THE DYNAMIC NATURE OF OUR MEMORIES

Experimental Investigations using functional Magnetic Resonance Imaging

DONDERS
SERIES

Wei Liu

THE DYNAMIC NATURE OF OUR MEMORIES

Experimental Investigations using

functional Magnetic Resonance Imaging

Wei Liu

The work described in this thesis was carried out at the Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands, financially supported by a Ph.D. fellowship from Chinese Scholarship Council (NO.201606990020) awarded to Mr. Wei Liu and Prof. dr. Guillén Fernández.

Wei Liu (刘威)
The dynamical nature of our memories: Experimental Investigations using functional Magnetic Resonance Imaging

ISBN: 978-94-6284-232-8

Cover design
Liaomo Design Company & Wei Liu (Inspired by ***The Persistence of Memory***, a Painting by ***Salvador Dalí***)

Layout
Puur*M Vorm & Idee

Print
Gildeprint

Paranimfen
Yan-nan Zhu & Yingjie Shi

© Copyright Wei Liu, 2020

THE DYNAMIC NATURE OF OUR MEMORIES

Experimental Investigations using functional Magnetic Resonance Imaging

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op woensdag 28 oktober 2020
om 13.30 uur precies

door

Wei Liu
geboren op 11 mei 1991
te Changsha (China)

Promotor:
Prof. dr. Guillén S.E Fernández

Copromotor:
Dr. Nils Kohn

Manuscriptcommissie:
Prof. dr. Francesco P. Battaglia
Prof. dr. Jaap M.J. Murre (Universiteit van Amsterdam)
Dr. Emma Sprooten

THE DYNAMIC NATURE OF OUR MEMORIES

Experimental Investigations using
functional Magnetic Resonance Imaging

Proefschrift

to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. J.H.J.M. van Krieken,
according to the decision of the Council of Deans
to be defended in public on Wednesday October 28, 2020
at 13.30 hours

by

Wei Liu
born on May 11, 1991
in Changsha (China)

Supervisor:
Prof. dr. Guillén S.E Fernández

Co-supervisor:
Dr. Nils Kohn

Doctoral Thesis Committee:
Prof. dr. Francesco P. Battaglia
Prof. dr. Jaap M.J. Murre (Universiteit van Amsterdam)
Dr. Emma Sprooten

TABLE OF CONTENTS

CHAPTER 1 General Introduction 11

CHAPTER 2 Hippocampal-medial prefrontal event segmentation and integration contribute to episodic memory formation 31

CHAPTER 3 The dynamic transition between neural states is associated with the flexible use of memory 69

CHAPTER 4 Probing the neural dynamics of mnemonic representations after the initial consolidation 105

CHAPTER 5 Common neural and transcriptional correlates of inhibitory control underlie emotion regulation and memory control 159

CHAPTER 6 General Discussion 195

APPENDIX
References 210
Research data management 232
Summary of thesis 234
List of publications 238
Academic portfolio DGCN 240
Curriculum vitae 242
Acknowledgments 244
Donders Graduate School for Cognitive Neuroscience 248

Chapter 1

General Introduction

In the current thesis, I set out to investigate the central question: *How are memories dynamically represented in our brains?* The introduction begins with an overview of what is known about the neural correlates of human episodic memories, such as the involved brain networks and mental operations. Then I will elaborate on two types of dynamic processes that we investigated in this thesis and will highlight the questions that motivated the experiments presented in **Chapters 2-5.** The critical methodological approaches that we used to answer these questions will be presented separately in **Box 1-4**. At the end of the introduction, I will give a summary and an outline of this thesis.

1.1. General Introduction

Historically, people hold the idea that memories are static entities. After a specific experience, some aspects of the brain endure an off-line, permanent, physical change (Schacter, 2012; Semon, 1923). This kind of memory representation in brains was described using the term "*engram*" or "*memory trace*." The idea is that these representations can be used as the foundations of memory retrieval (Josselyn et al., 2015; Tonegawa et al., 2015). Recent breakthroughs in neuroscience (e.g., optogenetics (Deisseroth, 2011; Fenno et al., 2011)) largely enhanced our understanding of "*engrams*" as the basic physical unit of memory (Josselyn & Tonegawa, 2020).

However, the investigations of engrams in rodents, although fascinating, ignored one of the critical features of episodic memory: its *dynamic nature*. Specifically, most engrams studies assumed that (1) each episodic event is stored in one specific engram for later retrieval. This one-to-one mapping between memory and engram allows memory processing (e.g., encoding and retrieval) of one particular event to avoid affecting another event. (2) physical changes (e.g., synaptic plasticity) during memory formation and consolidation determine whether specific memories can be retrieved later and how strong they are. Here, we challenged these assumptions and investigated two corresponding principles of episodic memories dynamics: (1) ***process dynamics*** (**Figure 1.1A**): we asked whether mnemonic processing can be viewed as processing units that are close in time and how interactions between these units are relevant for memory performance **(Chapters 2 and 3)**. (2) ***strength dynamics*** (Figure 1.1B): we probed whether the strength of each memory trace can be modified dynamically and if their perceived emotional intensity changes accordingly **(Chapters 3 and 4)**.

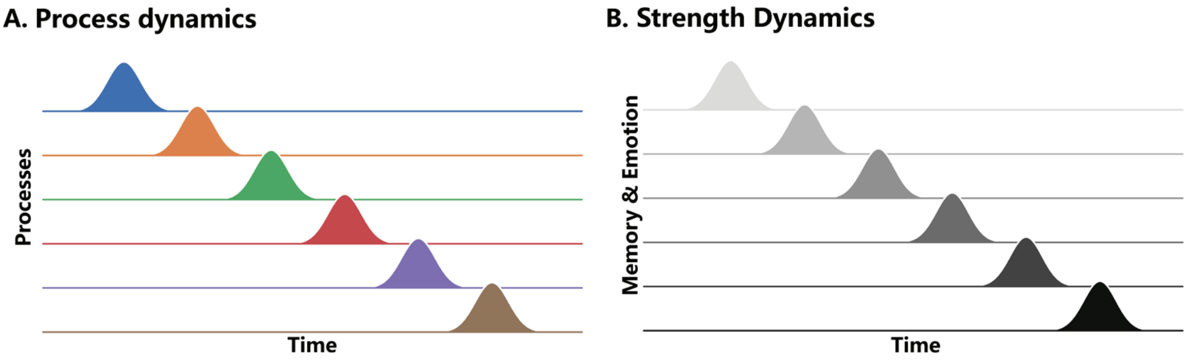


Figure 1.1 Schematic display of two types of memory dynamics. **(A)** Process dynamics: we argue that there are processing units within the temporal sequence of mnemonic processing. A unit can be a memory operation (e.g., encoding or retrieval), which is linked to a particular neural state and can interact with each other, leading to different behavioral consequences. **(B)** Strength dynamics: we propose that for each memory trace, its memory strength can still be modified after formation, and its emotional intensity can be adjusted during this dynamic process.

1.2 Episodic memory network in the human brain revealed by neuroimaging

In this thesis, we mainly used functional neuroimaging in humans to study the dynamics of episodic memory. Therefore, it is critical to first gain a big picture of the brain areas involved in episodic memory. Since the earliest days of functional neuroimaging investigation of human memory, the central scientific question is: how the brain builds memories (i.e., encoding), and later accesses these memories (i.e., retrieval). For example, if it is possible to predict whether a given experience would be remembered or forgotten based on neuroimaging measures during encoding (i.e., *subsequent memory effect* (Brewer et al., 1998; Wagner et al., 1998))? Can we detect the differences between successful and unsuccessful retrieval at the neural level (i.e., *retrieval success effect* (Buckner et al., 1998))? The answers to these questions depend on the type of the stimulus (e.g., visual or auditory) of the initial study (Gottlieb et al., 2010), and how the memory is tested (e.g., recognition or recall test) (Frithsen & Miller, 2014; Kim, 2013; McDermott et al., 2009; Otten, 2007; Yonelinas et al., 2005). In this section, we will only discuss the network of brain regions that show both encoding and retrieval-related activation and are consistently involved in different task designs and stimuli.

In prominent theories (Gilmore et al., 2015; Rugg & Vilberg, 2013; Wagner et al., 2005), human episodic memory network included three major sub-systems (**Figure 1.2**): (1) medial temporal lobe (MTL) system (e.g., hippocampus, parahippocampal cortex, and entorhinal cortex); (2) prefrontal system (e.g., inferior frontal gyrus (IFG) and medial prefrontal cortex; and (3) parietal system (e.g., inferior parietal lobule (IPL), precuneus, and angular gyrus (AG)). These regions were thought to work together as the memory circuit, but this hypothesis has not been investigated comprehensively until recently. Combining lesion data, resting-state fMRI data from healthy participants and patients with Alzheimer's disease, and information on the locations of brain stimulation sites reported to enhance memory, one study provided a circuit perspective into the human episodic network. Ferguson and co-workers suggested that the junction of the presubiculum and retrosplenial cortex acts as a hub of the human memory system to link three sub-systems together (Ferguson et al., 2019).

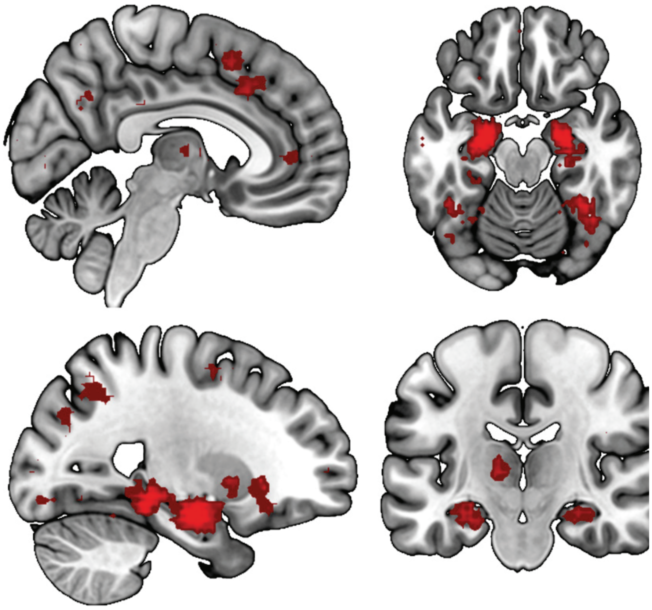


Figure 1.2. Episodic memory network in the human brain. The map was generated by the Neurosynth (neurosynth.org) based on 124 human fMRI studies. Keywords "encoding and/or retrieval" were used to search for relevant studies. The network mainly includes the medial temporal lobe, prefrontal, and parietal system.

1.3 Pattern reinstatement supports memory retrieval.

We already summarized *where* in the brain, the memory dynamics could happen, and now we focused on *how* these areas may support memory processing. One crucial neuro-cognitive process performed by the episodic memory network is *pattern reinstatement* (or *pattern reactivation*)(**Figure 1.3**)(Kent & Lamberts, 2008; Xue, 2018). Endel Tulving proposed memory retrieval as "mental time travel": during memory recall of a specific event, one is "transported" back to the situation in which that event took place (Tulving, 1984). Indeed, substantial neuroscientific evidence suggests that retrieving memories relies on the brain's ability to reactivate neural activity patterns that were present when the memory was initially experienced (Janice Chen et al., 2017; Kosslyn et al., 1997; S.-H. Lee et al., 2019; O'Craven & Kanwisher, 2000; Polyn et al., 2005; Wheeler et al., 2000; Maria Wimber et al., 2015) (**Figure 1.3A**). In humans, pattern reinstatement can be captured by high-resolution fMRI and Multi-Voxel Pattern Analysis (MVPA) (**Box 1**). Due to the spatial limitation of fMRI, the reinstatement of activity patterns can only be observed across voxels, which are not biologically-valid measurement units. Recent evidence from human single-neuron recording demonstrated a similar process in terms of temporal reinstatement, at least in the medial temporal lobe (Vaz et al., 2020). In rodents, memory-related individual neurons can be experimentally tagged, imaged, manipulated, and even ablated (Han et al., 2007, 2009; X. Liu et al., 2012; Ramirez et al., 2013, 2015; Vetere et al., 2019) **(Figure 1.3B)**. These memory-related neurons are not located within one brain region but distributed across multiple brain regions (Roy et al., 2019). They are also functionally connected and activated simultaneously by the same experience, which is consistent with the human data showing that pattern reinstatement happens in multiple regions of the memory network (Janice Chen et al., 2017; Xue, 2018).

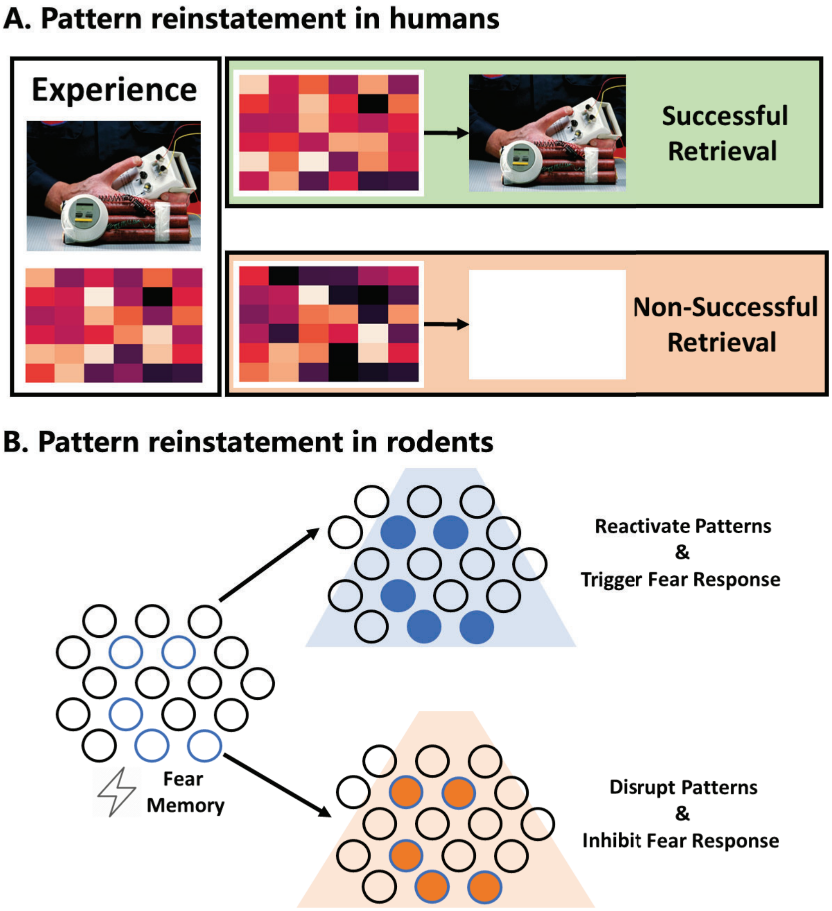


Figure 1.3 **Pattern reinstatement supports memory retrieval. (A)** Pattern reinstatement in humans. During the experience, brain regions such as the hippocampus may use a specific activity pattern to represent an individual memory. The same pattern will re-emerge in the same brain region during successful retrieval. The same pattern will not present during unsuccessful retrieval. **(B)** Pattern reinstatement in rodents. Memory-related neurons can be labeled during a specific experience (e.g., fear). Manipulation methods such as optogenetics can reactivate these neurons, leading to the retrieval of the experience or suppress these neurons, disrupting the retrieval process.

Box 1 Multivoxel Pattern Analysis (MVPA) of fMRI data in memory research

The early ***univariate analysis*** of fMRI data focused on the relationship between Blood-Oxygen-Level-Dependent (BOLD) signals within specific regions and mnemonic processing. This approach largely neglected the distributed nature of information representation over voxels within regions and even across different regions (i.e., functional brain networks). As opposed to the univariate analysis, ***Multivoxel Pattern Analysis (MVPA)*** considers finer-grained spatial patterns over voxels and tries to extract the information they encode jointly via machine-learning approaches (J. D. Cohen, Daw, Engelhardt, Hasson, Li, Niv, Norman, Pillow, Ramadge, Turk-Browne, & others, 2017).

The most important function of MVPA methods in memory research is that they provide a neural-based index, tracking memory-related *pattern reinstatement*. For example, MVPA can capture the cue-induced awake retrieval (Janice Chen et al., 2017; S.-H. Lee et al., 2019), targeted memory reactivation during sleep (Shanahan et al., 2018), post-encoding memory consolidation (Gerlicher et al., 2018; M. J. Gruber et al., 2016), and memory replay (Schuck & Niv, 2019).

There are two most common forms of MVPA: ***classifier-based MVPA*** and ***similarity-based MVPA***. ***Classifier-based MVPA*** is a machine learning-based approach (**Figure 1.4A**). Brain activation patterns and memory content (e.g., pictures of humans, animals, or houses…) are used to train classifiers and then applied on activation patterns related to new input. The trained classifier can be used to decode the mental memory trace for each given trial. For example, during the successful retrieval of specific visual images, content-specific classifiers can accurately predict the categories of retrieved images (Kerrén et al., 2018; Polyn et al., 2005). If the decoding can be performed at an accuracy higher than the chance level, then it is the neural evidence for *pattern reinstatement*. The second major type of MVPA (i.e., ***similarity-based MVPA***) focuses on the similarity measure of two or more activation patterns (**Figure 1.4B**). Brain activation patterns during perception or encoding of certain memory content are compared with the patterns during memory retrieval of that particular memory. There are different types of similarity measures that can be used in memory research. High ***perception-retrieval similarity*** or ***encoding-retrieval similarity*** (Xue, 2018) is regarded as the neural evidence for *pattern reinstatement*.

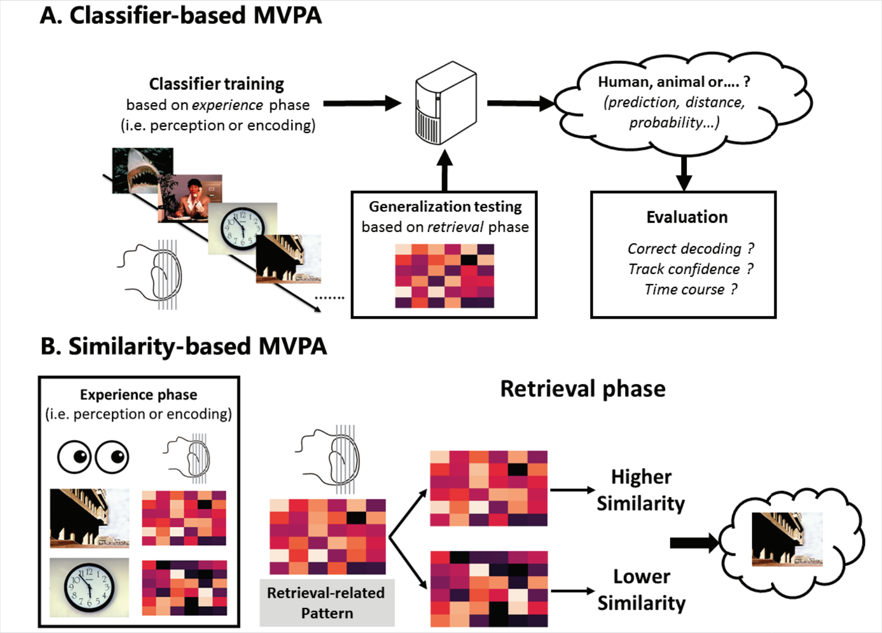


Figure 1.4 Applications of two types of MVPA in memory research. (A) Classifier-based MVPA. To detect pattern reinstatement, we can train multivariate classifiers on the activity patterns during the experience phase (e.g., perception or encoding) and test classifiers on activity patterns during memory retrieval. After seeing testing patterns, classifiers can then generate different outputs (e.g., category, decision distance, and probability). In the final evaluation phase, we can link these outputs with memory performance or subjective confidence. **(B)** Similarity-based MVPA. During the experience phase, we can estimate the activity pattern for each memory. During the retrieval, we extract the activity pattern when participants are instructed to recall. High pattern similarity between experience and retrieval is the neural evidence of pattern reinstatement and could suggest which memory the participants are recalling.

1.4 Modulations of memory traces between experience and retrieval

We have discussed *where* and *how* the brain represents information and accesses them when required. However, after initial formation and before the final memory retrieval, two types of dynamical changes can occur. First, neural traces of an episodic event change over time *spontaneously* via consolidation (Frankland & Bontempi, 2005; Takashima et al., 2006) and reconsolidation after reactivation (Nader et al., 2000; Schiller et al., 2010; Schwabe et al., 2014). Second, *external* factors can induce changes to the overall brain state or specific memory traces. For example, the administration of stress hormones (McGaugh & Roozendaal, 2002) or caffeine (Borota et al., 2014) can enhance memory performance in general. Behavioral tagging can improve performance for memories within the same conceptual space by changing its valence (Dunsmoor et al., 2015; Patil et al., 2017). Specific memory traces can be modulated as well. Methods such as targeted memory reactivation (Hu et al., 2020; Oudiette & Paller, 2013) and retrieval practice (Karpicke & Blunt, 2011; Roediger III & Butler, 2011) can be used to enhance memories and approaches, for instance, memory control (Michael C Anderson & Green, 2001; Michael C Anderson & Hanslmayr, 2014) and memory updating/competition (Jacques et al., 2013; Kuhl et al., 2010; Maria Wimber et al., 2015) can, in contrast, weaken them.

Both types of changes are likely to interact with each other and are often considered together **(Figure 1.5)** (Phelps & Hofmann, 2019). On one hand, during consolidation and reconsolidation memory traces are vulnerable to external modulations, and on the other hand, external factors can reactivate initially consolidated memory traces, triggering the reconsolidation (i.e., another vulnerable time window).

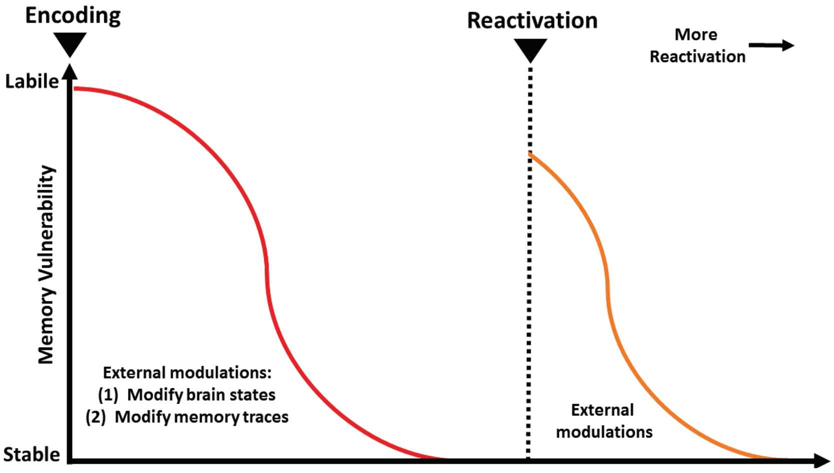


Figure 1.5 Vulnerable time window for memory modulations. Adapted from Phelps & Hofmann, 2019. The first time window when memories are vulnerable to external modulations is the consolidation after initial formation (RED). Memory consolidation is a long process which can take from hours to days or even months, but most related studies try to modulate memories before the overnight consolidation. Following reactivation (i.e., retrieval), memories could become vulnerable again, creating the second time window for modulation (ORANGE). Reactivated, vulnerable memories will be consolidated again, and this process is called reconsolidation.

1.5 Memory control

Among the different memory modulation approaches mentioned, **memory control** is a particular focus of this thesis because less is known about it compared to other memory operations. Memory control is when people can avoid unwanted memories via the inhibitory control mediated by the top-down regulation executed by the lateral prefrontal cortex on the hippocampus (Michael C Anderson & Hanslmayr, 2014). Since this process happens during memory retrieval, it is also called *retrieval suppression*. A typical way of studying memory control is the Think/No-Think paradigm (Details see **Chapters 2 and 3**). The typical finding using the Think/No-Think paradigm was impaired memory for items that were suppressed relative to unmanipulated baseline items. This difference is referred to as the *suppression-induced forgetting effect* and was observed across different types of memories (Michael C Anderson & Huddleston, 2012; Benoit et al., 2016; Bergström et al., 2013; Brendan E Depue et al., 2007; Gagnepain et al., 2014; Noreen & MacLeod, 2013).

Furthermore, expanding knowledge of memory control may have clinical relevance for affective disorders. Patients with affective disorders typically show deficits in the control of retrieving negative or traumatic memories in daily life (e.g., ruminations and flashbacks) (Bovy et al., 2019; Dillon & Pizzagalli, 2018; Kensinger & Ford, 2020). The Think/No-Think paradigm was used as a method to model this deficit in the laboratory setting (e.g., depression patients (Sacchet et al., 2017; Yang et al., 2020) and Post-traumatic stress disorder (PTSD) patients (Catarino et al., 2015). Importantly, the Think/No-Think paradigm based on trauma-unrelated materials was shown to predict resilience in the aftermath of a traumatic event (Mary et al., 2020), demonstrating its potential clinical value to identify high-risk individuals for developing related disorders.

1.6 Process dynamics in episodic memories

Our life experiences are usually perceived as continuous. However, one can identify separate “processes” within this continuous stream of experiences as basic processing units. Considering whether these units provide the computational underpinnings of the *same* or *different* mnemonic operations, they can interact with each other in two different ways, leading to distinct *process dynamics*. (1) If all units support the *same* mnemonic operations (e.g., encoding), it is optimal for the brain to separate them to prevent overlapping representations (i.e., *segmentation*) and at the same time, to establish, a link between them to form a coherent memory (i.e., *integration*). (2) If the units support *different* mnemonic operations, the brain has to switch dynamically within a given time window from one to another task demand when required (i.e., *task switching*) (Meiran, 2010; Monsell, 2003). Two process dynamics mentioned above may share similar neural underpinnings.

Segmentation and integration

Process dynamics, such as segmentation and integration, were usually ignored in previous neuroimaging investigations of memory encoding because to-be-remembered stimuli were presented as discrete trials (Kim, 2011). This design was a practical choice in the early days of functional neuroimaging because it enhances the statistical power to identify the neural effect of interest by averaging neural activity across trials. However, given the improved neuroimaging data acquisition and analytical tools, now it is possible to use naturalistic stimuli **(Box 2)**in the functional neuroimaging studies to probe *process dynamics*. For example, during the encoding of continuous experience, temporal information within the ongoing neural activity and the stimulus can be used to gain novel insights into the relationship between event processing and memory formation (Baldassano et al., 2017; Chien & Honey, 2020).

Using the naturalistic stimuli, we can study, at least, two fundamental *process dynamics* during continuous perception and memory formation: *event segmentation* and *event integration* (Benjamin J Griffiths & Fuentemilla, 2020; Williams et al., 2019; Zacks, 2020). How the human brain segments ongoing experience is not only critical for memory formation but also a range of other cognitive functions (e.g., language (Ding et al., 2016; Teng et al., 2020)). Here, we firstly focused on the *event segmentation* during episodic memory processing. Using MVPA (**Box1**), two recent studies demonstrated correlates of *event segmentation* in the neocortical regions, with distinct activity patterns representing different events within the continuous experience (Janice Chen et al., 2017). Interestingly, activity pattern-mediated event segmentation showed a different time scale along with a nested 'hierarchical memory system' (Baldassano et al., 2017). A recent study suggests that the same set of brain regions also encodes context information in a hierarchical manner (Chien & Honey, 2020).

Compared to *event segmentation*, the neural mechanism of *event integration* while engaging in a continuous experience is less studied. This process is particularly critical and unique to everyday episodic memory formation since separate event representations need to be integrated into a coherent narrative. The trial-based AB-AC paradigm (Backus et al., 2016; Zeithamova et al., 2012) is widely used to study memory integration and highlights the role of hippocampal-medial prefrontal interaction during memory integration (Preston & Eichenbaum, 2013). However, this trial-based paradigm can only reveal information integration of two or three items that are separate from each other temporally. Therefore, this method cannot reveal how events that are close in their temporal sequence interact with each other.

Taken together, neural correlates of event segmentation and integration are less understood in the context of episodic memory formation. This brings us to the first central question of this thesis (**Question 1**): *How do we transform continuous experience into discrete memories?* In **Chapter 2**, we described the answer to the question by discussing if we can reveal two complementary event processes (i.e., *segmentation* and *integration*) by probing the *process dynamics* of the ongoing neural activity, and by investigating how do they contribute to memory formation.

Box 2 Naturalistic stimuli to study human cognition

The classical ***trial-based paradigm*** is used widely in cognitive neuroimaging experiments. Recently, ***naturalistic stimuli,*** including but not limited to movies (Baldassano et al., 2017; Janice Chen et al., 2017), music (Ganesh et al., 2019), and poetry (Teng et al., 2020) are increasingly popular among neuroimaging experimenters. To better make use of these data, several old analytic approaches have been applied, and new approaches have been developed accordingly. For example, ***inter-subject correlation (ISC) analysis*** can be used to measure shared information representation across brains (Janice Chen et al., 2017; Hasson et al., 2004). ***Independent component analysis-based network analysis*** is applied to model brain-state shift induced by external, naturalistic stimuli (Hermans et al., 2011). ***Voxel-wise encoding models*** provide a sensitive method for generating a semantic atlas of the cerebral cortex (Huth et al., 2016). ***Hidden Markov Model (HMM)*** is used to detect the latent, separable states that are hidden within the brain response patterns measured during continuous cognitive processing (Baldassano et al., 2017).

Compared to the ***trial-based paradigm***, which provides precise experimental control to isolate different modulating factors and the possibility to average neural activity across trials, the naturalistic ***stimuli*** can largely enhance the ecological validity of the study design and related conclusions (Huk et al., 2018; Sonkusare et al., 2019). The issue of ecological validity in memory research is particularly critical since the trial-based paradigm cannot fully capture the characteristics of real-life memories. For example, contents are usually presented isolatedly in the trial-based paradigm, while in reality, we need to store the continuous, crowded stream of information into long-term memory. Together with machine learning approaches (e.g., computer vision, natural language processing), neuroimaging studies using naturalistic stimulus are starting to reveal neuro-computational principles of real-life memories (Baldassano et al., 2017; Janice Chen et al., 2017; Chien & Honey, 2020).

Task-switching

Task-switching is another form of *process dynamics*. The task demands within certain time windows can be viewed as different "*processing units*," which are the building blocks of a temporal sequence of task switching. *Task-switching* is usually studied when participants are required to switch between simple tasks in different cognitive domains, such as switching between size comparisons and object recognition. However, what if these "units" are opposite memory-related mental operations?

A recent study demonstrated the interactions between two memory operations: after participants performed a memory control trial, the following memory encoding is more likely to fail, suggesting the creation of an amnesic window. Neuroimaging results revealed that memory control had a lasting effect on reducing hippocampal activity, and this 'virtual lesion' was not recovered when hippocampal activity was required to support memory encoding (Hulbert et al., 2016). This study suggests that within the temporal sequence, the preceding unit has a lasting effect on the execution of the current unit. This switch cost could be even larger in cases where two different tasks involve overlapping brain areas, but different neural processes are needed (e.g., increasing hippocampal activity for encoding, but decreased hippocampal activity for memory control).

Because memory encoding and retrieval share overlapping neural resources, we reason that the dynamic interaction between memory control and retrieval might lead to, sometimes, postponed transitions between neural states. This effect may not be limited to the delayed recovery of hippocampal activity (Hulbert et al., 2016), but also at a large-scale network level. This brings us to the second central question of this thesis (**Question 2**): *How does the brain flexibly switch between memory retrieval and memory control*? Here, in **Chapter 3**, we investigated the task switching-induced *temporal dynamics* between retrieval and control. Time-resolved classifiers were used to track the fast transition of neural states between the memory system and the control system. Methods in this study are built on the recent developments enabling the characterization of dynamic neural reconfigurations (**Box 3**).

Box 3 Characterize dynamic neural reconfiguration

Classical analysis methods in fMRI have largely ignored the richness of temporal information within the data. Admittedly, the temporal resolution of the fMRI signal, even with the most advanced scanners and sequences (e.g., with the repetition time of 0.4-0.8 seconds), is incomparable to Electroencephalography (EEG) or Magnetoencephalography (MEG). With the new analytic methods and specific experimental designs, the temporal dynamics during both ***resting-state*** and ***task-based fMRI*** can open unique windows into the general organization of the human brain and its adaptation during cognitive tasks.

Resting-state fMRI has greatly advanced our understanding of the fundamental features of the brain, from the interaction between two brain regions to large-scale brain networks. However, the key method of these investigations (i.e., ***functional connectivity***) assumes that the statistical relationship between signals from two brain regions is constant throughout the entire resting-state period lasting several minutes to even hours (Buckner et al., 2013; Van Den Heuvel & Pol, 2010). Recent studies have begun to challenge this assumption: ***dynamic functional***

connectivity studies revealed the wealth of information contained along the temporal time-line of spontaneous fMRI signals (Hutchison et al., 2013) and their behavioral relevance (Liégeois et al., 2019).

Task-based fMRI also witnesses a similar development. Most early studies only generated one statistical map by averaging data from different trials, blocks, or runs. This approach ignored the neural adaption from early to the later stages of the same task, or the adaption between different tasks. For example, when participants are required to control over emotional memories, the prefrontal cortex demonstrated the time-specific two-phase control: at the initial attempts of memory control, inferior frontal gyrus is more involved, while medial frontal gyrus plays a more critical role at the later stage (Brendan E Depue et al., 2007). In a block-design fMRI study with reasoning, perception, and memory tasks, the coupling between the control network and the default network is stronger in the memory task, but not in the other two tasks (Westphal et al., 2017).

1.7 Strength dynamics in episodic memories

Memory strength is not a consistent concept and can be defined in different ways (e.g., availability, subjective confidence, vividness, or durability). Here, we used “strength” in a relatively abstract sense, by defining the higher strength leading to better retrieval performance, including all characteristics mentioned before. We propose that the strength of a memory trace, although static at a particular time point, can change towards two directions (i.e., become stronger or weaker) across time. We defined these changes as *strength dynamics*.

Retrieval practice and memory control

At the beginning of the introduction (*See Section 1.3*), we discussed several potential approaches that can induce *strength dynamics* by modulating the memory trace between memory formation and memory retrieval. Among them, retrieval practice and memory control are easy to study because no additional experimental setup is needed, and participants can actively follow instructions to modulate the strength of specific memories.

Although human neuroimaging studies revealed retrieval-related and control-related univariate activity changes in frontal, parietal, and temporal areas (Eriksson et al., 2011; Gagnepain et al., 2014; Kuhl et al., 2010; Nelson, Arnold, Gilmore, & McDermott, 2013; G. van den Broek et al., 2016; G. S. E. van den Broek et al., 2013; Maria Wimber et al., 2008, 2011; Wing et al., 2013; Wirebring et al., 2015), the precise neural mechanisms underlying these *strength dynamics* are still unknown. One possibility could be that these *strength dynamics* are achieved by modulating *pattern reinstatements* (*See section 1.4*). More specifically, during retrieval practice, memory-specific neural patterns are reinstated repeatedly, leading to the strengthening of item-specific memory representation. By contrast, during memory control, pattern reinstatements are prevented, and thus, when memory cues are presented, the link between the cue and corresponding neural pattern is weaker. This idea has not been tested thoroughly yet during both modulation and later retrieval within one experiment.

Furthermore, even though *strength dynamics* can be induced by memory modulation immediately after formation (i.e., the first vulnerable time window in *Figure 1.4*), whether and how memory modulations can also induce strength dynamics after an initial period of consolidation is currently not well understood. Because newly acquired memories are usually more labile compared to consolidated ones (Frankland & Bontempi, 2005) and mnemonic representations shift from the hippocampus to distributed neocortical regions following overnight sleep (Takashima et al., 2006, 2009), the same types of modulations may not produce *strength dynamics,* or the behavioral effects may be associated with different underlying neural changes.

This leads us to answer the third central question of this thesis (**Question 3**): *How does the memory modulation re-organize memory traces and change their memory strength after overnight consolidation?* In **Chapter 4,** we investigated whether memory traces, after their initial consolidation, are still flexible enough to change with repeated retrieval practice and retrieval suppression. How is the pattern reinstatement modified by retrieval practice and suppression? Is the modulation of newly acquired memories and initially consolidated memories associated with similar neural changes?

Dual modulation of memory and emotion

The *strength dynamics* of memories induced by memory modulations may be associated with changes in other characteristics of episodic memory (e.g., perceived emotional intensity). Although not strictly linear, the relationship between memory strength and emotional intensity may form a two-dimensional coordinate system. That is to say, if the strength of one memory is moved by the memory control along one axis, its emotional intensity, in another axis, may also change. This phenomenon of dual modulation was observed in memory control studies using the Think/No-think paradigm (De Vito & Fenske, 2017; Gagnepain et al., 2017). More specifically, compared to unmanipulated control memories, memory traces that underwent suppression do not only demonstrated lower availability (i.e., suppression-induced forgetting effect), but also show reduced unconscious influence (Gagnepain et al., 2014), and emotional impact (De Vito & Fenske, 2017; Gagnepain et al., 2017). Engen and Anderson recently reviewed related behavioral and neuroimaging studies in this field and proposed a conceptual link between emotion regulation and memory control (Engen & Anderson, 2018).

To reduce the emotional impact of a perceived stimulus is the central goal of emotion regulation (Gross, 2013). This ability can be critical in patients with affective disorders, especially for those who have memory-related symptoms, such as ruminations and flashbacks. However, at the same time, impaired emotion regulation is one of the key cognitive symptoms in affective disorders (Kring & Sloan, 2009), making the emotional impact of their negative memories more impactful. Thus, exposure therapy, which is thought to be based on fear memory extinction, is widely used to target traumatic memories in the clinical setting (Rothbaum & Schwartz, 2002). However, its long-term effect can be suboptimal due to the spontaneous recovery of memory traces after the therapy. Based on the research progress described here, memory control training may be used as an alternative method to weaken unwanted negative memories, in both memory strength and emotional intensity.

However, it is unknown how *strength dynamics* of memory relate to changes in emotional intensity. This brings us to the fourth central question of this thesis (**Question 4**): *Why are changes in memory strength also accompanied by alterations of emotional intensity?* In **Chapter 5**, we combined task fMRI data, neuroimaging meta-analytic approaches, and postmortem gene-expression data (**Box 4**) to explore the common neural and transcriptional correlates underlying memory control and emotion regulation. Understanding their common neurobiological correlates may pave the way for further developments to enhance memory control and emotion regulation ability by brain stimulation or pharmacological interventions.

Box 4 Gene expression-Neuroimaging association analysis

Due to the technical advances in molecular biology, a human brain atlas that maps expression levels of thousands of genes across the entire brain is openly available to neuroscientists around the world (Hawrylycz, Lein, Guillozet-Bongaarts, Shen, Ng, et al., 2012; Shen et al., 2012). This new type of data offers new opportunities for understanding the relationship between genes and brain structure and function beyond the conventional ***candidate gene studies*** (Hariri et al., 2002) or ***genome-wide association studies*** (Elliott et al., 2018). Understanding gene-function relationships sheds new light on not only the genetic correlates of fundamental brain architecture in the normal population but also the molecular underpinnings of particular brain disorders (Fornito et al., 2019).

The new set of approaches for linking gene expression to neural phenotypes measured by MRI can be divided into two major categories: ***connectivity-based association analysis*** and ***activation-based association analysis***. The pioneering study by Richiardi and colleagues provided evidence that two brain regions that are functionally connected (i.e., belong to the same functional network) tend to have more similar gene expression profiles (Richiardi et al., 2015). These results suggest a close relationship between the underlying molecular mechanisms and the functional role of a given brain region. Richiardi and colleagues used the ***connectivity-based analysis***, which calculates the transcriptional similarity of gene expression profiles between brain regions (i.e., the region-region transcriptional correlation across genes) (**Figure 1.6A**). Another pioneering work by Wang and colleagues showed that fractional amplitude of low-frequency fluctuations (fALFF), a region-specific resting-state fMRI activity measure, is selectively correlated with the spatial expression patterns of 38 genes among all genes (Wang et al., 2015). In this study, the ***activation-based association analysis*** was used, which focuses on the spatial pattern similarity between neuroimaging and transcriptional measures (i.e., activation-expression relation across tissues) (**Figure 1.6B**).

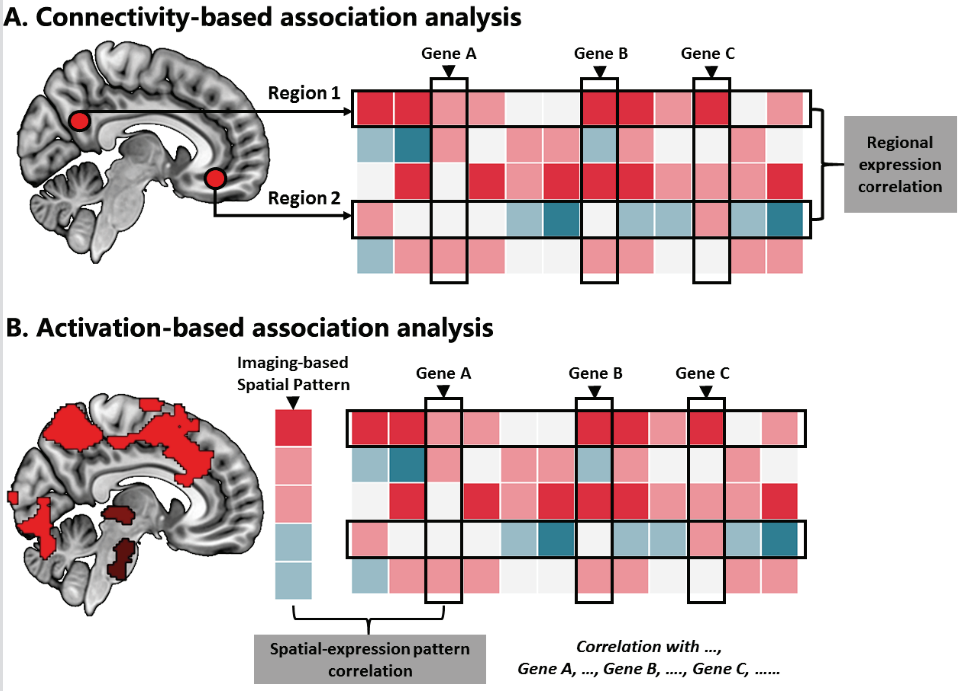

A. Connectivity-based association analysis
Gene A
Gene B
Gene C
Region 1
Region 2
Regional expression correlation
B. Activation-based association analysis
Imaging-based Spatial Pattern
Gene A
Gene B
Gene C
Spatial-expression pattern correlation
Correlation with ..., Gene A, ..., Gene B, ..., Gene C, ...

Figure 1.6 Two types of methods to link brain imaging with gene expression. (A) Connectivity-based association analysis. Genome-wide expression values can be extracted from two regions in the brain and correlated with each other, yielding a regional correlation value. Regions that belong to the same functional network or connected by the anatomical connection tend to demonstrate more similar transcriptional profiles compared to two unrelated regions. **(B)** Activation-based association analysis. A brain-wide spatial pattern from neuroimaging (e.g., structure or activity map) can be correlated with the spatial patterns of transcriptional regulation of different genes. Genes whose spatial patterns are similar to the neuroimaging patterns can be identified in a data-driven way.

These new approaches together revealed an association between connectome and transcriptome (Fornito et al., 2019), molecular correlates of neural changes in brain disorders (Grothe et al., 2018; McColgan et al., 2018; Morgan et al., 2019; Romero-Garcia et al., 2019; Romme et al., 2017), and transcriptional correlates of cognitive functions including memory (Berto et al., 2018), spatial navigation (X. Kong et al., 2017), and language (X.-Z. Kong et al., 2020).

Thesis outline

In this thesis, I describe research that aims to gain more insight into the *dynamic nature of human memory*. Specifically, the overarching question for the chapters in this thesis was: "*How does our brain support both temporal and strength dynamics of memory?*". The next four chapters **(Chapters 2-5)** presented four experimental investigations of testable research questions:

1. **How do we transform continuous experience into discrete memories?**
2. **How does the brain flexibly switch between memory retrieval and memory control?**
3. **How does memory modulation re-organize memory traces and change their memory strength after overnight consolidation?**
4. **Why are changes in memory strength also accompanied by alterations of emotional intensity?**

To answer the above questions, I used novel cognitive tasks and fMRI in healthy human subjects and combined methods from imaging genetics, neuroinformatics, and machine learning. Here, I will give a brief outline of the experimental and analytic approaches.

In **Chapter 2**, we were interested in the mechanism by which the hippocampus segments and integrates discrete events within continuous experiences and how these two processes relate to the subsequent memory retrieval. Human participants watched a movie while being scanned and afterward, they were instructed to recall the story of the movie freely. We quantified neural signals of event segmentation and event integration using two multi-voxel pattern similarity metrics in the hippocampus and searched for a similar process across the brain. We further associated these metrics with subsequent memory performance to demonstrate their role in memory formation.

