Memory formation, consolidation and transformation. Neurosci Biobehav Rev. 36:1640–1645.
Nadel L, Land C. 2000. Memory traces revisited. Nat Rev Neurosci. 1:209–212.
Nader K, Einarsson EO. 2010. Memory reconsolidation: an update. Ann N Y Acad Sci. 1191:27–41.
Nili H, Wingfield C, Walther A, Su L, Marslen-Wilson W, Kriegeskorte N. 2014. A toolbox for representational similarity analysis. PLoS Comput Biol. 10:1–11.
Preston AR, Eichenbaum H. 2013. Interplay of hippocampus and prefrontal cortex in memory. Curr Biol. 23:1–21.
Ritchey M, Montchal ME, Yonelinas AP, Ranganath C. 2015. Delay-dependent contributions of medial temporal lobe regions to episodic memory retrieval. Elife. 4:1–19.
Rossi S, Cappa SF, Babiloni C, Pasqualetti P, Miniussi C, Carducci F, Babiloni F, Rossini PM. 2001. Prefrontal cortex in long-term memory: an "interference" approach using magnetic stimulation. Nat Neurosci. 4:948–952.
Sandrini M, Brambilla M, Manenti R, Rosini S, Cohen LG, Cotelli M. 2014. Noninvasive stimulation of prefrontal cortex strengthens existing episodic memories and reduces forgetting in the elderly. Front Aging Neurosci. 6: 1–9.

Sandrini M, Censor N, Mishoe J, Cohen LG. 2013. Causal role of prefrontal cortex in strengthening of episodic memories through reconsolidation. Curr Biol. 23:2181–2184.

Schacter DL, Addis DR, Buckner RL. 2007. Remembering the past to imagine the future: the prospective brain. Nat Rev Neurosci. 8:657–661.

Schacter DL, Alpert NM, Savage CR, Rauch SL, Albert MS. 1996. Conscious recollection and the human hippocampal formation: evidence from positron emission tomography. Proc Natl Acad Sci USA. 93:321–325.

Schiller D, Monfils MH, Raio CM, Johnson DC, Ledoux JE, Phelps EA. 2010. Preventing the return of fear in humans using reconsolidation update mechanisms. Nature. 463:49–53.

Schlichting ML, Preston AR. 2016. Hippocampal-medial prefrontal circuit supports memory updating during learning and post-encoding rest. Neurobiol Learn Mem. 134:91–106.

Schwabe L, Nader K, Pruessner JC. 2014. Reconsolidation of human memory: brain mechanisms and clinical relevance. Biol Psychiat. 76:274–280.

Schwabe L, Nader K, Wolf OT, Beaudry T, Pruessner JC. 2012. Neural signature of reconsolidation impairments by propranolol in humans. Biol Psychiat. 71:380–386.

Simons JS, Spiers HJ. 2003. Prefrontal and medial temporal lobe interactions in long-term memory. Nat Rev Neurosci. 4: 637–648.

Squire LR, Zola-Morgan S. 1991. The medial temporal lobe network. Science. 253:1380–1386.

Stollhoff N, Menzel R, Eisenhardt D. 2008. One retrieval trial induces reconsolidation in an appetitive learning paradigm in honeybees (*Apis mellifera*). Neurobiol Learn Mem. 89: 419–425.

Szczepanski SM, Knight RT. 2014. Insights into human behavior from lesions to the prefrontal cortex. Neuron. 83:1002–1018.

Tulving E. 2002. Episodic Memory: From mind to brain. Annu Rev Psychol. 53:1–25.

van Kesteren MT, Ruiter DJ, Fernandez G, Henson RN. 2012. How schema and novelty augment memory formation. Trends Neurosci. 35:211–219.

van Veen V, Carter CS. 2002. The anterior cingulate as a conflict monitor: fMRI and ERP studies. Physiol Behav. 77:477–482.

Wagner IC, van Buuren M, Kroes MC, Gutteling TP, van der Linden M, Morris RG, Fernandez G. 2015. Schematic memory components converge within angular gyrus during retrieval. Elife. 4:1–28.

Weible AP. 2013. Remembering to attend: the anterior cingulate cortex and remote memory. Behav Brain Res. 245:63–75.

Wimmer EG, Shohamy D. 2012. Preference by association: how memory mechanisms in the hippocampus bias decisions. Science. 338:270–273.

Wolosin SM, Zeithamova D, Preston AR. 2013. Distributed hippocampal patterns that discriminate reward context are associated with enhanced associative binding. J Exp Psychol Gen. 142:1264–1276.

Yamaguchi S, Knight RT. 1991. Anterior and posterior association cortex contributions to the somatosensory P300. J Neurosci. 11:2039–2054.

Yassa MA, Stark CE. 2011. Pattern separation in the hippocampus. TrendsNeurosci. 34:515–525.

Zeithamova D, Dominick AL, Preston AR. 2012. Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. Neuron. 75: 168–179.

Appendix D

Study IV: Moderate post-encoding stress shows no influence on memory transformation rate after four weeks of systems consolidation

Unpublished Manuscript

Moderate post-encoding stress shows no influence on memory transformation rate after four weeks of systems consolidation

Lisa C. Dandolo

Abstract

Over time, episodic memories transform from detailed to gist-like memories. Whether this dynamic process is influenced by external factors such as stress has not yet been tested in humans so far. In general, post-encoding stress most often has an enhancing effect on subsequent memory performance. But does post-encoding stress also influence the time course of memory transformation? To investigate this question, we tested 160 participants on two experimental days, using an episodic picture memory task with negative and neutral pictures. Half of the participants performed a socially evaluated stress test shortly after memory encoding, while the other half performed a control task. Memory retrieval was tested either one day after encoding or four weeks later to capture memory transformation during prolonged systems consolidation. We found an enhancing effect of post-encoding stress on the subsequent recall of negative pictures in females. Most importantly, however, we did not find any effects of moderate post-encoding stress or cortisol increase on the degree of memory transformation tested after four weeks of systems consolidation.

1 Introduction

Memories are dynamic and can be influenced in various ways during learning, consolidation and retrieval processes. One factor which is often present in our everyday lives and has a considerable effect on learning and memory is stress (Roozendaal, 2002; Wolf, 2009). Acting through stress hormones such as gluccocorticoids and noradrenalin, stress can have either enhancing or impairing effects on learning and memory. The timing of the stressful experience in relation to the different memory phases is thereby one of the most decisive factors (Roozendaal, 2002, Schwabe, Joels, Roozendaal, Wolf, & Oitzl, 2012; Cadle & Zoladz, 2015). Stress shortly before memory retrieval has an impairing effect on memory performance (e.g. Kuhlmann, Piel, & Wolf, 2005; Buchanan, Tranel, & Adolphs, 2006; Smeets, Otgaar, Candel, & Wolf, 2008). On the other hand, studies which have targeted memory encoding and/or memory consolidation by administering the psychosocial stressor shortly before learning (affecting both encoding and consolidation) or shortly after memory encoding (affecting only memory consolidation) have mostly shown enhancing effects of stress on subsequent memory performances (e.g. Buchanan & Lovallo, 2001; Cahill, Gorski, & Le, 2003; Andreano & Cahill, 2006; Beckner, Tucker, Delville, & Mohr, 2006; Preuss & Wolf, 2009; McCullough & Yonelinas, 2013). To find these enhancing effects, subsequent memory has to be measured at a delay long enough to ensure that memory retrieval is not influenced by the stress hormones any longer (Schwabe et al., 2012). Yet, even when controlling for this factor, the enhancing effect of stress on memory encoding and consolidation is not always robustly found (for an opposite, impairing effect see Trammell and Clore (2014)) and often depends on other factors such as the emotional arousal of the learning material (Cahill et al., 2003; Payne et al., 2007; Felmingham, Tran, Fong, & Bryant, 2012), or the sex of the participant (Cahill, 2003; Cahill, 2006). It is also important to consider the strength of the stress response: it has been proposed that an inverted U-shape curve is best suited to relate the strength of the cortisol increase in response to the stressor to the memory scores: while low and very high cortisol responses do not affect or even impair memory performance, a moderate cortisol response leads to

the enhancing effect (Andreano & Cahill, 2006).

Most studies tested whether stress affects the overall memory performance, without, however, testing whether the quality of the memory is also affected by stress. There is one line of work, however, which has focussed on the different effects of post-encoding stress on familiarity and recollection based recognition (Yonelinas, Parks, Koen, Jorgenson, & Mendoza, 2011; McCullough & Yonelinas, 2013; McCullough, Ritchey, Ranganath, & Yonelinas, 2015). Recollection can be defined as re-experiencing an episode in a detailed way, thus remembering when and where one experienced the episode or learned a certain item, a process which has been shown to be hippocampus-dependent (Eichenbaum, Yonelinas, & Ranganath, 2007). While familiarity is based on knowing one has seen an item before without being able to recollect the context of when and where it was learned, a process which has been related to other, hippocampus surrounding medial temporal lobe areas (Eichenbaum et al., 2007). While early results suggest that stress only enhances familiarity based recognition without influencing recollection (Yonelinas et al., 2011; McCullough & Yonelinas, 2013), newer results suggest that recollection is also influenced by stress, but in a non-monotonic, inverted U-shape relationship with moderate increases in cortisol benefiting subsequent recollection and high cortisol increases impairing later recollection (McCullough et al., 2015). In a different approach, studies using false memory paradigms showed an increase in false memory scores after pre- or post-learning stress (Payne, Nadel, Allen, Thomas, & Jacobs, 2002; Payne et al., 2006; Payne et al., 2007; Pardilla-Delgado, Alger, Cunningham, Kinealy, & Payne, 2016). It has been suggested, that this increase in false recognitions is due to a deficit in retrieving detailed memories because of the negative effect of stress on hippocampal processing, and a subsequent over-reliance on gist-memory stored in neocortical regions (Payne et al., 2002; Pardilla-Delgado et al., 2016). Similarly, it has been suggested that stress and emotional arousal leads to a narrowing of attention, thereby impairing the ability to remember peripheral neutral details of an episode and simultaneously enhancing memories for the central emotional aspect, or gist of an episode (Kensinger, Garoff-Eaton, & Schacter, 2007; Cadle & Zoladz, 2015).

All these accounts imply an increased reliance on gist-like memories or familiarity components of recognition after exposure to (highly) stressful experiences, thus showing an influence on memory quality. According to the trace transformation theory (TTT, Winocur, Moscovitch, & Bontempi, 2010; Winocur & Moscovitch, 2011; Sekeres, Moscovitch, & Winocur, 2017), the process of systems consolidation is also accompanied by a change in memory quality over time, even without interfering factors. According to the TTT, origi-

nally detailed episodic hippocampus-dependent memories, transform into gist-like versions during the process of system consolidation. These gist-like memories can be retrieved from the neocortex alone, while detailed memories remain hippocampus-dependent. This memory transformation process naturally occurs over a time course of weeks to months, or even years (Sekeres et al., 2017). But, is the time course of memory transformation from detailed to gist-like memories influenced my the amount of stress present during initial encoding and early consolidation? This question has been addressed in an recent rodent study by Pedraza et al. (2016). They could show that stress, indeed, had an influence on the rate of memory transformation depending on the strength of the stress response during learning in an inverted U-shaped curve: moderate stress during learning led to a detailed representation of the memory which remained detailed for an extended period of time (up to 40 days). High stress during learning, on the other hand, led to an unusually fast rate of memory transformation, resulting in generalized, gist-like versions of the memory after only 2 weeks (Pedraza et al., 2016). Can we find the same pattern of results in a human study?

In the present study we tested whether post-encoding stress influenced the time course of memory transformation in humans. Participants encoded emotionally negative and neutral pictures on experimental day one, shortly before undergoing the standardized socially evaluated cold-pressor test (SECPT; Schwabe, Haddad, & Schachinger, 2008; Schwabe & Schachinger, 2018) or a warm water control condition. In a between-subject design, participants performed a free recall task and a recognition task either 1 day or 28 days after encoding. Critically, the recognition task not only contained the learned old pictures and novel pictures, but also related pictures showing the same gist as the respectively matched old picture but depicting different details. This enabled us to test whether a detailed version of the memory could still be retrieved during memory testing, or whether only a transformed, gist-like version was retrievable, leading to problems in differentiating between the learned pictures and the related lure pictures.

In an attempt to replicate former results, we first tested the effects of post-encoding stress on the overall memory performance. We expected to find an enhancing effect of stress on the overall memory strength after 1 and 28 days and expected this effect to be most pronounced for emotionally negative stimuli. We included sex as an additional factor, as former studies showed that results may differ between female and male participants. Most importantly, we then focussed on the effects of post-encoding stress on the time course of memory transformation. In accordance with the results from Pedraza et al. (2016), we ex-

pected to find an inverted U-shape relationship: less transformed memories for moderated amounts of stress compared to the control condition, but more transformed memories for high amounts of stress.

2 Methods

2.1 Participants

In total, we tested 162 healthy young adults (82 women, 80 men). Yet, two of these participants only completed experimental day 1 and were thus replaced during data collection, leading to a full sample of 160 participants (80 women, 80 men). Through prior telephone screenings we excluded participants with a history of psychiatric or neurological disease, medication intake or drug abuse. Women using hormonal contraceptives, habitual smokers and participants with a Body Mass Index below 18 or above 29 were excluded, as these factors may affect the endocrine stress response (Kirschbaum, Kudielka, Gaab, Schommer, & Hellhammer, 1999; Rohleder & Kirschbaum, 2006). All participants provided written informed consent and received monetary compensation. The study protocol was approved by the ethics committee of the German Psychological Society (DGPs). Participants were pseudo-randomly assigned to four different experimental groups: 1d-stress group, 28d-stress group, 1d-control group, 28d-control group (20 women and 20 men per group). All testing took place in the afternoon between 12.00 A.M. and 07.00 P.M. to control for diurnal variations of cortisol and participants were instructed not to eat meals, drink caffeine or perform severe physical exercise within the 2h before the experiment. After data collection, one participant was excluded due to a lack of sufficient memory encoding on experimental day 1 (below 25% correctly named pictures in the 3rd free recall run) and another participant was excluded due to missing cortisol data, thus leaving a final sample of 158 participants for analysis (mean age = 25.62 years, SD = 4.20 years; 1d-stress group = 20 women and 19 men, 28d-stress group = 19 women and 20 men, 1d-control group = 20 women and 20 men, 28d-control group = 20 women and 20 men).

2.2 Study design and experimental paradigm

Testing took place on two experimental days: experimental day 1 started with a memory encoding task immediately followed by either a stress or control manipulation depending

on the experimental group; experimental day 2 consisted of a free recall memory task, followed by a recognition memory task and a stimulus rating task. Critically, the time interval between memory encoding and memory testing was varied between the groups: for participants in the 1d-stress group or 1d-control group memory testing took place one day after encoding, while for participants in the 28d-stress group or the 28d-control group memory was tested 28 days, i.e. four weeks, after encoding.

2.3 Stress or control manipulation on experimental day 1

Participants in the stress condition underwent the standardized socially evaluated cold-pressor test (SECPT; Schwabe et al., 2008; Schwabe & Schachinger, 2018) approximately 5 mins after the end of the memory encoding task in a different room. During this 3 min long stress manipulation participants had to hold their left hand into ice cold water (0-2 °C) while being socially evaluated by a second experimenter who was newly introduced for this stress manipulation. This experimenter was wearing a white lab coat and was rather cold and distant during interactions and constantly made notes to give an impression of social evaluation. The participants received written instructions for this task, telling them to hold their hand into the water as long as possible (and only withdraw it if the pain gets unbearable) without telling them the duration of the task, thereby introducing a level of uncertainty. In addition, participants were videotaped and could see their own face on a TV screen behind the stress experimenter, introducing an element of self-monitoring. In the corresponding control manipulation, participants also entered a different room approximately 5 mins after the encoding task and had to hold their hand into warm water (35-37°C) for 3 minutes. Here, there was neither a newly introduced stress experimenter nor a video recording with TV screen (Schwabe et al., 2008; Schwabe & Schachinger, 2018)

2.4 Subjective and physiological measures

To measure the success of the stress manipulation we collected data corresponding to three different levels of the stress response (Schwabe & Schachinger, 2018). First, to assess the subjective feeling of being stressed, we asked participants to rate how difficult, painful, unpleasant and stressful they had experienced the tasks on a scale from 0 (“not at all”) to 100 (“very much”) immediately after the stress/control manipulation. Subjective feelings of stress were additionally measured using a German multidimensional mood questionnaire

(MDBF; three scales: elevated mood, wakefulness and calmness; Steyer, Schwenkmezger, Notz, & Eid, 1994), which participants filled in at five different time points (at the beginning of experimental day 1, directly after the memory encoding task, 5 and 40 mins after the onset of the stress/control manipulation and at the beginning of experimental day 2). Secondly, as an indicator of sympathetic nervous system activity, we measured participants' blood pressure and pulse at seven different time points (at the beginning of experimental day 1, directly after the memory encoding task, during the stress/control manipulation, 5, 20 and 40 mins after onset of the stress/control manipulation and at the beginning of experimental day 2) using a Dinamap system (Critikon, Florida). Finally, to assess the activation of the hypothalamus-pituitary-adrenal (HPA) axis we measured the cortisol response of the participants by collecting saliva samples using Salivettes (Sarstedt, Germany) at six different time points (at the beginning of experimental day 1, directly after the memory encoding task, 5, 20 and 40 mins after the onset of the stress/control manipulation and at the beginning of experimental day 2). We later analysed cortisol concentrations from saliva using a luminescence immunoassay (IBL, Germany).

2.5 Memory task

2.5.1 Stimulus material

As stimulus material we used the same set of picture stimuli as in a previous study from our lab (Dandolo & Schwabe, 2018), however we added some new pictures, increasing the total number of pictures from 180 to 240. The pictures were taken either from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 1997) or from open internet platforms.