In **Chapter 3**, we investigated the transition of neural states underlying the task-switching between memory retrieval and its control. We hypothesized that the switch-induced delay in transiting neural states causes behavioral switching-costs. Healthy participants underwent fMRI during a Think/No-Think task, which was specially designed to probe task switching between two memory-related behaviors. This design allowed us not only to induce memory-related switching costs but also to capture the dynamic transitions between neural states. Combining a time-resolved multivariate decoding analysis and a trial-by-trial task performance measure, we can observe the updating of task-related neural states and its behavioral relevance.

Next, in **Chapter 4**, we set out to investigate the neural dynamics of mnemonic representations during and after post-consolidation modulation. We asked how initially consolidated memories are dynamically modulated by retrieval and suppression 24 hours after learning. We hypothesized that repeated retrieval would promote episode-unique mnemonic representations in the neocortex, while memory control would have the opposite effect. Both the univariate activation analysis and multivariate pattern analysis were used to quantify the neural reactivation of memory traces during modulation and subsequent retrieval. We further investigated the dynamic changes of neural measures from these two analyses across repeated modulation attempts and the complementary relationship between the two neural measures.

Lastly, in **Chapter 5**, we explored the common role of inhibitory control in both memory control and emotion regulation. To reveal the neural and transcriptional commonalities underlying memory control and emotion regulation, we first performed a meta-analysis of fMRI studies across four inhibition-related task paradigms (i.e., memory control, emotion regulation, go/no-go, and strop-signal). Then, we linked these task-induced brain activity patterns to gene expression patterns in the Allen Human Brain Atlas (AHBA). The identified inhibition-related genes were further linked to biological processes and human diseases via a Gene Ontology enrichment analysis.

Chapter 2

Hippocampal-medial prefrontal event segmentation and integration contribute to episodic memory formation

This chapter is based on:
Liu, W., Shi, Y., Cousins, J. N., Kohn, N., & Fernandez, G. (2020). Hippocampal-medial prefrontal event segmentation and integration contribute to episodic memory formation. bioRxiv. (This preprint is *under revision* in a scientific journal)

Abstract

How do we encode our continuous life experiences for later retrieval? Theories of event segmentation and integration suggest that the hippocampus binds separately represented events into an ordered narrative. Using an open-access functional Magnetic Resonance Imaging (fMRI) movie watching-recall dataset, we quantified neural similarities between separate events during movie watching and related them to subsequent retrieval of events as well as retrieval of sequential order. We demonstrate that distinct *activation patterns* of the hippocampus and medial prefrontal cortex form event memories. By contrast, similar within-region *connectivity patterns* between events facilitate memory formation and are critical for the retention of events in the correct sequential order. These reported subsequent memory effects only existed when neural similarities were calculated based on actual event boundaries, but not shuffled event boundaries. We propose that distinct *activation patterns* represent neural segmentation of events while similar *connectivity patterns* act as the 'chunking code' for integration across events. Our results provide novel evidence for the role of hippocampal-medial prefrontal event segmentation and integration in episodic memory formation of real-life experience.

Keywords: subsequent memory effect; hippocampus; medial prefrontal cortex; event segmentation; event integration

Introduction

How we form memories of our life experiences is a fundamental scientific question with broad implications. In the past two decades, human neuroimaging and electrophysiology studies using the subsequent memory effect paradigm have implicated a distinct set of brain regions involved in successful memory formation (Brewer et al., 1998; Fernández et al., 1999; Kim, 2011; Wagner et al., 1998). In these subsequent memory studies, increased neural activity of the hippocampus, parahippocampal gyrus, and the prefrontal cortex during memory encoding is associated with successful subsequent retrieval. However, real-world memories are formed based on a continuous stream of information rather than the sequentially presented, isolated items used in most subsequent memory studies (Kim, 2011). Potentially, continuous sensory experience is segmented into distinct events (i.e., event segmentation) (Baldassano et al., 2017; Williams et al., 2019; Zacks, 2020) that are then bound together into a coherent narrative, preserving their sequential relationships (i.e., event integration) (Benjamin J Griffiths & Fuentemilla, 2020). To examine episodic memory formation of real-life-like experiences in humans, we analyzed brain activity using functional Magnetic Resonance Imaging (fMRI) while participants were watching a movie. Based on subsequent memory recall, we aimed at identifying brain regions and neural representational processes underlying event segmentation and integration during episodic memory formation.

Thanks to recent advances in the statistical analysis of ongoing neural activity (J. D. Cohen, Daw, Engelhardt, Hasson, Li, Niv, Norman, Pillow, Ramadge, Turk-Browne, & Willke, 2017; Hermans et al., 2011; Nastase et al., 2019; Xue, 2018), naturalistic stimuli (e.g., movie, spoken narratives, music) have been increasingly used in neuroscience (Hasson et al., 2004; Hermans et al., 2011; Huk et al., 2018; Sonkusare et al., 2019). This is especially valuable for memory research because naturalistic stimuli can greatly enhance the ecological validity of experimental studies (Baldassano et al., 2017; Janice Chen et al., 2017; Hasson et al., 2008; Montchal et al., 2019). Hasson and colleagues first investigated memory formation with cinematographic stimuli and demonstrated that brain activity was more correlated among participants for later remembered than forgotten events (Hasson et al., 2008). While that study uncovered regions that encode continuous experiences, the nature of representations in those regions remained unclear, particularly with regard to how episodes are segmented into separate events and then integrated into a coherent sequence.

Event segmentation theory suggests that continuous experiences need to be segmented into discrete event representations, and thereafter they can be better understood and encoded (Zacks, 2020; Zacks et al., 2001, 2007). Two recent studies provided novel perspectives into segmentation theory. Using Multi-Voxel Pattern Analysis (MVPA) and a movie watching-recall dataset, Chen and colleagues showed similar *activation patterns* of the same events across individuals and event-specific reinstatements of *activation patterns* between encoding and retrieval (Janice Chen et al., 2017). Following this, Baldassano and colleagues demonstrated a nested processing hierarchy of events ('hierarchical memory system', (Hasson et al., 2015)) from coarse segmentation in early sensory regions to fine-grained segmentation in regions of the higher-order default-mode network (e.g., medial prefrontal cortex (mPFC) and posterior medial cortex (PMC)). Importantly, boundaries of long events at the top of the hierarchy matched with event boundaries annotated by human observers and were coupled to increased hippocampal activity (Baldassano et al., 2017). These results demonstrated that human brains spontaneously used different *activation patterns* to represent events during continuous movie watching, and how these *activation patterns* reactivated during recall. Also, it may suggest that regions such as mPFC, PMC, and hippocampus encode events at the same level that we consciously perceive boundaries between events. However, it remains unclear how exactly this event segmentation at the neural level relates to subsequent memory recall.

Event segmentation alone is not sufficient for episodic memory formation of continuous real-life experiences. Temporal context theory suggests that it is essential to integrate segmented events into a coherent narrative via time, meaning, or other abstract features (Howard et al., 2005; Howard & Eichenbaum, 2013). Therefore, a non-exhaustive list of questions are: (1) what are the neural underpinnings of event integration during continuous memory formation, (2) does integration occur in the same brain regions as segmentation, and (3) how does integration relate to subsequent memory recall. A promising approach to answer these questions is to examine the local *connectivity pattern* (also called *multi-voxel correlation structure*), which may represent a brain signal that integrates events (Tambini & Davachi, 2019). This method was derived from rodent electrophysiology (Kudrimoti et al., 1999; Lansink et al., 2008; Qin et al., 1997) and has been used in human fMRI studies (Hermans et al., 2017; Tambini & Davachi, 2013) to quantify distributed memory representations in neuronal assemblies. Recently, Tambini and Davachi (Tambini & Davachi, 2019) proposed that *activation patterns* are the representations of specific perceptual inputs (e.g., stimuli), while local *connectivity patterns* reflect particular encoding contexts or states. However, the different mnemonic functions of *activity patterns* and *connectivity patterns* have yet to be compared empirically within a single study. If local *connectivity patterns* represent encoding context, they may facilitate integration across events. Examination of *connectivity patterns* alongside *activation patterns* would help to characterize how the brain simultaneously performs event segmentation and integration.

Recently, a hippocampal neural code (chunking code) that simultaneously tracked subdivisions of a continuous experience (i.e., events) and their sequential relationship was described in rodents' CA1 region (Sun et al., 2020). This 'chunking code' could be a fundamental neural code by which episodic experience is integrated, but has yet to be revealed in humans. Hippocampal activity was found to increase at the boundaries between two events during the continuous experience (Baldassano et al., 2017; Ben-Yakov et al., 2013; Ben-Yakov & Dudai, 2011; Ben-Yakov & Henson, 2018; DuBrow & Davachi, 2013; Williams et al., 2019), but what these hippocampal signals represent in terms of event segmentation and integration is not clear. Theoretical models proposed that increased hippocampal signal may reflect a rapid shift in mental representations (e.g., temporal and/or contextual information of an event) (DuBrow et al., 2017; DuBrow & Davachi, 2016; Ranganath & Ritchey, 2012). Therefore, it can be regarded as the neural signature of event segmentation. Alternatively, this increase may link to the integration of episodic memories across event boundaries, as suggested by scalp electrocorticography (EEG) studies (Silva et al., 2019; Sols et al., 2017) and the event conjunction framework (Benjamin J Griffiths & Fuentemilla, 2020). However, fMRI evidence for the role of hippocampal signals in integration across events is still limited.

The current study aimed to reveal the neural underpinnings of the two processes in question – event segmentation and event integration - during memory formation of naturalistic experiences. To that end, we used an existing dataset (Baldassano et al., 2017; Janice Chen et al., 2017) where participants watched a movie while being scanned (**Figure 2.1A**) and afterward were instructed to freely recall the story of the movie (**Figure 2.1B**). This design allowed us to associate different neural measures during episodic encoding with subsequent memory retrieval (**Figure 2.1C-D**). We extracted voxel-wise Blood Oxygenation Level Dependent (BOLD) time courses during movie watching (encoding) from six predefined regions-of-interest (ROI) in the 'hierarchical memory system' (Hasson et al., 2015) including early auditory and visual areas, posterior medial cortex, medial prefrontal cortex, hippocampus, and posterior parahippocampal gyrus (**Figure 2.2A**; **Figure S2.1**). To probe the role of a broader set of regions in event segmentation and integration, we repeated all analyses in each parcel of a neocortical parcellation (Schaefer et al., 2018) (**Figure 2.2B**). We first examined the relationship between ROI-based activity time courses and subsequent memory recall and replicated the classical subsequent memory effects (i.e., greater activation for *remembered* compared to *forgotten* events) in regions including the hippocampus as well as the posterior parahippocampal gyrus (**Figure S2.2-2.3**, details in **Supplementary Materials**). To dissociate the two event processes, we used voxel-wise activity (**Figure 2.2C**) from each ROI to quantify the similarity between neural representations of events by two different multivariate methods (i.e., *activation* and *connectivity patterns)* (**Figure 2.2D-E**). We reasoned that if the neural representation (*activation* or *connectivity pattern*) shows a large transition (i.e., negative neural similarity value) between two adjacent events, and if this dissimilarity associates with better subsequent memory for events, then this representation might be involved in event segmentation (**Figure 2.2E**). By contrast, if the neural representation remains stable (i.e., higher similarity) across two or more neighboring events, and this stability relates to event memory as well as retention of the correct order for those events (order memory), then this representation may underlie event integration (**Figure 2.2F**).

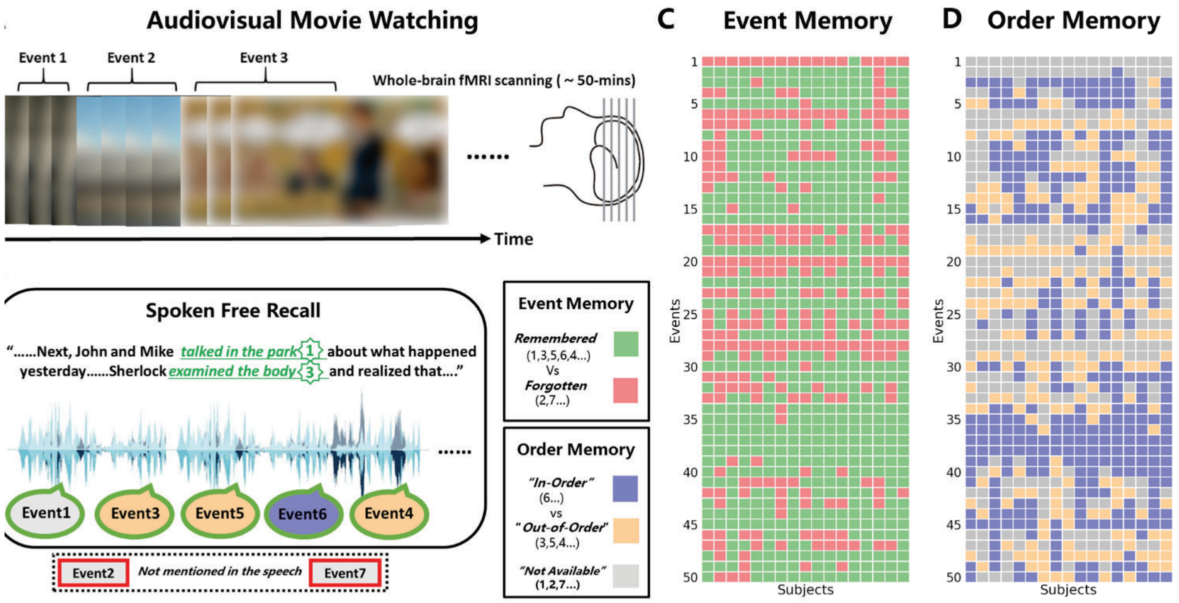

Audiovisual Movie Watching
Event 1
Event 2
Event 3
Whole-brain fMRI scanning (~ 50-mins)
......
Time
Spoken Free Recall
"......Next, John and Mike talked in the park 1 about what happened yesterday......Sherlock examined the body 3 and realized that...."
......
Event1
Event3
Event5
Event6
Event4
Event2
Not mentioned in the speech
Event7
Event Memory
Remembered (1,3,5,6,4...)
Vs
Forgotten (2,7...)
Order Memory
"In-Order" (6...)
vs
" Out-of-Order" (3,5,4...)
"Not Available" (1,2,7...)
C
Event Memory
D
Order Memory
Events
Subjects
1
5
10
15
20
25
30
35
40
45
50

Figure 2.1. Experimental procedure and behavioural performance. (A) Each participant watched a 50-min audiovisual movie, BBC's Sherlock (season 1, episode 1), while brain activity was recorded with fMRI. The movie was divided into 50 events based on major narrative shifts. Blurred images are shown here due to copyright reasons. However, the movie was shown in high resolution during the experiment. **(B)** Immediately after movie-watching, participants verbally recalled the movie content in as much detail as possible without any visual or auditory cues. Speech was recorded using a microphone and then transcribed. Critically, speech was also segmented into events and matched with the events segmented from the movie. All events mentioned in the speech were labeled as *remembered* while missing events were labeled as *forgotten*. In addition, among those remembered events, the ones that were recalled in the correct sequential order were labeled as *in-order* events (e.g., *event 6* was recalled after *event 5*). *Out-of-order* events were those that were recalled in an incorrect sequential order (e.g., *event 4* was recalled after *event 6*). We labeled the first recalled event and all *forgotten* events as *not available* because no sequential information can be accessed. **(C)** Illustration of all *remembere*d and *forgotten* events during movie-watching in all participants**. (D)** Illustration of all *in-orde*r and *out-of-order* events during movie watching in all participants. Each row of the heatmap is a different event, and each column represents a participant.

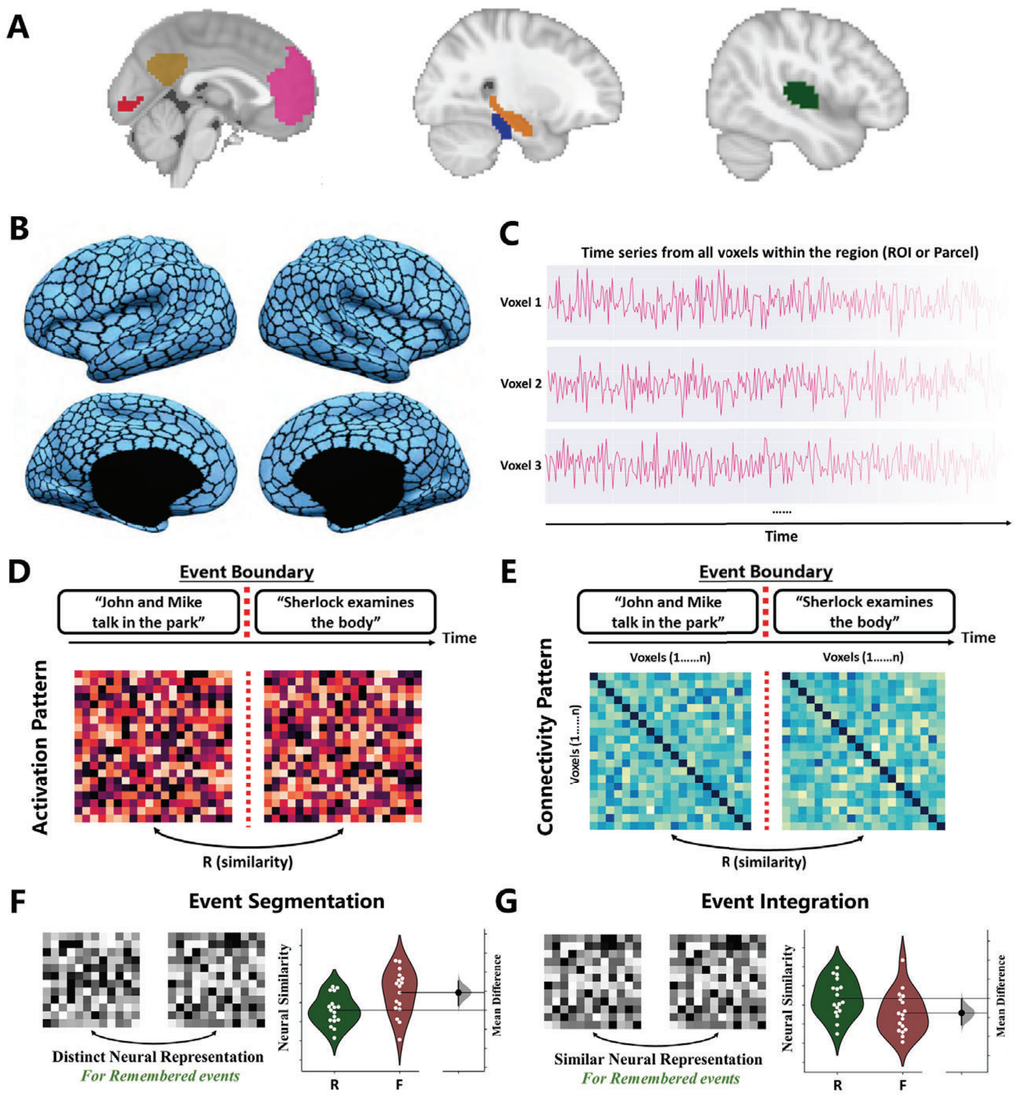


Figure 2.2. Neural similarities between separate events and their link with subsequent memory recall. **(A)** Six predefined regions-of-interest (ROIs): early auditory (green) and visual area (red), posterior medial cortex (brown), medial prefrontal cortex (pink), hippocampus (blue), and posterior parahippocampal gyrus (orange). See also **Supplementary Figure 1**. **(B)** Neocortical parcellation (1000 parcels) used in searchlight analysis. **(C)** For each region (ROI or parcel), voxel-wise signal during movie watching was extracted and then segmented into 50 events based on the event annotations. **(D)** We first generated event-specific *activation patterns* by averaging over all time points in that event. Then *activation pattern* similarity was calculated by Pearson's correlation between *activation patterns* of two sequential events. If a region encodes two events separately, we expect two distinct neural representations and therefore a negative correlation (i.e., lower than zero). **(E)** Event-specific within-region *connectivity patterns* were represented by voxel-by-voxel pairwise correlation matrices. *Connectivity pattern* similarity across event boundaries was also calculated using Pearson's r between two sequential events. Stable neural representations across two events should yield a positive correlation (i.e., higher than zero) in the corresponding region. **(F)** fMRI evidence for event segmentation. For a certain multivariate neural measure, if it can be found that two distinct neural representations are used to encode the adjacent events while the neural patterns for remembered *('R')* events are more dissimilar compared to forgotten *('F')* events, this measure is likely to be associated with event segmentation. **(G)** fMRI evidence for event integration. If the multivariate neural measure remains stable across the boundary of two neighboring events and remembered *('R')* events have higher neural similarity compared to forgotten *('F')* events, this measure may relate to event integration.

Methods

Participants and procedure

Participants

Twenty-two healthy young adults (10 females, age range 18-26, mean age 20.8 years) participated in the experiment. All participants were native English speakers and naïve to the BBC crime drama *Sherlock.* Data were discarded from participants with excessive motion ($>$ 1 voxel; n = 2), low recall duration ($<$ 10 min; n = 2), or sleeping during the experiment (n = 1). This leaves 17 participants in total for our analyses. Due to a technical problem, one participant (s5) is missing data for the last 75 s (part of event 49 and all of event 50) and the affected two events were excluded in the analyses.

Procedure

All our analyses are based on the Sherlock Movie Dataset (Baldassano et al., 2017; Janice Chen et al., 2017); see *Data availability* below) acquired and pre-processed at Princeton Neuroscience Institute. No similar analysis or results (excluding behavioural results of recall accuracy) have been reported in previous studies using this dataset.

Participants were informed that they would watch a movie and would later be required to recall its content. They were then presented with a 48-min segment of the first episode of the *Sherlock* series (encoding phase), split into two parts of approximately equal length (23 min and 25 min) and presented in two consecutive blocks. A 30 s introductory cartoon clip was prepended before each block. Immediately after the movie presentation, participants were instructed to verbally describe the movie in as much detail as they could and for as long as they wished (recall phase). They were asked to recall the episode in the correct sequential order but were permitted to return to earlier points in the narrative if they remembered further content. Audio was simultaneously recorded by a customized MR-compatible recording system throughout the recall phase.

Behavioural data analysis

Event annotations of the movie and verbal speech recording

The movie was segmented into 48 events by an independent observer, following major shifts in the narrative (e.g., director's cuts). Including the two introductory cartoon clips, 50 scenes were analyzed in total. The timestamps for both the onset and offset of identified scenes were recorded and aligned across all participants. Both the onset and offset are referred to as the boundaries of the respective event. This is a widely used method for event segmentation and has been validated by a data-driven approach (Baldassano et al., 2017). The length of the scenes ranges from 11 to 180s (Mean ± SD: 57.5 ± 41.7 s). Each subject's verbal speech was transcribed, segmented, and matched to the events that were recalled from the movie.

Event and order memory

For each participant, we first asked whether events were successfully recalled or not, as in the classical subsequent memory paradigm (Brewer et al., 1998; Fernández et al., 1999; Wagner et al., 1998). An event was labeled as 'remembered' if any part of the event was described during the recall. 'Forgotten' events are the ones that were not mentioned throughout the recall phase.

Secondly, *out-of-order* events were identified as a measure of sequential memory. Among all remembered events, an event was labeled as *out-of-order* if it was not described immediately after its preceding event in the original movie. For example, if *event 3* is described immediately after *event 1* without mentioning *event 2*, then *event 3* is an *out-of-order* event. By contrast, if a participant described *event 4, 5, 6* sequentially during the recall phase, since *event 5, 6* correctly followed their preceding event, *event 5, 6* were counted as *in-order* events. The first event verbally described in the recall phase was always labeled as 'not available' in the order memory analysis since it is not preceded by any event. It was possible that a single scene was mentioned multiple times (in different parts) during the recall, in which case the position of its first recall was used in the analyses.

fMRI data analysis

fMRI data acquisition and pre-processing

fMRI data were acquired using a T2*-weighted EPI sequence on a 3T Siemens Skyra scanner (20-channel head coil; TR 1,500 ms; TE 28 ms; flip angle 64, spatial resolution 3*3*4 mm^3). Only data from the encoding phase were analyzed and reported in the current study.

A standard pre-processing pipeline was followed using FSL (Jenkinson et al., 2012), which includes slice timing correction, motion correction, linear detrending, high-pass filtering (140 s cutoff), co-registration and affine transformation into 3 mm MNI standard space (Janice Chen et al., 2017). The time series were shifted 3 TRs (4.5 s) to account for the Haemodynamic response function (HRF). Data were z-scored across time at every voxel and a 6 mm smoothing kernel was applied.

All subsequent analyses were performed on the pre-processed voxel-wise BOLD signal, in units of functional volume (TR = 1.5 s). Custom MatLab (R2018, The Mathworks, Natick, MA) and Python (version 3.6) scripts were used for both Region of Interest and parcellation-based searchlight analysis.

Region of interest (ROI) selection

The six ROIs used in this study were independently defined by Chen and colleagues, in correspondence to the timescale hierarchy of the event segmentation model (Baldassano et al., 2017; Hasson et al., 2015). Early visual and early auditory cortex were functionally defined based on inter-subject correlation during an audio-visual movie and an audio narrative, respectively (Janice Chen et al., 2016; Simony et al., 2016). ROIs for medial prefrontal cortex (mPFC) and posterior medial cortex (PMC) were taken from the functional atlas derived from resting-state default mode network (https://findlab.stanford.edu/functional_ROIs.html) from FIND lab at Stanford University (Shirer et al., 2012). The hippocampus and posterior parahippocampal gyrus were anatomically defined from the probabilistic Harvard-Oxford Subcortical Structural Atlas (Desikan et al., 2006). Chen and colleagues manually adjust the threshold of around 50% to ensure better anatomical coverage during the visual check.

Whole-brain parcellation

Alongside the ROI-based analysis, we performed a parcel-based searchlight analysis on the basis of 1000 functionally parcellated cerebral regions (https://github.com/ThomasYeoLab/CBIG/tree/master/stable_projects/brain_parcellation/Schaefer2018_LocalGlobal). The parcellation was based on a gradient-weighted Markov Random Field (gwMRF) model, which integrated local gradient and global similarity approaches (Schaefer et al., 2018). Using both task and resting-state fMRI acquired from 1489 participants, parcels with functional and connectional homogeneity within the cerebral cortex were generated (hippocampus and subcortical regions were not included). In this fashion, each of these biologically meaningful and non-overlapping parcels can be treated in the same way as an independent region similar to an ROI in the following analyses.

fMRI-based neural responses to event boundaries

Univariate response

BOLD signals were first averaged for each TR across all voxels in an ROI. Then the time series were z-scored and segmented based on the event annotations mentioned above. The shortest event was 7 volumes (10.5 s), therefore we averaged 6 volumes at the beginning and end of all events in order to assess the change in activity between them.

Activation patterns

Voxel-wise BOLD time series from separate events were first extracted based on the onset and offset timestamps derived from the movie. Multivariate patterns of brain activation were generated for each event by averaging across all volumes within this event. To assess the similarity between two neighboring events, the *activation pattern* for each event of interest was correlated with its following event. The resulting Pearson’s correlation coefficient depicted the extent to which similar representational activity patterns were elicited by neighboring scenes. Lower similarity between two events represented a greater change in neural patterns across the event boundary.

Connectivity patterns

Intra-regional *connectivity pattern* analyses were conducted based on a method originally used in rodent electrophysiology studies to quantify the reactivation of sparsely distributed neuron assemblies (Lansink et al., 2008; Qin et al., 1997), and recently used in human fMRI (Hermans et al., 2017; Tambini & Davachi, 2013, 2019). For each event within each brain region, Pearson's correlations were performed on the extracted m*n (volumes*voxels) BOLD-fMRI time series, between each of the n voxel time series. This yielded an n-by-n pairwise correlation matrix (containing p values indicating the significance of the Pearson's correlations), representing the within-region connectivity structure for each scene. For two neighboring events, the Pearson's correlation coefficient of their correlation matrices was calculated to quantify the similarity for *connectivity patterns*. Lower similarity between two *connectivity patterns* represented a greater change in the intra-region connectivity patterns across the event boundary.

Relationship between neural responses during encoding and subsequent memory

Remembered and forgotten events comparisons

We first compared our neural pattern similarities (i.e., *activation pattern* similarity and *connectivity pattern* similarity) at the single-subject level explained above for each brain region (ROI or brain parcel). The similarity indices (Pearson's r between two matrices) for both *activation* and *connectivity patterns* were averaged for the two types of event pairs (*remembered* and *forgotten*) for each participant. If the first event of the pair was retrieved during the recall phase, the event pair was labeled as *remembered*. *Remembered* and *forgotten* event pairs were then compared in two separate *t*-tests for *activity* and *connectivity pattern* transitions (indexed by pattern similarity). We further examined the relationship between *connectivity pattern* transitions and order memory (i.e., temporal order of event recall). More specifically, *connectivity patterns* were averaged for another two types of event pairs (i.e., *In-order* or *Out-of-order*) for each participant. If the second event of the pair was recalled in an incorrect sequential order (e.g., *event 4* was recalled immediately after *event 6*), the event pair was labeled as *Out-of-order. Connectivity pattern* transitions for *In-order* and *Out-of-order* event pairs were then compared with *t*-tests.

Event-specific correlational analysis

Thus far we have examined the association between memory and neural pattern similarity in a within-participant fashion. We then examined whether the likelihood of an event being remembered correlated with neural responses across participants. The recall rate for an event was the proportion of participants that remembered it. At the same time, the pattern similarity of both *activation* and *connectivity patterns* was calculated and averaged across all participants, generating the neural transition indices across participants. Recall rates for all 50 events and their corresponding pattern transition measures were then correlated, providing a further indication of how subsequent memory related to pattern transitions across boundaries.

Relationship between hippocampal pattern similarity and event distance

The above analyses focused on neural pattern similarities between two neighboring events. Here, we examined the hippocampal pattern similarities between events with variable distances. Event distance was defined as the number of event boundaries between two events (the event distance between event 1 and event 3 is 2). For each event, we first calculated its *activation* and *connectivity pattern*. Then, we calculated the *activation* and *connectivity pattern* similarity between all possible combinations of event A-B pairs ('Event A' is the event which appeared earlier in the temporal sequence, and 'Event B' is the one presented later) within all 50 events. Finally, for each participant, and each event distance, two mean similarities for activation and connectivity pattern were calculated separately. Note that the number of available pairs decreases as the distance increases (e.g., events 1-50 are the only event pair with a distance of 49). To ensure a well-powered analysis for every event distance, we only compared event pairs with a distance less than or equal to 40, meaning at least 10 event pairs contributed to the event distance calculation. Analysis of all distances ($d \leq 49$) can be found in the **Supplementary Materials**.

First, one-sample *t*-tests were performed separately on each distance to test the difference between zero and the distance-specific activation and connectivity pattern similarities. All resulting *p* values were corrected for False Discovery Rate (FDR) based on the number of distances included (from $d_{min} = 1$ to $d_{max} = 40$). Next, we used linear regression to examine the relationship between pattern similarity and event distance. In addition, to investigate how the subsequent memory of the preceding event (event A) modulates the relationship between event distance and pattern similarity, we ran a two-way ANOVA (memory * event distance) using the memory performance (remembered or forgotten) of the preceding event and event distance (range from 1 to 40) as two independent variables.

Statistical analysis

For hypothesis tests involved in the fMRI data analyses, the significance level was set to $p = 0.05$ (two-tailed). Except for the permutation test for simulated event boundaries (see **Supplementary Materials**), *p* values were based on the parametric testing. To account for the multiple comparisons problem that comes with multiple ROIs or parcels, all reported *p* values in the main text were FDR-corrected (p_{FDR}) (Genovese et al., 2002) unless otherwise stated (p_{raw}). Specifically, this means correction was made for six tests in ROI analyses, and 1000 tests for the whole-brain analyses. All significant *p* values were reported together with the effect sizes (Cohen's d or partial η^2). The custom modified version of DABEST (https://github.com/ACCLAB/DABEST-python) was used to plot individual data points alongside bootstrapping-based resampled distributions of the mean difference between conditions (Ho et al., 2019).

Data and code availability

ROI data are available at http://datasets.datalad.org/?dir = /workshops/mind-2017/sherlock. Whole-brain neuroimaging data are available at https://dataspace.princeton.edu/jspui/handle/88435/dsp01nz8062179. Custom code used in this study will be publicly available via the Open Science Framework (OSF) (Link: https://osf.io/p68cv/?view_only=483703873dae4cfd-8b36e9d6df6b8c92) upon publication. Further requests for scripts should be directed to the corresponding author.

Results

Subsequent memory performance measured by spoken recall

The dataset (Baldassano et al., 2017; Janice Chen et al., 2017) is from an experiment in which 17 healthy participants watched a 50-min audio-visual movie (BBC's Sherlock) while undergoing an fMRI scan (**Figure 2.2A**). Immediately thereafter, participants were instructed to verbally recall the movie in as much detail as possible (**Figure 2.2B**). No visual or auditory cues were given during the retrieval session.

Similar to the previous experiments probing the subsequent memory effect (Brewer et al., 1998; Fernández et al., 1999; Wagner et al., 1998), the central purpose of our analyses was to identify brain regions and their response patterns that predict subsequent recall. To quantitatively analyze memory retrieval performance, the movie was divided into 50 events based on major narrative shifts (e.g., director's cuts). Each participant's spoken recall was transcribed and segmented into events matching those from the movie (**Figure 2.1B**) (Details in *Methods and Materials*). The current analyses used the same event annotations for the movie and spoken recall as the original studies (Baldassano et al., 2017; Janice Chen et al., 2017).

We first calculated recall accuracies for each participant. On average, 68.7% (*SD* = 12%, range 48% - 94%) of the 50 events (*Mean* = 34.4 events, *SD* = 6) were retrieved successfully (**Figure 2.2C**). Among these remembered events, we further defined *in-order* and *out-of-order* events based on whether they were recalled in the correct sequential order. On average, 58.8% (*SD* = 8%, range 40% - 71%) of the remembered events were *in-order* (**Figure 2.2D**).

Distinct activation pattern-mediated event segmentation is associated with subsequent retrieval success

We quantified neural similarities of event-specific *activation patterns* before and after event boundaries (i.e., two neighboring events). Specifically, we generated a voxel-wise *activation pattern* per event by averaging over all time points in that event. This time-averaged *activation* *pattern* of all voxels within an ROI for an event was compared to the pattern of its subsequent event using Pearson's correlation. A negative Pearson's r indicates two separateble activation patterns and thus distinct neural representations for two distinct events. We investigated whether *activation pattern* similarities relate to memory formation by contrasting the pattern similarities of remembered with forgotten events in six ROIs. That is, pattern similarity between two events was compared to subsequent memory for the first of those events. We found that subsequently remembered events were associated with lower *activation pattern* similarities than subsequently forgotten events in early auditory cortex (t = -3.56, p_{FDR} = 0.007, Cohen's d = 0.92, **Figure 2.3B**), hippocampus (t = -3.62, p_{FDR} = 0.007, Cohen's d = 0.92, **Figure 2.3E**), mPFC (t = -2.79, p_{FDR} = 0.01, Cohen's d = 0.80, **Figure 2.3C**) and posterior parahippocampal gyrus (pPHG) (t = -2.85, p_{FDR} = 0.01, Cohen's d = 0.89, **Figure 2.3F**). This finding suggests that distinct *activation patterns* for two sequential events are beneficial for the memory of the first event in that sequence. Early visual areas (t = -1.13, p_{FDR} = 0.27, Cohen's d = 0.35, **Figure 2.3A**) and PMC (t = -1.91, p_{FDR} = 0.08, Cohen's d = 0.65, **Figure 2.3D**) did not show this marked effect. In addition to the comparison between remembered and forgotten events, we generated shuffled event boundaries as the baseline. Our main results only existed for the actual even boundaries, but not shuffled boundaries (**See Supplementary Materials (Figure S2.4))**.

So far, within-participant comparisons between remembered and forgotten events revealed that differences in *activation pattern* similarities of several ROIs are related to subsequent memory. Next, we examined whether a similar relationship is evident across participants. Specifically, we investigated the relationship between the *event-specific recall rate* (the percentage of participants that successfully recalled a particular event) and the averaged *activation pattern* similarity for the corresponding event (the first one in the sequence) across all participants. Consistent with our main analyses, this analysis revealed that the recall rate negatively correlated with *activation pattern* similarity in the hippocampus (r = -0.292, p_{raw} = 0.042) and pPHG (r = -0.344, p_{raw} = 0.015), suggesting that events with lower *activation pattern* similarity were more likely to be recalled (**Figure S2.5**).

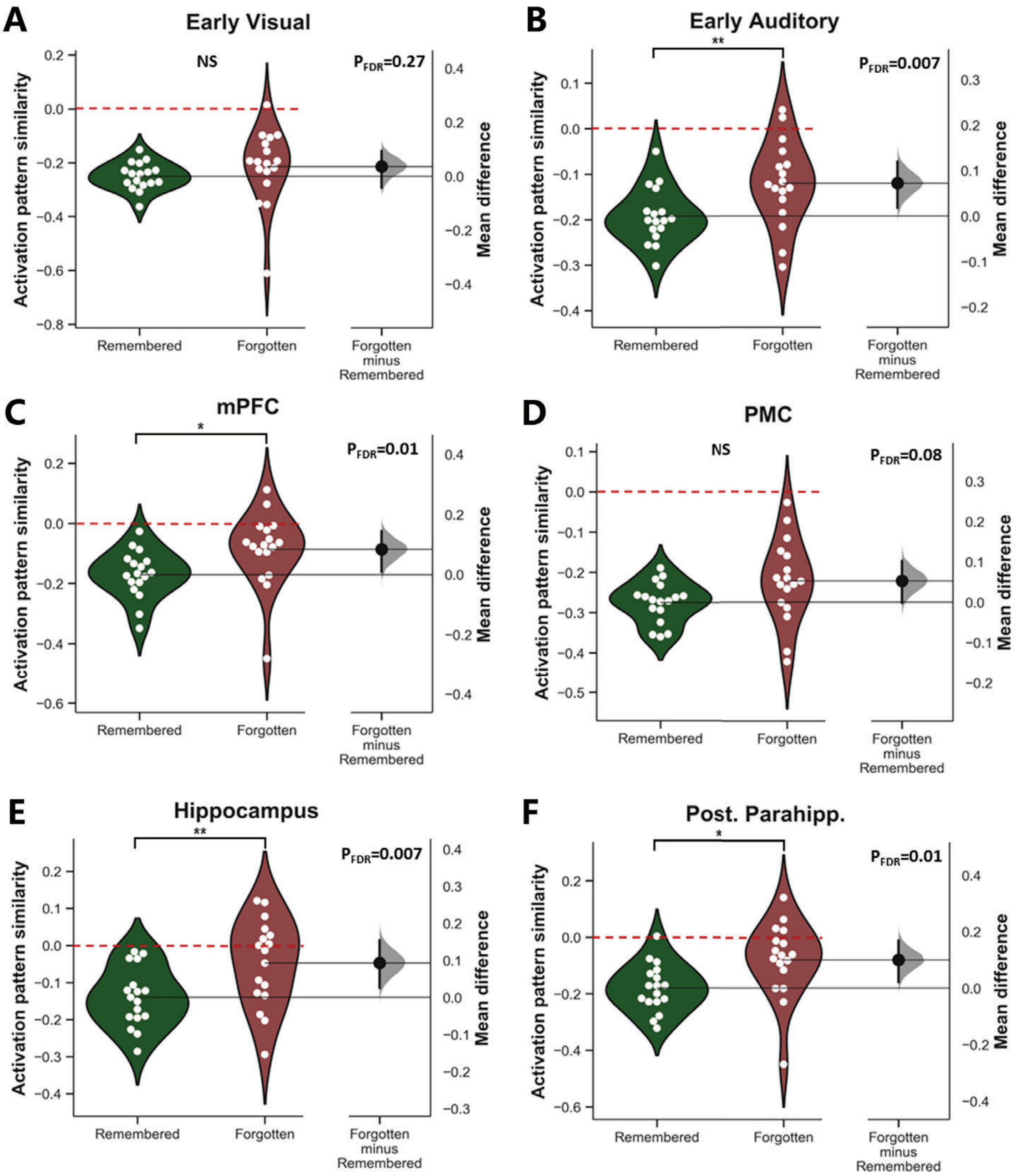


Figure 2.3. Association between activation pattern similarities of six ROIs and subsequent memory recall. We compared *activation pattern* similarities of sequential event pairs based on subsequent memory performance of the first event (*Remembered* vs. *Forgotten*) across six ROIs. For panel A-F, *activation pattern* similarities for *Remembered* events are displayed on the left (*green*), while similarities for *Forgotten* events are displayed on the right *(red).* For each comparison, a separate axis displays the *mean difference*. The curve (gray) indicates the resampled distribution of the *mean difference* generated via bootstrapping. The solid vertical line attached to the curve represents the *mean difference* as a 95% bootstrap confidence interval. We found significantly lower *activation pattern* similarity for *Remembered* vs. *Forgotten* event pairs in the early auditory area (t = -3.56, p_{FDR} = 0.007, Cohen's d = 0.92; panel **B**), mPFC (t = -2.79, p_{FDR} = 0.01, Cohen's d = 0.80; panel **C**), hippocampus (t = -3.62, p_{FDR} = 0.007, Cohen's d = 0.92; panel **E**), and pPHG (t = -2.85, p_{FDR} = 0.01, Cohen's d = 0.89; panel **F**). No significant differences were found in early visual areas (t = -1.13, p_{FDR} = 0.27, Cohen's d = 0.35; panel **A**) and PMC (t = -1.91, p_{FDR} = 0.08, Cohen's d = 0.65; panel **D**). NS=Not significant; ***** p_{FDR}<0.05; ****** p_{FDR}<0.01.

Similar connectivity pattern-mediated event integration is correlated with subsequent retrieval success

Next, we investigated the association between *connectivity patterns* – a different multivariate method to characterize neural representations – and subsequent memory retrieval. Within-region multi-voxel *connectivity patterns* were calculated by a voxel-by-voxel pairwise correlation matrix resulting from the correlations between time courses of all voxels within a given region. This represents the relative correlation structure between all voxels in a certain region during event processing. We first calculated the event-specific within-region *connectivity patterns* for two sequential events, and then we quantified the similarity between *connectivity patterns* across event boundaries also using Pearson's r. Contrasting similarities of *connectivity patterns* of subsequently remembered and forgotten events allowed us to examine how transitions in *connectivity patterns* contribute to memory formation. We found higher *connectivity pattern* similarity for subsequently remembered compared to forgotten events in the early auditory area ($t = 2.9$, $p_{FDR} = 0.02$, Cohen's d = 0.72, **Figure 2.4B**), visual areas ($t = 3.34$, $p_{FDR} = 0.01$, Cohen's d = 0.74, **Figure 2.4A**), hippocampus ($t = 3.39$, $p_{FDR} = 0.01$, Cohen's d = 0.73, **Figure 2.4E**), and PMC ($t = 2.79$, $p_{FDR} = 0.02$, Cohen's d = 0.47, **Figure 2.4D**). The same contrast was not significant for mPFC ($t = 1.22$, $p_{FDR} = 0.23$, Cohen's d = 0.25, **Figure 2.4C**) and pPHG ($t = 1.36$, $p_{FDR} = 0.22$, Cohen's d = 0.30, **Figure 2.4F**). A follow-up permutation test examining the specificity of subsequent memory effects to actual event boundaries (as opposed to randomly generated pseudo boundaries) can be found in the **Supplementary Materials (Figure S2.6)**.

The event-specific correlational analysis demonstrated that the recall rate positively correlated with *connectivity pattern* similarity in the early auditory area ($r = 0.327$, $p_{raw} = 0.022$), visual areas ($r = 0.35$, $p_{raw} = 0.014$), hippocampus ($r = 0.301$, $p_{raw} = 0.036$), PMC ($r = 0.341$, $p_{raw} = 0.017$), and pPHG ($r = 0.341$, $p_{raw} = 0.017$) (**Figure S2.7**). These results suggest that events with higher connectivity pattern similarity in these ROIs were more likely to be recalled.