Half of the pictures showed emotionally negative objects or scenes, while the other half showed neutral images. On experimental day 2, after the recognition test, we asked participants to rate all pictures on two scales: valence (scale from 0 = negative to 100 = positive, with 50 = neutral) and arousal (scale from 0 = not arousing to 100 = very arousing). The valence ratings showed that negative pictures (mean = 24.15, SEM = 0.42) were indeed perceived as more negative than neutral pictures (mean = 56.30, SEM = 0.32; paired t-test: $t(157) = -30.66$, $p < 2.2e\text{-}16$, Cohen's $d = -2.99$). Likewise, the arousal ratings showed that negative pictures (mean = 46.54, SEM = 0.97) elicited more arousal than neutral pictures (mean = 7.20, SEM = 0.47; paired t-test: $t(157) = 25.83$, $p < 2.2e\text{-}16$, Cohen's $d = 1.67$). We created three different lists (List A, List B and List C), each containing 40 negative and

40 neutral pictures: List A and List B were related in the sense that each picture in List A had a corresponding picture in List B showing the same gist (e.g. mowing tractor) but different details (e.g. different brand, color, perspective and background). List C consisted of novel pictures which were not related to the other two lists. For half of the participants List A was used for encoding and List B as related lures in the recognition task, while the other half of participants received List B during encoding and List A was used as related lures in the recognition task.

2.5.2 Experimental day 1 - memory encoding

All participants performed three encoding runs on experimental day 1, shortly before the stress or control manipulation. In each encoding run participants viewed the 80 pictures of their encoding list (either A or B) on a computer screen for 2 seconds each in a random order. After each picture a fixation cross was shown for 1 second. We used MATLAB (www.mathworks.com) and the Psychophysics Toolbox extensions (Brainard, 1997) for picture presentation. Memory encoding was intentional, thus participants were instructed to memorize as many pictures as possible and were asked to freely recall the remembered pictures immediately after each of the three encoding runs. During free recall, participants were instructed to name each picture using a short title or short description which allowed the investigator to check off the respective picture on a list. In cases were given titles or descriptions were inconclusive, investigators asked for a more detailed description. The entire encoding session (3 encoding runs + 3 immediate free recalls) was on average about 30 mins long.

2.5.3 Experimental day 2 - memory testing: free recall

On the second experimental day, either 1d or 28d after encoding, participants first performed another free recall task. However, in this free recall task participants were not only asked to name each picture using a short title as they did on experimental day 1, but were additionally instructed to describe each picture in as much detail as possible. The experimenter had a list of details for each picture and checked off each detailed named by the participant. This enabled us to test if participants only remembered the gist, e.g. the short title they had given the picture, or if they could actually remember some details of the picture.

2.5.4 Experimental day 2 - memory testing: recognition task

After the detailed free recall task participants performed a recognition task. In this task they saw all 240 pictures in random order on a computer screen. 80 pictures were the old pictures encoded on experimental day 1, 80 pictures were their respective lure pictures showing the same gist but different details and 80 pictures showed completely novel scenes and objects. For each picture participants had to indicate if they had seen it on experimental day one or not, by pressing either a button for "yes" or a button for "no". Each picture was shown until the participant pressed either button. A confidence rating followed in those cases that were answered with "yes". In this rating participants indicated how confident they were about their positive answer on a 4-point scale (not at all confident, slightly confident, quite confident or very confident). Between each picture trial a fixation cross was shown for 1 second. Importantly, participants were informed before the start of the recognition task that some pictures may be similar to the original ones and were instructed to only press "yes" if they thought the picture was exactly the same as on experimental day 1.

2.6 Statistical analysis

2.6.1 Subjective and physiological measures

We calculated repeated measurement ANOVAs with the repeated-measurement factor Time Points and the between-subject factors Treatment (stress vs control), Interval (1d vs 28d) and Sex (female vs male) for the MDBF scales, blood pressure, pulse and cortisol data. All measurement time points from experimental day 1 were included in the respective ANOVAs (4 time points for the MDBF scales, 5 time points for the cortisol data and 6 time points for blood pressure and pulse data). Greenhouse-Geisser corrections were applied in case of a violation of the sphericity assumption.

2.6.2 Free recall data

Because participants showed a high inter-subject variability during memory encoding on day 1 (free recall scores of encoding run 3 ranged from 28.75% to 96.25%), we used the ratio "Day2/Day1-Run3" as dependent variable for the experimental day 2 recall data analysis. Thus, this recall ratio captures memory consolidation effects independent of initial individual encoding performance.

During the recall task on day 2, we additionally asked participants to try and recall as many details of the picture as possible. From this we calculated for each participant the average number of details named per picture, for each emotion category separately.
Both dependant variables (recall ratio; average number of details per picture) were subjected to ANOVAs with the between-subject factors Treatment (stress vs control), Interval (1d vs 28d) and Sex (female vs male) and the within-subject factor Emotion (negative pictures vs neutral pictures).

2.6.3 Recognition test - hits and confidence ratings

We first looked at the overall percentage of hits and calculated an ANOVA for this dependent variable with the between-subject factors Treatment, Interval and Sex and the within-subject factor Emotion.
In a next step, we used the confidence ratings which participants were asked to make whenever they answered the question, if they had seen the respective picture before, with 'Yes'. Although we used a 4-point scale (not at all confident, slightly confident, quite confident or very confident) during testing, we later classified the confidence ratings into two categories "high confidence hits" and "low confidence hits". Only those hits which were rated with "very confident", were classified as "high confidence hits", as these ratings can be compared to "remember" ratings in a remember-know procedure and best capture the memory process of recollection in recognition tests (Yonelinas, 2001). We combined the other three ratings (not at all confident, slightly confident, quite confident) as "low confidence hits", as these can be related to the "know" ratings in remember-know procedures and have been shown to capture the memory process of familiarity (Yonelinas, 2001). After classifying the hits, we first calculated the probability of a high confidence response for each participant per emotion category (Number of High Confidence Hits/Number of Old Pictures). We then calculated the probability of a low confidence response to an old picture, given the opportunity to make such a response (Number of Low Confidence Hits/[Number of Old Pictures - Number of High Confidence Hits]), thus this is the probability of making a low confidence response in those cases in which one was not very confident. These probabilities were subjected to two separate ANOVAs (one for high confidence hits and one for low confidence hits) with the factors Interval, Treatment, Sex and Emotion.

2.6.4 Recognition test - false alarms to related and novel pictures

We divided the false alarms ('Yes' answers to new pictures) in the recognition test into two categories: false alarms to related lure pictures and false alarms to novel pictures. We then calculated an ANOVA on the percentage of false alarms with the between-subject factors Treatment (stress vs control), Interval (1d vs 28d) and Sex (female vs male) and the within-subject factors Emotion (negative pictures vs neutral pictures) and Picture Typ (related pictures vs novel pictures). This allowed us to differentiate between a general loss in memory strength across time and a memory transformation process. If memory only showed a loss of strength across time this would lead to a main effect of Interval which is not influenced by the factor Picture Type as both false alarms to related and novel pictures would be affected in the same way. However, if a transformation from detailed to gist-like memories occurs across time, this would lead to a Picture Type $\times$ Interval interaction in the above ANOVA, as participants would especially show an increase in false alarms to related pictures which show the same gist as the old pictures, but a smaller increase of false alarms to novel pictures across time.

2.6.5 Recognition test - picture pair analysis

Another way to capture the transformation process is to look at the matching picture pairs in more detail. A matching pair is defined as an old picture and the respective corresponding related lure picture, carrying the same gist. We characterized the picture pairs depending on the combination of answers to the two related pictures. If participants correctly identified the old picture as old (hit) and, at the same time, had no problem rejecting the related lure picture as new (correct rejection) than one can expect their memory of the old picture to be detailed, thus these pairs were classified as "detailed pairs". However, in cases in which participants answered "Yes" to the related lure picture (false alarm), one can expect the memory of the old picture to be gist-like, as they identified the gist but did not recognize the different details. Thus, these pairs were classified as "transformed pairs", irrespective of the response to the old picture (either hit or miss). We classified pairs in which both the old picture (miss) and the related picture (correct rejection) were answered with "No" as "forgotten pairs", as in these cases participants seemed to have lost their memory for the old picture entirely.

The percentage of picture pairs of each category (detailed, transformed and forgotten) were analysed in three separate ANOVAs with the factors Interval, Treatment, Sex and

Emotion.

2.6.6 Relation between subsequent memory scores and cortisol increase

Using regression analysis we additionally tested the relation between the salivary cortisol increase (subtracting the baseline cortisol shortly before the stress manipulation from the cortisol value 20 mins after the onset of the stress manipulation) and most of the described subsequent memory scores (recall ratio, average number of details per picture, percentage of hits, probability of high confidence responses, probability of low confidence responses, percentage of detailed pairs, percentage of transformed pairs, percentage of forgotten pairs). We calculated correlation coefficients (R^2) and tested for significance of the models for both linear and quadratic regression lines, as some former studies have reported inverted-U shaped relationships of cortisol increase to subsequent memory scores (Andreano & Cahill, 2006, McCullough et al., 2015). The regression analyses were performed only in the stress groups, as only stressed participants can show a meaningful stress-related cortisol increase. In addition, we performed the regression analysis separately for the negative and neutral pictures, as well as separately for the female and male participants.

3 Results

3.1 Overall effective stress manipulation

3.1.1 Subjective stress ratings

Subjective stress ratings showed that participants in the stress condition rated the SECPT as significantly more difficult (Welch Two Sample t-test: $t(95.56) = 16.47$, $p < 2.2e\text{-}16$, Cohen's $d = 2.65$), unpleasant (Welch Two Sample t-test: $t(129.84) = 13.85$, $p < 2.2e\text{-}16$, Cohen's $d = 2.22$), painful (Welch Two Sample t-test: $t(87.12) = 21.34$, $p < 2.2e\text{-}16$, Cohen's $d = 3.43$), and stressful (Welch Two Sample t-test: $t(87.50) = 12.32$, $p < 2.2e\text{-}16$, Cohen's $d = 1.98$) as control participants rated the control condition. Additionally, results from the MDBF showed significant Treatment (stress vs control) × Time Point interactions for the elevated mood scale ($F(3, 450) = 1.65$, $p = 1.2e\text{-}05$, generalized $\eta^2 = 0.009$) and the calmness scale ($F(2.48, 372.03) = 7.68$, $p = 0.0002$, generalized $\eta^2 = 0.013$), with participants from the stress group showing lower mood and calmness scores at the time point measured directly after the stress manipulation compared to the control participants. As to be expected, we found no Treatment × Time Point interaction effect for the MDBF wakefulness scale.

3.1.2 Indicators of sympathetic nervous system activity

For the blood pressure and pulse measurements, we found significant Treatment × Time Point interactions (systolic blood pressure: $F(4.40, 672.79) = 7.68$, $p = 2.9e\text{-}16$, generalized $\eta^2 = 0.0293$; diastolic blood pressure: $F(2.41, 368.91) = 33.96$, $p = 1.97e\text{-}16$, generalized $\eta^2 = 0.0729$ and pulse: $F(3.12, 477.30) = 16.71$, $p = 1.22e\text{-}10$, generalized $\eta^2 = 0.0153$). As seen in Figure 1 these blood pressure and pulse measurements differed significantly at the time point during the SECPT/control manipulation but not at the other time points, reflecting the fast reaction of the sympathetic nervous system activity.

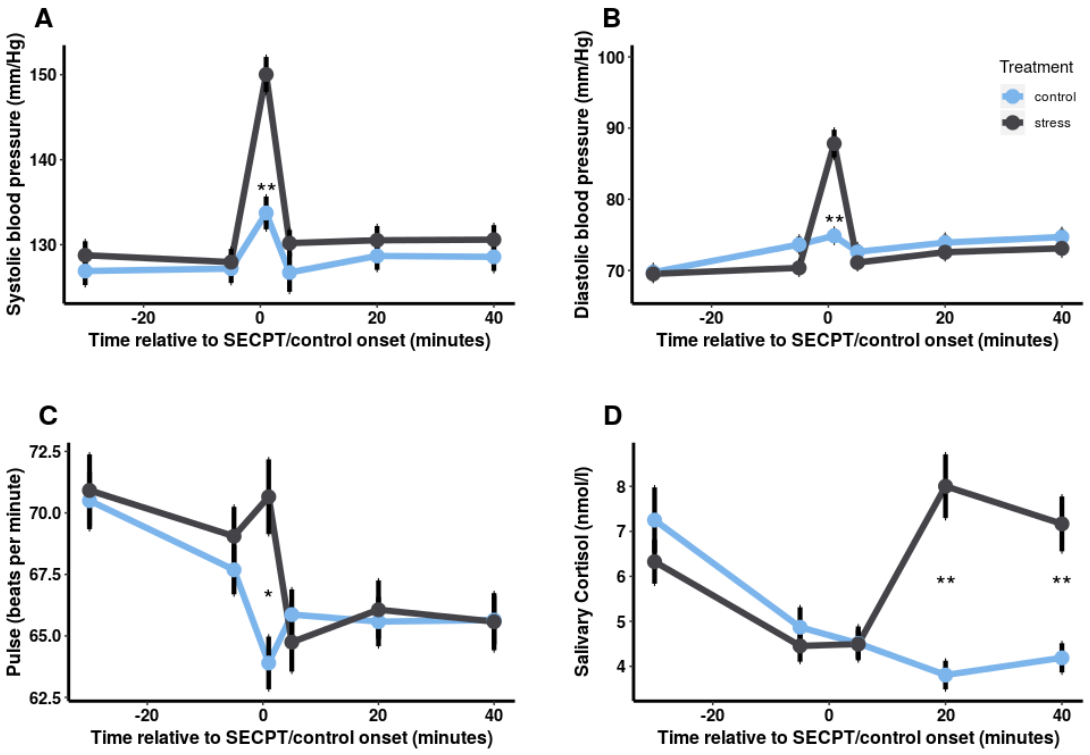


Figure 1: Stress Responses

3.1.3 Cortisol response

Importantly, there was also a Treatment × Time Point interaction effect for the salivary cortisol data ($F(1.57, 241.80) = 23.09$, $p = 2.13e\text{-}08$, generalized $\eta^2 = 0.0532$), with post-hoc tests indicating a significant difference between the cortisol levels of the stress group in comparison to the control group at the time points 20 mins and 40 mins after the beginning of the stress/control manipulation. Before and 5 mins after the treatment the cortisol levels did not differ between the groups, reflecting the slow reaction of the HPA axis.

3.1.4 Equally successful stress manipulation in both interval groups

Note, that we also included the factors Interval and Sex in all the above described ANOVAs. We did not find any main effects of Interval or interaction effects including the factor Interval (except for an Interval × Time Point interaction in the MDBF calmness scale, but post-hoc test were not significant), confirming that the stress manipulation was equally successful on experimental day 1 for both stress groups (1d-stress group and 28d-stress group). As to be expected, we found main effects of Sex for systolic blood pressure and cortisol levels, with overall higher values for male participant than female participants in

both measurements.

3.1.5 Cortisol responders and non-responders

In order to compare our stress response results to other studies, we classified participants into cortisol responders and non-responders, using a criterion established by Miller, Plessow, Kirschbaum, and Stalder (2013), according to which participants are labelled as responders, if they show a cortisol increase of at least 1.5 nmol/l from baseline to peak. As baseline we used the cortisol value measured shortly before the stress manipulation and as peak the cortisol value measurement 20 mins after the onset of the treatment. This classification showed that out of the 78 stress participants 38 were responders, equalling to a responder rate of 49%. Although, this responder rate is similar to some other studies using the SECPT, it is, nevertheless, slightly under the average responder rate reported for SECPT studies (about 60%; Schwabe & Schachinger, 2018). This could be related to the fact that, with our two cortisol measurements at about 20 mins and 40 mins after the SECPT, we might have slightly missed the actual peak of the cortisol response, which is expected 25-30 mins after the stressor onset (Schwabe & Schachinger, 2018, Kirschbaum, Pirke, & Hellhammer, 1993). Thus some of the participants who were just slightly under the cut-off of 1.5 nmol/l might have been classified as responders if we would have captured their actual peak response. Nevertheless, our more moderate responder rate benefits correlational analyses of the cortisol increase in the stress group with subsequent memory scores (Schwabe & Schachinger, 2018).

3.2 Successful memory encoding on experimental day 1

After three encoding runs, participants freely recalled in average 64.83% of the pictures with a standard deviation of 13.84%. We calculated an ANOVA with the factors Interval, Treatment, Sex and Emotion for the free recall data of the third run. There were no effects involving the factors Interval or Treatment, showing that, importantly, all four experimental groups (1d-stress group, 28d-stress group, 1d-control group, 28d-control group) learned the stimuli equally well on experimental day 1 before the stress/control manipulation. We did, however, find a main effect of Emotion ($F(1,150) = 81.35$, $p = 8.36e\text{-}16$, generalized $\eta^2 = 0.0870$) with participants recalling more negative (69.32%) than neutral (60.33%) pictures. There was also a main effect of Sex ($F(1,150) = 8.52$, $p = 4.05e\text{-}03$, generalized $\eta^2 = 0.0447$) with female participants (67.96%) recalling more pictures than male participants

(61.69%).

3.3 Post-encoding stress benefits the free recall of negative pictures in females

As expected there was a significant and large main effect of Interval ($F(1,150) = 426.34$, $p = 1.08e\text{-}45$, generalized $\eta^2 = 0.664$) for the free recall ratio, with participants in the 1d-groups recalling twice as many pictures as participants in the 28d-groups.