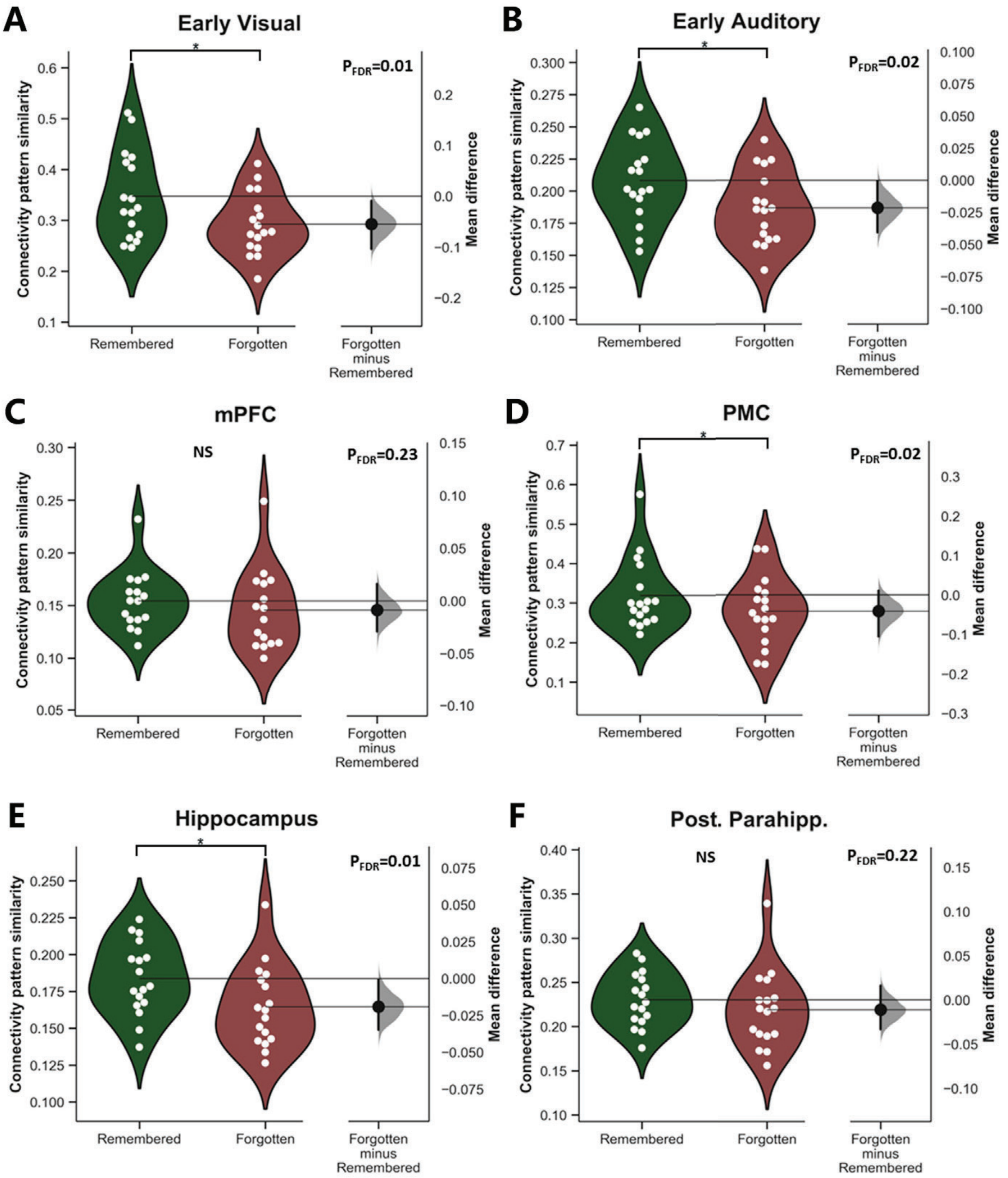


Figure 2.4. Association between connectivity pattern similarities of six ROIs and subsequent memory recall. We compared connectivity pattern similarities of sequential event pairs based on subsequent memory performance of the first event (*Remembered* vs. *Forgotten*) across six ROIs. For panel A-F, connectivity pattern similarities for *Remembered* events are displayed on the left (*green*), while similarities for *Forgotten* events are displayed on the right *(red).* For each comparison, a separate axis displays the *mean difference*. The curve (*gray*) indicates the resampled distribution of the *mean difference* generated via bootstrapping. The solid vertical line attached to the curve represents the *mean difference* as a 95% bootstrap confidence interval. We found significantly higher connectivity pattern similarity for *Remembered (green)* vs. *Forgotten (red)* event pairs in the early auditory area ($t = 2.9$, $p_{FDR} = 0.02$, Cohen's d = 0.72, panel **B**), visual areas ($t = 3.34$, $p_{FDR} = 0.01$, Cohen's d = 0.74, panel **A**), hippocampus ($t = 3.39$, $p_{FDR} = 0.01$, Cohen's d = 0.73, panel **E**), and PMC ($t = 2.79$, $p_{FDR} = 0.02$, Cohen's d = 0.47, panel **D**). No significant differences were found in mPFC ($t = 1.22$, $p_{FDR} = 0.23$, Cohen's d = 0.25, panel **C**) and pPHG ($t = 1.36$, $p_{FDR} = 0.22$, Cohen's d = 0.30, panel **F**). NS=Not significant; ***** $p_{FDR}<0.05$.

Similar connectivity pattern-mediated event integration preserves sequential order of events in later retrieval

So far we have shown the opposite association between our two multivariate neural pattern measures and subsequent memory performance: distinct *activation patterns*, but similar within-region *connectivity patterns* across events in the early auditory cortex and hippocampus predict retrieval success. This pattern of results suggests that the *connectivity pattern* may represent the 'chunking code' to integrate events into a continuous sequence. To directly test this 'chunking code' hypothesis, we examined the relationship between *connectivity pattern* similarity and sequential order of subsequent recall. We reasoned that if the *connectivity patterns* remain stable across event boundaries, events should tend to be recalled in the correct sequential order. We compared the mean *connectivity pattern* similarities for *in-order* and *out-of-order* events. Controlling for multiple comparisons, we found that *connectivity pattern* similarity in early visual cortex to be larger for *in-orde*r compared to *out-of-order* events ($t = 3.16$, $p_{FDR} = 0.03$, Cohen's d = 0.47, **Figure 2.5A**). Similar trends that did not survive correction for multiple comparisons were detected in the hippocampus ($t = -2.43$, $p_{raw} = 0.026$, $p_{FDR} = 0.08$, Cohen's d = 0.53, **Figure 2.5E**), auditory area ($t = -2.08$, $p_{raw} = 0.053$, $p_{FDR} = 0.084$, Cohen's d = 0.46, **Figure 2.5B**) and posterior parahippocampal gyrus ($t = -2.05$, $p_{raw} = 0.056$, $p_{FDR} = 0.084$, Cohen's d = 0.36, **Figure 2.5F**). No such effect was observed in the mPFC ($t = -1.35$, $p_{FDR} = 0.19$, Cohen's d = 0.19, **Figure 2.5C**), and PMC ($t = -2.05$, $p_{FDR} = 0.12$, Cohen's d = 0.33, **Figure 2.5D**).

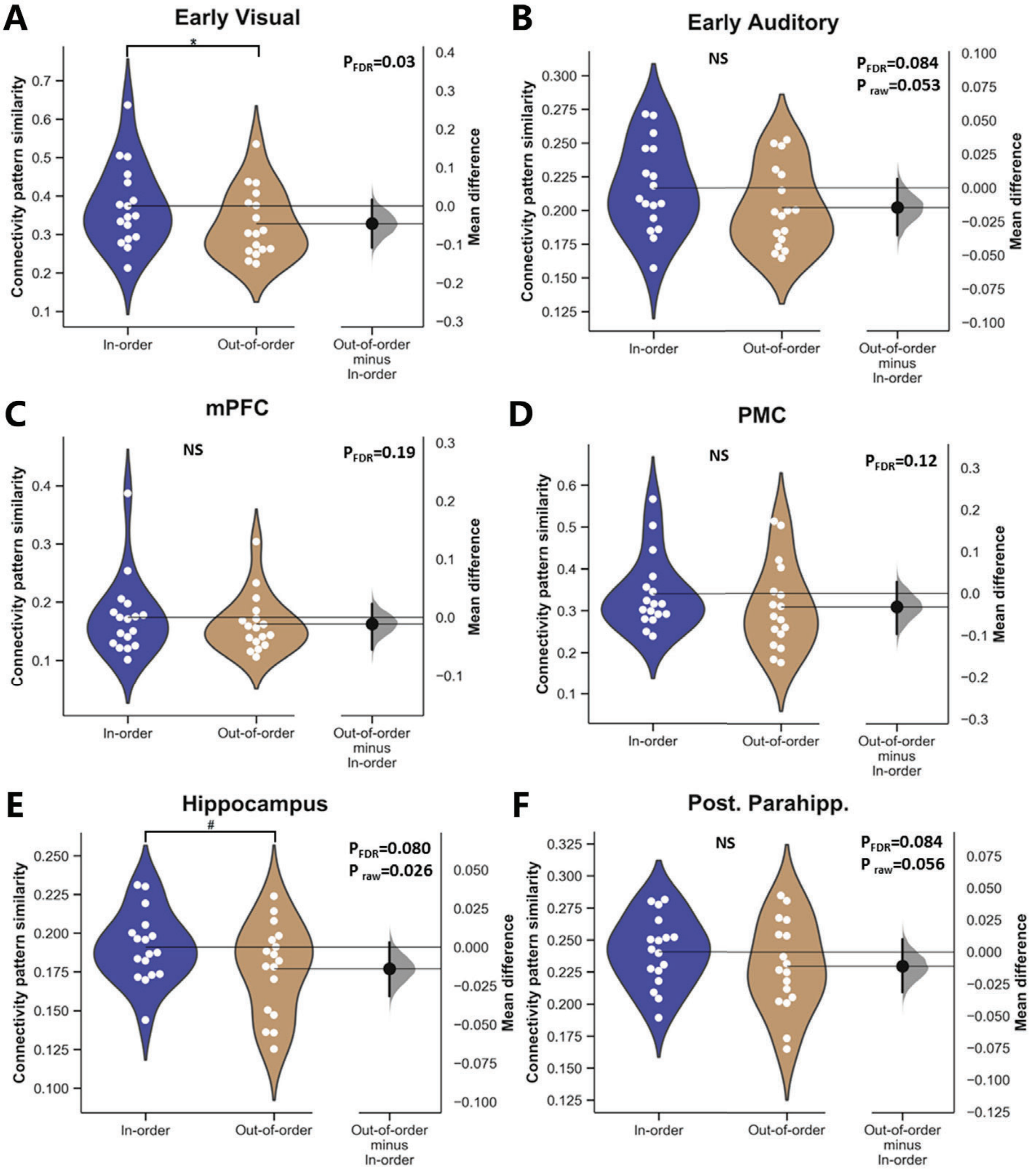


Figure 2.5. **Association between connectivity pattern similarities of six ROIs and sequential order of memory recall.** We compared connectivity pattern similarities of sequential event pairs (*In-order* vs. *Out-of-order*) based on sequential memory performance of the first event across six ROIs. For panel A-F, connectivity pattern similarities for *In-order* events are displayed on the left (*BLUE*), while similarities for *Out-of-order* events are displayed on the right *(BROWN)*. Early visual areas (t = 3.16, p_{FDR} = 0.03, Cohen's d = 0.47, panel **A**) demonstrated higher connectivity pattern similarity for the *In-order* events compared to *Out-of-order* events. A similar trend was also detected in the hippocampus (t = -2.43, p_{raw} = 0.026, Cohen's d = 0.53, panel **E**), but it did not survive FDR correction (p_{FDR} = 0.08). We also found modest, non-significant trends in the early auditory area (t = -2.08, p_{raw} = 0.053, p_{FDR} = 0.084, Cohen's d = 0.46, panel **B**) and posterior parahippocampal gyrus (t = -2.05, p_{raw} = 0.056, p_{FDR} = 0.084, Cohen's d = 0.36, panel **F**). No similar effects were detected in mPFC (t = -1.35, p_{FDR} = 0.19, Cohen's d = 0.19, panel **C**), and PMC (t = -2.05, p_{FDR} = 0.12, Cohen's d = 0.33, panel **D**). NS=Not significant; ***** $p_{FDR}<0.05$; **#** $p_{raw}<0.05$.

Hippocampal activation and connectivity patterns change differently with event distance

Among our six ROIs, we found converging evidence for a dissociation of event segmentation and integration in the hippocampus: lower *activation pattern* similarity, but higher *connectivity pattern* similarity was beneficial for memory formation. Building on these findings, we hypothesized that hippocampal *activation patterns* of neighboring events should be less similar than events that occur far apart. By contrast, hippocampal *connectivity patterns* of close events should be more similar than events with a long interval in between. Thus, we calculated the *activation* and *connectivity pattern* similarity between all possible combinations of event pairs ('Event A' and 'Event B') within all 50 events (**Figure 2.6A** and **2.6D**). For all pairs of events with the same event distance (e.g., separated by four events), we calculated the mean similarity measure for *activation pattern* and *connectivity pattern* separately. This calculation was repeated for all possible event distances. To ensure reliable estimations of pattern similarities, we only present the similarities of distances with at least ten event pairs ($d \leq 40$) in the main text. (Complete calculations can be found in **Figure S2.8**)

We analysed the hippocampal *activation* and *connectivity patterns* separately. First, our *activation pattern* analysis found that the shorter the event distance, the more distinct the hippocampal *activation patterns* ($r = 0.21$, $p_{raw} = 1.8 \times 10^{-8}$; **Figure 2.6B** and **S2.9A**). This positive correlation was largely driven by the negative similarity values between events that occurred close to each other: events separated by a distance of less than four were represented by two distinct (neural similarity significantly lower than 0) hippocampal *activation patterns* ($d = 1$, $t = -5.52$, $p_{FDR} = 0.0006$; $d = 2$, $t = 3.86 \times 10^{-11}$, $p_{FDR} = 1.5 \times 10^{-9}$; $d = 3$, $t = 6.75 \times 10^{-6}$, $p_{FDR} = 0.0001$; $d = 4$, $t = -2.98$, $p_{FDR} = 0.08$). Events with a distance larger than or equal to four did not show markedly distinct *activation patterns* (neural similarity not significantly different from 0) (**Figure 2.6B**). Furthermore, we found that subsequent memory recall of Event A modulated the relationship between event distance ($d = 1 - 4$) and *activation pattern* similarity (ANOVA with event A × distance interaction: $F(3,48) = 10.1$, $p < 0.001$; **Figure 2.6C**). That is, hippocampal *activation pattern* similarities increased as the event distance changes from 1 to 4, but only if event A was later recalled ($F_{remembered}(3,48) = 9.54$, $p < 0.001$; $F_{forgotten}(3,48) = 1.35$, $p = 0.268$).

Second, our *connectivity pattern* analysis found that the shorter the event distance, the more similar the hippocampal *connectivity patterns* ($r = -0.439$, $p_{raw} = 1.8 \times 10^{-33}$; **Figure 2.6E** and **S2.9B**). At the same time, across all event distances, the *connectivity pattern* similarities were consistently higher than zero (from $d = 1$, $t = 31.86$, $p_{FDR} = 2.29 \times 10^{-14}$ to $d = 40$, $t = 18.16$, $p_{FDR} = 4.4 \times 10^{-12}$; $p_{FDR} < 0.05$ for all d). Furthermore, we found a significant interaction between event A recall and distance ($F(19, 304) = 2.37$, $p = 0.001$), and a significant main effect of event A ($F(1, 16) = 7.53$, $p = 0.014$). That is, if event A was recalled later, its hippocampal connectivity pattern was more similar to any other event in the sequence, compared to when event A was not successfully recalled (**Figure 2.6F**). This suggests that if *connectivity patterns* between pairs of events are more similar, for both short and long distances, then events are more likely to be successfully encoded.

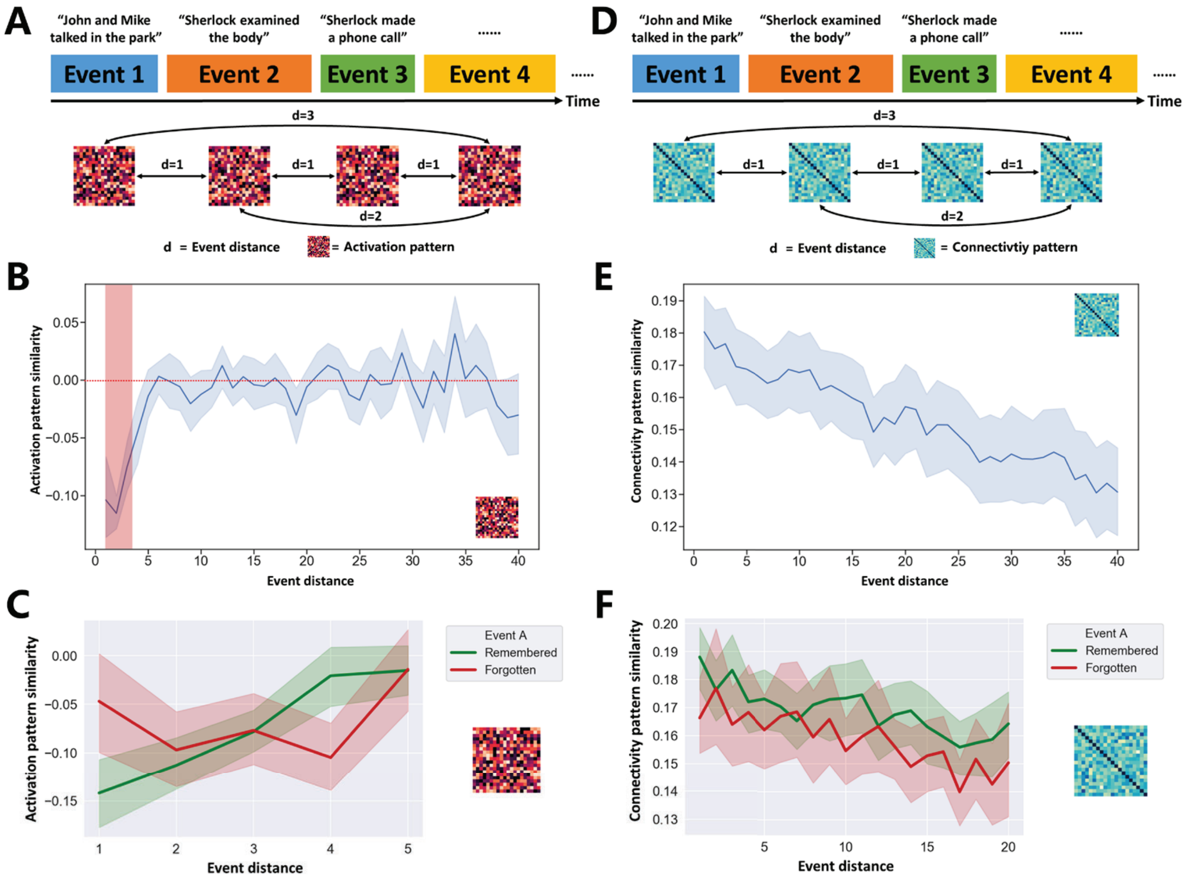


Figure 2.6. Hippocampal pattern similarity changes with event distance. (A) Hippocampal *activation patterns* were generated for all 50 events. We calculated *activation pattern* similarities between sequential events (event distance = 1) and all possible combinations of non-sequential event pairs (event distance > 1). **(B)** Hippocampal *activation patterns* between pairs of events were significantly dissimilar for events separated by a distance of less than 4 (red shadow). **(C)** Memory performance modulated the distance-*activation pattern* similarity relationship. If the first event (*Event A*) of the pair was successfully encoded, *activation pattern* similarities of the event pair increased with event distance (green line). **(D)** Hippocampal *connectivity patterns* were generated for all possible combinations of event pairs. **(E)** Event pairs with shorter event distance had more similar hippocampus *connectivity patterns*. At the same time, similarities of hippocampus *connectivity patterns* are higher than 0 regardless of event distance. **(F)** Memory performance modulated distance-*connectivity pattern* similarity relationship. If the first event (*Event A*) of the event pair was successfully encoded, *connectivity pattern* similarities of the event pair are enhanced regardless of their event distance. For panel B-F, error bands (i.e., light shadow around the solid line) represent the 95% confidence interval of the mean.

Subregions of the prefrontal cortex perform event segmentation and integration

Our ROI-level analyses found that (1) distinct hippocampal *activation patterns* were associated with better event memory; (2) similar hippocampal *connectivity patterns* were beneficial for event memory; (3) although not surviving multiple comparison correction, similar hippocampal *connectivity patterns* tended to preserve the sequential order of events (**Figure 2.7A**). To investigate whether these relationships are present in other brain regions beyond our six ROIs, we ran a parcel-based searchlight version of our pattern similarity analysis to identify overlapping event segmentation and integration computations across neocortical parcels. In sum, we investigated three potential relationships between neural pattern similarity and subsequent retrieval separately. First, we identified brain parcels whose lower *activation pattern* similarities across events were associated with retrieval success (**Figure S2.10A**). Next, we mapped the association between higher *connectivity pattern* similarities and retrieval success on each parcel (**Figure S2.10B**). Then, we identified the parcels, which demonstrated a positive association between *connectivity pattern* similarities and order memory (**Figure S2.10C**).

To identify brain parcels that may support all three neural computations, similar to the hippocampus, we overlapped spatial patterns for these three effects (all $p_{\mathrm{FDR}} < 0.05$). This revealed a set of brain regions including relatively large clusters (at least 50 voxels) in the mPFC, right inferior frontal gyrus (IFG), anterior/middle cingulate cortex and supplementary motor area (SMA), left inferior temporal gyrus (ITG) and left insular (**Figure 2.7B**). These results suggest that this network of cortical regions may use the same neural processes to perform event segmentation and integration as the hippocampus during continuous memory encoding.

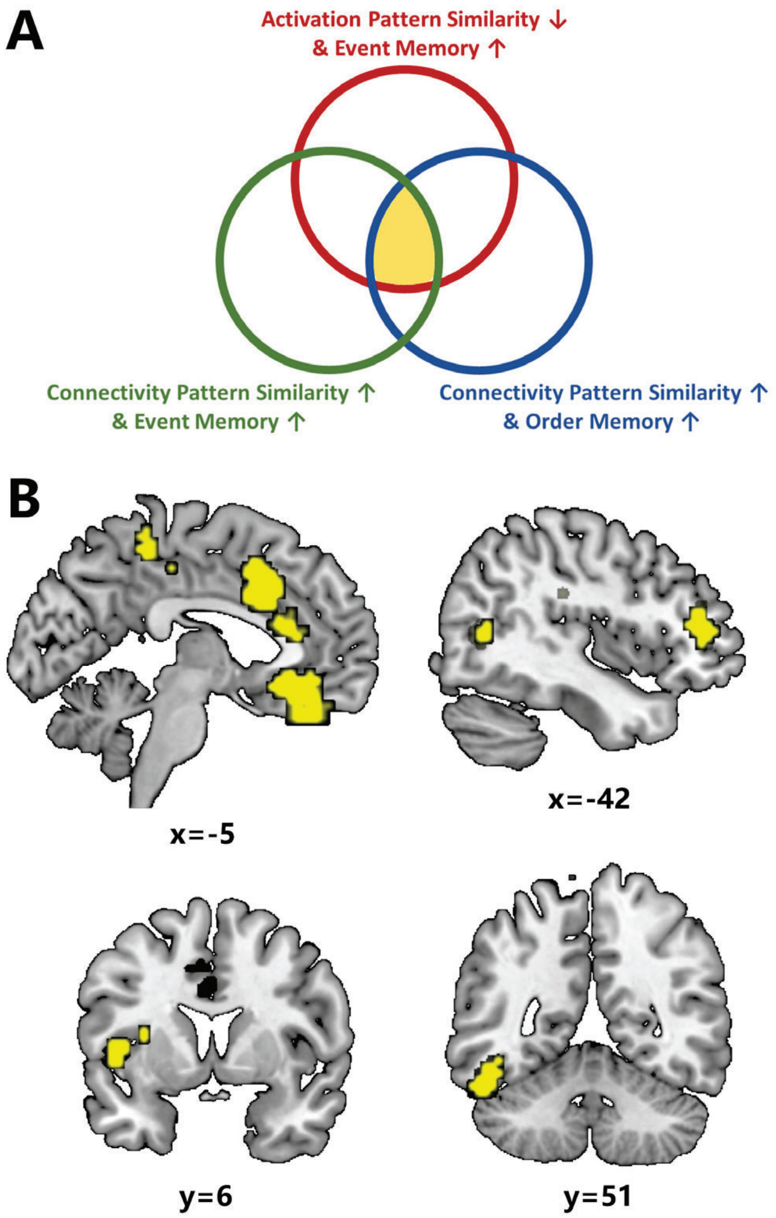

A
Activation Pattern Similarity ↓
& Event Memory ↑
Connectivity Pattern Similarity ↑
& Event Memory ↑
Connectivity Pattern Similarity ↑
& Order Memory ↑
B
x=-5
x=-42
y=6
y=51

Figure 2.7. Identifying overlapping event segmentation and integration computations across the neocortex. (A) We identified three relationships between neural pattern similarity and subsequent memory in the hippocampus. **(B)** Similar to the hippocampus, overlapping event segmentation and integration computations were found in a network of brain regions including the medial prefrontal cortex (mPFC), right inferior frontal gyrus (IFG), anterior/middle cingulate cortex and supplementary motor area (SMA), left inferior temporal gyrus (ITG), and left insular ($p_{FDR} < 0.05$ across 1000 parcels, cluster size >= 50).

Discussion

To successfully form memories of our life experiences, we need to segregate continuous experience into events (Baldassano et al., 2017; Williams et al., 2019), and integrate those events across their boundaries into a coherent narrative (Benjamin J Griffiths & Fuentemilla, 2020). Here we show that distinct hippocampal *activation patterns*, but similar hippocampal *connectivity patterns* across event boundaries, facilitate these two vital episodic memory functions. We propose that distinct *activation patterns* reflect event segmentation while similar *connectivity patterns* represent a 'chunking code' that integrates separately represented events into a narrative. Supporting this role of *connectivity patterns* for event integration, we found that similar hippocampal *connectivity patterns* were crucial for the correct sequential order of subsequent retrieval. Our whole-brain analysis demonstrates that similar neurocomputations were performed by a network of cortical regions, in particular for the mPFC. Overall, these results suggest that both hippocampal and medial prefrontal event segmentation and integration support memory formation of continuous experience.

Using multivoxel pattern analysis, we found that distinct local *activation patterns* across event boundaries in the early auditory area, mPFC, posterior parahippocampal gyrus, and hippocampus, were associated with better subsequent memory, indexed by more negative similarities of activation patterns between two adjacent events. The ability to segment continuous experience has been linked to successful memory encoding in a behavioural experiment (Sargent et al., 2013) and compelling evidence suggested that the hippocampus is activated around event boundaries (Ben-Yakov et al., 2013; Ben-Yakov & Dudai, 2011; Ben-Yakov & Henson, 2018; DuBrow & Davachi, 2013; Williams et al., 2019). This hippocampal activity has been proposed to be associated with a hippocampal segmentation process, but how the hippocampus represents two separate events, and whether the corresponding neural representations are relevant for memory remained unclear. Our findings suggest that the hippocampus and other brain regions (e.g., mPFC) segment events by representing them with two distinct patterns of activity. This is consistent with the role of the hippocampus in pattern separation: when similar experiences need to be discriminated against and encoded, the underlying hippocampal neural representations tend to be dissimilar (Bakker et al., 2008; Yassa & Stark, 2011). This has typically been studied to show how the brain separates perceptually similar stimuli (i.e., images), but our findings indicate that a similar separation occurs at the level of events and this determines subsequent memory. The episodic memory system may use 'orthogonalized' neural representations to encode two events for the purpose of event segmentation. Further, we show these 'orthogonalized' neural representations are potentially event-distance dependent: the hippocampus only generates consecutive dissimilar patterns when events occur relatively close in time. Taken together, this suggests the existence of a brain network (mainly hippocampus and mPFC) for the continuous segmentation of ongoing experience, and the degree of neural separation is relevant for memory formation.

Complementing this, we found that more similar within-region *connectivity patterns* of several regions across event boundaries, including again the early auditory area and hippocampus, were associated with the better subsequent recall. Compared to local *activation patterns* (J. D. Cohen, Daw, Engelhardt, Hasson, Li, Niv, Norman, Pillow, Ramadge, Turk-Browne, & Willke, 2017; Xue, 2018), within-region *connectivity patterns* are a less used multivariate approach. Recently, Tambini and Davachi proposed that both *activation* and *connectivity patterns* could be used to capture neural states during memory encoding and reactivation, but *connectivity patterns* tend to encode contexts or states instead of particular perceptual inputs (Tambini & Davachi, 2019). Our results support this notion, whereby *activation patterns* were more event-specific, while *connectivity patterns* were more associated with the temporal context of events. Therefore, the *connectivity pattern* acts as the ‘chunking code’ to integrate segmented and separately represented events into a coherent narrative. Previous evidence from invasive recordings of hippocampal neurons in rats (Manns et al., 2007) and patients with pharmacologically intractable epilepsy (Paz et al., 2010) suggested that the temporal context of events is hippocampally encoded. Specifically, Paz and colleagues found that neuronal activity in the hippocampus became more correlated across viewing repetitions of short movie clips, which suggests coding of the temporal context within events (Paz et al., 2010). Our *connectivity pattern* measure suggests that the hippocampus also codes temporal context across successive events, integrating them into a narrative.

In addition, we found that close event pairs tend to have more similar *connectivity patterns*, and that *connectivity pattern* similarities are lower for forgotten compared to remembered events. This holds for event pairs with both short and long distances, suggesting the relevance of similar *connectivity patterns* for memory formation across the entire narrative. Multi-voxel *connectivity pattern* analysis, as a less used multivariate neural measure, may be applied as an alternative approach to study how temporal sequences are neurally represented. Evidence suggests that neural activity in the hippocampal-entorhinal region, measured in both rats and humans, represents the temporal sequence of experience (Bellmund et al., 2019; Lositsky et al., 2016; MacDonald et al., 2011; Montchal et al., 2019; Pastalkova et al., 2008; Thavabalasingam et al., 2019; Tsao et al., 2018). Adding to this evidence, our findings suggest that a stable *connectivity pattern* across events appears to be a marker of this temporal sequence coding. Future studies are needed to further investigate the precise mnemonic functions of different neural measures (e.g., *activity pattern*, within-region *connectivity pattern*, and system-level interaction between regions) during memory formation (Tambini & Davachi, 2019).

Our ROI analysis highlights the two functions of the hippocampus in the separate representation of segmented events and the binding function that linked events into a narrative, and parcel-based searchlight analysis identified the role of subregions of the prefrontal cortex (e.g., mPFC, IFG), insular, and inferior temporal gyrus in event segmentation and integration during memory formation. The role of the mPFC in event integration is particularly thought-provoking. The mPFC is generally implicated in encoding and retrieval of episodic memories (Kim, 2010; Rugg & Vilberg, 2013). Among its variety of functions in learning and memory (Fernández & Morris, 2018), the online integration of events we observed here is consistent with its function in the facilitation of associative inference (Preston & Eichenbaum, 2013; Schlichting et al., 2014; Schlichting & Preston, 2015; Spalding et al., 2018; Zeithamova et al., 2012), accumulation of knowledge (Berkers et al., 2018; Kumaran et al., 2009), and integration of new and prior knowledge (van Kesteren et al., 2010, 2013, 2014). We propose that the general mnemonic function of mPFC is to establish links between separate elements across time and space. Taken together, we found that the hippocampus-mPFC circuit performs event segmentation and integration during memory formation of continuous experience. These findings demonstrate the contribution of two complementary event processing mechanisms and underlying neural representations in episodic memory formation. The hierarchical network model of event segmentation proposes that higher-order regions receive event representations from lower-order perceptual regions, and then transfer these representations to the hippocampus for storage (Baldassano et al., 2017; Hasson et al., 2015). Our study suggested that event integration is another key cognitive process involved in event memory by showing how distinct event representations are integrated by similar *connectivity patterns* of the hippocampus and mPFC.

Our study, together with previous studies also combining human fMRI with naturalistic stimuli (Baldassano et al., 2017; Janice Chen et al., 2017; Hasson et al., 2008), demonstrates the potential of this approach to advance our understanding of the human memory system, in particular for the formation of real-life memories. Similar paradigms and analyses can be easily adapted in clinical (e.g., memory and affective disorders) and developmental neuroimaging studies (e.g., children and older adults) to reveal changes related to disease or (mal)development. For example, fMRI-based event segmentation and integration measures could be used to probe how these processes are impaired in Alzheimer's disease and mild cognitive impairment, how they develop from childhood to adulthood and diminish in normal aging. In addition, *connectivity patterns* have the potential to inform our understanding of other cognitive operations that require integration of information, such as inferential reasoning (Preston & Eichenbaum, 2013). However, due to the low temporal resolution of fMRI, the directionality of information flow between the neocortical regions of the 'hierarchical memory system' (Hasson et al., 2015) and the hippocampus remains unclear. Future application of deep-source magnetoencephalography (MEG) (e.g., Backus, Schoffelen, Szebényi, Hanslmayr, & Doeller, 2016) or intracranial electroencephalography (iEEG) (e.g., Jafarpour, Griffin, Lin, & Knight, 2019) with naturalistic memory paradigms may bridge this gap.

In sum, we show that the hippocampus and mPFC may perform a dual function during naturalistic memory formation. Both regions segment events by representing them with distinct *activation patterns*, while also integrating those events by retaining similar *connectivity patterns* across events, enabling the representation of a coherent narrative. The ability to measure segmentation- and integration-related neural operations using fMRI opens new opportunities to investigate the mechanisms of memory encoding for real-life experience.

Acknowledgment

We thank Dr. Janice Chen for sharing the dataset and ROI masks and suggestions on the searchlight analysis of this manuscript; Kirsten Rittershofer for proofreading this manuscript; Members of the Hasson and Norman laboratories for their efforts during data acquisition and data sharing. This work was supported by the Chinese Scholarship Council (Ph.D. fellowship (201606990020)) to W.L), Ministerie van OCW (Sino-Dutch Bilateral Exchange Scholarship to YJS), and Nuffic Neso China (Orange Tulip Scholarship to YJS).

Supplementary Material for
Hippocampal-medial prefrontal event segmentation and integration contribute to episodic memory formation

This file includes:
Supplementary Text
Figure S2.1-S2.10

1. Activity time courses across event boundaries are related to the subsequent recall

To identify regions in which the univariate neural responses to event boundaries were related to subsequent memory, we directly compared the activity time courses around boundaries for *remembered* (*R*) with *forgotten* (*F*) events. We were interested in the activity time courses for an event that was associated with successful memory (*R* vs. *F;* **Figure S2.2**, highlighted in blue shading), but also how the activation changed in the following event. This gave us a 2*2 design with *Memory* (*R* and *F)* and *Event (current* and *next event)* as independent variables (**Figure S2.2**). For each of the six predefined ROIs (**Figure S1.2**), BOLD signals around event boundaries (from 6 volumes before to 6 volumes after each boundary) were extracted from the time series during movie watching, labeled (*R* vs. *F*) based on the memory performance for the first of the two events, averaged across all the available boundaries, and transformed into z-scores for each participant.

The interaction between *Memory* and *Event* was significant in early visual cortex ($F = 13.109$, $p_{FDR} = .006$, $\eta^2 = 0.45$; **Figure S2.2A**), early auditory cortex ($F = 10.852$, $p_{FDR} = .010$, $\eta^2 = 0.404$, **Figure S2.2B**), hippocampus ($F = 8.217$, $p_{FDR} = .017$, $\eta^2 = 0.339$, **Figure S2.2E**), and posterior parahippocampal gyrus (pPHG) ($F = 15.562$, $p_{FDR} = .006$, $\eta^2 = 0.493$, **Figure S2.2F**), but not in mPFC ($F = 3.036$, $p_{FDR} = .101$, $\eta^2 = 0.160$, **Figure S2.2C**) and PMC ($F = 4.718$, $p_{FDR} = .054$, $\eta^2 = 0.228$; **Figure S2.2D**). For the current (first) event, we found greater activation for *remembered* compared to *forgotten* events in early visual cortex (*R*: 0.143 ± 0.125 (mean ± SD), *F*: -0.090 ± 0.145 (mean ± SD); $t = 21.911$, $p_{FDR} < .001$), hippocampus (*R*: 0.134 ± 0.079 (mean ± SD), *F*: -0.019±0.185 (mean ± SD); $t = 11.801$, $p_{FDR} = .006$) and pPHG (*R*: 0.118 ± 0.083 (mean ± SD), *F*: -0.028 ± 0.155 (mean ± SD); $t = 13.264$, $p_{FDR} = .006$). This effect was reversed in early auditory cortex for the next event, which showed greater activity for *forgotten* compared to *remembered* events (*R*: -0.217 ± 0.155 (mean ± SD), *F*: -0.034 ± 0.136 (mean ± SD); $t = 10.803$, $p_{FDR} = .030$). Consistent with previous research, enhanced activity in the timeframe of the current event was associated with successful subsequent recall. To further take into account the memory performance of the *next event*, the response curves were split into four (*R-R*; *R-F*; *F-R*; *F-F*) based on the memory performance of two consecutive events (**Figure S2.3**).

2. Event boundary permutation analysis

To confirm that the subsequent memory effect on pattern similarity was only present for actual event boundaries (annotated by an independent rater based on the movie narrative), we shuffled the timestamps of the original annotated boundaries and tested whether the simulated boundaries were associated with memory. To retain the event structure (i.e., avoid cases where the pseudo-events are too long/short), we held onto the duration of the 50 events and only scrambled their order, which led to randomly placed event boundaries across the sequence with the original set of event durations. Our calculation for *activation* and *connectivity patterns* was then performed on these permutated boundaries, yielding simulated *activation,* and *connectivity pattern* similarities.

Paired *t*-tests were performed to test whether the pattern similarity indices were still associated with memory labels. This permutation procedure followed by the *t*-test was repeated 5000 times for both the *activation* and *connectivity patterns*, which gave us a null distribution of *t* and *p* values. The *p*-value from the analysis using genuine event boundaries was subsequently corrected against this distribution by calculating the proportion of sampled permutations where the *p* values were smaller than or equal to the real observation. The regions that showed significant differences in *activation pattern* similarities are the same ones as in the previous analysis using real event boundaries: early auditory area ($p = 0.001$), hippocampus ($p = 0.003$), mPFC ($p = 0.01$), and pPHG ($p = 0.006$) (**Figure S2.4**). For *connectivity patterns*, the SME only exists for actual event boundaries in the early auditory area ($p = 0.012$), visual areas ($p = 0.005$), hippocampus ($p = 0.005$), and PMC ($p = 0.012$) (**Figure S2.6**). These results confirmed the specificity of the SME to actual event boundaries.

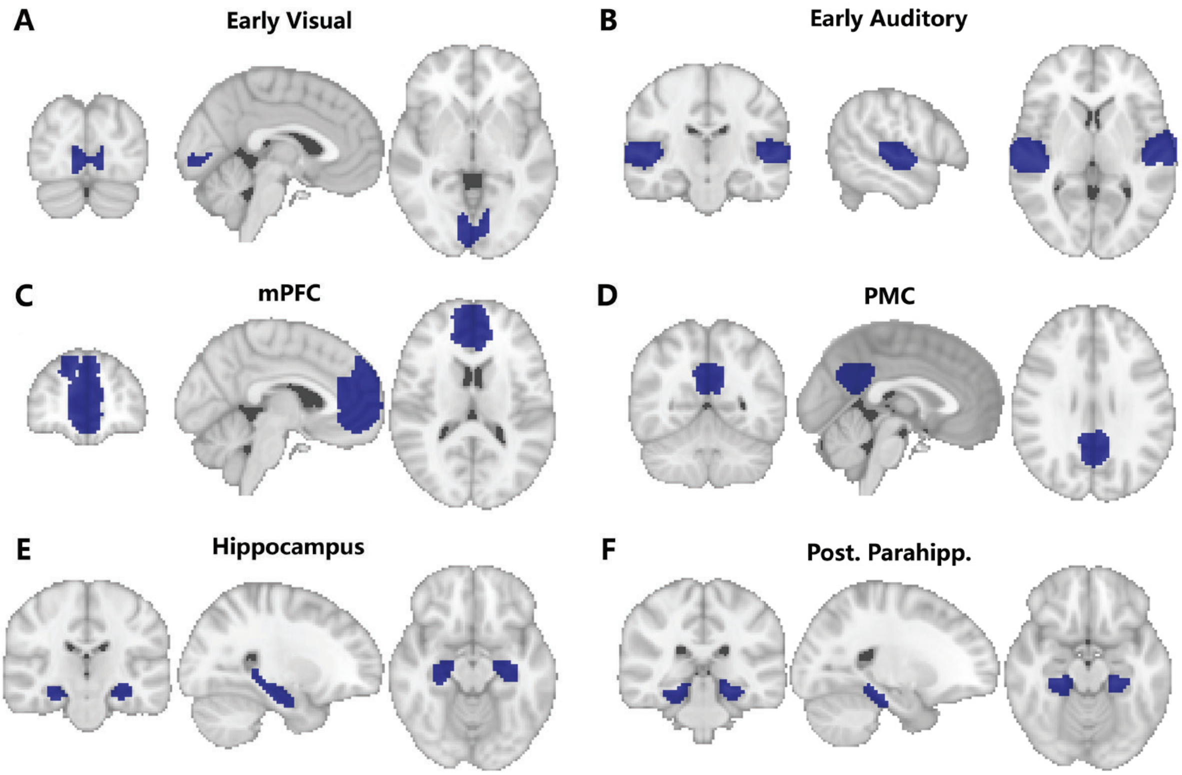


Figure S2.1. Six predefined regions-of-interest (ROIs) used. (A) (B) Early visual and auditory cortex were functionally defined in the literature using inter-subject correlation analysis. **(C) (D)** The medial prefrontal cortex (mPFC) and primary motor cortex (PMC) were defined in the functional atlas of the resting-state default mode network. **(E) (F)** The hippocampus and posterior parahippocampal gyrus were anatomically defined from the probabilistic *Harvard-Oxford Subcortical Structural Atlas* using the threshold of 50%.

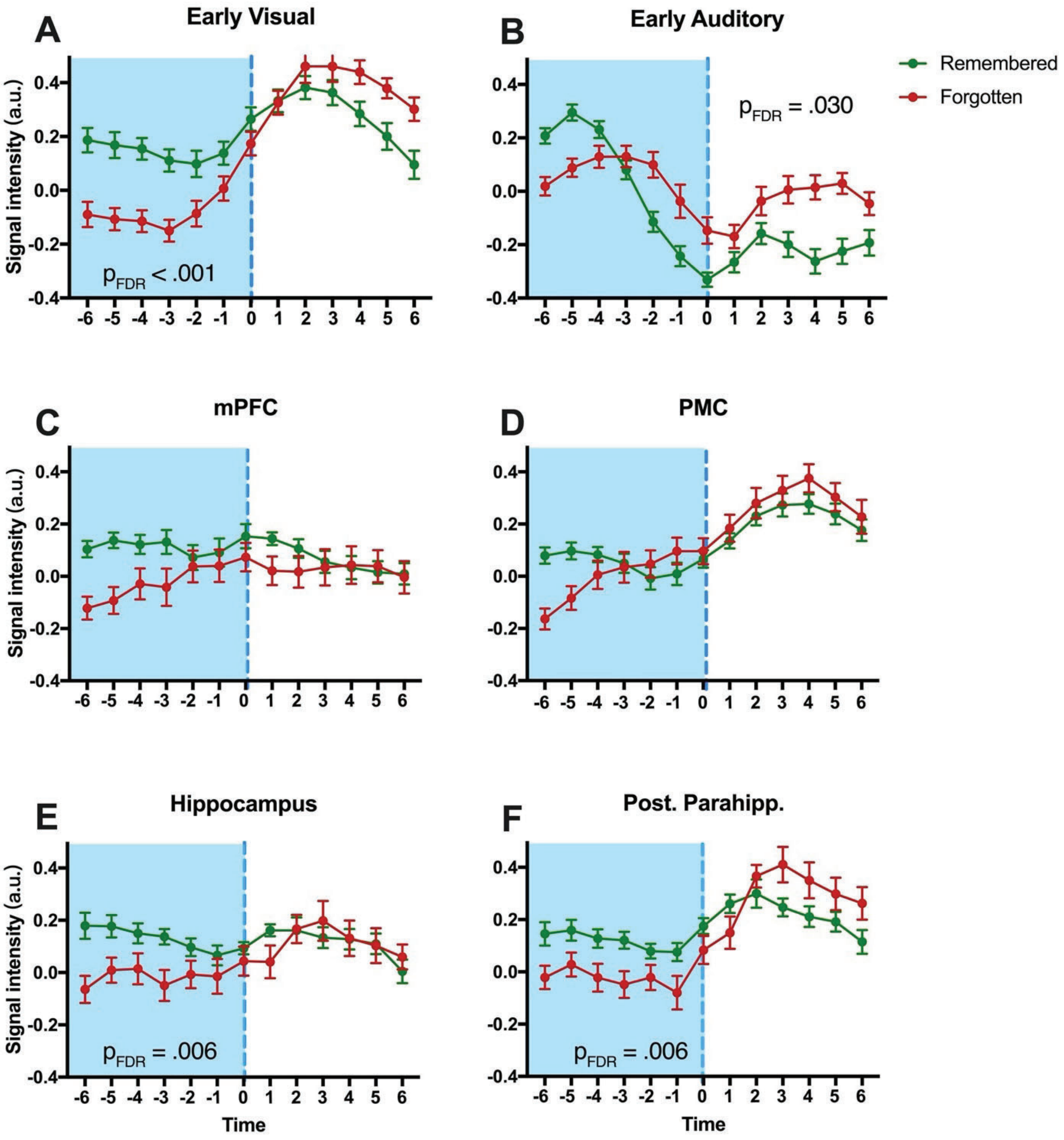


Figure S2.2. Activity time courses of *remembered* and *forgotten* events around event boundaries in six ROIs. We compared activity time courses of sequential event pairs (*Remembered* vs. *Forgotten*) based on subsequent memory performance of the first event across six ROIs using repeated two-way ANOVA (*memory*; *event*). The left side (blue) shows activity for an event that was later *remembered* or *forgotten*, while the right side (white) shows what happened to the signal after those events. The early visual cortex ($F = 13.109$, $p_{FDR} = .006$, $\eta^2 = 0.45$, panel **A**), hippocampus ($F = 8.217$, $p_{FDR} = .017$, $\eta^2 = 0.339$, panel **E**), and pPHG ($F = 15.562$, $p_{FDR} = .006$, $\eta^2 = 0.493$, panel **F**) demonstrated a significant interaction between *event* and *memory*. Post-hoc comparison revealed greater activation for *remembered* compared to *forgotten* events during the *current event*. Similar patterns were seen in PMC ($F = 4.718$, $p_{FDR} = .054$, $\eta^2 = 0.228$, panel **D**) and mPFC ($F = 3.036$, $p_{FDR} = .101$, $\eta^2 = 0.160$, panel **C**), but did not survive correction for multiple comparisons. The reverse effect (higher activation for *forgotten* vs. *remembered* events) was found in early auditory cortex ($F = 10.852$, $p_{FDR} = .010$, $\eta^2 = 0.404$, panel **B**).