Interestingly, there was also a significant Treatment × Sex × Emotion interaction effect ($F(1,150) = 4.22$, $p = 4.16e\text{-}02$, generalized $\eta^2 = 0.0085$) and corresponding main effects of Treatment (more recalled pictures for stress than control participants: $F(1,150) = 6.76$, $p = 1.02e\text{-}02$, generalized $\eta^2 = 0.0304$), Sex (more recalled pictures for female than male participants: $F(1,150) = 5.11$, $p = 2.52e\text{-}02$, generalized $\eta^2 = 0.0232$) and Emotion (more negative recalled pictures than neutral recalled pictures: $F(1,150) = 9.61$, $p = 2.31e\text{-}03$, generalized $\eta^2 = 0.0191$). Follow-up analyses of the interaction effect showed that the Treatment effect was only significant for negative pictures (ANOVA for only negative pictures; main effect Treatment: $F(1,150) = 8.52$, $p = 4.05e\text{-}03$, generalized $\eta^2 = 0.0537$) but not for neutral pictures (ANOVA for only neutral pictures; main effect Treatment: $F(1,150) = 2.36$, $p = 0.127$, generalized $\eta^2 = 0.0154$). As Figure 2 shows, the difference between the stress and the control group for negative pictures was, in addition, influenced by the factor Sex: only female participants showed significant post-hoc t-tests, but not male participants. Thus, for female participants stress after encoding led to a better delayed recall performance of negative picture stimuli when compared to a control task after encoding. This, importantly, was the case for both the 1d-group and the 28d-group.

Despite this significant Treatment effect for negative pictures in the female group, we did not find any significant linear or quadratic regression models for the relation between the free recall ratio and the cortisol increase in the stress group, neither for negative pictures in females or males, nor for neutral pictures in females or males (all multiple $R^2 < 0.056$, all $F < 1.53$, all $p > 0.222$). Thus the stress effect for females on negative pictures can not directly be related to the cortisol increase after stress.

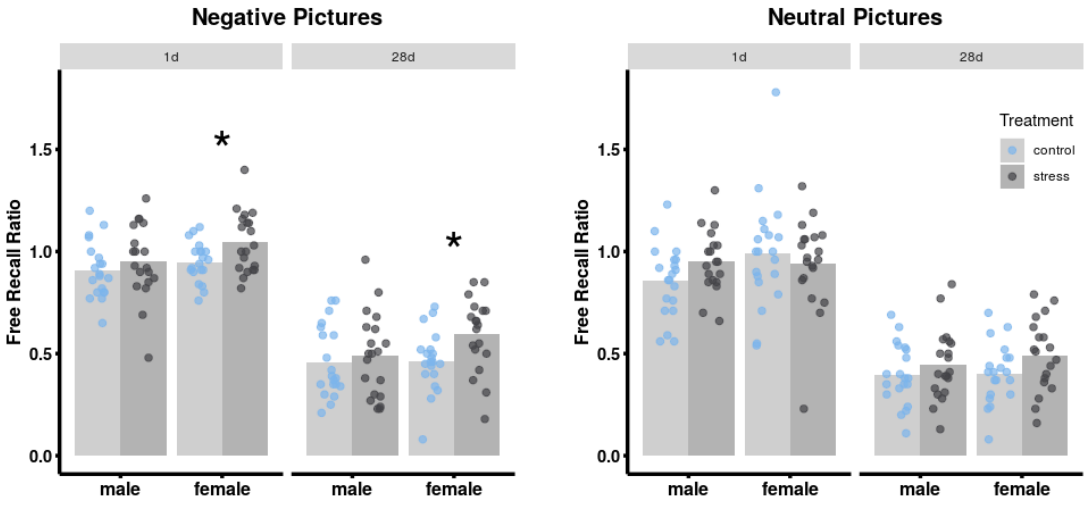


Figure 2: Ratio Hits

3.4 No effects of stress or cortisol increase on hits or recollection and familiarity scores in the recognition task

We did not find effects of stress or cortisol increase on the overall percentage of hits, nor did we find effects when dissociating the recognition memory into recollection and familiarity components based on confidence ratings.

3.4.1 Overall percentage of hits

From the recognition task data, we first analysed the overall percentage of hits for old pictures: we found a main effect of Interval (higher percentage of hits in the 1d-groups compared to the 28d-groups; $F(1,150) = 95.83$, $p = 8.34e\text{-}18$, generalized $\eta^2 = 0.3543$), a main effect of Emotion (higher percentage of hits for negative than for neutral pictures; $F(1,150) = 54.16$, $p = 1.13e\text{-}11$, generalized $\eta^2 = 0.0485$) and an Interval $\times$ Emotion interaction effect ($F(1,150) = 16.32$, $p = 8.52e\text{-}05$, generalized $\eta^2 = 0.0151$) which was explained by a somewhat larger emotion effect in the 28d-groups (paired t-test negative vs neutral: $t(78) = 6.43$, $p = 9.29e\text{-}09$, Cohen's $d = 0.72$) than in the 1d-groups (paired t-test negative vs neutral: $t(78) = 3.81$, $p = 0.00028$, Cohen's $d = 0.43$). Importantly, there were no effects of Treatment or Sex for the overall percentage of hits. There was also no significant linear or quadratic relationship between the overall percentage of hits and the cortisol increase in any of the performed regression analyses (all multiple $R^2 < 0.062$, all $F < 2.45$, all $p > 0.126$).

3.4.2 Probability of high confidence hits

For the probability of high confidence hits (a measure which can be related to recollection; Yonelinas, 2001), we found the same pattern of results as for the overall percentage of hits: a main effect of Interval ($F(1,150) = 144.79$, $p = 9.08e\text{-}24$, generalized $\eta^2 = 0.4652$), a main effect of Emotion ($F(1,150) = 75.73$, $p = 5.40e\text{-}15$, generalized $\eta^2 = 0.0475$) and an Interval × Emotion interaction effect ($F(1,150) = 5.58$, $p = 1.94e\text{-}02$, generalized $\eta^2 = 0.0037$) with post-hoc tests showing a larger emotion effect in the 28d-groups (paired t-test negative vs neutral: $t(78) = 7.61$, $p = 5.38e\text{-}11$, Cohen's $d = 0.86$) than the 1d-group (paired t-test: $t(78) = 4.50$, $p = 2.29e\text{-}05$, Cohen's $d = 0.51$). Importantly, we again did not find any effects of Treatment or Sex. And, once again, non of the linear or quadratic regression analyses of the probability of high confidence hits with the cortisol increase were significant (all multiple $R^2 < 0.101$, all $F < 3.31$, all $p > 0.077$).

3.4.3 Probability of low confidence hits

For the probability of low confidence hits (a measure which can be related to familiarity; Yonelinas, 2001), on the other hand, we only found an Interval × Emotion interaction effect ($F(1,150) = 7.71$, $p = 0.0061$, generalized $\eta^2 = 0.0120$), which was explained by significantly higher probabilities for negative than neutral low confidence hits in the 28d-groups (paired t-test: $t(78) = 3.34$, $p = 0.0013$, Cohen's $d = 0.38$) but no difference in the 1d-groups (paired t-test: $t(78) = -1.40$, $p = 0.1654$, Cohen's $d = -0.16$). However, for this variable we neither found a main effect of Interval, nor a Treatment effect, and Sex also showed no influence. Non of the regression analyses relating the probability of low confidence hits to the cortisol increase were significant (all multiple $R^2 < 0.047$, all $F < 1.03$, all $p > 0.317$).

3.5 No effects of post-encoding stress on the degree of memory transformation captured after four weeks

We captured the transformation process in three different ways. First, we asked participants to recall as many details as possible during the free recall task on experimental day 2, calculating the average number of details named per picture from this. Secondly, we compared false alarms of related pictures to false alarms of novel pictures. Third, we classified picture pairs into detailed, transformed and forgotten pairs (see methods).

3.5.1 Average number of details named in the free recall task

For the average number of details named, we found a main effect of Interval (more details named in the 1d-group compared to the 28d-group; $F(1,150) = 63.26$, $p = 4.05e\text{-}13$, generalized $\eta^2 = 0.2766$), a main effect of Sex (females named more details than males; $F(1,150) = 4.60$, $p = 3.36e\text{-}02$, generalized $\eta^2 = 0.0271$) and a main effect of Emotion (more details named for negative than for neutral pictures; $F(1,150) = 36.71$, $p = 1.06e\text{-}08$, generalized $\eta^2 = 0.0223$), indicating that memories became less detailed across time, and that this was also dependent on the emotion of the pictures and the sex of the participant.

Yet, for the average number of details there was no significant main effect of Treatment ($F(1,150) = 0.23$, $p = 0.635$, generalized $\eta^2 = 0.0014$). And, although there was a significant Interval × Treatment × Emotion interaction effect ($F(1,150) = 4.63$, $p = 0.0331$, generalized $\eta^2 = 0.0029$), we did not find Treatment effects in follow-up analyses in which we either calculated ANOVAs separately for negative and neutral pictures, or calculated ANOVAs separately for the 1d-group and the 28d-group.

There was also no linear or quadratic relationship between this average number of details and the cortisol increase in the stress group for any sex or either emotional category (all multiple $R^2 < 0.084$, all $F < 1.65$, all $p > 0.206$).

3.5.2 False alarms to related and novel pictures

For the false alarms we, importantly, found a significant Interval × Picture Type interaction effect ($F(1,150) = 55.29$, $p = 7.41e\text{-}12$, generalized $\eta^2 = 0.1103$; main effect Interval: $F(1,150) = 55.18$, $p = 7.72e\text{-}12$, generalized $\eta^2 = 0.1388$, main effect Picture Type: $F(1,150) = 272.74$, $p = 1.44e\text{-}35$, generalized $\eta^2 = 0.3796$). This replicates the finding from our former study using the same task (Dandolo & Schwabe, 2018) and again suggests, that across time, memories for the pictures are transformed from detailed memories to gist-like versions, as participants especially had problems identifying related pictures as new, but had fewer problems with novel pictures.

However, we found no main effect of Treatment for the false alarms. We did find a small but significant Interval × Treatment × Emotion × Picture Type interaction effect in the ANOVA: ($F(1,150) = 4.06$, $p = 4.57e\text{-}02$, generalized $\eta^2 = 0.0031$). As this four times interaction effect is hard to interpret we first calculated two follow up ANOVAs with the factors Interval, Treatment, Picture Type and Sex separately for negative and neutral

pictures. In both ANOVAs we found significant Interval × Picture Type interactions, main effects of Interval and main effects of Picture Type. However, we did not find any effects involving the factor Treatment. Thus, here, potential Treatment effects do not depend on the emotional category. We also calculated two follow-up ANOVAs separately for stressed and control participants, and found significant Interval × Picture Type effects in both groups. Therefore, the transformation effect, captured by this Interval × Picture Type interaction exists in both treatment groups, suggesting that it is not influenced by stress after encoding.

3.5.3 Detailed, forgotten and transformed picture pairs

Finally, we capture the transformation process with the picture pair analysis. We first looked at the percentage of detailed pairs and found a main effect of Interval ($F(1,150) = 146.95$, $p = 5.23e\text{-}24$, generalized $\eta^2 = 0.4593$), a main effect of Emotion ($F(1,150) = 69.05$, $p = 5.30e\text{-}14$, generalized $\eta^2 = 0.0577$) and an Interval × Emotion interaction effect ($F(1,150) = 11.38$, $p = 9.45e\text{-}04$, generalized $\eta^2 = 0.0100$) which was explained by a larger Emotion effect in the 28d-groups (negative pictures = 68.73%, neutral pictures = 59.81%; paired t-test: $t(78) = 6.99$, $p = 8.10e\text{-}10$, Cohen's $d = 0.79$) than in the 1d-groups (negative pictures = 89.93%, neutral pictures = 86.16%; paired t-test: $t(78) = 4.45$, $p = 2.87e\text{-}05$, Cohen's $d = 0.50$). Yet, importantly, we found no effects involving the factors Treatment or Sex.

For the forgotten pairs we found the same effects: main effect of Interval ($F(1,150) = 81.09$, $p = 9.08e\text{-}16$, generalized $\eta^2 = 0.3165$), main effect of Emotion ($F(1,150) = 44.36$, $p = 4.85e\text{-}10$, generalized $\eta^2 = 0.0408$), Interval × Emotion interaction effect ($F(1,150) = 16.89$, $p = 6.50e\text{-}05$, generalized $\eta^2 = 0.0159$) with a larger negative Emotion effect in the 28d-groups (negative pictures = 21.01%, neutral pictures = 29.43%; paired t-test: $t(78) = -6.07$, $p = 4.41e\text{-}08$, Cohen's $d = -0.68$) than the 1d-groups (negative pictures = 7.01%, neutral pictures = 9.00%; paired t-test: $t(78) = -2.99$, $p = 0.0037$, Cohen's $d = -0.34$). Again, there was no influence of Treatment or Sex.

Most importantly, for the transformed pairs we, once again, found a main effect of Interval ($F(1,150) = 55.59$, $p = 6.63e\text{-}12$, generalized $\eta^2 = 0.2254$) and a main effect of Emotion ($F(1,150) = 6.41$, $p = 1.24e\text{-}02$, generalized $\eta^2 = 0.0091$) and, this time, an Emotion × Sex interaction effect ($F(1,150) = 4.35$, $p = 3.88e\text{-}02$, generalized $\eta^2 = 0.0062$), with significantly more neutral transformed pairs than negative transformed pairs for female participants (paired t-test: $t(78) = -3.15$, $p = 0.0023$, Cohen's $d = -0.35$), but no effect

of Emotion for male participants (paired t-test: t(78) = -0.29, p = 0.7733, Cohen's d = -0.03). Yet, there were again no effects involving the factor Treatment.

We next performed linear and quadratic regression analyses in the stress group between the percentage of the different pair types (detailed, transformed, forgotten) and the cortisol increase, respectively. We did this separately for female and male participants and divided into neutral and negative picture pairs. We found a small but marginally significant positive linear relationship between the percentage of detailed pairs and the cortisol increase for male participants for negative picture pairs (multiple $R^2 = 0.10$, $F = 4.14$, $p = 0.049$). The stronger the cortisol increase in response to the stressor in males the more detailed the subsequent memory for neagtive pictures. Correspondingly, there was a significant negative linear relationship between the percentage of transformed pairs and the cortisol increase for male participants for negative picture pairs (multiple $R^2 = 0.12$, $F = 4.84$, $p = 0.034$). The stronger the cortisol increase the less were memories transformed. It is important to note here, that the significant result was only found when including male stress participants of both the 1d-stress group and the 28d-stress group. Performing the same analysis for the 1d-group and the 28d-group separately showed a trend in the 1d-group (multiple $R^2 = 0.19$, $F = 4.12$, $p = 0.058$) but not in the 28d-group (multiple $R^2 = 0.06$, $F = 1.07$, $p = 0.316$).

There were no significant regression models for neutral detailed or transformed picture pair percentages and cortisol increase in males (all multiple $R^2 < 0.062$, all $F < 1.96$, all $p > 0.170$). For females, non of the regression models for the picture pair percentages with the cortisol increase were significant (all multiple $R^2 < 0.030$, all $F < 0.94$, all $p > 0.337$). Also, non of the regression models with the percentage of forgotten pairs and cortisol increase were significant (all multiple $R^2 < 0.053$, all $F < 2.07$, all $p > 0.158$).

In summary, all three ways of capturing a possible transformation process (average number of details named in the recall task, comparison of false alarms in related picture and false alarms in novel pictures, picture pair analysis), suggest that, across time, memories are transformed from detailed versions to a gist-like versions. Importantly, we did not find an effect of the treatment group (stress vs. control) on this transformation process. Taking the varying amount of cortisol increase in the stress group into account, we found a small negative linear relationship between the cortisol increase and the transformation process for negative pictures in male participants. This effect, however, was only found in male participants of the 1d-stress group, but not the 28d-stress group.

3.6 Effects of emotional arousing stimuli on memory transformation

In our prior study using the same memory paradigm (Dandolo & Schwabe, 2018) we found an influence of emotional arousing stimuli on memory transformation: negative pictures were significantly more often transformed to gist-like versions than neutral pictures. On the other hand, neutral pictures were more often forgotten than negative ones. We interpreted this is in line with former accounts suggesting a trade-off between an overall superior memory for emotional material and reduced memory for contextual details of these emotional memories (Kensinger et al., 2007; Waring & Kensinger, 2011).

However, in the present study we could only replicate parts of this trade-off. We could replicate the effect, that neutral pictures were more often forgotten than negative pictures after 28 days (neutral forgotten pairs in the 28d-groups = 29.43%; negative forgotten pairs in the 28d-groups = 21.01%, paired t-test: t(78) = -6.07, p = 4.41e-08, Cohen's d = -0.68), thus the overall memory strength was again higher for emotional material. However, we did not find a higher number of transformed memories for negative pictures than neutral pictures after 28 days. In fact, for female participants we found the opposite effect with more neutral transformed pairs than negative transformed pairs (paired t-test: t(78) = -3.15, p = 0.0023, Cohen's d = -0.35), with no effect of Emotion for male participants. Thus, the effect which we found in Dandolo and Schwabe (2018), that negative pictures are more often transformed to gist-like versions than neutral pictures, could not be replicated here.

4 Discussion

4.1 Enhancement of memory strength by post-encoding stress

We found an enhancing effect of post-encoding stress on the subsequent recall of negative pictures in females. This results fits to the common opinion that stress after encoding benefits subsequent memory retrieval (Roozendaal, 2002, Schwabe et al., 2012; Cadle & Zoladz, 2015). However, our results also show that this stress effect is not robustly found in all circumstances but depends on many factors.