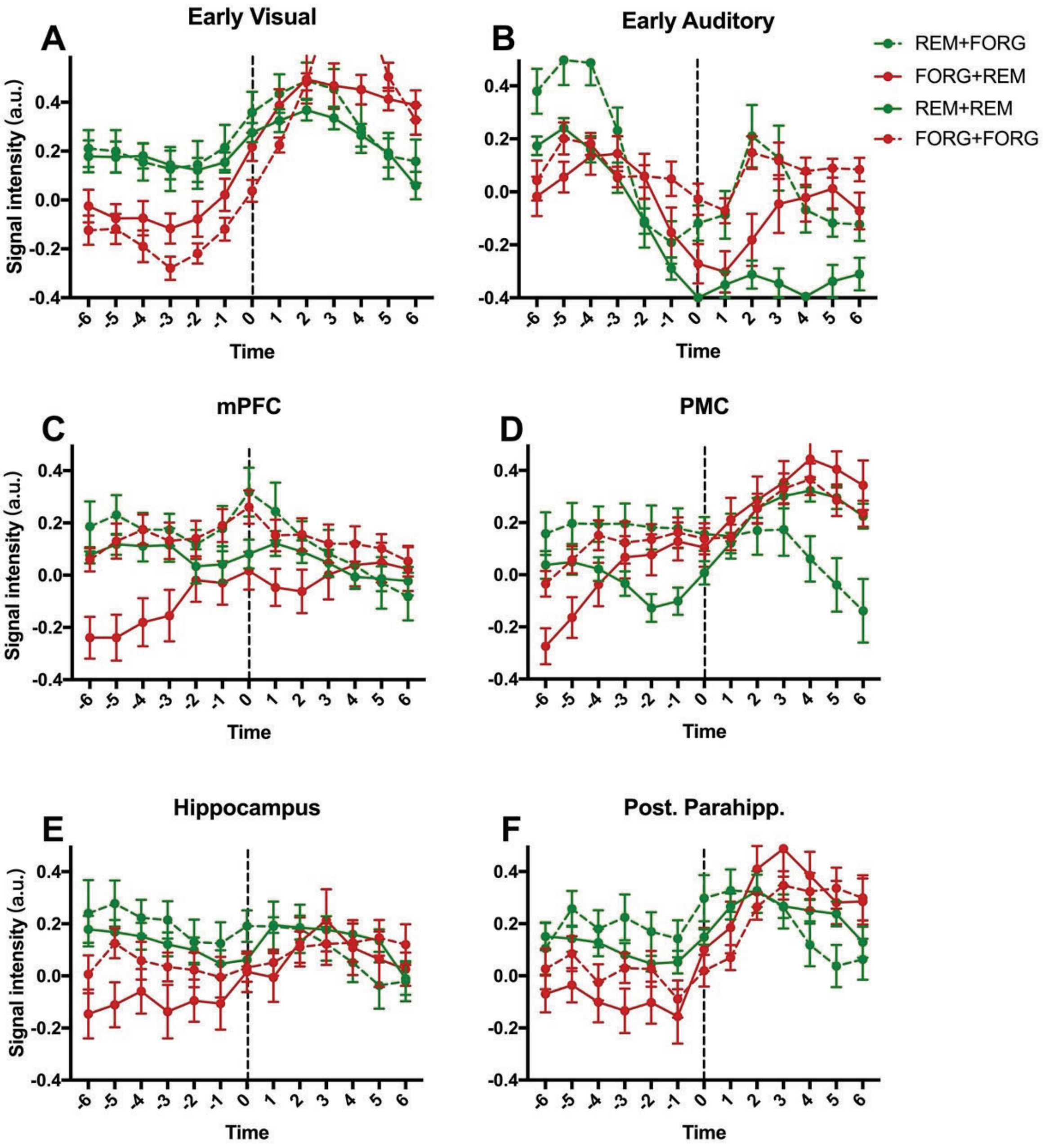


Figure S2.3. **Univariate response curves based on memory performance of two consecutive events.** BOLD signals around the event boundaries (6 TRs before and after event offset; 13 time points including the boundary itself) were extracted from the time series, averaged across all voxels, and z-scored within each region. Considering the memory label of the two consecutive events simultaneously, the time course segments were categorized into four types: R (green)/F (red) + R (solid)/F (dashed), and were plotted against the time (in TRs) relative to the event boundary.

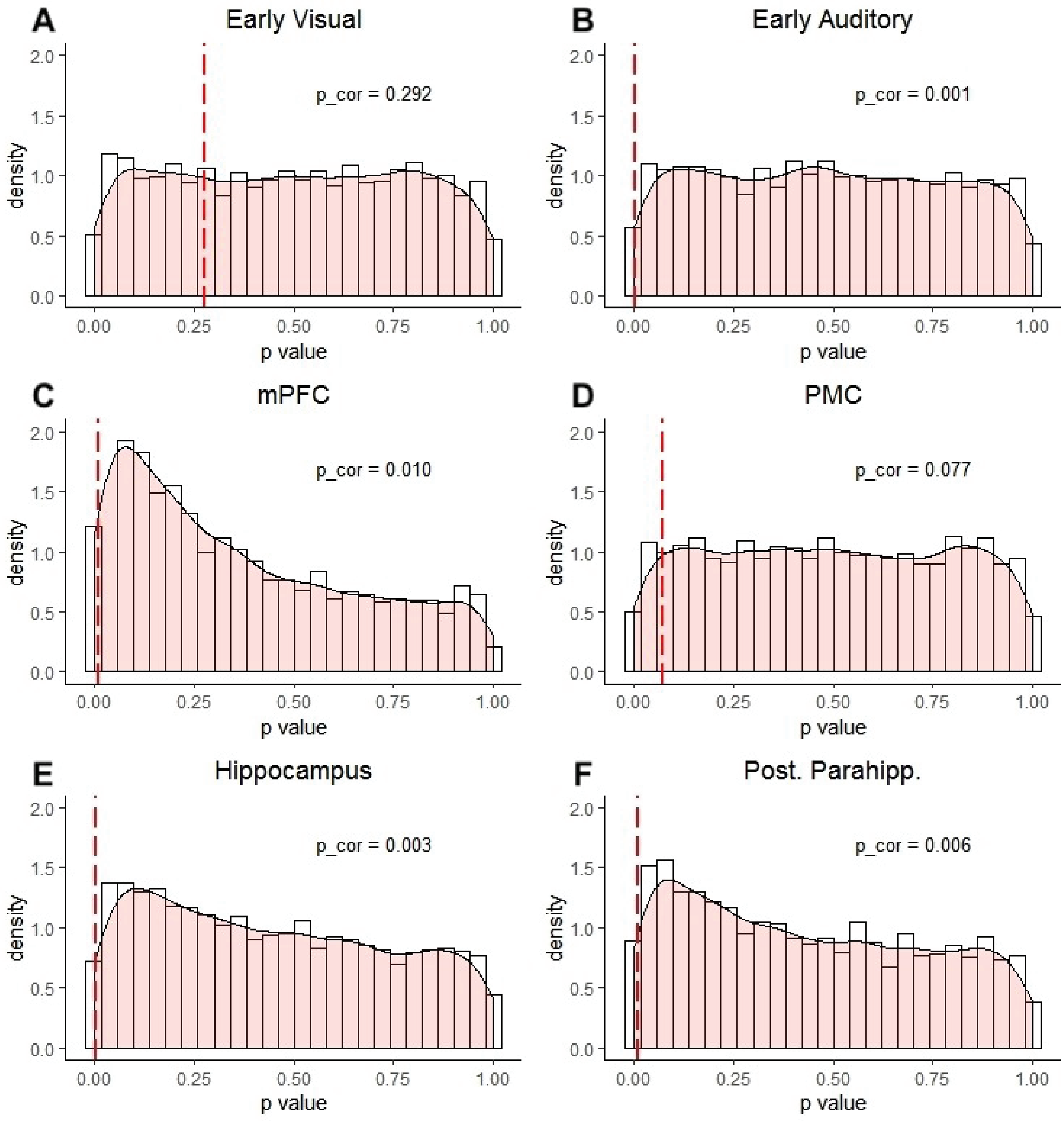


Figure S2.4. Event boundary permutation analysis for *activation pattern* similarity. The *p*-value obtained from the true event boundary similarity analysis (red dashed line) is corrected based on the null distribution (histogram) generated by random permutation of event intervals (5000 times). A smoothed density estimate calculated by *stat_density* with *ggplot2* and *R* is indicated by the semitransparent region. The corrected values for each ROI are shown in the corresponding subplots.

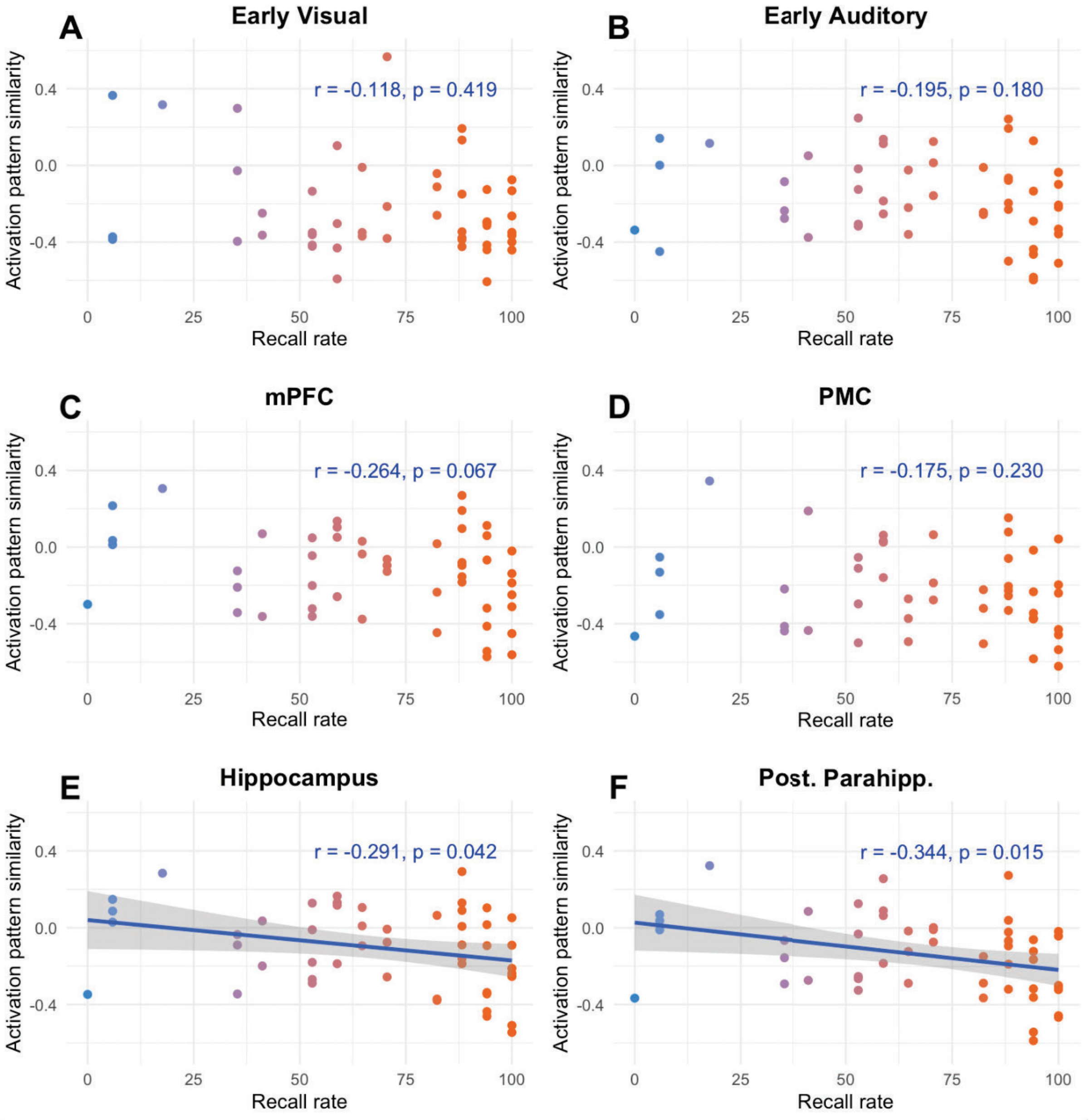


Figure S2.5. Event-specific correlational analysis between *activation pattern* similarity and memory performance. Memory performance per event is calculated by averaging the number of participants that successfully recalled the event divided by the total number of participants (recall rate). Pearson's correlation coefficient *r* and *p*-value (uncorrected) are shown for each ROI. For the regions that demonstrated a significant correlation (i.e., hippocampus and pPHG), a linear model based regression line (blue) is fitted to the data points with the confidence interval shown in gray. To better visualize the correlation, the color of the data points are set to change in the direction of the first principal component of the data.

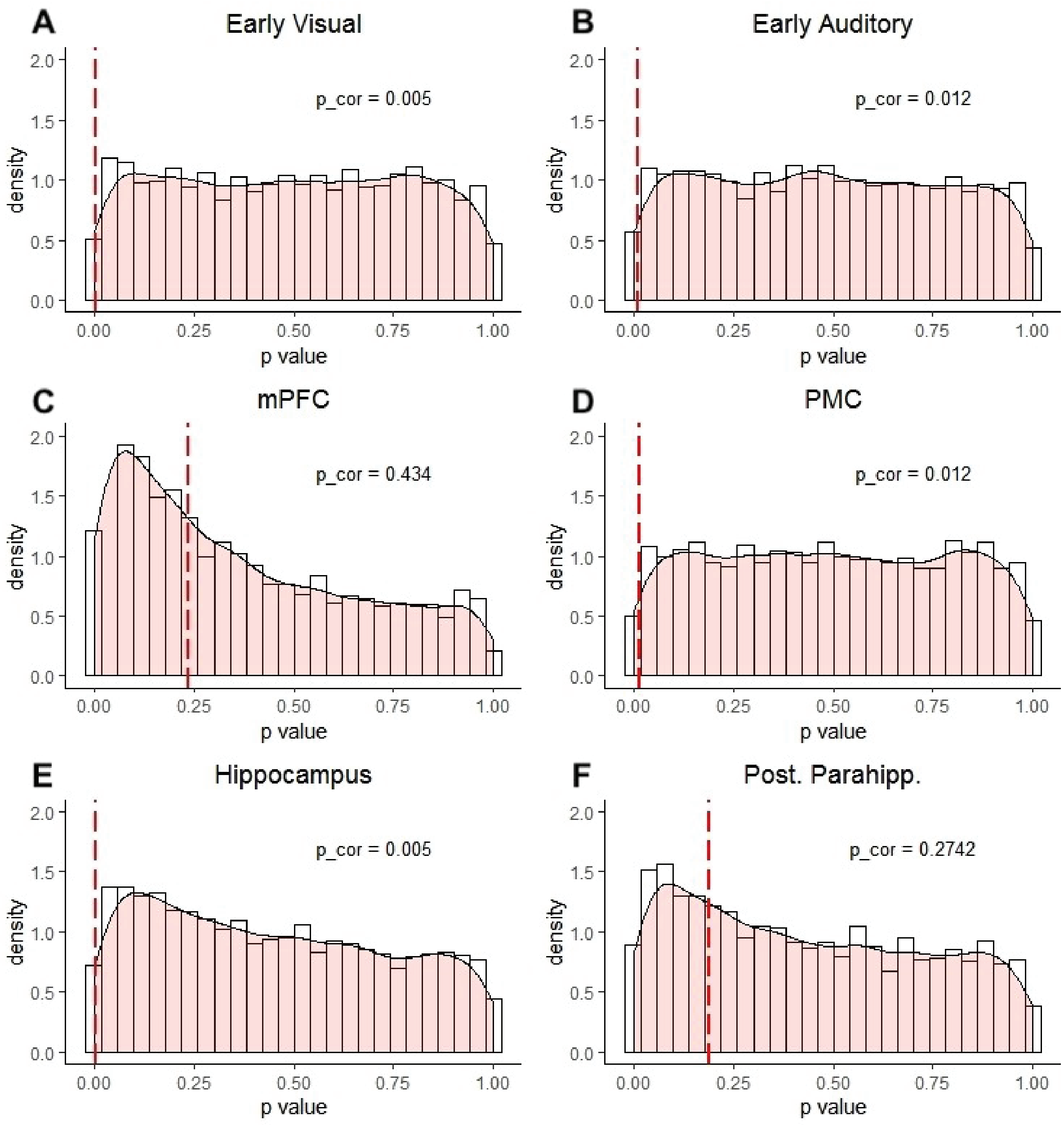


Figure S2.6. Event boundary permutation analysis for *connectivity pattern* similarity. Analyzing *connectivity pattern* similarity with "scrambled" event boundaries. The *p*-value obtained from the true event boundary similarity analysis (red dashed line) is corrected based on the null distribution (histogram) generated by random permutation of event intervals (5000 times). A smoothed density estimate calculated by *stat_density* with *ggplot2* and *R* is indicated by the semitransparent region. The corrected values for each ROI are shown in the corresponding subplots.

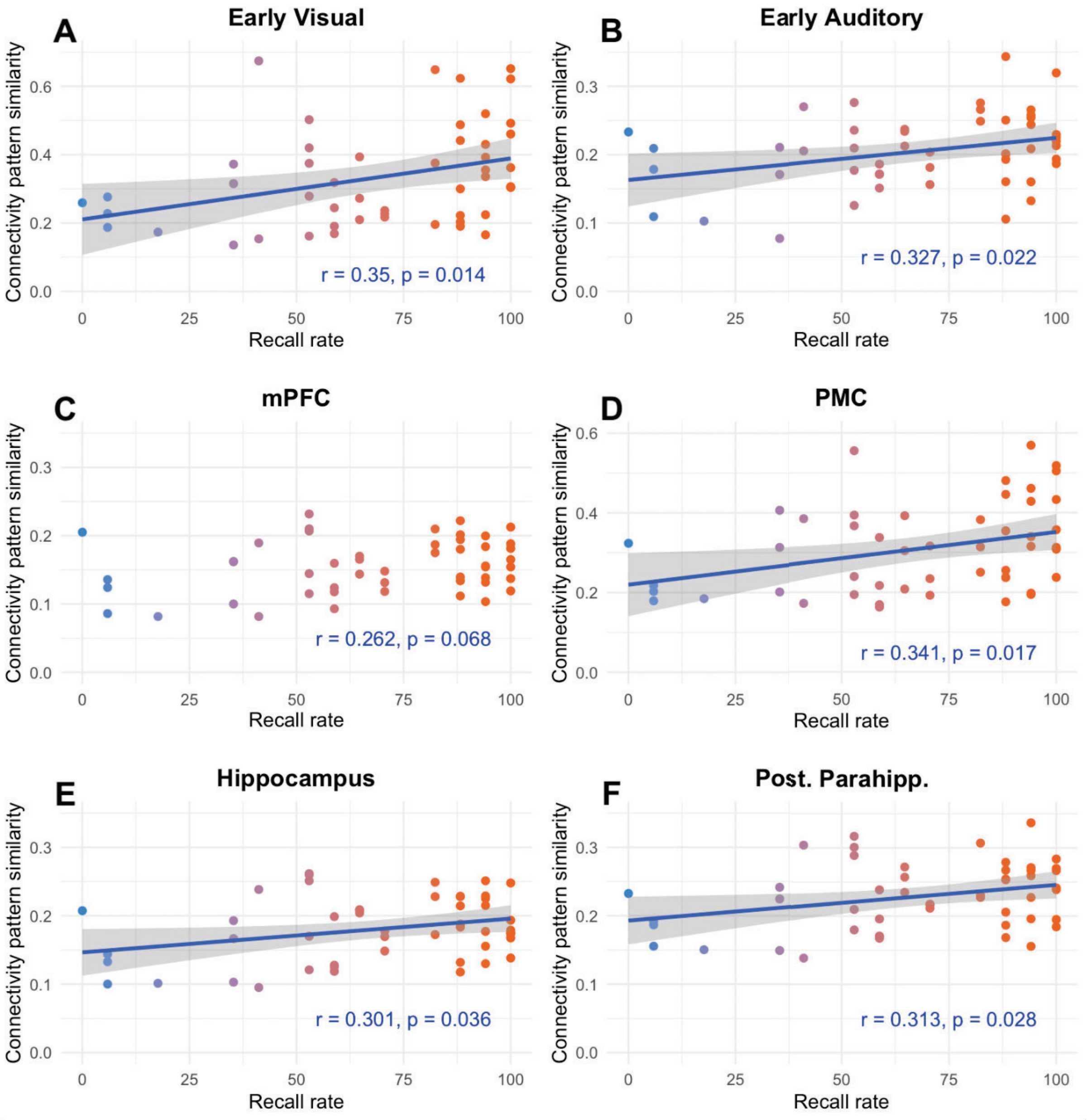


Figure S2.7. Event-specific correlational analysis between *connectivity pattern* similarity and memory performance. Memory performance per event is calculated by averaging the number of participants that successfully recalled the event divided by the total number of participants (recall rate). Pearson's correlation coefficient *r* and *p*-value (uncorrected) are shown for each ROI. For the regions that demonstrated a significant correlation (i.e., all six regions except for mPFC), a linear model based regression line (blue) is fitted to the data points with the confidence interval shown in gray. To better visualize the correlation, the color of the data points are set to change in the direction of the first principal component of the data.

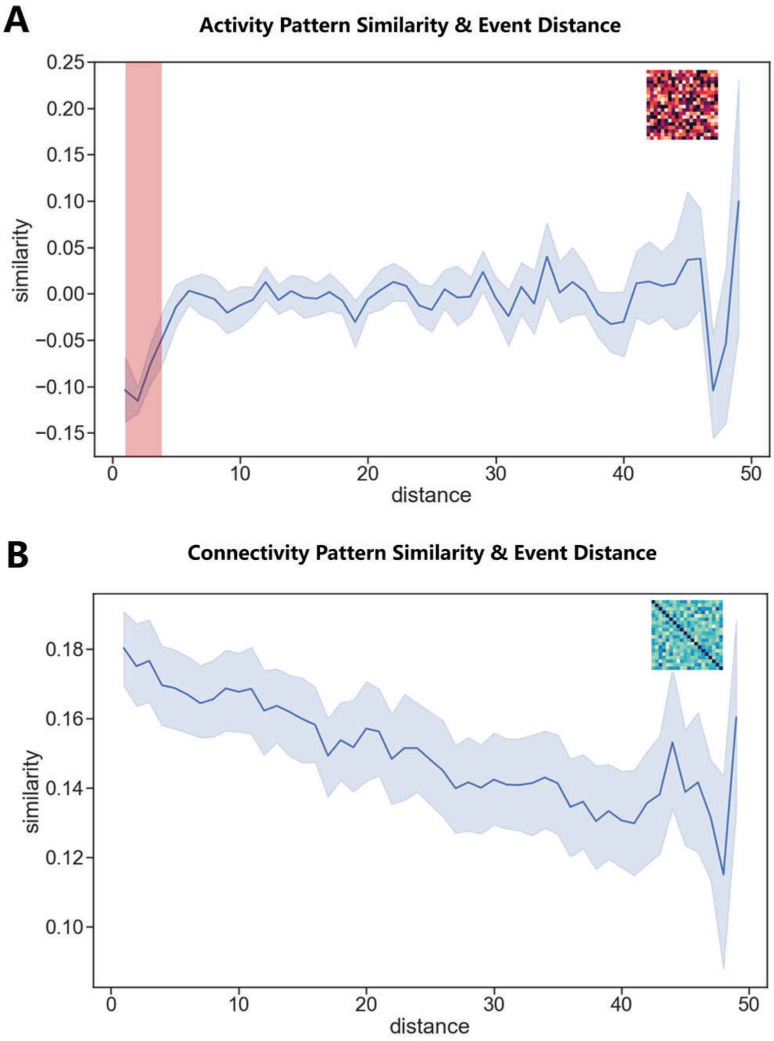


Figure S2.8. Relationship between all possible event distance and hippocampal neural similarities. (**A**) Relationship between hippocampal *activation pattern* similarities and event distance. Event distance ranges from 1 to 49. The shallow red bar indicates that the similarity is significantly lower than 0 after false discovery rate (FDR) correction. (**B**) Relationship between event distance and hippocampal *connectivity pattern* similarities. Event distance ranges from 1 to 49.

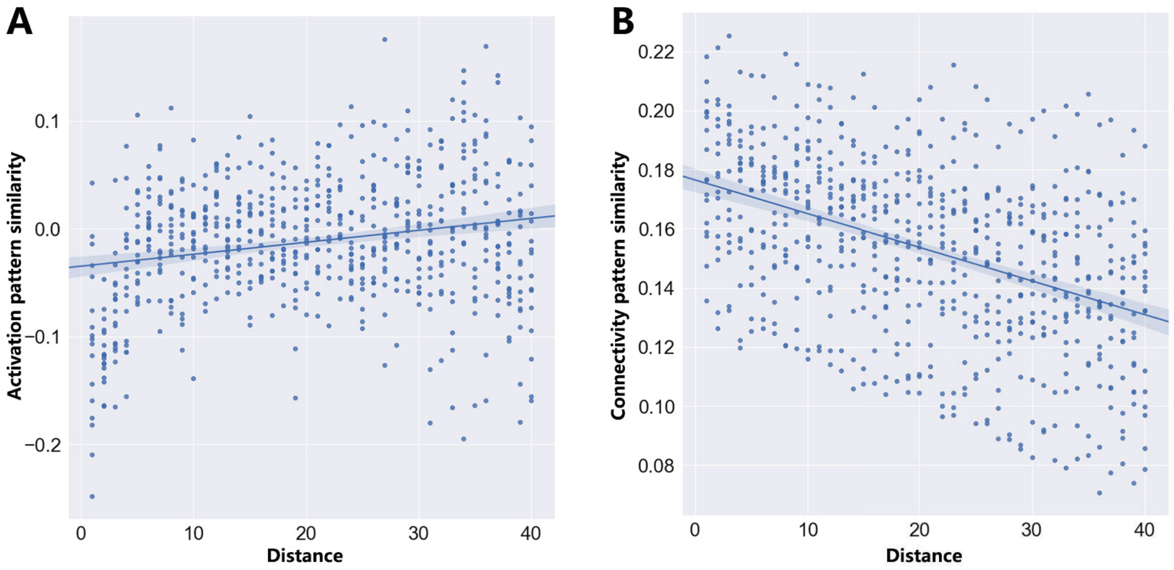


Figure S2.9. Relationship between event distance and hippocampal *activation* and *connectivity pattern* similarities. (A) Positive correlation between event distance (ranging from 1 to 40) and hippocampal *activation pattern* similarities ($r = 0.21$, $p_{raw} = 1.8 \times 10^{-8}$). **(B)** Negative correlation between event distance (ranging from 1 to 49) and hippocampal *connectivity pattern* similarities ($r = -0.439$, $p_{raw} = 1.8 \times 10^{-33}$).

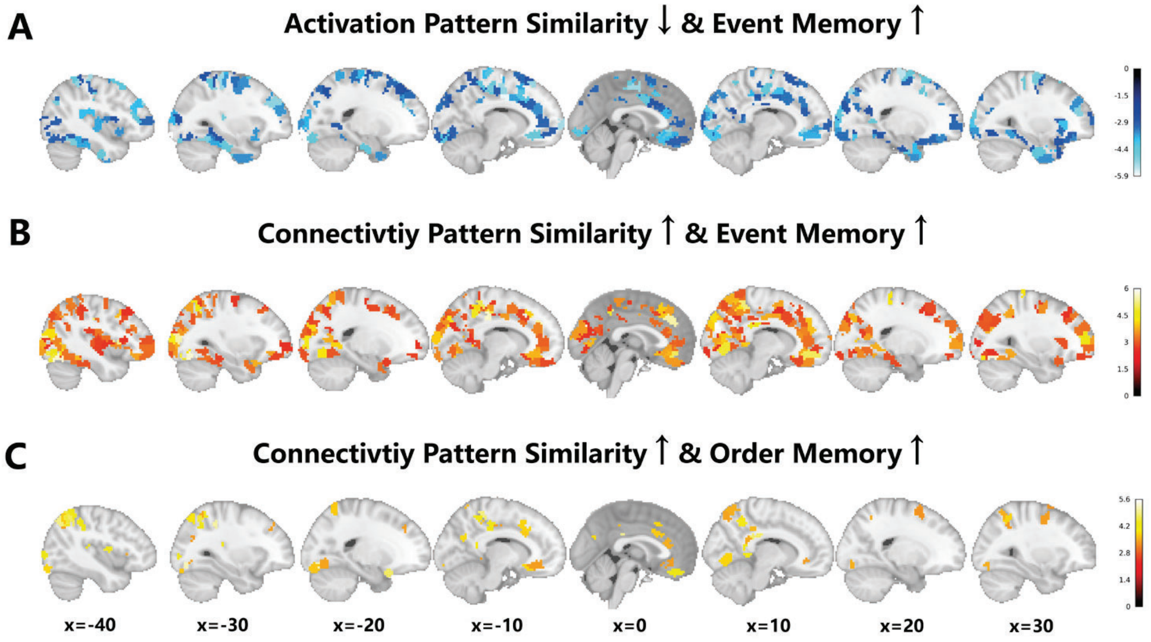

A
Activation Pattern Similarity ↓ & Event Memory ↑
0
-1.5
-2.9
-4.4
-5.9
B
Connectivtiy Pattern Similarity ↑ & Event Memory ↑
6
4.5
3
1.5
0
C
Connectivtiy Pattern Similarity ↑ & Order Memory ↑
5.6
4.2
2.8
1.4
0
x=-40
x=-30
x=-20
x=-10
x=0
x=10
x=20
x=30

Figure S2.10. Identifying event segmentation and integration computations separately across the neocortical parcellation. (**A**) Distinct *activation patterns* of these brain parcels across event boundaries correlate with better event memory. (**B**) Similar *connectivity patterns* of these brain parcels across event boundaries relate to better event memory. (**C**) Similar *connectivity patterns* of these brain parcels across event boundaries link to better order memory. All results are displayed at $p_{FDR} < 0.05$ across 1000 brain parcels.

Chapter 3

The dynamic transition between neural states is associated with the flexible use of memory

This chapter is based on:
Liu, W., Kohn, N., & Fernandez, G. (2020). The dynamic transition between neural states is associated with the flexible use of memory. bioRxiv. (This preprint is under review in a scientific journal)

Abstract

Flexible behavior requires switching between different task demands. It is known that such task-switching is associated with costs in terms of slowed reaction time, reduced accuracy, or both. The neural correlates of task-switching have usually been studied by requiring participants to switch between distinct tasks that recruit different brain networks. Here, we investigated the transition of neural states underlying switching between two memory-related processes with opposite task demands (i.e., memory retrieval and memory suppression). We investigated 26 healthy participants who performed a Think/No-Think task while being in the fMRI scanner. Behaviorally, we show that it was more difficult for participants to suppress unwanted memories when a No-Think was preceded by a Think trial instead of another No-Think trial. Neurally, we demonstrate that Think-to-No-Think switches were associated with an increase in control-related and a decrease in memory-related brain activity. Neural representations of task demand, assessed by decoding accuracy, were lower immediately after task switching compared to the non-switch transitions, suggesting a switch-induced delay in the neural transition towards the required task demand. This suggestion is corroborated by an association between demand-specific representational strength and demand-specific performance in switch trials. Taken together, we propose that the brain's delayed transition of neural states towards the task demand at hand is associated with a switch cost leading to less successful memory suppression.

Keywords: task switching; memory suppression; memory retrieval; cognitive control; fMRI

Introduction

In everyday life, we are continuously switching between different tasks (Monsell, 2003). Transitions between task demands have often been studied using task-switching paradigms in which participants are required to switch between two or more distinct tasks (Meiran, 2010). Usually, participants perform less accurately and/or more slowly immediately after switches (i.e., *switch costs*) (Goschke, 2000; Jersild, 1927; Rogers & Monsell, 1995; Spector & Biederman, 1976). Results from univariate fMRI studies suggested the involvement of prefrontal-parietal regions in task switching (Braver et al., 2003; Dove et al., 2000; O. Gruber et al., 2006; Richter & Yeung, 2014). Two studies (Loose et al., 2017; Waskom et al., 2014) used multivariate fMRI methods (J. D. Cohen, Daw, Engelhardt, Hasson, Li, Niv, Norman, Pillow, Ramadge, Turk-Browne, & Willke, 2017; Haynes, 2015) to investigate how task switching modulates task representations but reported mixed results. Waskom and colleagues reported stronger task representations after a shift in task demands (Waskom et al., 2014), while Loose and colleagues found no difference in task representations between the switch and the non-switch condition (Loose et al., 2017).

Previous experiments that investigated task switching were typically designed to minimize the perceptual differences between conditions, but not to maximize differences in underlying cognitive demands (Braver et al., 2003; Kiesel et al., 2010; Loose et al., 2017; Waskom et al., 2014). Behavioral and neural correlates of task-switching between two opposite tasks within one cognitive domain remain largely unexplored. Switching between two opposite task demands should be cognitively more challenging than between unrelated tasks, because they may be based on (partly) overlapping neural mechanisms. More importantly, it provides us with the opportunity to examine the fast-adaptive transition between task demands represented in the neural states of the same brain networks in response to switches. Here, we used a modified Think/No-Think paradigm (Michael C Anderson & Green, 2001; B. J. Levy & Anderson, 2012) to probe task-switching within the memory domain. Specifically, participants were instructed to switch between memory retrieval and memory suppression according to trial-specific instructions. We asked whether we can find behavioral *switch costs* when participants switch between two opposite memory tasks.

To detect the neural source of *switch costs*, we analyzed the dynamic transitions between neural states during task switching using time-resolved multivariate decoding. Our analyses focused on frontoparietal regions associated with control and a set of regions associated with memory retrieval. Cognitive and neural models of memory retrieval and suppression suggest that successful retrieval could be the result of cooperation between an inhibitory control network and an episodic retrieval network (Rugg & Vilberg, 2013), while effective suppression depends on top-down control of the inhibitory control network upon a general episodic retrieval network

(M. C. Anderson & Hanslmayr, 2014). Previous fMRI studies of memory suppression supported this idea by showing that compared to Think trials, No-Think trials are associated with stronger activation in control-related regions including the dorsolateral prefrontal cortex, ventrolateral prefrontal cortex, inferior parietal lobule, and supplementary motor area (Michael C Anderson, 2004; Guo et al., 2018; W. Liu, Peeters, et al., 2020). At the same time, these activity increases are accompanied by reduced activity in memory-related areas in the medial temporal lobe, including the hippocampus (M. C. Anderson & Hanslmayr, 2014).

We hypothesized that a delayed transition between neural states that represent task demands could be the neural underpinning of behavioral *switch costs* because failing to update neural states on time could result in a neural state that is optimal for the opposite (e.g., retrieval), but not the current (e.g., suppression), task demand. This assumption is built on the idea that the human brain can demonstrate diverse brain states during different cognitive tasks or environmental demands (Cocuzza et al., 2019; Gonzalez-Castillo et al., 2015; Hermans et al., 2011; Sadaghiani et al., 2015; Shine et al., 2016; Shine & Poldrack, 2018; Westphal et al., 2017). The task-switching paradigm is suitable to study such a rapid neural reconfiguration process because it allows us to compare directly how different task demands are represented in neural states and how transitions of neural states are associated with switch costs.

Results

Behavioral results

Our study used a modified think/no-think (TNT) paradigm (**Figure 3.1A**) with trial-by-trial reports of (in)voluntary memory retrieval (i.e., retrieval/intrusion frequency rating) (B. J. Levy & Anderson, 2012). As intended, most of the associations were successfully recalled in Think trials (1-mean $_{\text{P(Never)}}$=84.05%, SD=11.79 %, range from 56.25% to 100%; **Figure S3.1A**), while participants suppressed memory retrieval successfully in No-Think trials in about half of the trials (mean $_{\text{p(Never)}}$ =50.62%, SD=25.35%, range from 4% to 92.5%; **Figure S3.1B**).

Our central aim was to determine whether there were behavioral *switch costs* in the TNT task and to reveal their neural underpinnings. We defined each trial as a “switch” or “non-switch” trial considering both the task demand of the current trial and its predecessor (**Figure 3.1B**). Specifically, we identified “switch” trials if a preceding trial had the *opposite* task demand (e.g., previous trial: Think; current trial: No-Think). By contrast, if the current trial and the preceding trial had the *same* task demand, then the current one was a “non-switch” trial. We compared the trial-by-trial performance between “switch” and “non-switch” trials for the Think and the No-Think condition separately. Participants showed comparable performance for “switch” and "non-switch" trials in the Think condition (t(25)=0.348, p=0.731, Cohen's d=0.068; **Figure 3.1C**), while they reported more memory intrusions for "switch" trials compared to "non-switch" trials in the No-Think condition (t(25)=3.19, p=0.004, Cohen's d=0.627; **Figure 3.1D**), suggesting *switch costs* when the task demand switched from a previous Think trial to a current No-Think trial.

After the TNT task, participants performed a final memory test. Results from this task had been reported in another publication in detail (W. Liu, Kohn, et al., 2020) and *supplemental materials* (**Figure S3.2**) of this study. Here, we focused on participants' performance during the final memory test at the individual level. More specifically, we quantified individual differences in both subjective and objective *suppression-induced forgetting effects* and correlated them with fMRI measures (*see below*).

fMRI results

To replicate the univariate neural signature of memory suppression reported in prior studies (M. C. Anderson & Hanslmayr, 2014; Michael C Anderson, 2004), we first conducted a univariate analysis to contrast brain regions engaged in memory suppression and memory retrieval (i.e., No-Think vs. Think). We found an increased activity for No-Think trials in regions that are consistently involved in memory suppression, including the bilateral dorsolateral prefrontal cortex (DLPFC), bilateral insula, bilateral inferior parietal lobule (IPL), supplementary motor area (SMA), and middle cingulate gyrus (**Figure 3.1E; Table S3.1**). Additionally, we found higher activity in ventral visual areas and the right thalamus during No-Think compared to Think trials. Next, we contrasted the Think condition with the No-Think condition and found the increased activity for Think condition in a set of regions including the medial prefrontal cortex (mPFC), posterior cingulate cortex (PCC), hippocampus, inferior parietal lobule (IPL), precuneus, angular gyrus, and cerebellum (**Figure 3.1F; Table S3.2**). Together with the behavioral results from the final memory test, these results confirmed that participants in our experiment followed task instructions, leading to univariate neural signatures of memory retrieval and suppression consistent with prior findings (Michael C Anderson, 2004; Hulbert et al., 2016; B. J. Levy & Anderson, 2012)**,** as well as recent meta-analyses of memory suppression (Guo et al., 2018; W. Liu, Peeters, et al., 2020)**.**

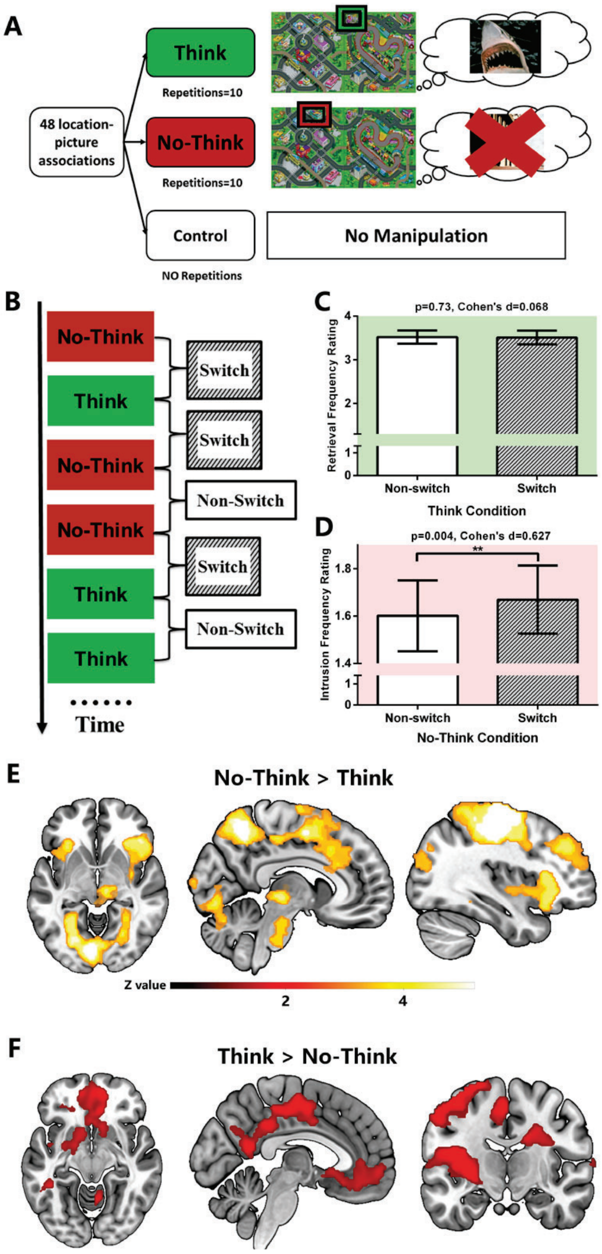


Figure 3.1 **(A)** After learning 48 location-picture associations, participants performed a Think/No-Think task while brain activity was measured by fMRI. During think trials, participants were instructed to retrieve associated pictures based on the highlighted locations as memory cues. By contrast, during no-think trials, participants were required to suppress the tendency to retrieval the associated pictures. **(B)** The sequence of trials was designed to probe task switching between two task demands (i.e., Think and No-Think). When the task demand of the current trial was the same as the previous trial, it was defined as the "Non-switch" trial. By contrast, while the task demand of the current differed from the previous trial, it was defined as the "switch" trial. **(C)** During Think trials, participants demonstrated comparable memory retrieval performance ($p=0.73$, Cohen's $d=0.068$) for both "switch" and "Non-switch" trials. **(D)** During No-Think trials, participants reported worse memory suppression performance, indexed by more memory intrusions for "switch" trials compared to "non-switch" trials ($p=0.004$, Cohen's $d=0.627$). **(E)** Brain regions showed increased activation during No-Think trials compared to Think trials. **(F)** Brain regions showed increased activation during Think trials compared to No-Think trials.

The transition of large-scale neural states from memory retrieval to memory suppression

Based on neurocognitive models of memory suppression (M. C. Anderson & Hanslmayr, 2014), we focused on the neural dynamics within the inhibitory control network and the memory retrieval network. First, we used *Neurosynth* (https://neurosynth.org/), an automatic meta-analysis tool of neuroimaging data (Yarkoni et al., 2011), to identify the inhibitory control network and memory retrieval network independently from our fMRI data. Using the term "*inhibitory control*" and "*memory retrieval*," we performed term-based meta-analyses to reveal two distinct brain networks of inhibitory control (**Figure S3.3A**) and memory retrieval (**Figure S3.3B**) separately. The two meta-analytic maps have *overlapping* areas, including the IFG, insular, SMA, inferior parietal lobule (**Figure S3.3C**). Interestingly, the latter areas are highly similar to a "task switching" map generated by *Neurosynth* using the term "*task switching*" (**Figure S3.3D**).

In the next step, to enable a regions-of-interest analysis, we separated the identified *inhibitory control*, *memory retrieval* and *overlapping* networks into regions-of-interest (ROIs) based on the combination of a connectivity-based neocortical parcellation (number of parcels=300) (Schaefer et al., 2018) and subcortical regions (number of regions=14) (*Details see Methods*). Using this approach, we identified 71 parcels as *memory-related regions*, 29 parcels were categorized as *control-related regions*, and 10 parcels were labeled as *overlapping regions* (**Figure 3.2A**). Finally, BOLD time series were extracted from each voxel, averaged within each ROI, and further processed.