First of all, we only found stress effects for negative pictures. This result is in line with many former studies which also only found stress effects for emotionally arousing materials (e.g. Buchanan & Lovallo, 2001; Cahill et al., 2003; Payne et al., 2006; Payne et al., 2007; Smeets et al., 2008; Felmingham et al., 2012). There have been circumstances in which the beneficial stress effect was also found for neutral learning material, however most of these neutral materials were more complex stimuli such as stories (Andreano & Cahill, 2006), films (Beckner et al., 2006), or pictures accompanied by narratives (Preuss & Wolf, 2009). In studies using simple unrelated stimuli such as words or pictures, as is the case in the present study, the effect seems to be most prominently found for emotional material (but see McCullough and Yonelinas (2013)). This emotion effect can be explained by the beneficial influence of an interaction of the arousal-induced increases in noradrenergic activity in the amygdala during the encoding of emotional material and the stress-induced increase in stress related hormones (Payne et al., 2006; Payne et al., 2007).

Secondly, the enhancing effect of stress on overall memory strength was only found in females. The sex of the participants has been shown to have substantial influences on the neurobiology of emotion and memory and therefore should be considered in stress and memory research (Cahill, 2003; Cahill, 2006). Our results are in line with some studies which also only found enhancing stress effects in female participants (Payne et al., 2006; Felmingham et al., 2012), and also fit to studies which, although not explicitly testing for

sex differences, found stress effects using predominately female samples (93.33% females in Smeets et al. (2008); 73.11% females in Nielson and Powless (2007); 70.83% females in Cahill et al. (2003)). However, there have also been a number of studies which showed stress effects only in male participants (Andreano & Cahill, 2006; Yonelinas et al., 2011; McCullough & Yonelinas, 2013). Thus, although sex seems to be an important factor, results so far have not been conclusive. This might be explained by the female menstrual cycle stage, which has been shown to influence stress effects on emotional memory in the sense that stress only enhanced the recall of emotionally arousing words in females in the follicular menstrual cycle stage but not in females in the luteal phase (Zoladz et al., 2015). Thus, stress effects in females might be found in studies which tested a majority of females in the follicular stage, while studies showing no effects in females might have tested a sample dominated by females in the luteal phase. This, however, still does not explain the varying results found in males. Thus, further research is needed to fully explain the effects of sex and sex hormones in the stress and memory research.

Thirdly, we only found the enhancing stress effect for the recall data and not for the percentage of hits in the recognition task. This is line with Zoladz et al. (2015) who also found stress effects in recall but not recognition scores and fits to studies which exclusively tested free recall (Andreano & Cahill, 2006; Felmingham et al., 2012; Cahill et al., 2003). One explanation might be that recognition tests are more prone to ceiling effects than free recall tasks. In the present study a ceiling effect could explain the missing stress effect in the 1d-group, as the average percentage of hits was 91.39% (SD = 9.74) in the 1d-groups leaving little room for improvement. However, data from the 28d-groups (72.28% hits; SD = 16.37) can not as easily be explained by a ceiling effect. Another difference between recall and recognition is that recognition can be based on either recollection or familiarity processes, while free recall mostly requires recollection (Eichenbaum et al., 2007). A recent study has shown that cortisol released after post-encoding stress influences familiarity in a linear relationship, while recollection is related to the stress response in a non-monotonic inverted-U-shape fashion (McCullough et al., 2015). Yet, even when separating our data into recollection and familiarity components we could not find any influences of stress or stress hormones, neither in linear nor non-monotonic relationships. One explanation could be, that our confidence ratings did not capture the difference between recollection and familiarity as reliably as a remember-know procedure would, although the two have been shown to be highly equivalent (Yonelinas, 2001). Further research, thus, might benefit from distinguishing recollection and familiarity in a more reliable way.

Lastly, although we found a significant difference between the stress group and the control group in the recall of negative pictures in females, there was no significant linear or quadratic correlation between the recall scores and the cortisol increase. Cahill et al. (2003) showed the same pattern of results and suggested that this might be due to the fact that they did not captured the peak of the cortisol response. This might also be an explanation for the present results, as we might have slightly missed the peak cortisol response. An alternative explanation might be that, as cortisol alone does not represent all dimensions of a psychosocial stressor, combinations of cortisol response, sympathetic nervous system activity and/or subjective feelings may show significant correlations with memory scores (see Abercrombie, Speck, & Monticelli, 2006 for a necessary combination of stress hormones and subjective feelings).

4.2 No clear effects of post-encoding stress on memory transformation after four weeks

In the present study we did not find any effects of post-encoding stress or cortisol increase on the degree of memory transformation tested after four weeks of systems consolidation. This is at odds with a recent rodent study which showed that moderate stress during encoding led to detailed memories for a long period of time while high amounts of stress, on the other hand, led to earlier memory generalization sometime between 1 to 2 weeks (Pedraza et al., 2016). It also does not align with human studies which, although not explicitly testing the time course of memory transformation, nevertheless showed an over-reliance on gist-like memories after pre -or post-encoding stress (Payne et al., 2002; Pardilla-Delgado et al., 2016).

It is important to note here, that the stress response to the SECPT in the present study was in general rather moderate in comparison to other stress studies. When classifying stress participants into stress responders and non-responders according to the commonly used threshold of 1.5 nmol/l cortisol increase from baseline to peak (Miller et al., 2013), only 49% were classified as responders. Although there were also a few participants with a strong cortisol response, the majority of stress participants showed a moderate cortisol response, or even no response. Thus, when testing for quadratic relationships between the cortisol increase and the memory transformation scores, the small amount of participants with high stress responses might have limited the power to find an inverted-U-shape relationship as suggested in Pedraza et al. (2016). Therefore, future studies might profit from stressors

eliciting higher stress responses (e.g. the Trier Social Stress Test, Kirschbaum et al., 1993) or from a pharmacological administration of cortisol.

We did find a significant linear relationship between the cortisol increase and the percentage of transformed negative memories for male participants in the 1d-group, but not in the 28d-group. The direction of the linear relationship was negative, meaning that the higher the cortisol response in reaction to the stressor the less gist-like (and more detailed) the memories were after 24h hours. This might be related to the suggestion of Pedraza et al. (2016), that moderate stress will create a detailed memory. Yet, it does not fit to their reported time course, as they suggested that this detailed memory created by the influence of moderate stress would persist over a long period of time. We did, however, not find this relationship in the 28d-group. Thus, further research is needed to interpret this finding in a broader framework.

Future studies might also benefit from testing the time course of memory transformation at more different time points. It could be that our interval of 28 days does not capture potential variations in the time course of memory transformation induced by moderate (or high) stress. For example, it could be that stress effects are observable at later time intervals when memories encoded under normal circumstances, i.e. without stress, start fading more rapidly.

4.3 Conclusion

All in all, our study, firstly, contributes to the general discussion about the influence of post-encoding stress on subsequent memory. Although the overall dominant opinion has been that post-encoding stress enhances subsequent memory (Schwabe et al., 2012), there have been contradictory findings (e.g. Trammell & Clore, 2014). Our data suggests that it is important to take several factors into consideration, such as the sex of the participants in combination with the emotionality of the material. It is also important to consider the complexity of the material (simple stimuli vs. complex narratives) and the typ of memory retrieval (recall vs recognition), as well as the strength of the stress response. It requires further studies to understand under which circumstances and conditions the enhancing post-encoding stress effect on memory can be reliably found and replicated.

Secondly, our study shows that stress does not have a general robust effect on the time course of memory transformation, as we did not find any clear results regarding different memory transformation measures. However, we can only draw conclusions for moderate

stress, as we did not have enough participants with high stress responses, and only for a time interval of four weeks. I suggest to add more time points in future studies as well as repeating the study with a more robust stress test or pharmacological manipulations.

References

Abercrombie, H. C., Speck, N. S., & Monticelli, R. M. (2006). Endogenous cortisol elevations are related to memory facilitation only in individuals who are emotionally aroused. *Psychoneuroendocrinology*, *31*, 187-196.

Andreano, J. M., & Cahill, L. (2006). Glucocorticoid release and memory consolidation in men and women. *Psychological Science*, *17*, 466-470.

Beckner, V. E., Tucker, D. M., Delville, Y., & Mohr, D. C. (2006). Stress facilitates consolidation of verbal memory for a film but does not affect retrieval. *Behavioral Neuroscience*, *120*, 518-527.

Brainard, D. H. (1997). The psychophysics toolbox. *Spat Vis*, *10*, 433-436.

Buchanan, T. W., & Lovallo, W. R. (2001). Enhanced memory for emotional material following stress-level cortisol treatment in humans. *Psychoneuroendocrinology*, *26*, 307-317.

Buchanan, T. W., Tranel, D., & Adolphs, R. (2006). Impaired memory retrieval correlates with individual differences in cortisol response but not autonomic response. *Learning and Memory*, *13*, 382-387.

Cadle, C. E., & Zoladz, P. R. (2015). Stress time-dependently influences the acquisition and retrieval of unrelated information by producing a memory of its own. *Frontiers in Psychology*, *6*, 910.

Cahill, L. (2003). Sex-related influences on the neurobiology of emotionally influenced memory. *Annals of the New York Academy of Sciences*, *985*, 163-173.

Cahill, L. (2006). Why sex matters for neuroscience. *Nature Reviews Neuroscience*, *7*, 477-484.

Cahill, L., Gorski, L., & Le, K. (2003). Enhanced human memory consolidation with post-learning stress: interaction with the degree of arousal at encoding. *Learning and Memory*, *10*, 270-274.

Dandolo, L. C., & Schwabe, L. (2018). Time-dependent memory transformation along the hippocampal anterior-posterior axis. *Nature Communications*, *9*, 1205.

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, *30*, 123-152.

Felmingham, K. L., Tran, T. P., Fong, W. C., & Bryant, R. A. (2012). Sex differences in emotional memory consolidation: the effect of stress-induced salivary alpha-amylase and cortisol. *Biological Psychology*, *89*, 539-544.

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2007). Effects of emotion on memory specificity: Memory trade-offs elicited by negative visually arousing stimuli. *Journal of Memory and Language*, *56*, 575-591.

Kirschbaum, C., Kudielka, B. M., Gaab, J., Schommer, N. C., & Hellhammer, D. H. (1999). Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus-pituitary-adrenal axis. *Psychosomatic Medicine*, *61*, 154-162.

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The 'trier social stress test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, *28*, 76-81.

Kuhlmann, S., Piel, M., & Wolf, O. T. (2005). Impaired memory retrieval after psychosocial stress in healthy young men. *The Journal of Neuroscience*, *25*, 2977-2982.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). *International affective picture system (iaps): technical manual and affective ratings* [Book]. Gainesville: National Institute of Mental Health Center for the Study of Emotion and Attention.

McCullough, A. M., Ritchey, M., Ranganath, C., & Yonelinas, A. P. (2015). Differential effects of stress-induced cortisol responses on recollection and familiarity-based recognition memory. *Neurobiology of Learning and Memory*, *123*, 1-10.

McCullough, A. M., & Yonelinas, A. P. (2013). Cold-pressor stress after learning enhances familiarity-based recognition memory in men. *Neurobiology of Learning and Memory*, *106*, 11-17.

Miller, R., Plessow, F., Kirschbaum, C., & Stalder, T. (2013). Classification criteria for distinguishing cortisol responders from nonresponders to psychosocial stress: evaluation of salivary cortisol pulse detection in panel designs. *Psychosomatic Medicine*, *75*, 832-840.

Nielson, K. A., & Powless, M. (2007). Positive and negative sources of emotional arousal enhance long-term word-list retention when induced as long as 30 min after learning. *Neurobiology of Learning and Memory*, *88*, 40-47.

Pardilla-Delgado, E., Alger, S. E., Cunningham, T. J., Kinealy, B., & Payne, J. D. (2016). Effects of post-encoding stress on performance in the drm false memory paradigm.

Learning and Memory, *23*, 46-50.

Payne, J. D., Jackson, E. D., Hoscheidt, S., Ryan, L., Jacobs, W. J., & Nadel, L. (2007). Stress administered prior to encoding impairs neutral but enhances emotional long-term episodic memories. *Learning and Memory*, *14*, 861-868.

Payne, J. D., Jackson, E. D., Ryan, L., Hoscheidt, S., Jacobs, J. W., & Nadel, L. (2006). The impact of stress on neutral and emotional aspects of episodic memory. *Memory*, *14*, 1-16.

Payne, J. D., Nadel, L., Allen, J. J., Thomas, K. G., & Jacobs, W. J. (2002). The effects of experimentally induced stress on false recognition. *Memory*, *10*, 1-6.

Pedraza, L. K., Sierra, R. O., Boos, F. Z., Haubrich, J., Quillfeldt, J. A., & Alvares, L. D. (2016). The dynamic nature of systems consolidation: Stress during learning as a switch guiding the rate of the hippocampal dependency and memory quality. *Hippocampus*, *26*, 362-371.

Preuss, D., & Wolf, O. T. (2009). Post-learning psychosocial stress enhances consolidation of neutral stimuli. *Neurobiology of Learning and Memory*, *92*, 318-326.

Rohleder, N., & Kirschbaum, C. (2006). The hypothalamic-pituitary-adrenal (hpa) axis in habitual smokers. *International Journal of Psychophysiology*, *59*, 236-243.

Roozendaal, B. (2002). Stress and memory: opposing effects of glucocorticoids on memory consolidation and memory retrieval. *Neurobiology of Learning and Memory*, *78*, 578-595.

Schwabe, L., Haddad, L., & Schachinger, H. (2008). Hpa axis activation by a socially evaluated cold-pressor test. *Psychoneuroendocrinology*, *33*, 890-895.

Schwabe, L., Joels, M., Roozendaal, B., Wolf, O. T., & Oitzl, M. S. (2012). Stress effects on memory: an update and integration. *Neuroscience and Biobehavioral Reviews*, *36*, 1740-1749.

Schwabe, L., & Schachinger, H. (2018). Ten years of research with the socially evaluated cold pressor test: Data from the past and guidelines for the future. *Psychoneuroendocrinology*, *92*, 155-161.

Sekeres, M. J., Moscovitch, M., & Winocur, G. (2017). Mechanisms of memory consolidation and transformation. In N. Axmacher & B. Rasch (Eds.), *Cognitive neuroscience of memory consolidation* (p. 17-44). Cham: Springer International Publishing.

Smeets, T., Otgaar, H., Candel, I., & Wolf, O. T. (2008). True or false? memory is differentially affected by stress-induced cortisol elevations and sympathetic activity at consolidation and retrieval. *Psychoneuroendocrinology*, *33*, 1378-1386.

Steyer, R., Schwenkmezger, P., Notz, P., & Eid, M. (1994). Testtheoretische analysen des mehrdimensionalen befindlichkeitsfragebogens (mdbf). *Diagnostica*, *40*, 320-328.

Trammell, J. P., & Clore, G. L. (2014). Does stress enhance or impair memory consolidation? *Cognition and Emotion*, *28*, 361-374.

Waring, J. D., & Kensinger, E. A. (2011). How emotion leads to selective memory: Neuroimaging evidence. *Neuropsychologia*, *49*, 1831-1842.

Winocur, G., & Moscovitch, M. (2011). Memory transformation and systems consolidation. *Journal of the International Neuropsychological Society*, *17*, 766-780.

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

Wolf, O. T. (2009). Stress and memory in humans: twelve years of progress? *Brain Research*, *1293*, 142-154.

Yonelinas, A. P. (2001). Consciousness, control, and confidence: the 3 cs of recognition memory. *Journal of experimental psychology. General*, *130*, 361-379.

Yonelinas, A. P., Parks, C. M., Koen, J. D., Jorgenson, J., & Mendoza, S. P. (2011). The effects of post-encoding stress on recognition memory: examining the impact of skydiving in young men and women. *Stress*, *14*, 136-144.

Zoladz, P. R., Peters, D. M., Cadle, C. E., Kalchik, A. E., Aufdenkampe, R. L., Dailey, A. M., ... Rorabaugh, B. R. (2015). Post-learning stress enhances long-term memory and differentially influences memory in females depending on menstrual stage. *Acta Psychologica*, *160*, 127-133.

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR PSYCHOLOGIE UND
BEWEGUNGSWISSENSCHAFT
Institut für Bewegungswissenschaft
Institut für Psychologie

Erklärung gemäß *(bitte Zutreffendes ankreuzen)*

☐ **§ 4 (1c) der Promotionsordnung des Instituts für Bewegungswissenschaft der Universität Hamburg vom 18.08.2010**

☐ **§ 5 (4d) der Promotionsordnung des Instituts für Psychologie der Universität Hamburg vom 20.08.2003**

Hiermit erkläre ich,

_____ (Vorname, Nachname),

dass ich mich an einer anderen Universität oder Fakultät noch keiner Doktorprüfung unterzogen oder mich um Zulassung zu einer Doktorprüfung bemüht habe.

_____ _____

Ort, Datum Unterschrift

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR PSYCHOLOGIE UND
BEWEGUNGSWISSENSCHAFT
Institut für Bewegungswissenschaft
Institut für Psychologie

Eidesstattliche Erklärung nach *(bitte Zutreffendes ankreuzen)*

☐ **§ 7 (4) der Promotionsordnung des Instituts für Bewegungswissenschaft der Universität Hamburg vom 18.08.2010**

☐ **§ 9 (1c und 1d) der Promotionsordnung des Instituts für Psychologie der Universität Hamburg vom 20.08.2003**

Hiermit erkläre ich an Eides statt,

1. dass die von mir vorgelegte Dissertation nicht Gegenstand eines anderen Prüfungsverfahrens gewesen oder in einem solchen Verfahren als ungenügend beurteilt worden ist.
2. dass ich die von mir vorgelegte Dissertation selbst verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und keine kommerzielle Promotionsberatung in Anspruch genommen habe. Die wörtlich oder inhaltlich übernommenen Stellen habe ich als solche kenntlich gemacht.

_____ _____

Ort, Datum Unterschrift