Using these time series, we characterized the group-average transition of neural states when the task demand changed from Think to No-Think trials (**Figure 3.2B**). Based on the task instruction, the time series were firstly split for the Think and No-Think conditions separately and then concatenated across all runs of all participants. For each task demand, all ROIs were ranked based on their state-specific averaged neural activity across runs (the highest activity was ranked first) to represent their relative dominance during that neural state (i.e., Think or No-Think). For control purposes, the same analysis was repeated for raw signal intensities (**Figure S3.4B**) and their Z-values (**Figure S3.4C**) and yielded highly similar patterns. A Kruskal-Wallis test showed that during Think trials, *memory-related regions*, *control-related regions*, and *overlapping regions* differed in their ranks ($H(2)=40.48$, $p<0.001$). Post-hoc Mann-Whitney tests using a Bonferroni-adjusted alpha level of 0.017 (0.05/3) were used to compare all group pairs. *Memory-related regions* ($\text{mean}_{\text{memory}}=40.22$, $\text{SD}_{\text{memory}}=27.39$) ranked higher than *control-related regions* ($\text{mean}_{\text{control}}=82.34$, $\text{SD}_{\text{control}}=22.28$) and *overlapping regions* ($\text{mean}_{\text{overlap}}=75.10$, $\text{SD}_{\text{overlap}}=19.08$) (memory-related vs. control-related: $U=263$, $p<0.001$; memory-related vs. overlapping: $U=108$, $p<0.001$). *Control-related regions* and *overlapping regions* did not differ significantly in their ranks ($U=104$, $p=0.096$). Three types of regions also differed in their ranks during No-Think trials ($H(2)=36.60$, $p<0.001$). *Memory-related regions* ($\text{mean}_{\text{memory}}=67.96$, SD

$_{memory}$=27.94) ranked lower than *control-related regions* (mean $_{control}$=27.24, SD $_{control}$=23.13) and *overlapping regions* (mean $_{overlap}$=37.80, SD $_{overlap}$=21..10) (memory-related vs. control-related: U=238, $p<0.001$; memory-related vs. overlapping: U=144, p=0.001). *Control-related regions* and *overlapping regions* did not differ significantly in their ranks (U=101, p=0.081). All comparisons between *memory-related* and *control-related/overlapping regions* were significant after Bonferroni-adjustment (all ps $\leq$ 0.001). Furthermore, we performed additional analyses of neural state transition by dividing all ROIs into three groups (i.e., *increased group, stable group, and decreased group*) based on their relative changes in rank (See *Supplemental Material*).

The transition of neural states during the TNT task is associated with subsequent suppression-induced forgetting

We already demonstrated that the activity of *control-related regions* increased, and the activity of *memory-related regions* decreased when the task demands switched from Think to No-Think. To assess if these regional brain activity changes are associated with the behavioral consequence of memory suppression (i.e., suppression-induced forgetting effect), we quantified this neural state transition at the individual level and examined whether individual differences in the transition predict individual subsequent suppression-induced forgetting measures (i.e., *subjective and objective suppression score*). The subjective and objective *suppression score* was calculated by subtracting the memory measure (i.e., confidence rating or recall accuracy) of suppression associations (i.e., “No-Think” items) from the control association separately.

Based on the group-level fMRI results, we calculated a *state transition index* to represent the degree of neural transition during the TNT task for each participant. The state transition index was calculated by adding up the averaged relative decreases in ranks of all *memory-related regions* and the averaged relative increase in rank of all *control-related regions*. The *state transition index* tended to be positively correlated with individual differences in *objective suppression scores* (r=0.36, p=0.06; **Figure 3.2C**), and *subjective suppression scores* (r=0.38, p=0.05; **Figure 3.2D**). For validation purposes (not an independent analysis), we used an alternative method (i.e., *state transition index Version2(V2))* to measure neural state transitions for each participant. This method was based on additional analyses of Think-to-No-Think neural state transition (See *Supplemental Materials*): all ROIs were divided into three groups (i.e., *increased group, stable group, and decreased group*) based on their relative changes in ranks. The state transition index V2 was calculated as the sum of the percentage of *memory-related nodes* within the *decreased group* and percentage of *control-related* nodes within the *increased group.* We also found the same significant correlations between *state transition index V2* and both *objective* and *subjective suppression scores* (**Figure S3.5**). These results suggested that the transition of neural states during the TNT task is relevant for the subsequent suppression-induced forgetting measured in the final memory test.

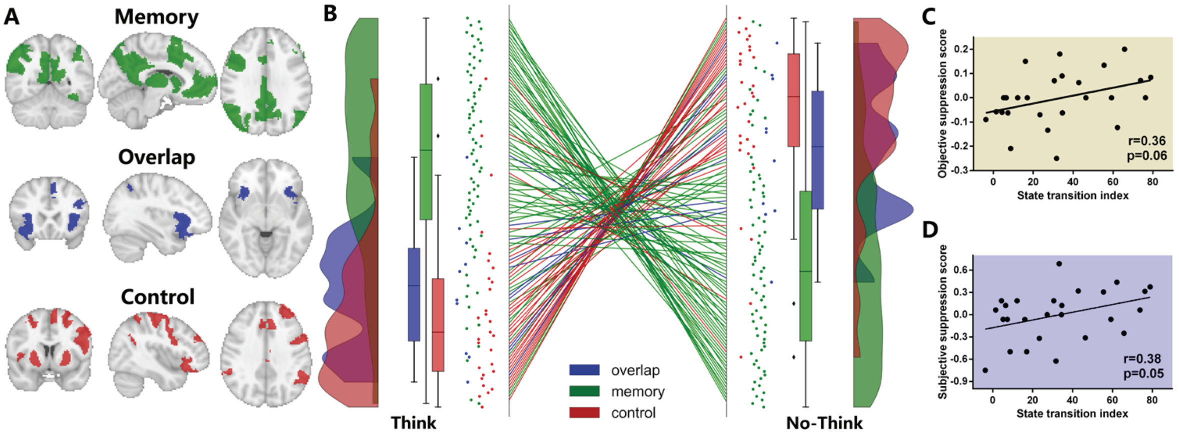


Figure 3.2 **(A)** Memory retrieval network (GREEN) and inhibitory control network (RED) was defined using the *Neurosynth* independent of fMRI data analyzed in this study. The overlap between the two brain networks was defined as the overlapping network (BLUE). **(B)** When the task demand switched from Think to No-Think, the activity of brain regions within the inhibitory control network increased, while the activity of brain regions within the memory retrieval network decreased. **(C)** Individual differences in the neural state transition (*state transition index details see Methods*) tended to correlate with the objective suppression-induced forgetting effect during subsequent memory retrieval task (r=0.36, p=0.06). **(D)** The same index also tended to correlate with the subjective suppression-induced forgetting effect (r=0.38, p=0.05).

Switch of task demand is accompanied by the delayed transition between neural states

To reveal how neural representations of task demands change during task switching, we used a multivariate decoding method to track the dynamics of neural state transitions on the volume-by-volume basis (**Figure 3.3A**). Support Vector Classification (SVC) was used to classify the underlying neural states (i.e., Think vs. No-Think) based on the fMRI activity intensity of all 110 ROIs at each given time point. Participant-specific classifiers were fitted on neural and task demand data from N-1 runs (i.e., four runs) and tested on the one remaining test run. Then the decoding accuracy was evaluated for each TNT run by comparing the decoded task demands with the actual demands. Averaged across runs, we were able to decode task demands based on ROIs' neural activity with the mean accuracy of 59.5% (SD=3.9% range from 52.5% to 67.1%) (**Figure 3.3B**). This accuracy is significantly higher than the chance level (i.e., 50%) (t(25)=12.5, p<0.001, Cohen's d=2.453). We generated the confusion matrix of our decoding analysis to quantify all types of correct and incorrect classifications (**Figure 3.3C**). 57.9% (SD=4.1%, range from 50.6% to 65.7%) of Think time points were correctly classified as Think. Among all No-Think time points, 61.1% (SD=3.9%, range from 53.9% to 68.4%) of them were correctly classified as No-Think. To reveal the relative contribution of each ROI to this decoding performance, we visualized the neural state-predictive pattern (i.e., SVC discriminating weights) in **Figure 3.3D**, which revealed a frontoparietal network of strong task demand representation, including the dorsal anterior cingulate cortex (dACC), DLPFC, IFG, superior, and inferior parietal lobule (**Table S3.3**). These regions were largely similar to the *overlapping network* (**Figure S3.6**).

To reveal how the switch of task demands affected underlying neural state transitions, we calculated the decoding accuracy for “switch” and “non-switch” time points separately. Higher decoding accuracy represented a timely update of neural states according to the current task demand, thus stronger neural representation of task demand. Compared to “switch” time points, task demand of “non-switch” time points can be decoded more accurately ($t(25)=3.93$, $p<0.001$, Cohen’s $d=0.77$; **Figure 3.3E**). That is to say, time points within No-Think trials following a Think trial were more often misclassified as Think trials compared to a No-Think trial following another No-Think trial. This pattern of results was also observed for Think trials. These findings suggest a delayed neural state transition immediately after the task switching.

Mismatches between task demand and underlying neural state relate to switch costs

Unlike most of the decoding analyses, which usually focused on the accuracy of the classifier, here we were particularly interested in the relationship between misclassified moments and *switch costs.* We already demonstrated that these misclassifications were largely induced by task switching, and we predicted that this mismatch could be the neural source of behavioral *switch costs*. To test this idea, we averaged the trial-by-trial performance measures (i.e., retrieval frequency rating for Think trials and intrusion frequency rating for No-Think trials) for four situations at issue (i.e., Think-Correct classification, Think-Incorrect classification, No-Think-Correct classification and No-Think- Incorrect classification) within switch trials.

We found that participants’ behavioral performance was impaired during these mismatch moments (i.e., incorrect classifications ($\text{mean}_{\text{incorrect}}=43.9\%$; $\text{SD}_{\text{incorrect}}=2.6\%$; ranging from 38.5% to 49.8% of all classifications)) immediately after task switches. Specifically, during the Think condition, when neural states were mistakenly classified as No-Think, the retrieval frequency rating was lower ($t(25)=3.57$, $p=0.001$, $d=0.701$; **Figure 3.3F**) compared to the situation in which task demands matched with the neural state. During No-Think trials, if neural states were erroneously decoded as Think, participants reported higher intrusion frequency rating ($t(25)=-3.08$, $p=0.005$, $d=-0.606$; **Figure 3.3G**) compared to the situation in which classifications were correct.

In our exploratory analyses, we found that such mismatch moments not only occurred during the task-switching but was also observed (but less frequently) during non-switch trials. Using the same decoding method, but focusing on non-switch time points, we found a similar detrimental effect of mismatch on behavioral performance (**Figure S3.7**). These findings suggested that spontaneous, uninstructed neural state transitions that do not fit current task demands also have behavioral impacts.

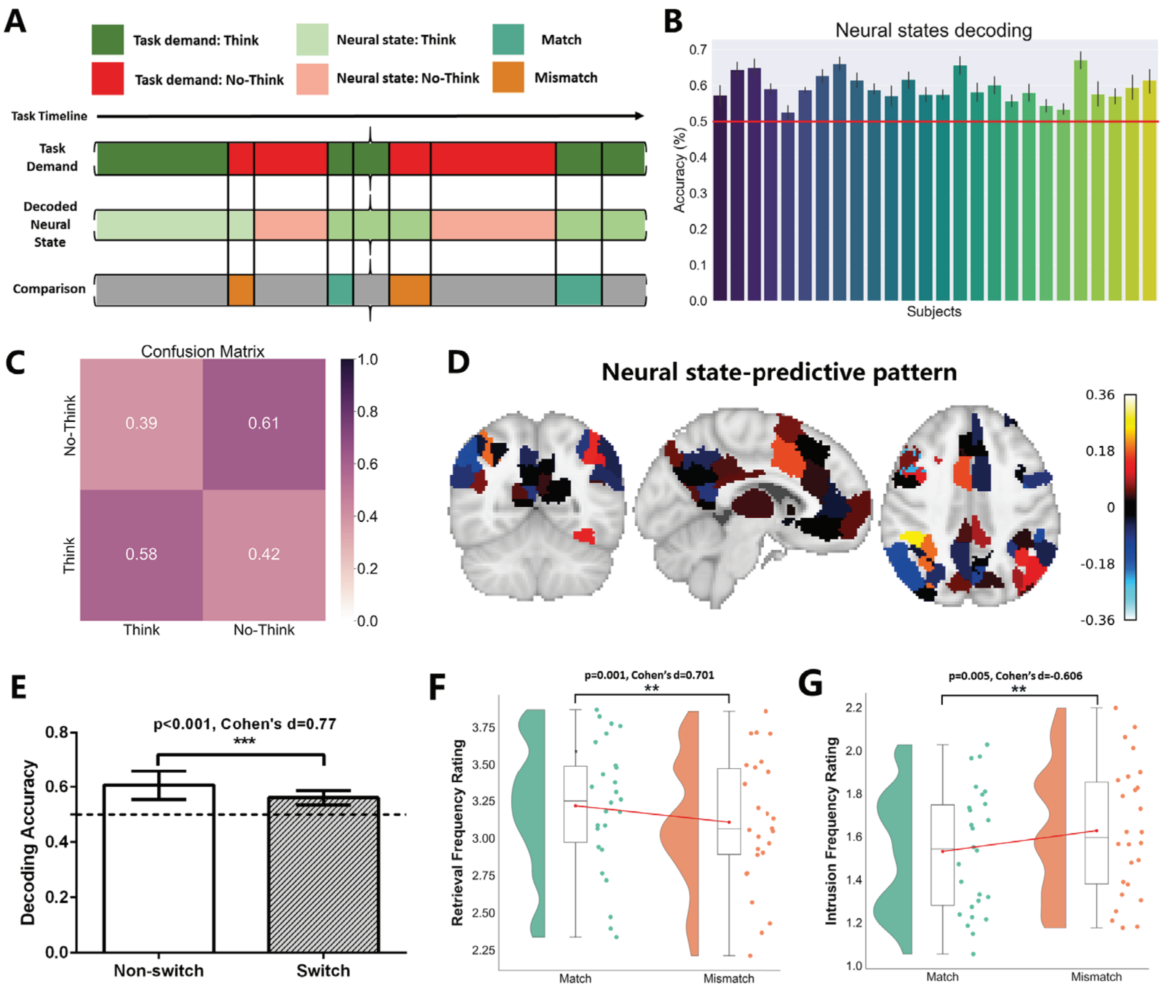


Figure 3.3 **(A)** Neural state decoding analysis. We trained the decoder based on large-scale brain network activity to classifier the task demand represented in the brain. We hypothesized that immediately after the switch of the task demand, the transition of the underlying neural state could be delayed. Therefore, the task demand could be misclassified as the opposite by the decoder. The real task demand was compared with the decoded neural state. The correctly decoded moments were labeled as "match" (e.g., Think as Think), while incorrectly decoded moments were defined as "mismatch" (e.g., Think as No-Think). **(B)** Decoding accuracies were presented for each participant and against the chance level of 50%. **(C)** Confusion matrix for four types of classification results. On average, 39% of the Think moments were labeled as No-Think, and 42% of the No-Think moments were regarded as Think moments by the decoder. These were the so-called "mismatch" moments depicted in Figure 3A. **(D)** The contribution of different brain regions during the decoding. This predictive pattern mainly includes the dACC, DLPFC, IFG, superior, and inferior parietal lobule. **(E)** More "mismatch" moments were found immediately after the task switching, indexed by the lower decoding accuracies during switch compared to non-switch moments ($p<0.001$, Cohen's $d=0.77$). **(F)** After the task switching, when the decoded neural state did not match with the task demand (i.e., Think decoded as No-Think), participants reported worse memory retrieval performance during Think trials ($p<0.001$, Cohen's $d=0.70$). **(G)** When the neural decoder misclassified No-Think moments as Think, participants reported more memory intrusions during No-Think trials ($p=0.005$, Cohen's $d=0.606$).

Neural state transitions are not the results of differences in head motion

We observed large-scale neural state transitions during the TNT task. Individual differences in these transitions were associated with the subsequent suppression-induced forgetting effect. There is a possibility that these neural state transitions are based on artifacts caused by different levels of participants' head motion (P. Huang et al., 2018; Siegel et al., 2017) between Think and No-Think trials since more inhibitory control resource was required for No-Think trials compared to Think trials (M. C. Anderson & Hanslmayr, 2014; Michael C Anderson & Green, 2001). Therefore, we examined the relationship between head motion, neural state transitions, and behaviors to rule out this alternative explanation.

We analyzed the time series of head motion (i.e., framewise displacement (FD) (Power et al., 2012)) during the TNT task. First, there is literally no difference in mean FD values between Think and No-Think trials (FD_{Think}=0.149 (SD=0.047); $FD_{No\text{-}Think}$=0.149 (SD=0.046); t(25)=0.30, p=0.76, Cohen's d=0.06). Second, for each participant, we calculated differences between the head motion of Think and No-Think trials (i.e., FD_{Think}-$FD_{No\text{-}Think}$) and found no correlation between these differences and *state transition indices* (r=-0.3, p=0.12)*, objective suppression scores* (r=-0.14, p=0.46) *or subjective suppression scores* (r=-0.23, p=0.25). Third, we asked whether head motion could affect our neural state decoding analysis. The head motion level tended to be lower (t(25)=-1.96, p=0.06, d=-0.38) for correct decoding ($FD_{correct}$=0.147; SD=0.045), compared to incorrect decoding ($FD_{incorrect}$=0.151; SD=0.048). This difference raised the question of whether the lower decoding accuracy for switch compared to non-switch condition resulted from higher head motion instead of differences in the neural representation related to task demands. Therefore, we also compared head motions between the switch and the non-switch conditions. In fact, we found that head motion is even lower in the switch condition (FD_{switch}=0.145 (SD=0.048); $FD_{non\text{-}switch}$=0.150 (SD=0.047); t(25)=3.35, p=0.003, d=0.65). This result ruled the head motion out as an alternative explanation for the lower decoding accuracy for the switch condition. If the lower decoding accuracy during switching were driven by excessive head motion dominantly, we would observe relatively higher instead of lower head motion. In sum, analyses of head motion suggest that our neuroimaging results are not likely to be the consequences of variations in head motions.

Discussion

Task switching is a crucial cognitive ability that has been intensively studied using behavioral and neuroimaging methods (Meiran, 2010; Richter & Yeung, 2014; Ruge et al., 2013). Here, we investigated the task switching process between memory retrieval and suppression and demonstrated that memory suppression is more difficult when the task demand for the participants just switched from retrieval to suppression. Applying multivariate decoding methods to human fMRI data, we revealed that immediately after the switch, task demands were weakly represented by the inhibitory control and memory retrieval networks, indexed by the lower decoding accuracy, compared to non-switch trials. Importantly, during the switching, when the neural representation of task demand cannot be updated in time to match the current demand, participants reported more memory intrusions in No-Think trials and less memory retrieval in Think trials. Together, we propose a novel mechanism that explains behavioral *switch costs* during task switching between retrieval and suppression: delayed transition of task-related neural states is associated with behavioral switch costs. That is to say, if the neural state cannot be timely updated after the switch of task demand, behavioral performance is compromised.

In the current study, participants were instructed to perform one of two opposite memory-related tasks (i.e., memory retrieval and memory suppression), with the task demand staying the same or switching between consecutive trials. Similar to what was reported in the classical task-switching paradigms (Jersild, 1927; Spector & Biederman, 1976), we found *switch costs* that are specific to memory suppression. Participants reported more memory intrusions when the current No-Think trial followed a Think trial, suggesting a higher demand for cognitive control over the tendency to retrieve during switch trials compared to non-switch trials. This lasting effect of memory retrieval on the subsequent memory suppression has not been reported before, but Hulbert and colleagues reported a lasting effect of memory suppression on the subsequent memory formation (Hulbert et al., 2016): when healthy participants suppressed unwanted memories, they were more likely to fail to encode information that was presented after a suppression trial. It was proposed that memory suppression created an amnesic time window, preventing the experience within the window being transformed into long-term memory. Evidence from fMRI supported this model by showing the reduction of hippocampal activity during memory suppression trials, and the positive correlation between individual differences in decreased hippocampal activity and the extent of memory impairment across participants (Hulbert et al., 2016).

Our finding of more memory intrusions during the No-Think trials that followed a Think trial could result from a similar mechanism: the preceding Think trials creates a time window in which the hippocampus remains active to support retrieval. However, if the transition of the neural state is delayed, the following No-Think trials are still located within this window, and therefore more prefrontal control resources are needed to down-regulate hippocampal activity. We tested this prediction beyond the hippocampus: large-scale neural activity of the inhibitory control and memory retrieval networks were analyzed by multivariate decoding methods to track the adaptive neural state transitions. We first characterized the transitions in neural states of memory retrieval and inhibitory control networks between Think and No-Think trials. Consistent with previous models of memory suppression (M. C. Anderson & Hanslmayr, 2014), our results showed that when the task demand switched from retrieval to suppression, memory retrieval-related regions, mainly including the hippocampus and regions of DMN, decreased their neural activity, while inhibitory control-related regions, such as dACC and LPFC, increased their activity. We also examined the relationship between individual differences in the efficiency of neural state transitions and the *suppression-induced forgetting effect* measured in the subsequent final memory test and found a positive correlation between them. There are two possible explanations for how neural state transitions are related to the effect of memory suppression: either larger or smaller state transitions are associated with stronger suppression effect. The more intuitive explanation is that larger transitions are beneficial for suppression; however, our data suggested the opposite: participants who demonstrated less neural reconfiguration showed stronger memory suppression effect in the following final memory test. This finding is nevertheless consistent with a previous study, which demonstrated that higher intelligence is associated with less task-related neural reconfiguration (Schultz & Cole, 2016). Our data, together with this study, may suggest that less neural reconfigurations could reflect optimization for efficient (i.e., less) state updates, reducing processing demands (Schultz & Cole, 2016). This optimal task-related neural reconfiguration could then be beneficial for memory suppression.

Recent human fMRI studies revealed task representations using multivariate decoding methods. Brain regions such as the parietal cortex, medial, and lateral PFC encode the current task demands (Bode & Haynes, 2009; Cole et al., 2011; Etzel et al., 2016; Gilbert, 2011; Momennejad & Haynes, 2013; Waskom et al., 2014; Wisniewski et al., 2015; Woolgar et al., 2011) and our study provided further support for this idea by showing that neural activity patterns of these regions largely contributed to successful discrimination between two kinds of visually highly similar trials with opposite task demands (i.e., memory retrieval and suppression). These identified regions have been previously associated with cognitive processing such as retrieval, maintenance, the process of rules or demands during task switching (Bunge et al., 2003; Gilbert, 2011; Reverberi et al., 2012; Sakai & Passingham, 2003; Woolgar et al., 2011). Beyond that, memory-related areas such as the hippocampus and regions within the DMN also contributed to the successful decoding in our study because the retrieval-demand and its associated neural activity significantly differed between Think and No-Think trials. However, whether these task representations can be modulated by external experimental manipulations and detected by fMRI signals is an ongoing debate. Task representations are modulated by factors including rule complexity (Woolgar et al., 2015), rewards (Etzel et al., 2016), and skill acquisition (Jimura et al., 2014), but not by variables such as task novelty (Cole et al., 2011), difficulty (Wisniewski et al., 2015), or intention (Wisniewski et al., 2016; Zhang et al., 2013). Two studies directly investigated whether and how cognitive control processes during task switching modulate the neural representation of task demands. Waskom and colleagues found that task representations are enhanced after switches, indexed by the higher decoding accuracy (Waskom et al., 2014). However, they did not find evidence for behavioral switch costs in their sample; thus, the relationship between higher decoding accuracy and task representation is unclear. Loose and colleagues did find the behavioral switch costs, but no modulation effect in the task representations (i.e., comparable decoding accuracy between the switch and non-switch trials) (Loose et al., 2017). They, therefore, proposed the switch-independent neural representations of task demands. Compared to the mentioned two studies mentioned here, our study found behavioral switch costs for memory suppression and lower decoding accuracies for task representations in switch trials. Critically, our time-resolved decoding approach revealed the relationship between task representation and behavioral performance on a trial-by-trial basis. Specifically, we showed that in switch trials, if the underlying neural state matched the external task demand, behavioral performance remained intact, while if the neural state was incorrectly represented, task performance was compromised. This pattern of results may explain why higher decoding accuracy was reported together with limited behavioral switch costs in Waskom's study (Waskom et al., 2014). As the adaptive coding hypothesis suggests (Duncan, 2001, 2010; Waskom et al., 2014), our findings using the time-resolved decoding approach demonstrated the dynamic adjustment of task-specific neural representations and was able to associate delayed neural transitions with behavioral *switch costs*.

In summary, our results provide novel insights into the switch between memory retrieval and memory suppression. We found evidence for *switch costs* in memory suppression: it is more difficult to suppress unwanted memories immediately after memory retrieval. During switching between retrieval and suppression, we observed delayed transitions of neural states that each of them separately represents current task demand. Delayed neural transitions were associated with *switch costs* (i.e., unsuccessful suppression and retrieval). These results provide insight into the critical role of dynamically adjusted neural reconfigurations in supporting flexible memory suppression and the broader neural mechanisms by which humans can flexibly adjust their behavior in ever-changing environments.

Materials and Methods

Participants

In total, thirty-two right-handed, healthy young participants recruited from the Radboud Research Participation System finished all of the experimental procedures. All of them are native Dutch speakers. Six participants were excluded from data analyses due to low memory performance (i.e., lower the chance level) (n=2), or excessive head motion (n=4). We used the motion outlier detection program within the FSL (i.e., FSLMotionOutliers) to detect timepoints with large motion (threshold=0.9). There are at least 20 spikes detected in these excluded participants with the largest displacement ranging from 2.6 to 4.3, while participants included had less than ten spikes. Finally, 26 participants (15 females, age=19-30, mean=23.51, SD=3.30) were included in the behavioral and neuroimaging analysis reported in this study. Due to the reconstruction error during the data acquisition, one run of one participant is not complete (20-30 images were missing). Therefore, that run was not included in our analysis of time series. But unaffected acquired images of that run were used in our univariate activation analysis. No participants reported any neurological and psychiatric disorders. We further used the Dutch-version of the Beck Depression Inventory (BDI) (Roelofs et al., 2013) and State-Trait Anxiety Inventory (STAI) (van der Bij et al., 2003) to measure the participants' depression and anxiety level during scanning days. No participant showed a sign of emotional problems (i.e., their BDI and STAI scores are within the normal range). The experiment was approved by and conducted in accordance with requirements of the local ethics committee (Commissie Mensgebonden Onderzoek region Arnhem-Nijmegen, The Netherlands) and the declaration of Helsinki, including the requirement of written informed consent from each participant before the beginning of the experiment. Each participant got 10 euros/hour for their participating.

Experiment design

This experiment is a two-day fMRI study, with 24 hours delay between two sessions (Figure S8). fMRI data of the day2 final memory test has been published in another publication (W. Liu, Kohn, et al., 2020), and the comprehensive reports of the experimental materials and design can be found there. Because all of the behavioral and neuroimaging data included in this study came from the Day2 session, we just presented a brief description of the Day1 session. On day1, we instructed participants to memorize a series of sequentially presented location-picture associations, for which 48 distinct photographs were presented together with 48 specific locations on two cartoon maps. All photographs can be assigned into one of the four categories, including animal, human, scene (e.g., train station), and object (e.g., pen and notebooks). Therefore, objective memory performance could be assessed within the scanner by instructing participants to indicate the picture's category when cued by the map location. During this study phase, each location-picture association was presented twice, and the learning was confirmed by two typing tests outside the scanner. During the typing tests, participants were required to describe the photograph associated with the memory cue in one or two sentences. Immediately after the study phase (Day1), 88.01% of the associated pictures were described correctly (SD= 10.87%; range from 52% to 100%).

On Day2, participants first performed the second typing test, and still recalled 82.15% of all associations (SD = 13.87%; range from 50% to 100%). Then, they performed the Think/No-Think (TNT) task, and final memory test insider the scanner. We used the TNT task with trial-by-trial performance rating to monitor the retrieval or suppression of each trial. Compared to the original TNT task (Michael C Anderson, 2004), the additional self-report did not affect the underlying memory suppression process and also was used in a neuroimaging experiment before (B. J. Levy & Anderson, 2012). Forty-eight picture-location associations were divided into three conditions (i.e., "think or retrieval," "no-think or suppression," and "baseline or control" condition) in a counterbalanced way, therefore, for each association, the possibility of belonging to one of the three conditions is equal. During the retrieval condition, locations were highlighted with the GREEN frame for 3s, and participants were instructed to recall the associated picture quickly and actively and to keep it in mind until the map disappeared from the screen. By contrast, during the suppression condition, locations were highlighted with the RED frame for 3s, and our instruction for participants was to prevent the potential memory retrieval and try to keep an empty mind. We gave additional instructions for the suppression condition: "*when you see a location, highlighted with a RED frame, you should NOT think about the associated picture. Instead, you should try to keep an empty mind during this stage. It is a difficult task, and it is totally fine that sometimes you still think about the associated picture. But please do NOT close your eyes, focus on something outside the screen, or think about something else in your life. These strategies, although useful, could negatively affect the brain activity that we are interested in ……*". After each trial, participants had a maximum 3s to press the button on the response box to indicate whether and how often the associated picture entered their mind during Think or No-Think trials. Specifically, they rated their experience from 1-4 representing from No Recall (i.e., Never) to Always Recall. Responses during Think trials were used as retrieval frequency ratings, while responses during No-Think trials were regarded as intrusion frequency ratings. Associations which belong to the control condition were not presented during this phase. The TNT task included five functional runs, with 32 retrieval trials and 32 suppression trials per run. All "retrieval" or "suppression" associations were presented twice within one run, but not next to each other. Therefore, they were presented ten times during the entire TNT task. Between each trial, fixation was presented for 1-4s (mean=2s, exponential model) as the inter-trial intervals (ITI).

To investigate the task switching within the TNT task, for each run of each participant, we predefined the sequence of task demand to form "blocks" of memory retrieval or suppression with the length range from 1 trial to 4 trials (mean=1.9 trials, std=1.01 trials, $P_{\text{one-trial block}}$=46.875%, $P_{\text{two-trials block}}$=25%, $P_{\text{three-trials block}}$=18.75%, $P_{\text{four-trials block}}$=9.375%). In this sequence, the task demand of the current trial can be the same as the previous trial ("non-switch" trial) or differ from the previous trial ("switch" trial). Within one run of a total of 64 trials, 31 trials were "non-switch" trials, 32 trials were "switch" trials, and the first trial cannot be labeled as "non-switch" trials or "switch" trials because it has no predecessor. The "non-switch" trials and "switch" trials both accounted for around 50% of the "retrieval" and "suppression" trials. After determining the sequence of task demand, specific location-picture associations from retrieval or suppression condition were randomly selected for each trial.

After the TNT task, a final memory test was performed by participants within the scanner to evaluate the effect of different modulations on memory. All 48 memory cues (i.e., locations) were presented again with the duration of 4s by highlighting a certain part of the map with a BLUE frame. Participants were instructed to recall the associated picture as vividly as possible during the presentation and then give the responses on two multiple-choice questions within 7s (3.5s for each question). The first one is the measure of subjective memory: "how confident are you about the retrieval?". Participants had to rate from 1 to 4 representing "Cannot recall, low confident, middle confident and high confident" separately. The second one is the measure of objective memory: "Please indicate the category of the picture you were recalling." They needed to choose from four categories (i.e., Animal, Human, Scene, and Object). It is notable that we only analyzed the behavioral data from this within-scanner memory test; the neural activity during this test is not the focus of this study.

Behavioral data analysis

Behavioral results of this project were comprehensively reported in another study of our lab with the focus on the final memory test (W. Liu, Kohn, et al., 2020). No results of tasks witching (i.e., *switch costs*) were reported in that study, and task switching is the central scientific question of this study. First, we analyzed the behavioral performance during the TNT task. Trial-by-trial performance reports from each participant were used to calculate the percentage of successful recall chosen across 160 retrieval trials and successful suppression across 160 suppression trials. Following previous studies (B. J. Levy & Anderson, 2012; W. Liu, Kohn, et al., 2020), performance reports from suppression trials were used to quantify individual differences in memory suppression efficiency ("*intrusion slope score*"). To account for the individual differences in memory performance before the TNT, we restricted the analysis of suppression into the associations for which participants can still remember during the second typing test ("remembered associations"). We used linear regression to model the relationship between intrusion frequency ratings of "remembered associations" and the number of repetitions of suppression at the individual level. Participants with more negative slope scores are better at downregulating memory intrusions than those with less negative slope scores. Furthermore, we labeled each trial as the "non-switch" trial or "switch" trial based on whether the task demand of the current trial is the same as the previous trial. Trial-by-trial performance between "switch" and "non-switch" trials during retrieval or suppression was compared using paired t-tests.

We also quantified the individual differences in *suppression-induced forgetting effect* based on two types of participants' performance (i.e., recall accuracy and confidence rating) during the final memory test. For each participant, recall accuracy (objective memory measure) and confidence rating (subjective memory measure) were calculated for No-Think associations and control associations separately. Then objective and subjective *suppression scores* were computed separately by subtracting the accuracy and confidence of No-Think associations from the control associations. The more negative a *suppression score* is, the stronger the *suppression-induced forgetting effect* is. The memory suppression score was used to correlate with the "*intrusion slope score*" and transition of neural states during the TNT.

MRI data acquisition and preprocessing

We used a 3.0 T Siemens PrismaFit scanner (Siemens Medical, Erlangen, Germany) and a 32 channel head coil system at the Donders Institute, Centre for Cognitive Neuroimaging in Nijmegen, the Netherlands to acquire MRI data. For each participant, MRI data were acquired on two MRI sessions (around 1 hour for each session) with 24 hours' interval. In this study, we only used the data from the day2 session. Specifically, we acquired a 3D magnetization-prepared rapid gradient echo (MPRAGE) anatomical T1-weighted scan for the registration purpose with the following parameters: 1 mm isotropic, TE = 3.03 ms, TR = 2300 ms, flip angle = 8 deg, FOV = 256 × 256 × 256 mm. All functional runs were acquired with Echo-planar imaging (EPI)-based multi-band sequence (acceleration factor=4) with the following parameters: 68 slices (multi-slice mode, interleaved), voxel size 2 mm isotropic, TR = 1500 ms, TE = 39 ms, flip angle =75 deg, FOV = 210 × 210 × 210 mm. In addition, to correct for distortions, magnitude and phase images were also collected (voxel size of 2 × 2 × 2 mm, TR = 1,020 ms, TE = 12 ms, flip angle = 90 deg).

We used the FEAT (FMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl) (Jenkinson et al., 2012) together with Automatic Removal of Motion Artifacts (ICA-AROMA) (Pruim et al., 2015) to perform our preprocessing. This pipeline was based on procedures suggested by Mumford and colleagues (http://mumfordbrainstats.tumblr.com) and the article that introduced the ICA-AROMA (Pruim et al., 2015). Specifically, we first removed the first four volumes of each run from the 4D sequences for the stabilization of the scanner and then applied the following pre-statistics processing: (1) motion correction using MCFLIRT (Jenkinson et al., 2002); (2) field inhomogeneities were corrected using B0 Unwarping in FEAT; (3) non-brain removal using BET (S. M. Smith, 2002); (4) grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor; (5) spatial smoothing (6mm kernel). ICA-AROMA was used to further remove motion-related spurious noise. We chose to conduct "non-aggressive denoising" and applied highpass temporal filtering (Gaussian-weighted least-squares straight-line fitting with sigma=50.0s) before the following analyses.

All of the mentioned preprocessing steps were performed in native space. We used the following steps to perform the registration between native space, participant's high-resolution T1 space, and standard space. Firstly, we used the Boundary Based Registration (BBR) (Greve & Fischl, 2009) to register functional data to the participant's high-resolution structural image. Next, registration of high resolution structural to standard space was carried out using FLIRT (Jenkinson et al., 2002; Jenkinson & Smith, 2001) and was then further refined using FNIRT nonlinear registration (Andersson et al., 2007). Resulting parameters were used to align processed functional images from native-space to standard space for the following signal extraction.

Univariate General Linear Model (GLM) analyses

We ran the voxel-wise GLM analyses of the TNT task to identify brain regions that are more active during memory suppression compared to memory retrieval (i.e., No-Think VS Think). All time-series statistical analysis was carried out using FILM with local autocorrelation correction (Woolrich et al., 2001) using FEAT. In total, three regressors were included in the model. We modeled the presentation of memory cues (locations) as two kinds of regressors (duration=4s) (i.e., suppression trials and retrieval trials). To account for the effect of unsuccessful memory retrieval, we separately modeled the location-picture associations, which participants cannot recall before the TNT as a separate regressor. All the trials were convolved double gamma hemodynamic response function (HRF) within the FSL.

We conducted the two contrasts-of-interest (i.e., No-Think VS Think and Think VS No-Think) first at the native space and then aligned resulting statistical maps to MNI space using the parameters from the registration. These aligned maps were first used for participant-level averaging across five TNT runs, and then the group-level analyses. The group-level statistical map was corrected for multiple comparisons using default cluster-level correction within FEAT (voxelwise $Z>3.1$, cluster-level $p < .05$ FWER corrected).

Networks-of-interest identification

To identify our networks-of-interest (i.e., inhibitory control network and memory retrieval network), we performed several term-based meta-analyses using the *Neurosynth* (https://neurosynth.org/) (Yarkoni et al., 2011). “Inhibitory control” and “memory retrieval” were used as terms separately to search for all studies in the *Neurosynth* database whose abstracts include the input term at least once. Then, all identified studies were combined separately for each term to generate the corresponding statistical map. We used uniformity test maps in our study. This method tested whether the proportion of studies that report activation at a given voxel differs from the rate that would be expected if activations were uniformly distributed throughout the grey matter. Voxel-wise Z-score from the one-way ANOVA testing was saved in a statistical map. Each map was thresholded to correct for multiple comparisons using a false discovery rate (FDR)($p<0.01$). It is notable that due to the continuous update of the *Neurosynth* database, the number of studies included in the analyses could be slightly different for each search, the maps we used can be found in our Neurovalut repository (https://identifiers.org/neurovault.collection:7731). Similar network identification was also performed using *BrainMap* (Angela R Laird et al., 2005) as a confirmation. The two methods of meta-analysis yielded highly similar maps of network-of-interests (**Figure S3.9**), and we used the maps generated by the *Neurosynth* in our main text.

We used the thresholded ($p_{FDR}<0.01$) spatial maps of “inhibitory control” and “memory retrieval” to general three masks of networks-of-interest. The areas which belong to both the “inhibitory control” and “memory retrieval” masks were labeled as *overlap regions*, the areas which only belong to the “inhibitory control” mask were labeled as *control-related regions*, and the areas which only belong to the “memory retrieval” masks were labeled as *memory-related regions*.

Brain parcels for the extraction of time series

We combined a parcellation of cerebral regions (N=300) (Schaefer et al., 2018) and all subcortical regions (N=14) from the probabilistic Harvard-Oxford Subcortical Structural Atlas (Desikan et al., 2006) as a whole-brain parcellation. The parcellation of cerebral regions was based on a gradient-weighted Markov Random Field (gwMRF) model, which integrated local gradient and global similarity approaches (Schaefer et al., 2018). Based on both task fMRI and resting-state fMRI acquired from 1489 participants, parcels with functional and connectional homogeneity within the cerebral cortex were generated. Each parcel is one of the seven large-scale functional brain networks, including *Visual, Somatomotor, Dorsal Attention, Ventral Attention, Limbic, Frontoparietal, Default network* (Yeo et al., 2011). Subcortical regions included bilateral thalamus, caudate, putamen, globus pallidus, hippocampus, amygdala, and ventral striatum. Details of each parcel (e.g., name, coordinates, hemisphere) within the whole brain parcellation can be found in our OSF folder (https://osf.io/cq96h/).

For each of the 314 parcels of the whole-brain parcellation, we compared it with the mask of *overlap regions*, *control-related regions*, and *memory-related regions* and identified the mask in which the parcel shared the highest percentage of common voxels. The parcel was assigned to that category if the highest percentage is higher than 10%. If the highest percentage of common voxels is lower than 10%, the parcel was not assigned to any category. After this procedure, 110 out of the 314 parcels were assigned to one of the categories. Specifically, 71 parcels were considered as *memory-related regions*, 29 parcels were categorized as *control-related regions*, and 10 parcels were labeled as *overlap regions* in our following analysis.

Extraction of time series from parcels

We additionally removed nuisance time series (cerebrospinal fluid (CSF) signals, white matter signals, motion, and event-related activity) using a method based on a projection on the orthogonal of the signal space (Friston et al., 1994; Lindquist et al., 2019). We generated confounding time series (CSF, white matter, the six rigid-body motion parameters (three translations and three rotations), and framewise displacement (FD)) for each run of each participant. Event-related activity time series were estimated by a finite impulse response (FIR) function. A recent study has shown that removal of event-related activity based on FIR modeling is an important step for the preprocessing of time series during a task (Cole et al., 2019). The signal from each parcel was extracted and z-scored, and all nuisance time series were removed simultaneously using the *nilearn.signal.clean* function. All cleaned time series were shifted 3 TRs (4.5 s) to account for the HRF delay and then aligned with the task demand (i.e., retrieval or suppression) at that moment.

The transition of neural states analysis

First, we characterized the transition of neural states at the group level. Extracted time series from each run of each participant were split according to the task instruction (i.e., memory retrieval or memory suppression) and concatenated. Second, two kinds of time series were further concatenated across five TNT runs within that participant (*except for one participant, only four complete TNT runs were included*). Third, time series were concatenated across all participants. Fourth, two time-series were averaged across all time points to represent mean activity intensity for that parcel during retrieval or suppression.

To estimate the relative dominance of each parcel during two neural states (i.e., Think and No-Think), we ranked the mean activity intensity of each parcel (the highest activity was ranked first). We then calculated the changes in ranks when the task switched from Think to No-Think by subtracting the rank during Think from the rank during No-Think. The same analyses were conducted with raw signal intensity and Z-values. Related results can be found in the *Supplemental Materials*. The negative change suggested an increase in relative dominance, while the positive change represented the opposite. We calculated two neural indexes ("state transition index" and "state transition index Version 2 (V2)") to quantify the transition of neural states at the individual level and associate this individual difference with the subsequent suppression-induced forgetting effect. The state transition index was calculated by adding up the averaged relative decreases in rank values all *memory-related regions,* and the averaged relative increase in rank values of all *control-related regions*. The calculation and results of "state transition index Version 2 (V2)" can be found in the *Supplemental Material*. It is notable that although transition index and transition index V2 were calculated using different methods, they were based on the same set of data. Therefore, the data analyses of index V2 should not be regarded as independent analysis. These two state transition indices were used to correlate with the *objective suppression score* and *subjective suppression score* calculated based on behavioral performance during the final memory test after TNT.

Neural states decoding analysis

Before the decoding analysis, we generated the labels of task demand for each time point within the trial based on its instruction (i.e., Think or No-Think). For example, if the trial is a Think trial, time points started from the presentation of memory cues of this trial to the presentation of memory cues of the next trial were labeled as "Think." We performed the time-resolved multivariate decoding analysis based on the brain activity of all 114 ROIs and corresponding labels of task demand during each time point. This decoding analysis allowed us to generate the predicted label of task demand for each time point, thus revealing the fast dynamics of the neural state transition induced by the switch of task demand. Specifically, decoding analysis via the Support Vector Classification (SVC), the C-Support Vector Machine within the scikit-learn package (https://scikit-learn.org/stable/). We used default parameters of the function (regularization (C)=1, radial basis function kernel with degree=3). The classification of neural states was performed separately for each time point using a leave-one-run-out cross-validation approach within each participant. This procedure resulted in a decoded task demand for each time point of each participant. These predictions were evaluated by comparing these decoded task demands with actual task demand. To separate all types of correct and incorrect classification for the following analyses, we generated the confusion matrix for each participant. This confusion matrix contained the percentage of all four situations based on the task demand and if the prediction matches the task instruction (i.e., Think-Correct classification, Think- Incorrect classification, No-Think- Correct classification, and No-Think-Incorrect classification). We extracted all SVC discriminating weights assigned to the features during the participant-specific decoding and averaged them across all participants to generate the neural state-predictive pattern. The brain parcels with higher absolute values contributed more to decoding models.

To test for possible differences in neural representations of task demand induced by the task-switching, we performed the described decoding analyses for switch time points and non-switch time points separately. The switch time points were defined as the presentation time (3TRs; 4.5s) of first memory cue after the switch of task demand. The two decoding analyses yielded decoding accuracies for switch time points and for non-switch time points for each participant. We compared these two types of decoding accuracies using the paired t-test. Less accurate decoding was described as the evidence for the weaker representation of the current task demand in the literature (Loose et al., 2017; Waskom et al., 2014). Also, because we only have two task demands, less accurate decoding reflects the unsuccessful transition from the previous demand to the current demand according to the instruction.

Next, we aimed to investigate the behavioral relevance of the mismatch (i.e., incorrect classification) between task demand and the underlying neural state. Because we were mainly interested in the switch-induced mismatch, we first restricted our analyses to these switch time points and then extended to the non-switch time points as an exploratory analysis. For each participant, we averaged the trial-by-trial behavioral performance during the TNT task based on whether the actual task demand matches with decoded task demands. This yielded retrieval performance and suppression performance for match and mismatch conditions. Paired t-tests were performed to examine the effect of mismatch on the performance of memory retrieval and memory suppression separately. The performance calculations and comparisons described above were repeated for non-switch time points as well.

Relationship between head motion, neural state transitions, and behaviors

To explicitly assess how head motion could potentially affect our results, we derived a volume-by-volume measure of head motion, framewise displacement (FD) (Power et al., 2012), during the TNT task. FD is defined as the sum (in mm) of rotational and translational displacements from the current volume to the next volume. We aligned the time-series of FD with task structure and behaviors in a way similar to the analyses of time series of fMRI signals but did not consider the HRF. The following contrasts were performed to compare head motion between conditions: (1) difference in FD between Think trials and No-Think trials; (2) difference in FD between correct neural state decoding and incorrect neural state decoding; (3) difference in FD between the switch and non-switch condition. Correlations analyses were performed between individual differences in head motion between Think and No-Think trials (i.e., FD_{Think}-$FD_{No\text{-}Think}$), *state transition index*, *objective/subjective suppression score*.

Data and code availability

Custom scripts used in this study, immediate data (i.e., preprocessed single-trial activation patterns used for reinstatement analyses) as well as raw data were uploaded to the Donders Repository (https://data.donders.ru.nl/). The project was named as *Tracking the involuntary retrieval of unwanted memory in the human brain with functional MRI* in the Repository (https://

doi.org/10.34973/5afg-7r41). Some data, such as statistical maps and brain parcels of interest were shared via the Neurovault Repository (https://identifiers.org/neurovault.collection:7731). Supplemental Material can be found in OSF ((https://osf.io/cq96h/).

Behavioral data were analyzed by *JASP* (https://jasp-stats.org/). For the term-based meta-analysis of neuroimaging studies, we used the *Neurosynth* (https://neurosynth.org/), and *BrainMap* (http://www.brainmap.org/). Preprocessing of neuroimaging data was performed by *FSL* (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki), *ICA-AROMA* (https://github.com/maartenmennes/ICA-AROMA), and *fMRIPrep* (https://fmriprep.readthedocs.io/en/stable/). Python packages, including *Nilearn* (https://nilearn.github.io/), *Nistats* (https://nistats.github.io/), *Pandas* (https://pandas.pydata.org/), and *Nunpy* (https://numpy.org/) were used for the analyses of time series. Machine learning algorithms were based on *scikit-learn* (https://scikit-learn.org/) and implemented via *Nilearn* (https://nilearn.github.io/). *Anaconda* (https://www.anaconda.com/) Python 3.6 was used as the platform for all the programming and statistical analyses. Custom Python scripts were written to perform all analyses described based on the mentioned Python packages; all code is available from the authors upon request and will be released via our OSF repository (https://osf.io/cq96h/) upon publication.

Supplementary Material for

The dynamic transition between neural states is associated with the flexible use of memory

Wei Liu[1], Nils Kohn[1], Guillén Fernández[1]

1. Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, The Netherlands

This file includes:

Supplementary Text
Figure S3.1-S3.9
Table S3.1-S3.3

1. Behavioral performance during the final memory test

During the final memory test, each memory cue was presented again, and the participant was instructed to rate the confidence of their memory for this association, and then classified the category of the associated picture. We examined the effect of memory retrieval and suppression during the TNT on the subsequent subjective (confidence rating) and objective (if they selected the correct category) memory. Three kinds of associations (i.e. retrieval association, suppression association, and control association) did not differ in their objective recall accuracy

(F [2,26] =0.524, p=0.595, η^2 =0.02; **Figure S3.2A**). Our experiment design may explain the lack of the main effect of modulation on objective memory at the group level: all associations underwent overnight consolidation, thus difficult to be modulated (W. Liu, Kohn, et al., 2020; Y. Liu et al., 2016). Crucially, replicating the previous study (B. J. Levy & Anderson, 2012), we found the suppression-induced forgetting at the individual level, and this effect was determined by individual differences in the efficiency of suppression during the TNT task. Specifically, participants who were more effective in suppressing intrusions (more negative *intrusion slope score*) during the TNT phase were the ones who show larger suppression-induced forgetting effects (r=0.411, p=0.03; **Figure S3.2C**). Next, analyzing the subjective memory, we found a significant effect of modulation on subjective memory (F [2,26] =5.928, p=0.005, η^2 =0.186; **Figure S3.2B**). Participants reported higher confidence for retrieval associations compared to control associations (t=3.35, p_{holm}=0.007) and a trend towards higher confidence compared to suppression associations (t=2.172, p_{holm}=0.07). Finally, we asked if modulation affected retrieval speed indexed by the RT during the final test. Even though we did not find a significant main effect of modulation (F [2,26]=2.905, p=0.06, η^2 =0.03; **Figure S3.2D**), recall of *RETRIEVAL ASSOCIATIONS* was faster compared to the recall of *CONTROL ASSOCIATIONS* (t(26)=-2.486, p=0.02, Cohen's d=-0.47).

2. Additional analyses of Think-to-NoThink neural state transition

To describe further the reconfiguration, we divided all ROIs into three groups (*increased group, stable group, and decreased group*) based on their relative changes in rank. When the task demand changed from Think to No-Think, 47.88% of the *memory-related regions* showed the top one-third decrease in relative rank values and therefore belonged to the decreased group. Another 39.43% of the *memory-related regions* did not change extensively during the transition (*stable group*), and 12.67% of the regions showed increases in their ranks (*increase group*). *Control-related regions* demonstrated the opposite neural changes: 75.86% of them belonged to the *increased group*, with 17.24% and 6.89% of their regions belonged to a *stable group* or *decreased group* separately. For *overlap regions*, 50% of them belonged to the *increased group*, 40% of them belonged to the stable group, and only one region belonged to the decreased group. We further looked at the proportion of *memory-related regions, control-related regions, and overlap regions* within *increased, decreased*, and *stable* groups separately. A chi-square test of independence was performed to examine the relations between their functions (i.e., *memory-related, control-related, or overlap*) and which change group they belong to. The relation between these variables was significant (X^2 =41.38, p<0.001). *Control-related regions* were more likely to be allocated to the increased group, while *memory-related regions* were more likely to be assigned to the decreased group. Specifically, among the increased group with a total of 36 regions (around 33.3% of all 110 ROIs analyzed), 61% of the regions were *control-related regions* (expected percentage=26.4%, p<0.001, one-side binomial test).

By contrast, within the decreased group, 92% of the regions were *memory-related regions* (expected percentage=64.5%, $p<0.001$, one-side binomial test). Taken together, we found that during Think condition, *memory-related regions* showed relatively high neural activity compared to *control-related regions* and *overlap regions*. When the task demand changed from Think to No-Think, *memory-related regions* showed decreases in their relative contribution, while *control-related regions* demonstrated an increase in their activity ranks. These patterns of changes were not only presented when we analyzed the rank of activity among ROIs (**Figure S3.4A**) but also existed when we analyzed their raw (**Figure S3.4B**) or Z score (**Figure S3.4C**) of activity intensity.

3. An alternative method to quantify individual differences in neural state transitions

In the main text, we presented significant correlations between the state transition index and *objective/subjective suppression effect* (See Figure 2C and D). However, statistical tests were just around a significant level of $p=0.05$. To further validate the relationship between neural state transitions during the TNT and the subsequent forgetting effect, we used an alternative method (i.e., state transition index Version2 (V2)) to quantify individual differences in neural state transitions and performed the correlation again. This method is based on the additional analysis of the Think-to-NoThink neural state transition above. For each participant, all 110 ROIs were divided into three groups (i.e., *increased group, stable group, and decreased group*) based on their relative changes in rank**.** The state transition index V2 was defined as the sum of the percentage of *memory-related nodes* within the *decreased group* and percentage of *control-related nodes* within the *increased group*. As shown in **Figure S3.5** (*right panels*), state transition index V2 positively associated with the individual differences in objective suppression score ($r=0.43$, $p=0.02$), and tended to associate with subjective suppression score ($r=0.36$, $p=0.06$).

References

Levy, B.J., Anderson, M.C. (2012). Purging of Memories from Conscious Awareness Tracked in the Human Brain. Journal of Neuroscience, 32, 16785–94

Liu, Y., Lin, W., Liu, C., et al. (2016). Memory consolidation reconfigures neural pathways involved in the suppression of emotional memories. Nature communications, 7, 13375

Liu, W., Kohn, N., Fernández, G. (2020). Probing the neural dynamics of mnemonic representations after the initial consolidation. Neuroimage.

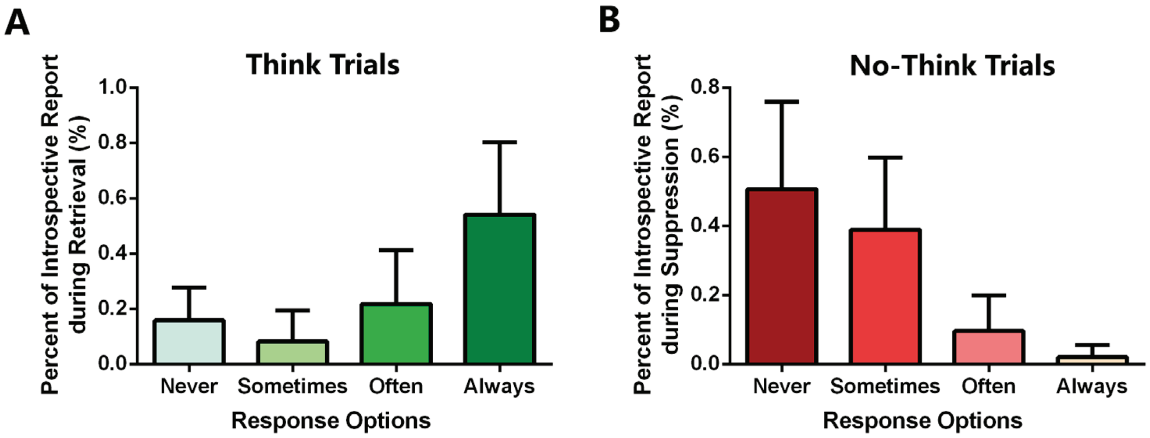


Figure S3.1 Behavioral performance during the Think/No-Think task. (A) Percentage of the trial-by-trial introspective report during the Think trials. For most of the Think trials, associated pictures were successfully recalled (1-P_{never}: mean=84.05%, SD=11.79 %). **(B)** Percentage of the trial-by-trial introspective report during the No-Think trials. During half of the No-Think trials, participants successfully suppressed the tendency to recall the associated pictures (P_{never}: mean=50.62%, SD=25.35%).

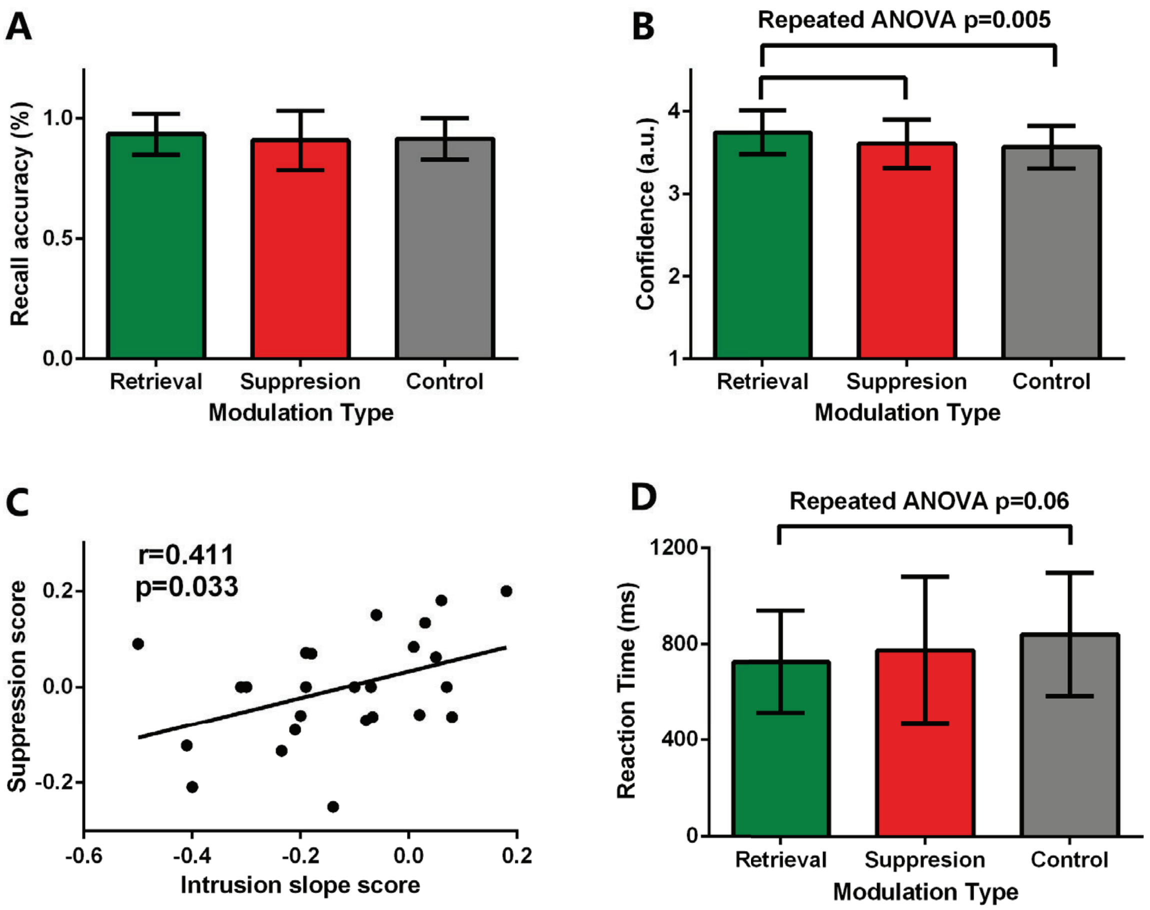


Figure S3.2 Behavioral performance during the final memory test. (A) There is no effect of retrieval or suppression on the accuracy of the categorization during the final test (p=0.595). **(B)** For *RETRIEVAL ASSOCIATIONS*, participants reported higher subjective confidence compared to *SUPPRESSION ASSOCIATIONS* (t(26)=2.172, p_{holm}=0.07, Cohen's d=0.41) , and *CONTROL ASSOCIATIONS* (t(26)=3.35, P_{holm}=0.007, Cohen's d=0.64). **(C)** Participants who are more effective in reducing suppression failures (more negative the *Intrusion Slope Score*) were the ones who show more evidence suppression-induced forgetting (more negative the *Suppression Score*). **(D)** For *RETRIEVAL ASSOCIATIONS*, participants spent less time during categorization compared to the *CONTROL ASSOCIATIONS* (t(26)=-2.486, p=0.02, Cohen's d=-0.47), and the effect between three conditions tend to be significant (F [2,26]=2.905, p=0.06, □^{2} =0.03). a.u= arbitrary unit.

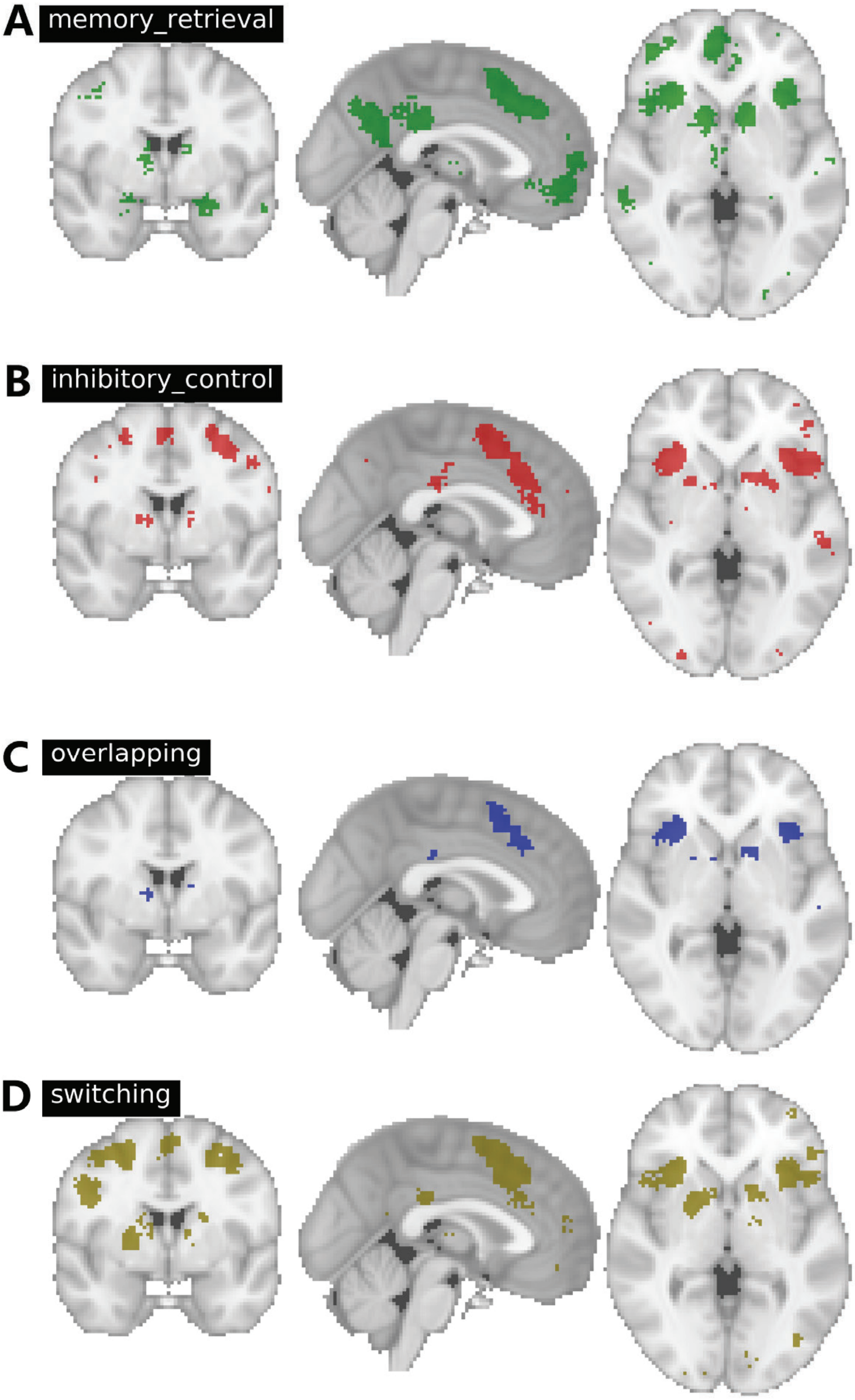


Figure S3.3 Brain networks defined by the Neurosynth-based meta-analyses. (A) Memory retrieval network generated using the term “memory retrieval.” **(B)** Inhibitory control network generated used the term “inhibitory control.” **(C)** Voxels that belong to both memory retrieval network and inhibitory control network. **(D)** Switching network generated used the term “task switching.” All raw statistical maps can be found in our Neurovault repository (https://identifiers.org/neurovault.collection:7731).

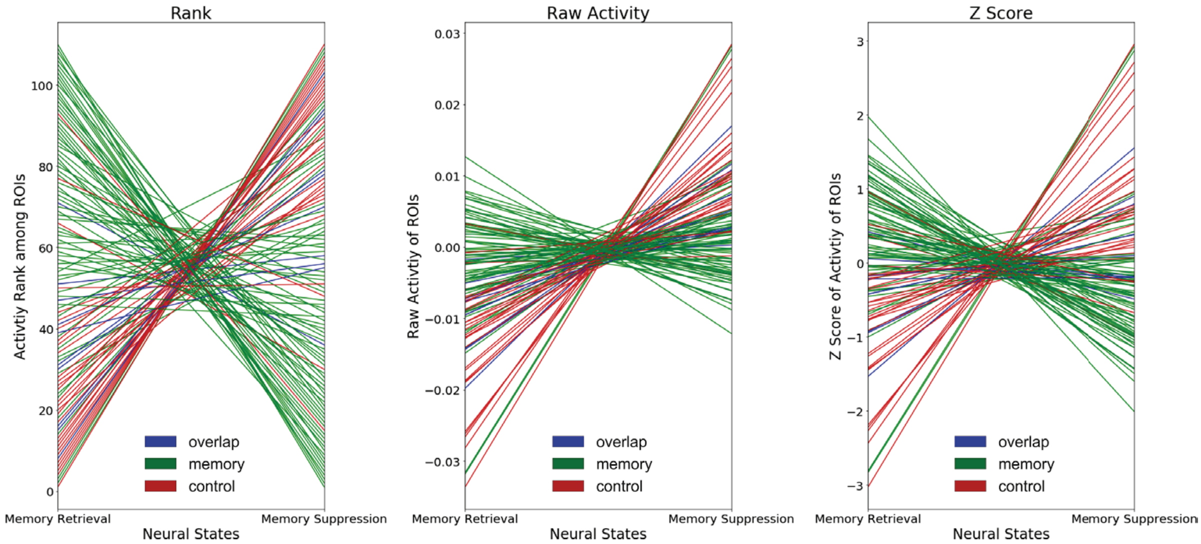


Figure S3.4. Neural state reconfiguration during the Think-to-NoThink transition. (A) Visualized based on the rank of activity intensity among ROIs. **(B)** Visualized based on raw activity intensities of ROIs. **(C)** Visualized based on the Z score of activity intensity of ROIs.

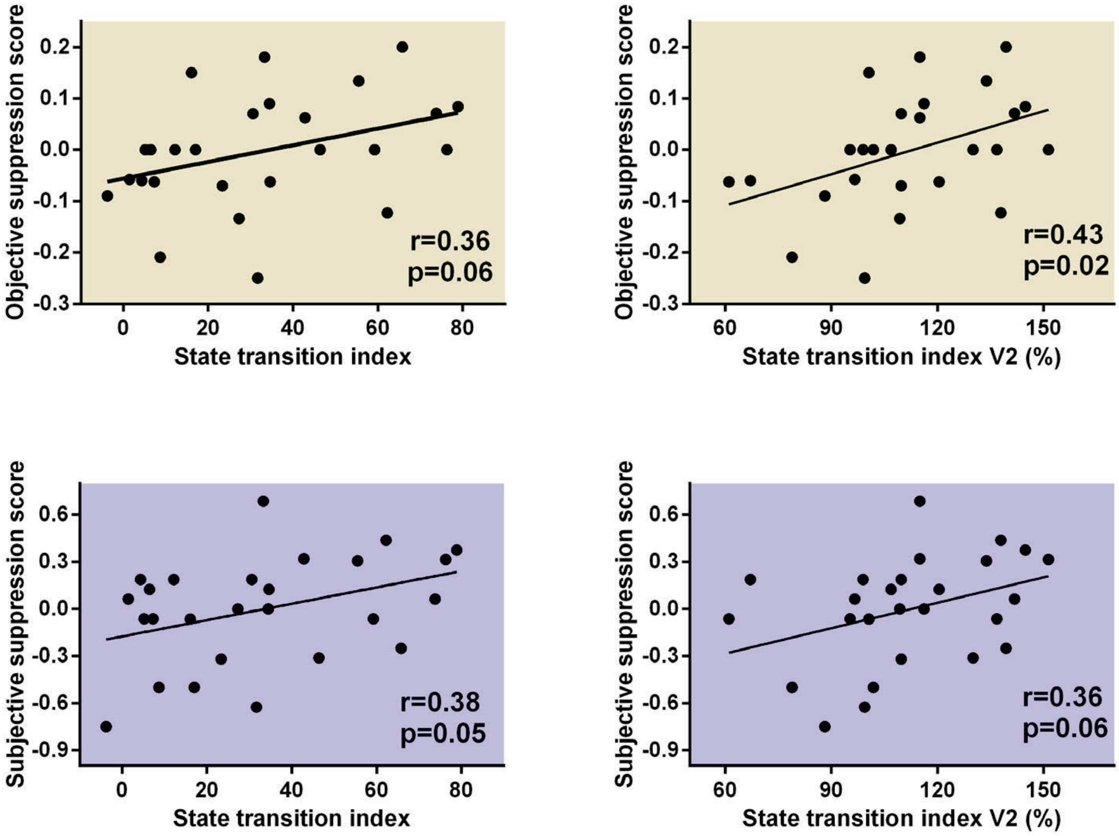


Figure S3.5 Individual differences in two kinds of state transition index are correlated with both objective and subjective suppression scores. The results of the state transition index (i.e., left two panels) were presented in the *main text*. The results of state transition index V2 (i.e., right two panels) were presented in the *Supplemental Text Section3*.

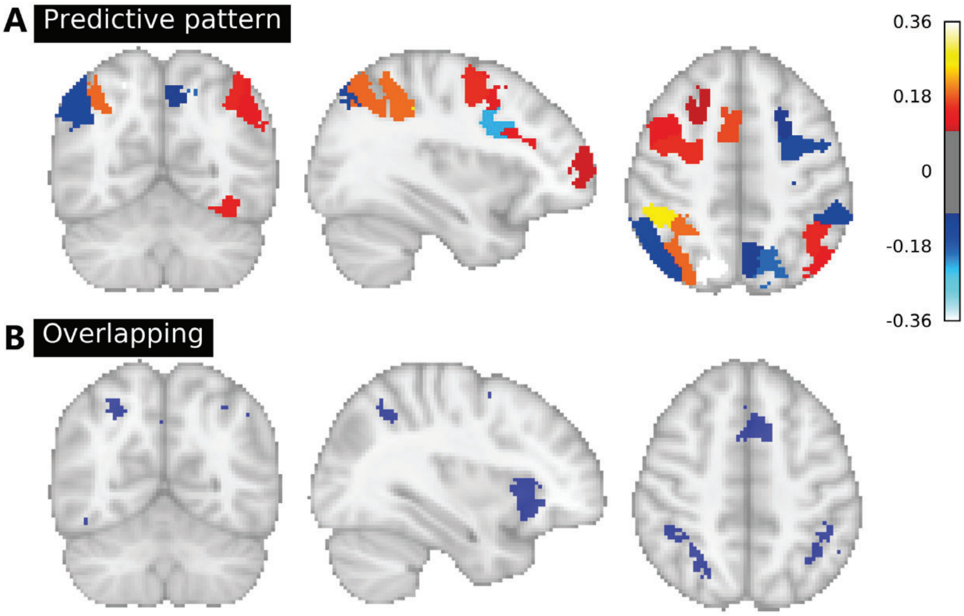


Figure S3.6 Neural state-predictive pattern and overlapping network share similar spatial patterns. (A) The contribution of different brain regions during the decoding. The map was visualized using an arbitrary threshold of 0.10. Higher absolute values of voxelwise statistical results represent a larger contribution during decoding. **(B)** Voxels that belong to both memory retrieval network and inhibitory control network (Neurosynth-based).

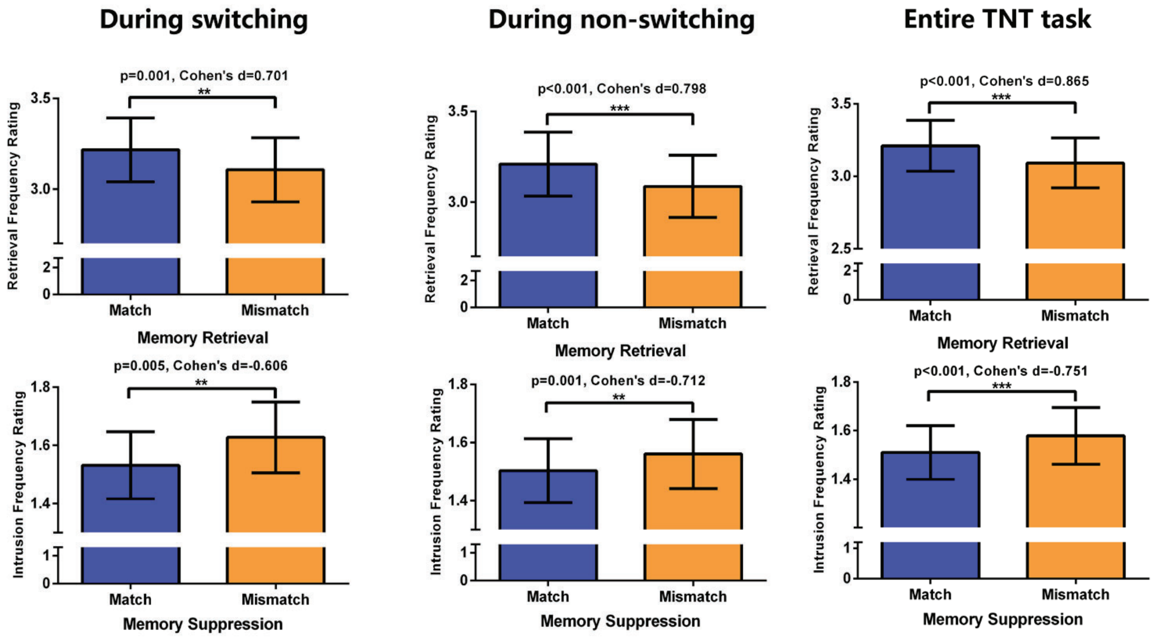


Figure S3.7 Behavioral consequences of the mismatch between neural state and current task demand. When the decoded neural state did not match with the task demand (i.e., Think decoded as No-Think), participants reported worse memory retrieval performance during Think trials. When the neural decoder misclassified No-Think moments as Think, participants reported more memory intrusions during No-Think trials. These two effects can be detected during the *switching* period (i.e., left panel), *non-switching* period (i.e., middle panel), and all time points within the *entire Think/No-Think* task (i.e., right panel)

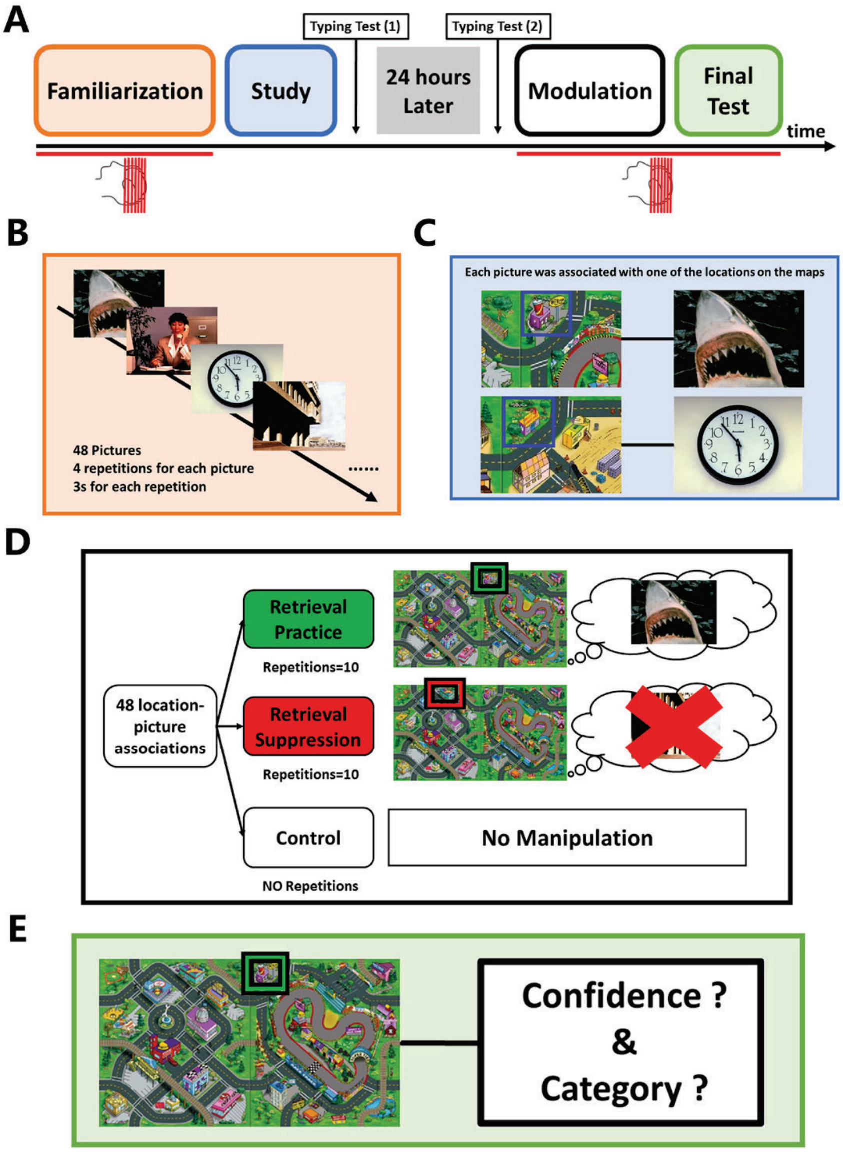
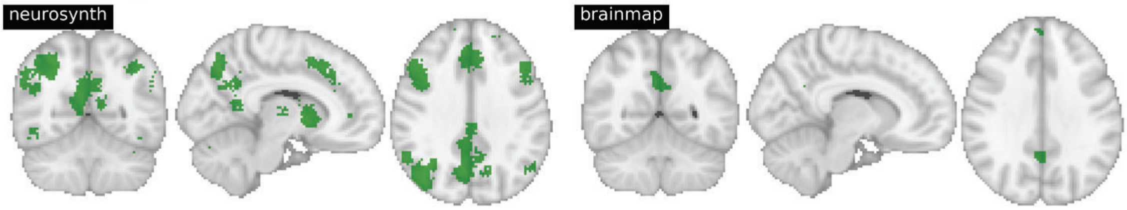


Figure S3.8 Schematic of the experiment design. (A) Timeline of the two-day experimental procedures. Red lines below the timeline indicate the tasks in the MRI scanner. **(B)** During the familiarization phase, all of the pictures of the to-be-remembered associations were randomly presented four times for the familiarization and estimation of picture-specific activation patterns. To keep participants focused, on each trial, they were instructed to categorize the picture shown as an animal, human, location, or object. **(C)** Study phase. Participants were trained to associate memory cues with presented pictures. **(D)** Modulation phase. After 24 hours, we used the Think/No-Think paradigm to modulate consolidated associative memories. Participants were instructed to actively retrieve associated pictures in mind (“retrieval”) or suppress the tendency to recall them (“suppression”) according to the colors of the frames (GREEN: retrieval; RED: suppression) around locations. **(E)** Final memory test phase. Participants performed the final memory test after the modulation. For each of the 48 location-picture associations, locations were presented again, and participants were instructed to report the memory confidence and categorize the picture that came to mind.



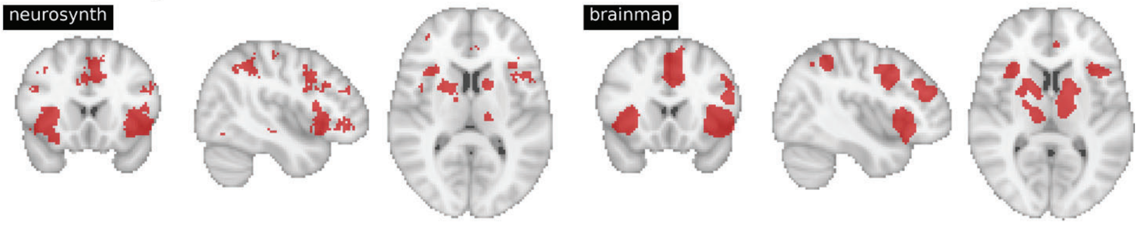


Figure S3.9 Comparison between brain networks generated by the Neurosynth and Brainmap. (A) Memory retrieval network. The network generated by the Neurosynth using the term "*memory retrieval*" (left panel). The network generated by the Brainmap based on activation-based mapping in healthy participants who perform the paradigm of "*episodic recall*" (right panel). **(B)** Inhibitory control network. The network generated by the Neurosynth using the term "*inhibitory control*" (left panel). The network generated by the Brainmap based on activation-based mapping in healthy participants who perform the paradigm of "*go/no-go*" (right panel)

Table S3.1 **Significant activated clusters during No-Think trials compared to Think trials**

Contrast	Brain region	Hemisphere	MNI coordinates	Cluster size	Peak voxel Value
No-Think > Think	IFG/Insula	L	-38 22 8	920	5.5
	IFG/Insula	R	46 18 12	2397	5.5
	DLPFC	R	22 46 24	1681	4.5
	DLPFC	L	-22 44 18	160	4.5
	IPL	R	56 -42 34	1305	5.5
	IPL	L	-58 -54 42	226	4.4
	Thalamus	R	16 -22 2	677	6.6
	Precuneus	R/L	14 -60 56	NA	5.4
	Postcentral gyrus	R	45 -20 55	NA	7.4
	SMA	R/L	8 0 53	NA	5.5
	dACC	R/L	8 20 37	NA	4.4

IFG=Inferior Frontal Gyrus; DLPFC=Dorsolateral Prefrontal Cortex; IPL=Inferior Parietal Lobule; SMA=Supplementary Motor Area; dACC= dorsal Anterior Cingulate Cortex

Table S3.2 Significant activated clusters during Think trials compared to No-Think trials

Contrast	Brain region	Hemisphere	MNI coordinates	Cluster size	Peak voxelValue
Think > No-Think	mPFC	R/L	0 -44 -10	6573	4.03
	Insula	L	-42 -16 18		7.49
	ITG	L	-42 -42 -10	178	4.54
	STG/IPL/Precunues	R	54 -30 10	2925	5.36
	Hippocampus	R	30 -40 4		3.74
	Precunues/PCC/ SMA	L	-4 -40 34	6220	4.85
	Precentral/ Postcentral Gyrus	L	-40 -36 60		7.32
	Posterior Cerebellum	R	22 -60 -46	169	5.30
	Anterior Cerebellum	R	16 -54 -20	1065	7.28

mPFC= medial Prefrontal Cortex; ITG= Inferior Temporal Gyrus; STG= Superior Temporal Gyrus; PCC= Posterior Cingulate Cortex; SMA= Supplementary Motor Area

Table S3.3 Top 10 brain parcels with high classification weights during the neural state prediction

Anatomical Label	X	Y	Z	Hemisphere	Network	Classification Weight
SPL	-34	-48	46	LH	DorsAttn	0.189181458
LOC	-16	-72	54	LH	DorsAttn	-0.358755616
Middle frontal gyrus	-32	-4	52	LH	DorsAttn	0.151391694
dACC	-6	10	40	LH	SalVentAttn	0.169217997
IPL	-34	-66	48	LH	Cont	0.1866069
IPL	-44	-42	46	LH	Cont	0.255550192
Middle frontal gyrus	-42	12	34	LH	Cont	-0.229290806
IFG	-54	20	12	LH	Default	-0.179694589
SPL	16	-72	54	RH	DorsAttn	-0.190727893
Putamen	NA	NA	NA	RH	Subcortical	-0.161437796

DorsAttn=dorsal attention network; SalVentAttn=salience ventral attention network; Cont=frontal parietal control network; Default=default network; SPL=superior parietal lobule; LOC= lateral occipital cortex; dACC= dorsal Anterior Cingulate Cortex; IPL= Inferior Parietal Lobule; IFG=Inferior Frontal Gyrus

Chapter 4

Probing the neural dynamics of mnemonic representations after the initial consolidation

This chapter is based on:
Liu, W., Kohn, N., & Fernandez, G. (2020). Probing the neural dynamics of mnemonic representations after the initial consolidation. **NeuroImage**

Abstract

Memories are not stored as static engrams, but as dynamic representations affected by processes occurring after initial encoding. Previous studies revealed changes in activity and mnemonic representations in visual processing areas, parietal lobe, and hippocampus underlying repeated retrieval and suppression. However, these neural changes are usually induced by memory modulation immediately after memory formation. Here, we investigated 27 healthy participants with a two-day functional Magnetic Resonance Imaging study design to probe how established memories are dynamically modulated by retrieval and suppression 24 hours after learning. Behaviorally, we demonstrated that established memories can still be strengthened by repeated retrieval. By contrast, repeated suppression had a modest negative effect, and suppression-induced forgetting was associated with individual suppression efficacy. Neurally, we demonstrated item-specific pattern reinstatements in visual processing areas, parietal lobe, and hippocampus. Then, we showed that repeated retrieval reduced activity amplitude in the ventral visual cortex and hippocampus, but enhanced the distinctiveness of activity patterns in the ventral visual cortex and parietal lobe. Critically, reduced activity was associated with enhanced representation of idiosyncratic memory traces in the ventral visual cortex and precuneus. In contrast, repeated memory suppression was associated with reduced lateral prefrontal activity, but relative intact mnemonic representations. Our results replicated most of the neural changes induced by memory retrieval and suppression immediately after learning and extended those findings to established memories after initial consolidation. Active retrieval seems to promote episode-unique mnemonic representations in the neocortex after initial encoding but also consolidation.

Keywords: episodic memory, memory retrieval, memory suppression, consolidation, pattern reinstatement

Introduction

Historically, memories were seen as more or less stable traces or engrams. After initial formation, memory traces are affected by consolidation leading to stabilization and weakening, leading to forgetting (Ebbinghaus, 1885; Lashley, 1950; Müller & Pilzecker, 1900). However, contemporary research has provided ample evidence showing that memories continue to be dynamically adapted after initial encoding and, thus, can be modified by external factors throughout their existence. For instance, retrieval practice can reinforce memory traces (Karpicke & Roediger, 2008), promote meaningful learning (Karpicke & Blunt, 2011), and protect memory retrieval against acute stress (A. M. Smith et al., 2016). In contrast, retrieval suppression can prevent unwanted memories to be retrieved (Michael C Anderson & Green, 2001), and reduce their emotional impact (Gagnepain et al., 2017).

Previous neuroimaging studies identified several neural changes that could explain the retrieval-mediated memory enhancement: after repeated retrieval, several studies reported decreased or increased univariate activity in frontal, parietal areas, and temporal gyrus (Eriksson et al., 2011; Gagnepain et al., 2014; Kuhl et al., 2010; Nelson, Arnold, Gilmore, & Mcdermott, 2013; G. van den Broek et al., 2016; G. S. E. van den Broek et al., 2013; M. Wimber et al., 2011; Maria Wimber et al., 2008; Wing et al., 2013; Wirebring et al., 2015). More direct evidence for retrieval-induced changes in mnemonic representations came from studies that applied multivariate pattern analysis. Karlsson Wirebring and colleagues reported that less similar activity patterns in the posterior parietal region across retrieval trials are associated with subsequent better memory (Wirebring et al., 2015). Wimber and colleagues founded that targeted activity patterns are increasingly reinstated over repeated retrieval in visual areas during memory competition (Maria Wimber et al., 2015). Most recently, Ferreira and colleagues reported retrieval-induced generalized and episode-unique representations in parietal areas (Ferreira et al., 2019). Ye and colleagues demonstrated that retrieval practice facilitated the rapid formation of memory representations in the medial prefrontal cortex (mPFC) (Ye et al., 2020). Regarding neural changes underlying suppression-induced forgetting, compelling evidence suggested the role of prefrontal top-down regulation of the hippocampus during suppression (M. C. Anderson, 2004; Michael C Anderson & Hanslmayr, 2014). However, only a few studies investigated neural changes in activity and/or activity patterns across repeated suppression. Depue and colleagues showed the time-specific involvement of inferior frontal gyrus and medial frontal gyrus during the suppression of emotional memory (Brendan E Depue et al., 2007). Gagnepain and colleagues demonstrated the effect of suppression on visual memories may be achieved by targeted cortical inhibition of visual-related activity and activity patterns (Gagnepain et al., 2014).

Although these studies shed light upon neural changes underlying memory retrieval and suppression, all of them were based on memory modulation (i.e., retrieval and suppression) immediately after initial memory formation, except for one study that included repeated retrieval on two consecutive days (Ferreira et al., 2019). How the modulation of memory traces after initial consolidation is reflected in the neural activity and mnemonic representation, as assessed by activation patterns during subsequent retrieval is currently not well understood. Studying the neural changes underlying the modulation of initially consolidated memories can provide complementary and critical understandings of the dynamic nature of human memory. Because newly acquired memories are usually more labile compared to consolidated ones (Frankland & Bontempi, 2005) and mnemonic representations shift from the hippocampus to distributed neocortical regions following overnight sleep (Takashima et al., 2006, 2009), the effectiveness of memory modulation could be decreased, and the underlying neural changes could be different. For example, a study showed that suppression of aversive memories after overnight consolidation is harder, and involved reconfigured neural pathways during suppression (Y. Liu et al., 2016). Also, modulation of consolidated memories may provide a clear focus on the changes of long-term memory representation, because previously reported immediate effects (i.e., changes in activity amplitude and activity patterns) can still be caused by short-term changes in related processes such as executive control or attention. Here, we used a two-day functional Magnetic Resonance Imaging (fMRI) design to characterize neural dynamics of initially consolidated memory. After overnight consolidation, memories were in one condition reinforced by repeated memory retrieval and in the other, weakened by repeated memory suppression. We analyzed the neuroimaging data from both the modulation and the subsequent memory retrieval phase to examine neural changes at the moment when specific memory was modulated and in the final memory test in which the aftereffects of the modulation can be measured.

Based on neural findings of memory reinstatement (Janice Chen et al., 2017; Kosslyn et al., 1997; Kuhl et al., 2010; S.-H. Lee et al., 2019; O'Craven & Kanwisher, 2000; Polyn et al., 2005; Shohamy & Wagner, 2008; Wheeler et al., 2000; Maria Wimber et al., 2015; Xue, 2018), we used both the levels of activity amplitude (i.e., univariate analysis) and activation patterns (i.e., multivariate pattern analysis) of visual area, parietal lobe, and hippocampus to characterize memory traces during memory retrieval and further examined the linear relationship between the two neural changes within the same regions. Furthermore, we adopted a novel design to disentangle perception-related neural activities associated with memory cues presented at the test and retrieval-related neural reactivation associated with reactivated mental images. One method to separate these two processes is to use two perceptual modalities (e.g., sounds as memory cues and pictures as information to be retrieved)(Bosch et al., 2014). Here, we used highly similar visual memory cues across different memory associations. Thus, item-specific neural patterns (at least in visual areas) during retrieval more likely to be caused by retrieval-related memory reactivation instead of visual processing of memory cues.

To sum up, our primary goal is to reveal if two behavioral techniques (i.e., retrieval and suppression) can modulate initial consolidated associative memories, and if such modulation results in altered activity and/or activity patterns detected by fMRI. We first investigated the possibility that associative memories can still be modulated after 24 hours. Behaviorally, we asked whether repeated retrieval and memory suppression would oppositely strengthen or weaken original memory traces. Next, using fMRI, we examined whether retrieval and suppression would modify neural measures of memory reactivation (i.e., activity amplitude and activity pattern similarity) oppositely.

Materials and Methods

Participants

Thirty-two right-handed, healthy young participants aged 18-35 years who were recruited from the Radboud Research Participation System finished two sessions of our experiment. They all had corrected-to-normal or normal vision and reported no history of psychiatric or neurological disease. All of them are native Dutch speakers. Two participants were excluded from further analyses due to memory performance at the chance level. Three additional participants were excluded because of excessive head motion during scanning. We used the motion outlier detection program within the FSL (i.e., FSLMotionOutliers) to detect timepoints with large motion (threshold=0.9). There are at least 20 spikes detected in these excluded participants with the largest displacement ranging from 2.6 to 4.3, while participants included had less than ten spikes. Neuroimaging data of one additional participant was partly used: she was excluded from the analysis of the modulation phase (Think/No-Think paradigm) due to head motion (in total 53 spike, largest displacement=5.7) only during this task, while his/her data during the other tasks were included in the analyses. Thus, data of 27 participants (16 females, age=19-30, mean=23.41, SD=3.30) were included in the analyses of the final test phase, and data of 26 participants (15 females, age=19-30, mean=23.51, SD=3.30) were included in the analyses of the modulation phase. All participants scored within normal levels when applying Dutch-versions of the Beck Depression Inventory (BDI) (Roelofs et al., 2013) and the State-Trait Anxiety Inventory (STAI) (van der Bij et al., 2003). Furthermore, because of the two-session design (24 h' interval), we used an adapted Dutch version of the Pittsburgh sleep quality index (PSQI) (Buysse et al., 1989) to assess the quality of sleep between the two scanning sessions. Questions for last night's sleep were added to the original version. We compared participants' sleep quality/duration for the last night and the average across the previous four weeks. No participants reported abnormal sleep-related behaviors during the night between two fMRI sessions (i.e., more than two hours of differences in sleep time, time to go to bed, or time to wake up between the last night and the previous four weeks). The experiment was approved by, and conducted in accordance with requirements of the local ethics committee (Commissie Mensgebonden Onderzoek region Arnhem-Nijmegen, The Netherlands) and the declaration of Helsinki, including the requirement of written informed consent from each participant before the beginning of the experiment.

Materials

Locations and maps

We used 48 distinctive locations (e.g., buildings, bridges) drawn on two cartoon maps as memory cues. The maps are not corresponding to the layout of any real city in the world, and participants have never been exposed to the maps before the experiment. During the task, the whole map was presented with sequentially highlighting specific locations by colored frames as memory cues. By doing this, we kept visual processes during memory tasks largely consistent.

Pictures

Forty-eight pictures (24 neutral and 24 negative pictures) from the International Affective Picture System (IAPS) (Lang et al., 1997) were used in this study, and these pictures can be categorized into one of four groups: animal (e.g., cat), human (e.g., reading girl), object (e.g., clock) or scene (e.g., train station). Category information was used for the following memory-based category judgment test. All images were converted to the same size and resolution for the experiment.

Picture-location associations

Each picture was paired with one of the 48 map locations to form specific picture-location associations. We (W.L and J.V) carefully screened all the associations to prevent the explicit semantic relationship between picture and location (e.g., lighter at the- fire department). All 48 picture-location associations were divided into three groups for different types of modulation (See Modulation Phase). For each map, 24 locations were paired 6 pictures from each category. One-third of associations (8 associations; 2 pictures from each category) on that map were retrieval associations (i.e. "think" associations), one-third of associations were suppression associations (i.e., "no-think" associations), and remaining one-third are control associations.

Experiment design

Overview of the design

This study is a two-session fMRI experiment, with the 24 hours interval between two sessions (**Figure 4.1A**). Day1 session consists of the familiarization phase (**Figure 4.1B**), the study phase (**Figure 4.1C**), and the immediate typing test. The Day2 session consists of the second typing test, the modulation phase (**Figure 4.1D**), and the final memory test (**Figure 4.1E**). Among these phases, the familiarization, modulation, and the final memory test phase were performed in the scanner, while the study phase and two typing tests were performed in the behavioral lab. The trial structure and timing are depicted in **Figure S4.1**. Stimuli were presented while participants were scanned projecting on to a translucent screen (diameter=598mm; maximum projection size=369 × 277 mm) mounted at the end of the scanner's bore and visible via a mirror mounted at the head coil and during behavioral sessions using a 24-inch LED monitor. During the MRI scanning, the distance between the visual surface mirror and the projection screen was around 85.5cm. Moreover, to keep the visual presentation as consistent as possible, we set the resolution as at 1280x1024 for both set-ups.

Familiarization phase

To obtain the picture-specific brain responses to all 48 pictures, we instructed participants to perform the familiarization phase while being scanned (**Figure 4.1B**). The second purpose of the task is to let participants become familiar with the pictures to be associated with locations later. Each picture (resolution=400 x 400) was shown four times at the center of the screen with a visual angle of 7 degrees for 3s and was distributed over in total of four functional runs. The order of the presentation was pseudorandom and pre-generated by self-programmed Python code. The dependencies between the orders of different runs were minimized to prevent potential sequence-based memory encoding. To keep participants focused during the task, we instructed them to categorize the presented picture via the multiple-choice question with four options (animal, human, object, and scene). We used an exponential inter-trial intervals (ITI) model (mean=2s, minimum=1s, maximum=4s) to generate the ITIs between trials. Participants' responses were recorded by an MRI-compatible response box.

Study phase

Each picture-location association was presented twice in two separate runs (**Figure 4.1C**). During each study trial, the entire map (resolution=1024 x 768) was first presented for 2.5s, then a BLUE frame was added to a layer on the top of the entire map to highlight one of the 48 locations, for 3s, and finally, the picture and its associated location were presented side-by-side together for 6s. We pre-generated a pseudorandom order of the trials to minimize the similarity between the orders in familiarization and the study phase.

Typing test phase

Immediately after the study phase, participants performed a typing test (day1) assessing picture-location association memory. Each location was presented again (4s) in an order that differed from the study phase, and participants had maximally 60s to describe the associated picture by typing its name/description on a standard keyboard. Twenty-four hours later (day2), participants performed the typing test again in the same behavioral lab. The procedure was identical to the immediate typing test, but with a different trial order.

Modulation phase

The modulation phase is the first task participants performed during the Day2 MRI session. We used the think/no-think (TNT) paradigm with trial-by-trial self-report measures to modulate initially consolidated memories (**Figure 4.1D**). The same paradigm has been used in previous neuroimaging studies, and the self-report does not affect the underlying memory control process (M. C. Anderson, 2004; B. J. Levy & Anderson, 2012). Forty-eight picture-location associations were divided into three conditions. One-third of the associations (16 associations) were assigned to the retrieval condition ("Think"), one-third of the associations were assigned to the suppression condition ("No-Think"), and the remaining one-third of the associations were assigned to the control condition. The assignment process was counterbalanced between participants. Therefore, at the group level, for each picture-location association, the possibility of belonging to one of the three modulation conditions is around 33.3%. Associations that belong to different conditions underwent different types of modulation during this phase. Locations which belong to the control condition were not presented during this phase. For a retrieval trial, the entire map was presented (visual angle=18 degrees) with one particular location, highlighted with a GREEN frame for 3s, and participants were instructed to recall the associated picture quickly and actively and to keep it in mind until the map disappeared from the screen. For a suppression trial, one location was highlighted with a RED frame for 3s, and participants were instructed that "*when you see a location, highlighted with a RED frame, you should NOT think about the associated picture. Instead, you should try to keep an empty mind during this stage. It is a difficult task, and it is totally fine that sometimes you still think about the associated picture. But please do NOT close your eyes, focus on something outside the screen, or think about something else in your life. These strategies, although useful, could negatively affect the brain activity that we are interested in ……*" After each retrieval or suppression trial, participants had up to a maximum of 3s to report their experience during the cue presentation. Specifically, they answered a multiple-choice question with four response options (*Never, Sometimes, Often, and Always*) by pressing the button on the response box to indicate whether the associated picture entered their mind during that particular trial or not and the relative frequency.

The modulation phase consisted of five functional runs (64 trials per run). In each run, 32 locations (half retrieval trials, and half suppression trials) were presented twice. Therefore, each memory cue that did not belong to the control condition was presented ten times during the entire modulation phase. Again, we pre-generated the presentation orders to prevent similar order sequences across five modulation runs. Between each trial, fixation was presented for 1-4s (mean=2s, exponential model) as ITI.

The final memory test phase

After the modulation phase, participants performed the final memory test within the scanner (**Figure 4.1E**). All 48 locations (including both the retrieval/suppression associations as well as control associations) were highlighted one-by-one while showing the entire map again with a BLUE frame. During its presentation (4s), participants were instructed to recall the associated picture covertly but as vividly as possible and keep the mental image in their mind. Critically, visual input during this phase was highly similar across trials because entire maps were always presented, just with different locations highlighted. Next, participants were asked to give the responses on two multiple-choice questions within 7s (3.5s for each question): (1) “how confident are you about the retrieval?” They responded with one of the four following response options: Cannot recall, low confidence, middle confidence, and high confidence. (2) “Please indicate the category of the picture you were recalling?” They also had four options to choose from (Animal, Human, Object, and Scene).

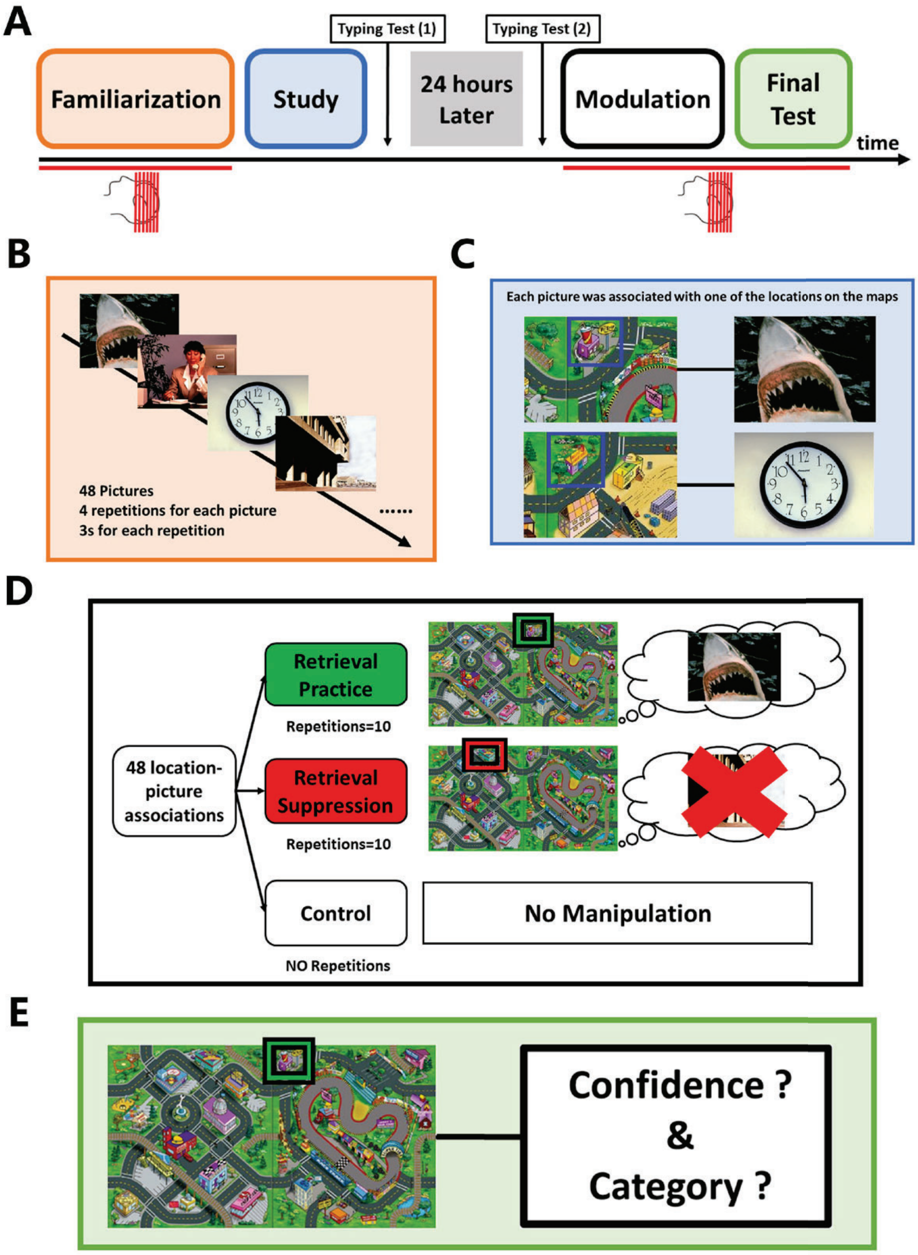


Figure 4.1 **Schematic of the experiment design. (A)** Timeline of the two-day experimental procedures. Red lines below the timeline indicate the tasks in the MRI scanner. The trial structure with exact timing was depicted in *Figure S4.1*. **(B)** During the familiarization phase, all of the pictures of the to-be-remembered associations were randomly presented four times for the familiarization and estimation of picture-specific activation patterns. To keep participants focused, on each trial, they were instructed to categorize the picture shown as an animal, human, location, or object. **(C)** Study phase. Participants were trained to associate memory cues with presented pictures. **(D)** Modulation phase. After 24 hours, we used the Think/No-Think paradigm to modulate consolidated associative memories. Participants were instructed to actively retrieve associated pictures in mind ("retrieval"), or suppress the tendency to recall them ("suppression") according to the colors of the frames (GREEN: retrieval; RED: suppression) around locations. **(E)** Final memory test phase. Participants performed the final memory test after the modulation. For each of the 48 location-picture associations, locations were presented again, and participants were instructed to report the memory confidence and categorize the picture that came to mind.

Behavioral data analysis

Familiarization phase

We did not calculate the accuracy of the category judgment during the familiarization phase because the categorization of a picture could be a rather subjective decision, and it is not relevant for the aim of this study. However, we used individual responses to control for subjective category categorization for the following memory performance evaluation. Specifically, if a participant consistently labeled a given picture across four repetitions as a different category compared to our predefined labels, we generated an individual-specific category label and used this category label for this picture to evaluate the responses in the final test. Otherwise, we used predefined labels to evaluate the responses.

Typing test

Participants' answers were evaluated by two native Dutch experimenters (S.M and J.V) independently. The general principle is that if the answer contains enough specific information (e.g., a little black cat), to allow the experimenter to identify the picture from the 48 pictures used, it was labeled as correct. In contrast, if the answer is not specific enough (e.g., a small animal), then it was labeled as incorrect. We used Cohen's kappa coefficient (κ) to measure inter-rater reliability. In general, κ lager than 0.81 suggests almost perfect reliability. If two accessors had different evaluations, the third accessor (W.L) determined the final result (i.e., correct or incorrect). After the immediate typing test, we only invited participants with at least 50% accuracy to the Day2 experiment. Three out of 35 recruited participants did not continue on Day2 due to low performance on Day1. For the typing test 24 hours later, participants' responses were evaluated by the same experimenters again. Based on the participants' responses in this typing test, we identified picture-location associations that the given participant did not learn or already forgot. These associations were not considered in the following behavioral and neuroimaging analyses, because participants have no memory associations to be modulated. We calculated the average accuracies for the immediate typing test and typing test 24 hours later and investigated the delay-related decline in memory performance using a paired t-test.

Modulation phase

Responses during the modulation phase were analyzed separately for retrieval trials and suppression trials. We first calculated the percentage of each option (never, sometimes, often, and always) chosen across 160 retrieval trials and 160 suppression trials for each participant. Next, we quantified the dynamic changes in task performance across repetitions (runs). Before the following analyses, we coded the original categorical variable using numbers (Never-1; Sometimes-2; Often-3; Always-4). For all the established picture-location associations, we calculated their average retrieval frequency rating (based on retrieval trials) and intrusion frequency rating (based on suppression trials) on each repetition. We used a repeated-measures

ANOVA to model changes in retrieval and intrusion frequencies rating across repetitions to test if the repeated attempt to retrieve or suppress a memory trace would strengthen or weaken the associations, respectively. Additionally, to quantify individual differences in memory suppression efficiency (B. J. Levy & Anderson, 2012), we calculated the *intrusion slope score* for each participant. Using all the intrusion rating for suppression trials, we used linear regression to calculate the slope of intrusion ratings across the ten repetitions for each participant. Participants with more negative slope scores are better at downregulating memory intrusions than those with less negative slope scores.

The final memory test phase

For each trial of the final memory test, we calculated both a subjective memory measure based on the confidence rating (1,2,3,4) and an objective memory measure based on the category judgment (correct/incorrect). Also, we recorded the reaction times (RT) for category judgments to estimate the speed of memory retrieval. To investigate the effect of types of modulation on the subjective, objective memory, and retrieval speed, we performed a repeated-measure ANOVA to detect within-participants' differences between *RETRIEVAL ASSOCIATIONS, SUPPRESSION ASSOCIATIONS*, and *CONTROL ASSOCIATIONS*. To assess individual differences in suppression-induced forgetting, we calculated the *suppression score* by subtracting the objective memory measure of retrieval suppression associations ("no-think" items) from the control association. Participants showed more forgetting as the result of suppression had more negative suppression scores.

Combinatory analysis of modulation and final test phase

To replicate the relationship between memory suppression efficiency during the TNT task and suppression-induced forgetting during the final memory test reported before (B. J. Levy & Anderson, 2012), we correlated suppression scores with intrusion slope scores across all participants. Notably, sample size (N=26) of this cross-participant correlational analysis is modest, but it is just a replication analysis of the previous study and the check for the memory suppression manipulation.

fMRI data acquisition and pre-processing

Acquisition

MRI data were acquired using a 3.0 T Siemens PrismaFit scanner (Siemens Medical, Erlangen, Germany) and a 32 channel head coil system at the Donders Institute, Centre for Cognitive Neuroimaging in Nijmegen, the Netherlands. For each participant, MRI data were acquired in two MRI sessions (around 1 h for each session) with 24 h' interval. We used three types of sequences in this study: (1) a 3D magnetization-prepared rapid gradient echo (MPRAGE) anatomical T1-weighted sequence with the following parameters: 1 mm isotropic, TE = 3.03

ms, TR = 2300 ms, flip angle = 8 deg, FOV = 256 × 256 × 256 mm; (2) Echo-planar imaging (EPI)-based multi-band sequence (acceleration factor=4) with the following parameters: 68 slices (multi-slice mode, interleaved), voxel size 2 mm isotropic, TR = 1500 ms, TE = 39 ms, flip angle =75 deg, FOV = 210 × 210 × 210 mm; (3) field map sequence (i.e. magnitude and phase images) were collected to correct for distortions (voxel size of 2 × 2 × 2 mm, TR = 1,020 ms, TE = 12 ms, flip angle = 90 deg).

During the day1 session, anatomical T1 image was acquired firstly, followed by the field map sequence. Before the four EPI-based pattern localization runs, 8 minutes of resting-state data were acquired from each participant using the same sequence parameters. Day2 session began with the field map sequence. Thereafter, we acquired six EPI-based task-fMRI runs (five runs of the modulation phase and one run of the final test phase).

Preprocessing of neuroimaging data

All functional runs underwent the same preprocessing steps using FEAT (FMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl)(Jenkinson et al., 2012). In general, the pipeline was based on procedures suggested by Mumford and colleagues (http://mumfordbrainstats.tumblr.com) and the suggestions for Automatic Removal of Motion Artifacts (ICA-AROMA) (Pruim et al., 2015). The first four volumes of each run were removed from the 4D sequences for scanner stabilization. The following preprocessing was applied; Motion correction using MCFLIRT (Jenkinson et al., 2002); field inhomogeneities were corrected using B0 Unwarping in FEAT; non-brain removal using BET (S. M. Smith, 2002); grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor. We used different spatial smoothing strategies based on the type of analysis. For data used in univariate analyses, we applied a 6mm kernel. In contrast, for data used in multivariate pattern analyses, no spatial smoothing was performed to keep the voxel-wise pattern information. In addition to the default FSL motion correction algorithm, we used ICA-AROMA to further remove the motion-related spurious noise and chose the results from the "non-aggressive denoising" algorithm for the following analyses. Prior to time-series statistical analyses, highpass temporal filtering (Gaussian-weighted least-squares straight-line fitting with sigma=50.0s) was applied.

Registration between all functional data, high-resolution structural data, and standard space was performed using the following steps. First, we used the Boundary Based Registration (BBR) (Greve & Fischl, 2009) to register functional data to the participant's high-resolution structural image. Next, registration of high resolution structural to standard space was carried out using FLIRT (Jenkinson et al., 2002; Jenkinson & Smith, 2001) and was then further refined using FNIRT nonlinear registration (Andersson et al., 2007). Resulting parameters were used to align maps between native-space and standard space and back-projected region-of-interests into native space.

Anatomical Region-of-Interest (ROI) in fMRI analyses

Based on previous pattern reinstatement studies (Jonker et al., 2018; H. Lee et al., 2017; S.-H. Lee et al., 2019; Polyn et al., 2005; Maria Wimber et al., 2015), we hypothesized that ventral visual cortex (VVC), parietal lobe and hippocampus might carry picture-specific and category-specific information of the memory contents during retrieval. Therefore, we chose them as the ROIs in our fMRI analyses. All ROIs were first defined in the common space and back-projected into the participant's native space for within-participant analyses using parameters obtained from FSL during registration.

We defined anatomical VVC ROI based on the Automated Anatomical Labeling (AAL) human atlas, which is implemented in the WFU pickatlas software (http://fmri.wfubmc.edu/software/PickAtlas). The procedure was used before in a previous neural reactivation study conducted by Wimber and colleagues (Maria Wimber et al., 2015). Brain regions, including bilateral inferior occipital lobe, parahippocampal gyrus, fusiform gyrus, and lingual gyrus were extracted from the AAL atlas and combined to the VVC mask. The VVC mask was mainly used as the boundary to locate visual-related voxels in the following activity pattern analyses.

The ROIs of the hippocampus and parietal lobe (including angular gyrus (AG), supramarginal gyrus (SMG), and precuneus) were defined using a bilateral mask within the AAL provided by WFU pickatlas software. To yield better coverage of participants' anatomy, we extended the original mask by two voxels in each direction (i.e., dilated by a factor of 2 in the software).

Univariate Generalized Linear Model (GLM) analyses of response amplitude

GLM analyses of neuroimaging data from the final test phase

To investigate how different modulations (retrieval/suppression) affect the subsequent univariate activation, we ran voxel-wise GLM analyses of the final test run. All time-series statistical analysis was carried out using FILM with local autocorrelation correction (Woolrich et al., 2001) using FEAT. In total, six regressors were included in the model. We modeled the presentation of memory cues (locations) as three kinds of regressors (duration=4s) based on their modulation history (retrieval, suppression, or control). To account for the effect of unsuccessful memory retrieval, we separately modeled the location-picture associations that participants could not recall as a separate regressor. Lastly, button presses were modeled as two independent regressors (confidence and category judgment). All trials were convolved with the default hemodynamic response function (HRF) within the FSL.

We conducted two planned contrasts (retrieval vs. control and suppression vs. control) first at the native space and then aligned, resulting in statistical maps to MNI space using the parameters from the registration. These aligned maps were used for the group-level analyses and corrected for multiple comparisons using default cluster-level correction within FEAT (voxelwise Z>3.1, cluster-level $p < .05$ FWER corrected). All of the contrasts were first conducted at the whole-brain level. Then, for the ROI analyses, we extracted beta values of these ROIs from the final test and compared them for the same contrasts (retrieval vs. control and suppression vs. control).

GLM analyses of neuroimaging data from the modulation phase

We ran the voxel-wise GLM analyses for each modulation run separately. In total, three regressors were included in the model. We modeled the presentation of the memory cues (location) as two kinds of regressors (duration=3s) according to their modulation instruction (retrieval or suppression). Button press was modeled as one independent regressor. Also, if applicable, location-picture associations that our participants could not recall were modeled as a regressor. For ROI analyses, we extracted beta values of these ROIs from whole-brain maps of each modulation run separately. We investigated repetition-related changes in beta values using the Repeated ANOVA for retrieval and suppression separately.

Multivariate pattern analyses of brain activation patterns

Activity pattern estimation

All preprocessed (unsmoothed) familiarization, modulation, and final test functional runs were modeled in separate GLMs in each participant's native space. For each trial within familiarization, we generated a separate regressor using the onset of picture presentation and 3s as the duration. At the same time, we generated one regressor for all button presses of the category judgment to control for the motor-related brain activity. In total, 49 regressors were included in the model. This procedure led to a separate statistical map (t-values) for each trial. Similarly, for each modulation and final test run, we generated a separate regressor using the onset of the presentation of location (memory cue) and 3s as the duration. However, button presses were not included in the model because they may potentially carry ongoing memory-related information. Also, we got a separate t map for each modulation or test trial.

Searchlight analysis of picture-sensitive voxels

For each participant, brain data on the familiarization phase (i.e., pattern localization phase) was analyzed using the searchlight method (Kriegeskorte et al., 2006, 2008) across the entire brain. More specifically, for each searchlight (centered at every voxel in the brain, a sphere with the radius of 5mm) of each participant, we trained Support Vector Classification (SVC) classifier to differentiate the activity patterns elicited by each picture (or each category) and tested its predictive power using the leave-one-run-out cross-validation. SVC was implemented using the C-Support Vector Machine within the scikit-learn package (https://scikit-learn.org/stable/) (Pedregosa et al., 2011). The multiclass classification was handled according to a one-vs.-one scheme. We used default parameters of the function (regularization (C)=1, radial basis function kernel with degree=3). The same setting was applied for all classification described below. Specifically, for each trial, activity patterns within the searchlight were extracted. Since each picture was presented four times during four pattern localization runs, in total, we got four activity patterns within the searchlight for each picture. The *within-participant classification* was performed using the leave-one-run-out cross-validation: activity patterns of one particular run were left out as the testing dataset, and the remaining three runs were used as the training dataset to train the SVC classifier. After all the training-testing procedures, our analyses resulted in one accuracy value to represent the overall predictive power of the activity patterns within this particular searchlight. The searchlight walked through the entire brain of each participant. After the searchlight procedure, each participant yielded a classification accuracy map, and each voxel within the map stored the classification accuracy of that particular searchlight sphere. To allow the group inferences of the brain regions, we performed one-sample t-tests on all of the classification accuracy maps and tested them against chance (chance level=1/48, 2%). Since we would like to identify picture-sensitive voxels within the VVC, we overlapped the voxels identified by the searchlight ($p_{uncorrected}<0.001$) with the anatomical VVC mask. Because choosing the $p_{uncorrected}<0.001$ as the threshold is arbitrary, we also used other thresholds ($p_{uncorrected}<0.05$ and $p_{uncorrected}<0.01$) to define the significant voxels and further validated our results using different threshold-dependent masks.

We already used the *within-participant* searchlight analysis to localize stimuli-sensitive voxels in visual areas. We validated these identified VVC voxels in a *cross-participant* procedure. By doing this, we explored whether visual perception-related activation patterns of these voxels are shared across participants. Specifically, instead of performing the leave-one-run-out cross-validation within each participant, we used the three-fold cross-validation within the entire sample. Firstly, t maps for each picture, and each run were transformed from native space to standard space to enable the cross-participant predictive model training and testing. Then, the identified voxels within the VVC were used as a mask to extract spatial patterns of activation. Finally, data from 2/3 participants was used to train the SVC model, and the remaining 1/3 participants were used to assess the model. It is notable that cross-participant classification is just the confirmatory analysis of the searchlight classification and should not be regarded as independent analysis. The cross-participant classification was also repeated in three clusters of VVC voxels under different thresholds ($p_{uncorrected}<0.05$, $p_{uncorrected}<0.01$, and $p_{uncorrected}<0.001$).

Pattern reinstatement analysis

The VVC voxels identified by searchlight analysis and other anatomical-defined masks (including hippocampus, AG, SMG, and precuneus) were used as the mask in the cross-task classification of memory contents. For each trial's t-map estimated based on the final test run, we transformed it from native space to standard space. ROI-based activity patterns from both the pattern localization and final memory test phase were extracted using ROI masks. We performed cross-task three-fold cross-validation to reveal the shared neural representation of the perception and retrieval of the same visual stimulus. Activity patterns estimated based on the pattern localization of the 2/3 participants (i.e., training sample) were used to train the SVC predictive model. We used the activity pattern during the final memory test evoked by the corresponding location (memory cure) of the remaining 1/3 participants (i.e., testing sample), together with the trained SVC model to predict the memory content on a trial-by-trial basis. Critically, the SVC model was trained solely on the localizer data (day1), and it was applied to the final memory test (day2) without further model fitting. Moreover, during the final memory test, visual input is highly similar across trials because we just highlighted each location on an identical map as the memory cue. Therefore, if a given classifier can significantly predict memory content, the classification is unlikely based on the neural responses to the memory cue only. For each ROI, we first calculated the average decoding accuracy for each participant across all trials. A common way to evaluate the significance of classification accuracies is to compare them with theoretical chance level (i.e., 1/number of categories). However, previous work has shown that this approach may overestimate the classification significance (Combrisson & Jerbi, 2015; Jamalabadi et al., 2016; Kowalczyk & Chapelle, 2005). We used an alternative method to control for this potential bias. For each decoding analysis, we generated an empirical null distribution of accuracies by repeating our decoding analyses with classifiers training on randomly shuffled labels (N=1000). Only accuracies whose values are larger than the 95th percentile of this null distribution were considered significant. Values that were larger than the maximum accuracy within this null distribution were assigned a p-value of <0.001.

ROI-based trial-by-trial pattern similarity analysis on the modulation and final memory test data

Representation similarity analysis (RSA) (J. D. Cohen, Daw, Engelhardt, Hasson, Li, Niv, Norman, Pillow, Ramadge, Turk-Browne, & others, 2017) was used to calculate trial-by-trial pattern similarity within particular types of test trials (e.g., recall of associations belongs to the *RETRIEVAL ASSOCIATIONS*). Given the nature of the within-participant analysis and to improve the pattern similarity estimation, we based all calculations on activity patterns in the native space. Firstly, we analyzed the multivariate activation patterns of the final test. The identified VVC voxels (**Figure 4.2A**) were transformed from standard space to native space and then used as a mask to extract 3D single-trial activity patterns to 2D vectors and z-scored for the latter correlational analysis. Activation patterns of the hippocampus (**Figure 4.2B**), angular gyrus (**Figure 4.2C**), supramarginal gyrus (**Figure 4.2D**), and precuneus (**Figure 4.2E**) were extracted in the same way. For each participant, after excluding all trials with incorrect memory-based category judgment, we divided the remaining trials into three conditions based on their modulation history (e.g., retrieval practice or retrieval suppression). Next, for activity patterns of trials within the same condition, we calculated neural pattern similarity using Pearson correlations between all possible pairs of trials within the condition (**Figure 4.2F**). The calculations led for each participant to three separate correlation matrices, one for each type of test trials for each participant. Finally, we used the mean value of all of the r-values located at the left-triangle of one participant-specific correlation matrix to represent the neural pattern similarity of that condition (the higher the r-value, the lower the pattern similarity). After repeating these steps for each participant separately, three kinds of pattern similarity values were generated for the statistical test. All mean r-values were Fisher-r-to-z transformed before the following statistical analyses. To investigate if different modulations have different effects on memory representation during the final test, we performed two planned within-participant comparisons: (1) *RETRIEVAL ASSOCIATIONS* vs. *CONTROL ASSOCIATIONS;* (2) *SUPPRESSION ASSOCIATIONS* vs. *CONTROL ASSOCIATIONS.*

Next, we used the same approach to analyze the modulation data. For each presented location, activity patterns were extracted using the same mask from five modulation runs. Similarly, within-condition (retrieval or suppression) trial-by-trial pattern similarity was calculated for each condition and each run. The dynamic change was modeled using the condition by run interaction using the ANOVA analysis.

Statistical analysis

When comparing continuous variables (e.g., reaction time) between experimental conditions, we used repeated Analysis of variance (ANOVA) or paired t-test. A significant main effect in an ANOVA was followed by post hoc tests, in which multiple comparisons were corrected by the Holm–Bonferroni method. Notably, classification accuracies were not normally distributed. Therefore, we used non-parametric methods (i.e., *Friedman Test*) to compare accuracies between experimental conditions. To evaluate the significances of classification accuracy, instead of comparing with theoretical chance levels, we compared real accuracies with an empirical null distribution of accuracies (*See Pattern reinstatement analysis above*). Accuracies were considered significant when they were at least higher than the 95th percentile of the corresponding null distribution. For ordinal responses (e.g., “never,” “sometimes”), the percentage of each option was calculated, and then percentages were compared across repetitions. To account for the number of comparisons that come with multiple ROIs (n=9), we applied False Discovery Rate correction based on the Benjamini-Hochberg procedure (Thissen et al., 2002). For all statistical tests that involved multiple ROIs, FDR-corrected p values (p_{FDR}) are reported along with raw p values (p_{raw}) and effect sizes (e.g., Cohen’s d, partial η^2).

Data and code availability.

Custom scripts used in this study, immediate data (i.e., preprocessed single-trial activation patterns used for reinstatement analyses) as well as raw data were uploaded to the Donders

Repository (https://data.donders.ru.nl/). The project was named as *Tracking the involuntary retrieval of unwanted memory in the human brain with functional MRI* in the Repository (https://doi.org/10.34973/5afg-7r41).

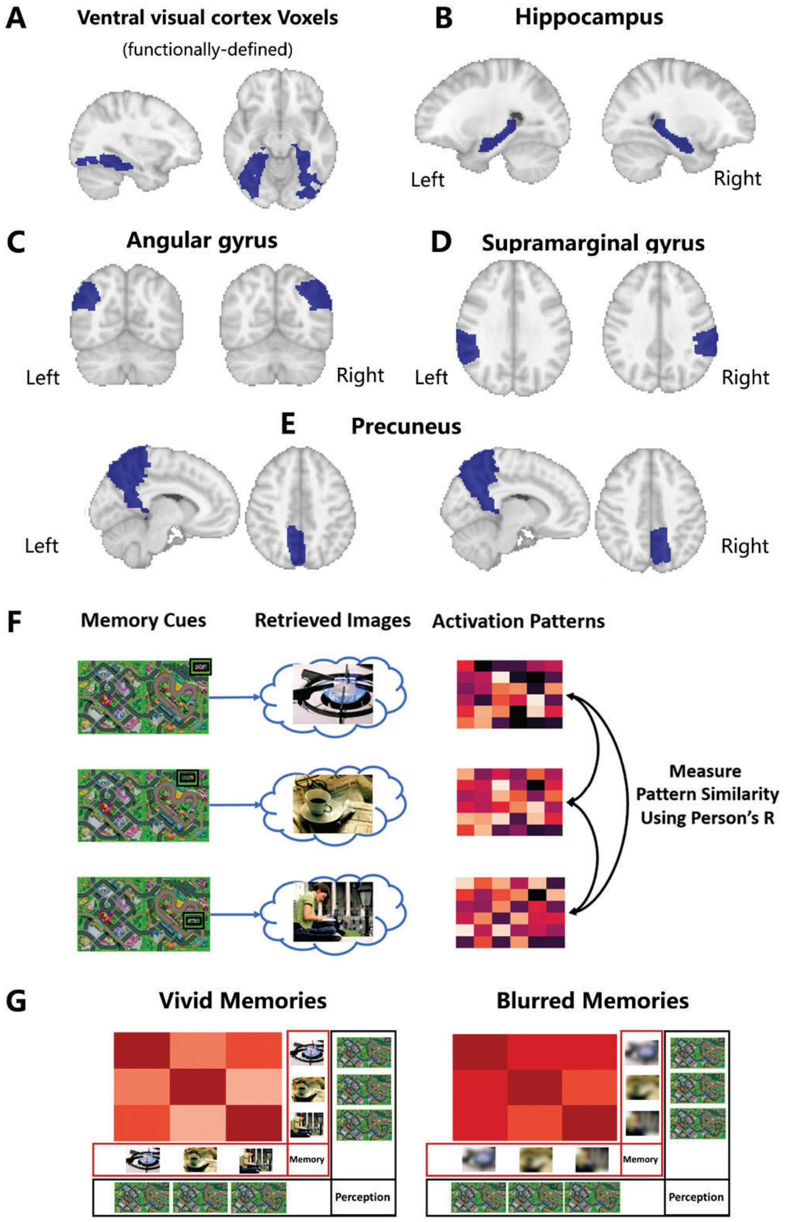


Figure 4.2 Regions-of-interest (ROI) and rationale of the pattern similarity analysis. (A) Functionally-defined voxels within the ventral visual cortex (VVC). We identified voxels whose activity patterns can be used to differentiate pictures that were processed during the familiarization phase and were reactivated during successful memory retrieval during the final test. **(B)** Anatomically-defined bilateral hippocampus ROI. **(C)** Anatomically-defined bilateral angular gyrus ROI. **(D)** Anatomically-defined bilateral supramarginal gyrus ROI. **(E)** Anatomically-defined bilateral precuneus ROI. **(F)** During the final test, "mental images" were retrieved based on highly similar memory cues (different locations within maps were cued). We derived activation patterns for each memory retrieval trials based on fMRI data, and then quantify the cross-item pattern similarity using Pearson's r. **(G)** Considering the highly similar perceptional processing, vivid "mental images" during memory retrieval should be reflected in lower activity pattern similarity.

Results

Behavioral results

Pre-scan memory performance immediately after study and 24 hours later

During the immediate typing test (day1), 88.01% of the associated pictures were described correctly (SD= 10.87%; range from 52% to 100%). Twenty-four hours later, participants still recalled 82.15% of all associations in the second typing test (SD = 13.87%; range from 50% to 100%). Although we observed less accurate memory 24 hours later ($t(26) = 4.73$, $p < 0.001$, Cohen's $d = 0.912$) (**Figure S4.2**), participants could still remember most location-picture associations well.

Behavioral performance during the modulation phase

During retrieval trials, participants reported that most associated pictures were successfully recalled ($1-P_{never}$: mean=84.05%, SD=11.79 %, range from 56.25% to 100%; **Figure 4.3A**). This number is close to the accuracy of the second typing test immediately before the modulation phase. Critically, we observed that with repeated attempts to retrieve, the percentage of the four types of trial-by-trial retrieval frequency ratings changed differently over repetitions (Choice×Repetition: $F[27,702]=3.4$, $p<0.001$, $\eta^2 = 0.01$; **Figure 4.3B**). More precisely, the percentage of reporting "always" increased ($F[9,234]=5.3$, $p<0.001$, $\eta^2 = 0.02$), while the percentage of reporting "often" ($F[9,234]=2.1$, $p=0.02$, $\eta^2 = 0.01$) and "sometimes" decreased ($F[9,234]=2.0$, $p=0.03$, $\eta^2 = 0.02$).

For the analyses of suppression trials, we excluded all location-picture associations which the participant could not describe correctly immediately before the modulation phase (i.e., Typing Test Day2). This approach controlled for individual differences in memory that could interfere with the analysis of memory suppression. On suppression trials, participants reported that they successfully suppressed the tendency to recall the associated pictures in about half of the trials (P_{never}: mean=50.62%, SD=25.35%, range from 4% to 92.5%; **Figure 4.3C**). As shown before in the think/no-think literature before (B. J. Levy & Anderson, 2012), the percentage of the four types of trial-by-trial intrusion reports changed differently from the first to the tenth repetition (Choice×Repetition: $F[27,702]=3.4$, $p<0.001$, $\eta^2 = 0.01$; **Figure 4.3D**). Specifically, the percentage of reporting "never" increased ($F[9,234]=5.4$, $p<0.001$, $\eta^2 = 0.04$), while the percentage of reporting "sometimes" ($F[9,234]=2.5$, $p=0.008$, $\eta^2 = 0.02$) decreased over repetitions. These results together suggest that participants were successful at retrieving or suppressing memory traces according to task instructions.

Memory performance during the final memory test

During the final test, participants selected, on average, the correct category (chance level=1/4) for the associated picture on 91.82% (SD = 6.05%; range from 70.83% to 100%) of the successfully recalled associations of the typing test on day2 (mean=39.43). We then examined how repeated retrieval and suppression affected memory performance. First, we compared recall accuracies between three kinds of associations (i.e., *RETRIEVAL ASSOCIATIONS, SUPPRESSION ASSOCIATIONS, and CONTROL ASSOCIATIONS*). Analysis of objective recall accuracy after modulation showed no significant main effect of *modulation* (F [2,26]=0.524, p=0.595, η^2 =0.013; **Figure 4.3E**). Due to the lack of suppression-induced forgetting effect (lower accuracy for *SUPPRESSION ASSOCIATIONS* compared to *CONTROL ASSOCIATIONS*) at the group level, we performed a correlational analysis to associate performance during memory suppression and the final memory test. We found that participants who were more effective in suppressing intrusions (higher *intrusion slope score*) during the modulation phase were the ones who showed larger suppression-induced forgetting effects (r=0.411, p=0.03; **Figure 4.3F**), suggesting that successful retrieval suppression was subsequently associated with suppression-induced forgetting. This correlation was also reported before in the think/no-think literature (B. J. Levy & Anderson, 2012). Additionally, we investigated the effect of *modulation* on memory confidence and found a significant main effect (F [2,26]=5.928, p=0.005, η^2 =0.07; Figure 3G). Post-hoc analyses revealed higher recall confidence for *RETRIEVAL ASSOCIATIONS* compared to the *CONTROL ASSOCIATIONS* (t(26)=3.35, p_{holm}=0.007, Cohen's d=0.64) and a trend towards higher confidence compared to *SUPPRESSION ASSOCIATIONS* that just failed to reach our threshold for statistical significance (t(26)=2.172, p_{holm}=0.07, Cohen's d=0.41). Finally, we asked if modulation affected retrieval speed indexed by the RT during the final test. Even though we did not find a significant main effect of modulation (F [2,26]=2.905, p=0.06, η^2 =0.03; Figure 3H), recall of *RETRIEVAL ASSOCIATIONS* was faster compared to the recall of *CONTROL ASSOCIATIONS* (t(26)=-2.486, p=0.02, Cohen's d=-0.47).

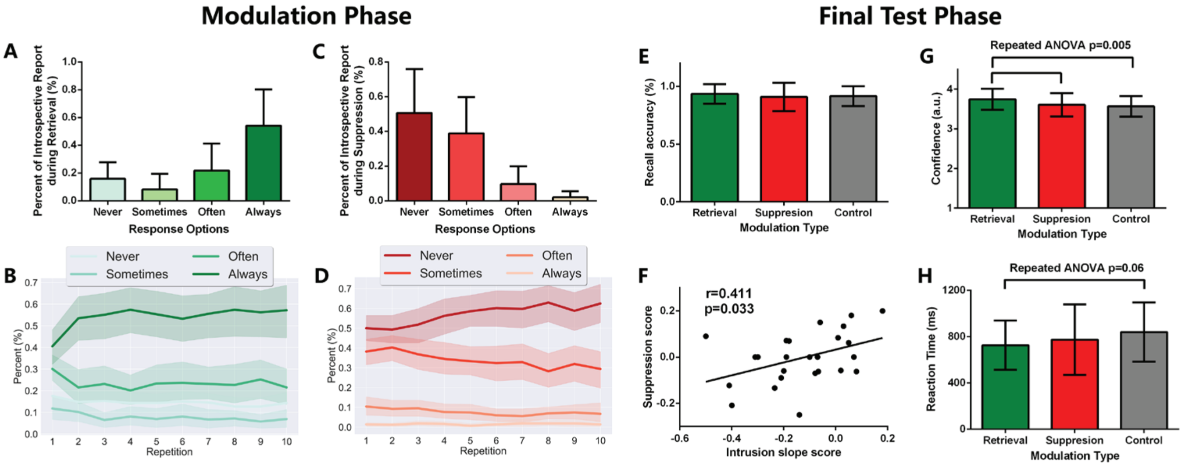


Figure 4.3 Behavioral performance during modulation and final memory test phase. (A) Percentage of the trial-by-trial introspective report during the retrieval trials. For most of the retrieval trials, associated pictures were successfully recalled (1-P_{never}: mean=84.05%, SD=11.79 %). **(B)** With repeated retrieval attempts, associated pictures were more likely to “always” stay in mind (P_{always}: F [9,234]=5.3, p<0.001, η^2 =0.02). **(C)** Percentage of the trial-by-trial introspective report during the suppression trials. During half of the suppression trials, participants successfully suppressed the tendency to recall the associated pictures (P_{never}: mean=50.62%, SD=25.35%). **(D)** As the number of repetitions of suppression increase, the possibility of successful suppression increased (F [9,234]=5.4, p<0.001, η^2 =0.04). **(E)** There is no effect of retrieval or suppression on the accuracy of the categorization during the final test (p=0.595). **(F)** Participants who are more effective in reducing suppression failures (more negative the *Intrusion Slope Score*) were the ones who show more evidence suppression-induced forgetting (more negative the *Suppression Score*). **(G)** For *RETRIEVAL ASSOCIATIONS*, participants reported higher subjective confidence compared to *SUPPRESSION ASSOCIATIONS* (t(26)=2.172, p_{holm}=0.07, Cohen’s d=0.41) , and *CONTROL ASSOCIATIONS* (t(26)=3.35, P_{holm}=0.007, Cohen’s d=0.64). **(H)** For *RETRIEVAL ASSOCIATIONS*, participants spent less time during categorisation compared to the *CONTROL ASSOCIATIONS* (t(26)=-2.486, p=0.02, Cohen’s d=-0.47), and the effect between three conditions tend to be significant (F [2,26]=2.905, p=0.06, η^2 =0.03). a.u= arbitrary unit.

fMRI results

Measuring the pattern reinstatement of individual memory during retrieval

The Support Vector Classification (SVC)-based searchlight analysis revealed brain regions including the lateral occipital cortex, fusiform gyrus, lingual gyrus, and calcarine cortex, which showed picture-specific activation patterns during the perception (uncorrected p_{voxel}<0.001, **Figure 4.4A**). We restricted our following activation pattern analyses to these voxels within the anatomical VVC boundary (**Figure 4.4B**). Next, we confirmed that activation patterns of these voxels could be used for cross-participant classification of the visual stimulus during perception. We trained the SVC based on activation patterns of two-thirds of all participants and tested the model using the remaining one-third. Results from the three-fold cross-validation confirmed these VVC voxels do enable cross-participant picture classification (mean accuracy=61.88%, SD=17.71%, shuffled accuracy $_{max}$=3.2%, p<0.001, **Figure 4.4D**).

The preceding results established that activity patterns of voxels within the VVC carry picture-specific information during perception, we next examined if we can detect the pattern reinstatements of memory traces within the same area during the final memory test. We trained the SVC model based on the neuroimaging data from the pattern localization phase to classify the trial-by-trial memory content in the final test (**Figure 4.4C**). Results showed that the classifiers could decode memory content based on activity patterns during the final test with an accuracy (mean accuracy=43.13%, SD=16.52%, shuffled accuracy $_{max}$=3.3%, p<0.001, **Figure 4.4E**), although the accuracy is significantly lower than the within-task classification of the perceived visual stimulus (t(26)=-3.97, p<0.001, Cohen's d=-0.76, **Figure 4.4F**).

We ran two control analyses to test the robustness of observed pattern reinstatement in the VVC during retrieval. We first examined the effect of arbitrary thresholds used in cluster formation on the subsequent classification of memory contents. Specifically, we used the two additional thresholds (uncorrected p $_{voxel}$=0.01 and 0.05) to identify picture-sensitive voxels during the whole-brain searchlight analysis and confirmed that the classifications could also be performed based on picture-sensitive voxels under other thresholds (0.01 and 0.05) (**Figure S4.3**). In addition, beyond picture-specific classifications, we investigated the possibility of category-specific classifications based on brain activity patterns. All of the pictures to be associated can be categorized as one of the four following groups: animal, human, object, or location. Similarly, we localized category-sensitive voxels within the VVC (**Figure S4.4D**) and confirmed that these voxels also carry category-specific information during perception (mean accuracy=69.13%, SD=9.67%, shuffled accuracy $_{max}$=29.6%, p<0.001, **Figure S4.4E**). Also, activity patterns of these category-sensitive voxels during memory retrieval could enable cross-participant, cross-task classification of the category during final memory test (mean accuracy=44.29%, SD=8.9%, shuffled accuracy $_{max}$=30.4%, p<0.001, **Figure S4.4E**).

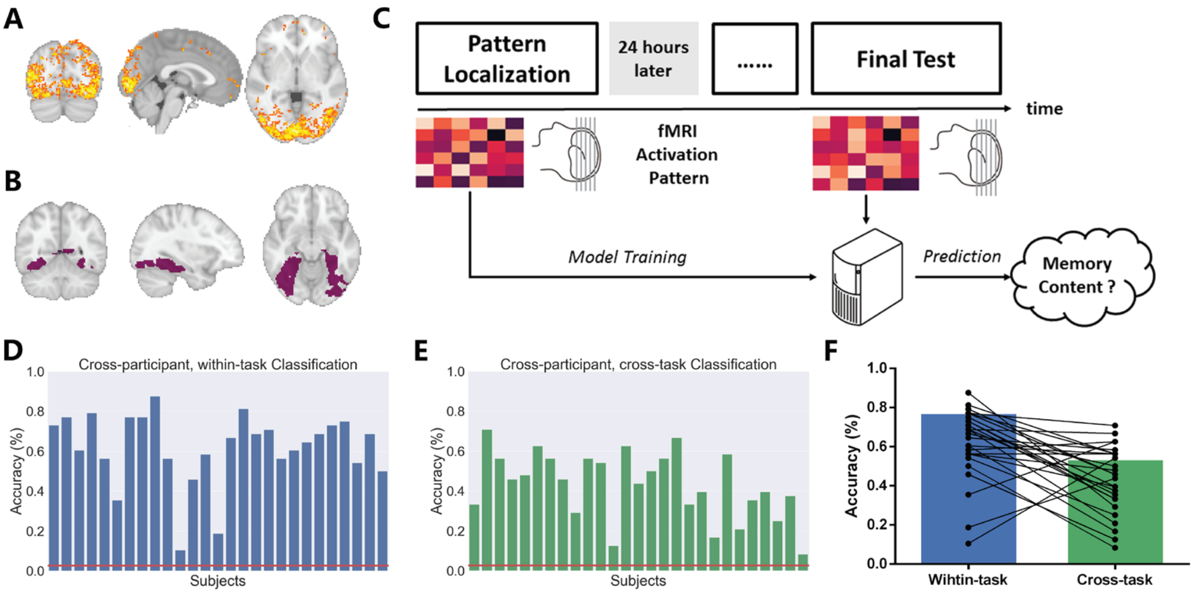


Figure 4.4 Identify picture-sensitive voxels and measure pattern reinstatement in the ventral visual cortex. (A) Using the searchlight method, we localized picture-sensitive voxels in brain regions included lateral occipital cortex, fusiform gyrus, lingual gyrus, calcarine cortex, postcentral and precentral gyrus, supplementary motor area, and small clusters within the medial and inferior prefrontal cortex. These voxels showed picture-specific activation patterns during the perception (uncorrected $p_{voxel}<0.001$). **(B)** We restricted our following pattern analyses into these voxels within the ventral visual cortex (VVC) boundary by overlapping the searchlight accuracy map and anatomical-defined VVC. **(C)** fMRI activation patterns of these voxels during pattern localization were extracted to train a classifier. The activity patterns of these voxels during the final test were further extracted and used as inputs for the classifier for different pictures. **(D)** The classifier was first validated in a cross-participant, within-task procedure. We demonstrated that picture-sensitive voxels voxels could enable the cross-participant picture classification during perception (mean accuracy=61.88%, SD=17.71%, p<0.001). **(E)** The same classifier, without further model training, was used for the decoding of memory contents based on activity patterns during retrieval. Results showed that the classifier can decode the memory contents with the accuracy higher than shuffled decoding models (mean accuracy=43.13%, SD=16.52%, p<0.001). **(F)** We observed the significant lower classification accuracies for cross-task classification compared to the within-task classification (t(26)=-3.97, p<0.001). The red line represents the 95th percentile of the accuracy within 1000 randomly label-shuffled null distribution.

Given the role of the hippocampus and parietal lobe in memory retrieval, we also performed the same pattern reinstatement pipeline (shown in **Figure 4.4C**) in these regions. We trained the classifier based on activity patterns of the hippocampus, angular gyrus, supramarginal gyrus, and precuneus during perception and applied it to decode memory content during retrieval. Activity patterns in these regions enabled us to perform picture-specific classification, but less accurately compared to visual areas (left hippocampus: mean accuracy=7.1%, SD=4.3%, p<0.001; right hippocampus: mean accuracy=6.5%, SD=2.8%, p<0.001; left AG: mean accuracy=10.9%, SD=7.2%, p<0.001; right AG: mean accuracy=10.9%, SD=9.1%, p<0.001; left SMG: mean accuracy=8.9%, SD=6.9%, p<0.001; right SMG: mean accuracy=13.5%, SD=12.8%, p<0.001; left precuneus: mean accuracy=16.2%, SD=8.9%, p<0.001; right precuneus: mean accuracy=18.1%, SD=7.4%, p<0.001; **Figure S4.5**).

Based on the same decoding pipeline, we performed a control pattern reinstatement analysis on activation patterns within the premotor cortex (**Figure S4.6A**), which, according to the reinstatement model, is not expected to represent memory content during retrieval (*details see Supplemental Texts; Section4*). Even for the category-based decoding, which requires less information than the item-based decoding, activation patterns of this area during retrieval could not be used to classify memory contents (**Figure S4.6B**).

Without considering the modulation of each association (i.e., retrieval, suppression, or control), we demonstrated pattern reinstatement of individual memories during retrieval after 24 hours delay. Based on the differences in RT and confidence, we tested whether different modulations have different effects on the evidence (i.e., decoding accuracy or decision value (Linde-Domingo et al., 2019)) of memory reactivation. For example, if repeated retrieval increased the reactivation evidence, while suppression decreased the evidence). We performed these analyses based on classifier training in both cross-participant and within-participant manner. These analyses yielded no significant results between different modulations in all ROIs investigated (*Details in Supplemental Materials;* **Table S4.1-S4.4**).

In sum, we identified picture-specific voxels within the VVC and demonstrated the pattern reinstatements of individual memory traces in these voxels during retrieval. The same pattern reinstatements were detected in anatomical-defined hippocampus, AG, SMG, and precuneus. These results are the foundations of our following multivariate pattern analysis: the pattern reinstatements 24 hours after initial learning suggested that activity patterns of these regions during retrieval carry mnemonic representations.

Repeated retrieval leads to reduced activity amplitude, but more distinct activity patterns

Repeated retrieval dynamically reduces the activity amplitude in the visual cortex and hippocampus: compared to *CONTROL ASSOCIATIONS*, retrieval of *RETRIEVAL ASSOCIATIONS* was associated with less activation in medial occipital cortex, fusiform gyrus, supplementary motor area (SMA), anterior/medial cingulate cortex (MCC), left precentral gyrus, precuneus, bilateral insula, and bilateral inferior frontal gyrus (IFG) (voxelwise $P_{uncorrected}<0.001$, $p_{FWE\text{-}cluster}<0.05$; **Figure S4.7A**; **Table S4.5**). The VVC cluster revealed by the whole-brain analysis largely overlapped with our functional-defined VVC voxels (see Figure S7 for comparison). Our ROI analysis of these functionally-defined picture-sensitive voxels confirmed the observation: we found a reduced activity amplitude of picture-sensitive voxels for *RETRIEVAL ASSOCIATIONS* compared to *CONTROL ASSOCIATIONS* ($t(26)=-4.8$, $p_{raw}<0.001$, $p_{FDR}<0.001$, Cohen's $d=-0.92$; **Figure 4.5A**). The whole-brain analysis did not show an effect of retrieval on the activity amplitude in hippocampal voxels under the same threshold. However, ROI-based analysis of hippocampal signal found reduced activity when retrieving *RETRIEVAL ASSOCIATIONS* compared to *CONTROL*

ASSOCIATIONS (left hippocampus: t(26)=-2.43, p_{raw}=0.022, p_{FDR}=0.06, Cohen's d=-0.46; **Figure 4.5E**; right hippocampus: t(26)=-2.18, p_{raw}=0.038, p_{FDR}=0.06, Cohen's d=-0.42; **Figure 4.5G**). For six ROIs of the parietal lobe, we only found a similar retrieval-related activity reduction in the right AG (t(26)=-2.68, p_{raw}=0.012, p_{FDR}=0.05, Cohen's d=-0.51; **Figure 4.6C**) and the right precuneus (t(26)=-2.33, p_{raw}=0.027, p_{FDR}=0.06, Cohen's d=-0.45; **Figure 4.6K**), but which was not significant in the left AG (t(26)=-1.57, p_{raw}=0.12, p_{FDR}=0.13, Cohen's d=-0.30; **Figure 4.6A**), left SMG (t(26)=1.65, p_{raw}=0.11, p_{FDR}=0.13, Cohen's d=0.31; **Figure 4.6E**), right SMG (t(26)=1.32, p_{raw}=0.19, p_{FDR}=0.19, Cohen's d=0.25; **Figure 4.6G**), or left precuneus (t(26)=-1.91, p_{raw}=0.067, p_{FDR}=0.08, Cohen's d=-0.36; **Figure 4.6I**).

Next, we confirmed that the observed activity reduction is related to a linear decrease in activity with repeated retrieval using the data from the modulation phase. Specifically, we extracted the beta coefficient from these clusters for each run of the modulation phase and tested for the change in activity amplitude across runs. We found reduced VVC activity over repeated retrieval attempts (F [4, 25]=5.95, p<0.001, η^2 =0.174). Similarly, for the bilateral hippocampus, we observed a trend toward a gradual decrease of hippocampal signal across repetitions (left hippocampus: F [4, 25]=2.39, p=0.056, η^2 =0.087 ; right hippocampus: F [4, 25]=2.22, p=0.072, η^2 =0.082). Even though we found the retrieval-related activity reduction in right AG and precuneus during the final test, we did not find the corresponding gradual decrease during modulation (right AG: F [4, 25]=0.734, p=0.571, η^2 =0.02; right precuneus: F [4, 25]=1.88, p=0.12, η^2 =0.05).

Repeated retrieval dynamically enhances the distinctiveness of activity patterns in the visual cortex, but not hippocampus: focusing on the identified VVC voxels, parietal lobe and hippocampus, we calculated the trial-by-trial activity pattern similarity for *RETRIEVAL ASSOCIATIONS* and *CONTROL ASSOCIATIONS* separately. Results show that retrieval-related activity patterns for *RETRIEVAL ASSOCIATIONS* have decreased similarity in VVC compared to *CONTROL ASSOCIATIONS* (t(26)=-2.3, p_{raw}=0.029, p_{FDR}=0.08, Cohen's d=-0.44; **Figure 4.4C**). To test the robustness of decreased pattern similarity for *RETRIEVAL ASSOCIATIONS* in the VVC*,* we performed the same contrast based on (1) all associations instead of only remembered association, the VVC areas defined by (2) different thresholds and (3) category-sensitive voxels instead of picture-sensitive voxels. All control analyses yield the same result as the reported main analysis (***Figure S4.8-S4.10***). However, we did not observe a similar effect in the hippocampus (left hippocampus: t(26)=-0.91, p_{raw}=0.36, p_{FDR}=0.40, Cohen's d=-0.177, Figure 4F; right hippocampus: t(26)=-0.456, p_{raw}=0.65, p_{FDR}=0.65, Cohen's d=-0.088 ; **Figure 4.4H**). For six ROIs of the parietal lobe, retrieval-related decreases in pattern similarities were found in right AG (t(26)=-2.148, p_{raw}=0.04, p_{FDR}=0.08, Cohen's d=-0.413; **Figure 4.6D**), left SMG (t(26)=-2.1, p_{raw}=0.045, p_{FDR}=0.08, Cohen's d=-0.406; **Figure 4.6F**), left precuneus (t(26)=-2.2, p_{raw}=0.035,

p_{FDR}=0.08, Cohen's d=-0.428; **Figure 4.6J**) and right precuneus (t(26)=-2.8, p_{raw}=0.009, p_{FDR}=0.08, Cohen's d=-0.54; **Figure 4.6M**). Similar trend was found in left AG (t(26)=-1.8, p_{raw}=0.07, p_{FDR}=0.10, Cohen's d=-0.36; **Figure 4.6B**) and right SMG (t(26)=-1.79, p_{raw}=0.08, p_{FDR}=0.10, Cohen's d=-0.34; **Figure 4.6H**), but failed to reach significance.

Our ROI analyses already found reduced activity amplitude, but more distinct activity patterns in VVC, right AG, and precuneus. Then we performed the correlational analysis to explore the relationship between changes in activity amplitude and changes in pattern similarity across participants. We found that participants who showed a larger reduction in VVC's activity amplitude were more likely to show a larger decrease in VVC pattern similarity (r=0.610, p<0.001; **Figure 4.5C**). This correlation is also significant for right precuneus (r=0.427, p=0.026), but not for right AG (r=-0.051, p=0.799).

To characterize the dynamic modulation of pattern similarity in the VVC, we further applied the same variability analysis to each run of the modulation phase and analyzed these pattern similarity values using a 2×5 ANOVA (*modulation*; *repetition*). We saw a significant main effect of *run*, reflecting that pattern similarity of the VVC decreased with repetitions (F [4, 100]=10.55, p<0.001, η^2 =0.028). We also saw a main effect of *modulation*, reflecting that pattern similarity of the *RETRIEVAL ASSOCIATIONS* is consistently lower than the similarity of *SUPPRESSION ASSOCIATIONS* (F [1, 25]=23.77, p<0.001, η^2 =0.028). The interaction between *modulation* and *runs* just failed to be significant (F [4, 100]=2.427, p=0.053, η^2 =0.001; **Figure 4.5D**). This pattern of results suggests that decreased pattern similarity is not only the result of repetition: even though memory cues of *SUPPRESSION ASSOCIATIONS* have also been presented ten times during the modulation, repeated retrieval more effectively enhanced pattern distinctiveness compared to suppression. We applied the same dynamic modulation analysis to the ROIs, which demonstrated lower cross-item pattern similarity for *RETRIEVAL ASSOCIATIONS* (i.e., right AG, left SMG, and bilateral precuneus) during the final memory test phase, but we found no evidence for an interaction between *modulation* and *runs* (right AG: F [4, 100]=1.42, p=0.23, η^2 =0.001; left SMG: F [4, 100]=0.23, p=0.92, η^2 =0; left precuneus: F [4, 100]=2.13, p=0.08, η^2 =0.002; right precuneus: F [4, 100]=0.51, p=0.72, η^2 =0.002).

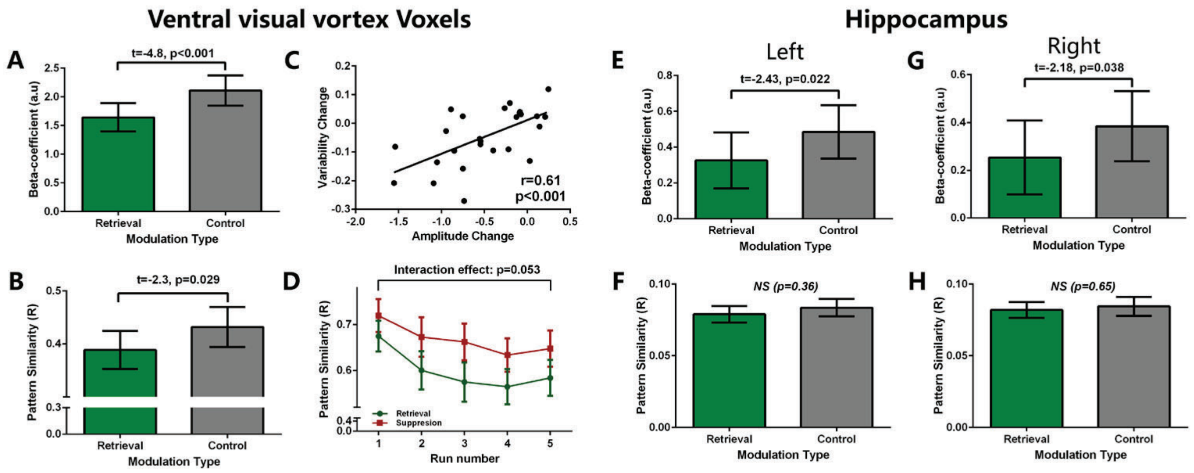


Figure 4.5 **Repeated retrieval dynamically modulated activity amplitude and pattern similarity. (A)** During the final test, compared to *CONTROL ASSOCIATIONS*, *RETRIEVAL ASSOCIATIONS* was associated with lower activity amplitude in voxels within the ventral visual cortex identified in the pattern reinstatement analysis. **(B)** Lower pattern similarity in these VVC voxels for *RETRIEVAL ASSOCIATIONS* compared to the *CONTROL ASSOCIATIONS* during the final test. **(C)** Across participants, the extent of activity amplitude reduction positively correlated with enhancement in pattern distinctiveness. **(D)** Dynamically decreased pattern similarity in the VVC. For both *RETRIEVAL ASSOCIATIONS* and *SUPPRESSION ASSOCIATIONS*, VVC's pattern similarity increased over repetitions during the modulation. However, repeated retrieval tends to more effectively decrease pattern similarity compared to suppression. **(E)** Reduced left hippocampal activity amplitude for *RETRIEVAL ASSOCIATIONS* compared to *CONTROL ASSOCIATIONS* during the final test. **(F)** No differences in left hippocampal activity pattern similarity between *RETRIEVAL ASSOCIATIONS* and *CONTROL ASSOCIATIONS* during the final test. **(G)** Reduced right hippocampal activity amplitude for *RETRIEVAL ASSOCIATIONS* compared to *CONTROL ASSOCIATIONS* during the final test. **(H)** No differences in right hippocampal activity pattern similarity between *RETRIEVAL ASSOCIATIONS* and *CONTROL ASSOCIATIONS* during the final test.

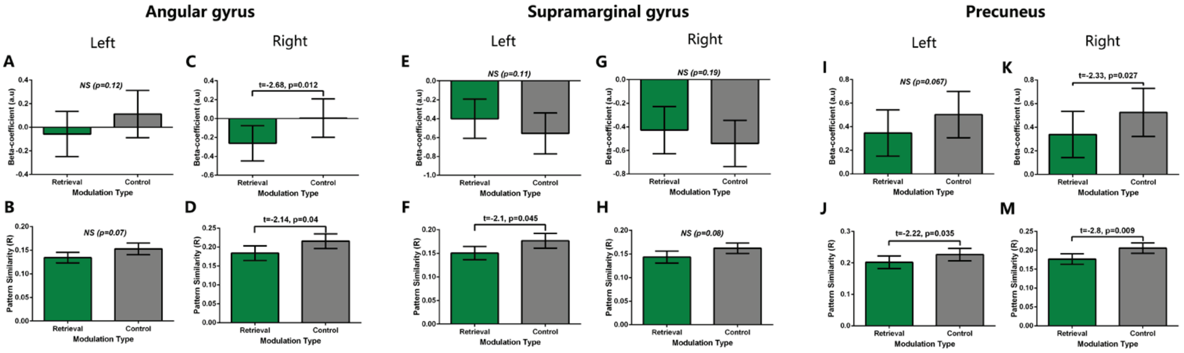


Figure 4.6. Effect of repeated retrieval on activity amplitude and patterns similarity of the parietal lobe. **(A)** No differences in activity amplitude of left angular gyrus. **(B)** No differences in activity pattern similarity of the left angular gyrus. **(C)** Lower pattern similarity of right angular gyrus for *RETRIEVAL ASSOCIATIONS* compared to the *CONTROL ASSOCIATIONS*. **(E)** No differences in activity amplitude of left supramarginal gyrus. **(F)** Lower pattern similarity of left supramarginal gyrus for *RETRIEVAL ASSOCIATIONS* compared to the *CONTROL ASSOCIATIONS*. **(G)** No differences in activity amplitude of right supramarginal gyrus. **(H)** No differences in activity pattern similarity of right supramarginal gyrus. **(I)** No differences in activity amplitude of left precuneus. **(J)** Lower pattern similarity of left supramarginal gyrus for *RETRIEVAL ASSOCIATIONS* compared to the *CONTROL ASSOCIATIONS*. **(K)** Reduced activity amplitude of right precuneus for *RETRIEVAL ASSOCIATIONS* compared to *CONTROL ASSOCIATIONS*. **(M)** Lower pattern similarity of left supramarginal gyrus for *RETRIEVAL ASSOCIATIONS* compared to the *CONTROL ASSOCIATIONS*.

Retrieval suppression was associated with reduced lateral prefrontal activity

Weaker lateral prefrontal cortex (LPFC) activation as the result of retrieval suppression: the contrast between retrieval of *SUPPRESSION ASSOCIATIONS* and *CONTROL ASSOCIATIONS* during the final test revealed decreased activation oin one cluster in the left LPFC (x=−52,y=38, z=16, Z_{peak}=4.09, size=1320 mm^3; **Figure 4.7A**). We did not find any significant effect of retrieval suppression on hippocampal activity amplitude in the whole-brain or the ROI analysis (left hippocampus: t(26)=-1.14, p=0.26, Cohen's d=-0.22; right hippocampus: t(26)=-0.81, p=0.43, Cohen's d=-0.15). Also, repeated retrieval suppression was associated with reduced activity in the right AG (t(26)=-2.07, p=0.048, Cohen's d=-0.40), but not left AG (t(26)=-0.865, p=0.395, Cohen's d=-0.16), left SMG (t(26)=1.214, p=0.236, Cohen's d=0.23), right SMG (t(26)=0.867, p=0.394, Cohen's d=0.16), left precuneus (t(26)=-0.77, p=0.44, Cohen's d=-0.15) or right precuneus (t(26)=-1.13, p=0.26, Cohen's d=-0.21).

To characterize dynamical activity changes in the left LPFC, we extracted beta values from the cluster for each modulation run and did not find decreased activity from the first to the fifth run during suppression (F [4, 25]=2.03, p=0.09, η^2 =0.056; **Figure 4.7B**). Subsequently, we performed an exploratory analysis to restrict analysis within the first four runs and found a gradually decreased activity in the left lPFC (F [3, 25]=2.98, p=0.036, η^2 =0.078).

Intact neural representations after memory suppression: next, we examined if retrieval suppression modulated activity patterns in the VVC, hippocampus, or parietal lobe. Pattern similarity analysis revealed no significant difference between *SUPPRESSION ASSOCIATIONS* and *CONTROL ASSOCIATIONS* in all regions investigated (VVC: t(26)=-1.035, p=0.31, Cohen's d=-0.19; left hippocampus: t(26)=-0.78, p=0.43, Cohen's d=-0.15; right hippocampus: t(26)=-.010, p=0.92, Cohen's d=-0.02; left AG: t(26)=0.44, p=0.663, Cohen's d=-0.08; right AG: t(26)=-0.48, p=0.63, Cohen's d=-0.09; left SMG: t(26)=-1.29, p=0.206, Cohen's d=-0.25; right SMG: t(26)=-1.15, p=0.26, Cohen's d=-0.22; left precuneus: t(26)=-0.47, p=0.63, Cohen's d=-0.09; right precuneus: t(26)=-1.29, p=0.2, Cohen's d=-0.25). Give the modest effect of memory suppression on final memory performance, but the strong correlation between the intrusion slope and suppression-induced forgetting, we further investigated suppression-induced changes in pattern similarity among participants who showed strong negative intrusion slopes and (by correlation) more suppression-induced forgetting. More specifically, we used the median split method to divide the data of all participants into two groups (strong suppression group vs. weak suppression group) according to their intrusion slope value and compared changes in pattern similarity between groups. Our results suggested that both groups did not demonstrate differential suppression-induced changes in changes in pattern similarity for all ROIs investigated (**Table S4.6**).

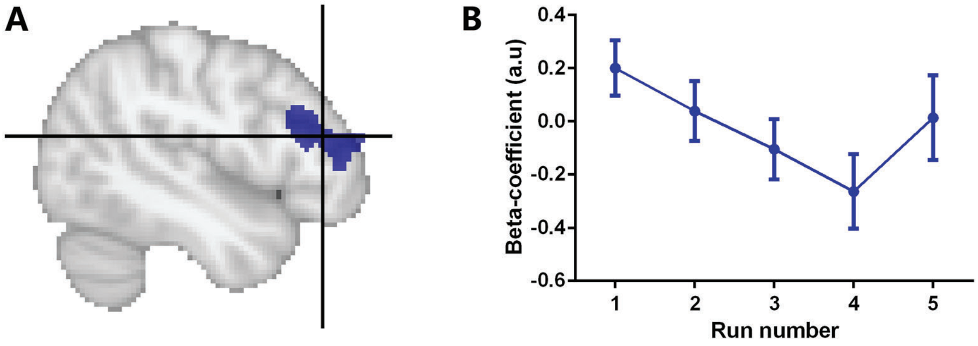


Figure 4.7 Repeated suppression disengaged lateral prefrontal cortex (LPFC) during subsequent memory retrieval. (A) During the final memory test, we found lower activity amplitude in the left LPFC for *SUPPRESSION ASSOCIATIONS* compared to *CONTROL ASSOCIATIONS*. (B) During the modulation, the activity amplitude of the same LPFC cluster decreased over repetitions from the first to fourth run (F [3, 25]=2.98, p=0.036, η^2 =0.078), but failed to be significant from the first to the fifth run (F [4, 25]=2.03, p=0.09, η^2 =0.056).

Discussion

Active memory retrieval is known to be a powerful memory enhancer, while memory suppression tends to prevent unwanted memories from further retrieval. Previous neuroimaging investigations of the neural effect of repeated retrieval and suppression revealed corresponding neural changes in both univariate activity analysis and multivariate activity patterns analysis. Building on these findings, we tested whether similar neural changes can be detected when modulation is delayed by 24 hours (i.e., newly acquired memories have undergone the initial consolidation). Also, because we collected fMRI data from both the modulation phase and the final memory test, this design allowed us to perform dynamic analysis on whether the neural changes seen in the final memory test are accompanied by gradual changes during the modulation phase. Similar to previous literature (Ferreira et al., 2019), our results demonstrated that repeated retrieval of consolidated memories was associated with enhanced episode-unique mnemonic representations in the parietal lobe. Critically, our dynamic analysis provided converging evidence for the adaption of stronger mnemonic representations in visual processing areas, which were involved in the initial perception. Our results suggested that repeated retrieval of newly acquired memory and initially consolidated memory may be associated with similar neural changes.

Repeated retrieval strengthened consolidated memories. Behaviorally, our results demonstrate that, after an initial delay of 24 hours, repeated retrieval strengthened memories further, indexed by higher recall confidence and shorter reaction times. The beneficial effect of retrieval practice on the subsequent retrieval is well established (Karpicke & Blunt, 2011; Karpicke & Roediger, 2008; Karpicke & Roediger III, 2007; A. M. Smith et al., 2016). In our study, memory accuracy was already near the ceiling level, and thus we did not find higher recall accuracy of *RETRIEVAL ASSOCIATIONS* compared to *CONTROL ASSOCIATIONS*. Corroborating the behavioral effect during the final memory test, we also found that repeated retrieval of certain memories increased their tendency to remain stable in mind during the modulation phase.

Repeated retrieval is associated with subsequent decreasing activity amplitude. Our whole-brain univariate analysis revealed a set of brain regions, including frontal, parietal (mainly precuneus), and ventral visual areas that showed decreasing activity amplitude with repeated retrieval. Activity changes in frontal and parietal areas have been reported frequently in the literature of retrieval-mediated learning/forgetting, but the directions of the reported changes are mixed. Some of the reports have found similar univariate decreases in frontal or parietal areas (Kuhl et al., 2010; M. Wimber et al., 2011; Maria Wimber et al., 2008), but others reported activity increases in these areas (Himmer et al., 2019; Nelson, Arnold, Gilmore, & Mcdermott, 2013; G. van den Broek et al., 2016; Wirebring et al., 2015). In addition to the whole-brain analysis, our ROI analysis further showed decreased activity in the right angular gyrus. In sum, our study mainly found decreased activity in frontal and parietal areas after repeated retrieval of initially consolidated memories. Moreover, decreased activity in ventral visual areas is a novel finding. Previous studies usually used words as materials to be remembered (Nelson, Arnold, Gilmore, & Mcdermott, 2013; M. Wimber et al., 2011; Maria Wimber et al., 2008; Wirebring et al., 2015), while we used pictures. One other study also used pictures and the TNT paradigm but did not reveal reliable activity changes for retrieved pictures compared to the controlled pictures (Gagnepain et al., 2014). To test the fast-consolidation hypothesis of retrieval-mediated learning (Antony et al., 2017), we further examined changes in hippocampal activity during modulation and final test. Similar to a recent report of slow hippocampal disengagement during repeated retrieval (Ferreira et al., 2019), we found dynamically decreasing hippocampal activity across repeated retrieval for initially consolidated memories. Our results, together with findings of Ferreira and colleagues, are consistent with decreasing retrieval-related hippocampal activity over the course of consolidation (Takashima et al., 2006, 2009).

Repeated retrieval enhanced episodic-unique cortical representations. Our multivariate pattern analysis showed that compared to controls, repeated retrieval led to less similar activity patterns in ventral visual areas, and almost all parietal ROIs, including AG, SMG, and precuneus. Using a conceptually similar method, Ferreira and colleagues also reported increased item-unique activity patterns in parietal regions across two days (Ferreira et al., 2019). Ye and colleagues further showed how retrieval practice led to memory updating by differentiating activity patterns in the mPFC (Ye et al., 2020). These results together may suggest the interaction between the effect of repeated retrieval and episodic-unique neural representations during the fast formation of cortical memories. Similar representational dissimilarity analysis has been used to analyze patterns of activity during retrieval suppression (Gagnepain et al., 2014). However, after the modulation, participants of this study only performed a visual perception task, which measures repetition priming instead of a direct measure of memory. Therefore, it is impossible to directly compare the trial-by-trial pattern similarity during retrieval between *RETRIEVAL* and *CONTROL* associations.

One novel aspect of our findings is that after repeated retrieval, we found the decreased retrieval-related activity amplitude correlated with enhanced distinctiveness of activity patterns in ventral visual areas and precuneus. Our dynamic analysis of these two neural measures during modulation and subsequent memory test confirmed further that the neural changes observed during the later test are associated with dynamic adaptation of activity amplitude and pattern similarity during modulation in the ventral visual areas. However, this is not true for the precuneus. In general, this pattern of results is in line with our knowledge about how preexisting associative memory shapes brain responses. Prior information about upcoming stimuli is often associated with overall lower activity amplitude, a phenomenon termed "*expectation suppression*"(Summerfield et al., 2008; Summerfield & de Lange, 2014). At the same time, underlying activity patterns carry more visual information (de Lange et al., 2018; Kok et al., 2012). By correlating these two neural changes in the same regions, our study reported a similar phenomenon during episodic memory retrieval. This finding suggests that the inverse relationship between overall activity amplitude and pattern-based information representation holds not only for low-level perceptual memory but also for episodic memory retrieval. Moreover, the correlation between the activity amplitude and pattern similarity may also be understood from a "*noise correlations*" perspective in information processing (Averbeck et al., 2006; M. R. Cohen & Kohn, 2011). A recent simultaneous EEG-fMRI study found that decreased alpha/beta power, as a potential marker of the reduced noise correlations, was associated with increased stimulus-specific activation patterns measured by representation similarity analysis (Benjamin James Griffiths et al., 2019). We speculate that retrieval practice might not directly enhance memory representations, but affect them by reducing their noise correlations. During retrieval of strengthened memories, redundant ongoing neuronal activity (i.e., noise) may be suppressed. Therefore, we observed lower overall activity amplitude and, at the same time, reduced "noise correlation," boosting the signal-to-noise ratio. Thus, stimulus-specific neural patterns are reinstated with more specificity, demonstrating lower pattern similarity across distinct trials.

Retrieval suppression inhibited lateral prefrontal activity during subsequent retrieval. For *SUPPRESSION ASSOCIATIONS*, we observed lower LPFC activity amplitude, but relatively intact activity patterns in visual areas, parietal lobe, and hippocampus during subsequent retrieval. Active memory suppression during retrieval is proposed to be partially supported by inhibitory control mechanisms mediated by the lateral prefrontal cortex (Michael C Anderson & Hanslmayr, 2014; Guo et al., 2018). During retrieval suppression, LPFC is typically activated (M. C. Anderson, 2004; Guo et al., 2018; B. J. Levy & Anderson, 2012), but it showed gradually decreasing activity amplitudes from early suppression attempts to the later trials of suppression (Brendan E Depue et al., 2007). Consistent with this pattern, we found a similar decrease in LPFC activity amplitude across suppression attempts during the modulation phase and lower activity amplitude during the subsequent retrieval. Together with the trial-by-trial intrusion frequency rating during modulation, this activity decrease across suppression attempts may suggest less inhibitory control demands when suppressing increasingly weakened memories. The observed reduction in LPFC activity during the subsequent retrieval might be a long-lasting effect of this reduced activity amplitude and suggests that modulated cognitive control allocation hampers retrieval. Another interesting observation is that we found weak evidence for suppression-induced changes in pattern reinstatement during the final memory test. Even though the involvement of the LPFC-hippocampal circuit in suppression has been examined (Michael C Anderson & Hanslmayr, 2014; Guo et al., 2018), the changes in neural representations of individual memory trace underlying suppression-induced forgetting remain less well studied. One study measured the effect of retrieval suppression on newly acquired visual memories via cortical inhibition (Gagnepain et al., 2014) and this study found that retrieval suppression reduced activity amplitude in the fusiform gyrus compared to retrieval, but the pattern was opposite to the one found in the lateral occipital complex. Effective connectivity and pattern similarity analysis suggested that top-down control mediated by the middle frontal gyrus suppressed perceptual memory traces in the visual cortex. Our study did find the comparable suppression-induced changes in activity amplitude but not mnemonic representations in the visual cortex. This may relate to the modest behavioral effects or less labile consolidated memory traces. Future studies with stronger suppression-induced forgetting effects can directly compare activity patterns between still-remembered associations and forgotten associations.

Limitations. Our study has a few limiting aspects that should be mentioned. Firstly, given that we only found a modest effect of suppression-induced forgetting, it is difficult to interpret repeated suppression-related fMRI results. There are at least two possible reasons for this modest effect: first, due to extensive training during encoding and/or the nature of our picture-location tasks, recall accuracy for all conditions was close to the ceiling level. Second, the suppression-induced forgetting effect is much smaller when memories have been consolidated (Y. Liu et al., 2016). Thus, in line with previous studies, suppression-induced forgetting may not have emerged as the group level (Gagnepain et al., 2017; Y. Liu et al., 2016). Nevertheless, we replicated two findings, confirming that our memory suppression modulation was still effective. First, when unwanted memories were suppressed repeatedly, their tendency to intrude was reduced during the TNT phase (Benoit et al., 2015; Gagnepain et al., 2017; Hellerstedt et al., 2016; B. J. Levy & Anderson, 2012; van Schie & Anderson, 2017). Second, the extent of this reduction (i.e., intrusion slope) correlated with subsequent suppression-induced forgetting effect across participants (B. J. Levy & Anderson, 2012). Given this correlation, we further compared suppression-induced neural changes between a strong and a weak suppression group, but still did not find an effect of suppression on mnemonic representations. These results may suggest that even for participants who showed suppression-induced forgetting, the underlying mnemonic representations remain intact. A second potential limitation of our study is that we only found the effect of repeated retrieval on trial-by-trial pattern similarity instead of the more direct measure of memory reactivation, such as decoding accuracy or decision value (Linde-Domingo et al., 2019). Therefore, the relationship between the reduction in univariate activity and enhanced multivariate representation can be interpreted from two different perspectives. On the one hand, it can be explained as the enhanced unique cortical memory representations. On the other hand, the reduction in across-item pattern similarity could be due to factors, for example, the reduced memory unrelated "*noise correlations*". It is noticeable that our pattern reinstatement analysis demonstrated that, based on activity patterns in our ROIs, the individual picture can be decoded when the classifier was trained on the localizer data (day1) before testing it on the final memory test (day2). This reinstatement laid the groundwork for our pattern similarity calculation because there is evidence that these activity patterns used in the variability analysis carry item-specific mnemonic information during retrieval. However, when we divided all associations into three groups (i.e., retrieval, suppression, and control), we did not find the evidence for the idea that retrieval or suppression can separately modulate decoding accuracies or d values, but that all three kinds of associations showed comparable decodability during retrieval. This result ruled out the possibility that could fully explain the differences in our pattern similarity measure. These results may suggest that decoding accuracies or d values used here were not sensitive enough after initial consolidation, because perceptual information might already be based on the transformed representation (Xiao et al., 2017). In addition, decoding outcomes and pattern similarity may associate with different aspects of mnemonic representations. Sensitive decoding depends on the reinstatement of the original representation related to the perceptual input, while pattern similarity reflects episode-unique activity patterns across retrieved "mental images". Enhanced episode-unique representations after repeated retrieval, particularly in the visual processing areas, support the following notion. Given that our memory cues (i.e., highlighted locations) are visually very similar, the changes in pattern similarity in visual areas are more likely to be the result of enhanced mnemonic reinstatements instead of variability induced by visual features of memory cues. Thirdly, when using a conservative correction for the number of ROIs tested, contrasts of parietal areas only showed only considerable trends toward significance, although the individual test is significant. We believe that trends in parietal areas could be caused by the definitions of our ROIs are based on the coarse atlas at the group level. That is to say, for each participant, maybe only part of the parietal ROIs is involved in the retrieval processing.

Conclusion. Taken together, our study probed the effects of repeated retrieval and suppression on initially consolidated memories. We showed that repeated retrieval dynamically reduces the activity amplitude in the visual cortex and hippocampus while enhances the distinctiveness of activity patterns in the visual cortex and parietal lobe. Moreover, reduction in activity amplitude correlated with the enhancement of episode-unique mnemonic representations in visual areas and precuneus. By contrast, repeated suppression, as done here, was associated with the reduced lateral prefrontal activity, but intact mnemonic representations. These findings extended our understanding of neural changes underlying memory modulations from newly acquired memories to initially consolidated memories and suggested that active retrieval may strengthen episode-unique information neocortically after initial encoding and also consolidation.

Supplementary Material for

Probing the neural dynamics of mnemonic representations after the initial consolidation

This file includes:

Supplementary Text
Figure S4.1-S4.10
Table S4.1-S4.6

1. Robustness of neural reinstatement of individual memory in the ventral visual cortex

1.1 Effect of arbitrary thresholds for cluster formation on the subsequent classifications

During the one-sample test on all of the classification accuracy maps resulting from the searchlight analyses, we used the arbitrary threshold (uncorrected $p_{\text{voxel}}<0.001$) for the cluster formation and all of the following classification analyses were based on thresholded voxels. To account for the effect of arbitrary thresholds during the cluster formation on the following analyses, we used the two additional thresholds (uncorrected $p_{\text{voxel}}=0.01$ and 0.05) to identify picture-sensitive voxels. Repeating the cross-participant, within-task classification and cross-participant, cross-task classification, we confirmed that the classifications could also be performed based on picture-sensitive voxels under other thresholds (0.01 and 0.05)(Figure S2)

1.2 Possibility of category-specific classifications.

Beyond picture-specific classifications, we investigated the possibility of category-specific classifications based on brain activation patterns. All of the pictures to be associated can be categorized as one of the four following groups: animal, human, object, or location. Similarly, we localised category-sensitive voxels within the ventral visual cortex (VVC) (**Figure S4.4D**) and confirmed that these voxels also carry category-specific information during perception (mean accuracy=73.5%, SD=8.6%, one-sample t-test: t=29.41, p<0.001, **Figure S4.4E**). Critically, activation patterns of these *category-sensitive voxels* during memory retrieval could enable cross-participant, cross-task classification of categories of the memory (mean accuracy=44.4%, SD=10.1%, one-sample t-test: t=10.03, p<0.001, **Figure S4.3**).

2. Comparisons of evidence for memory reactivation between *RETRIEVAL, SUPPRESSION*, and *CONTROL* associations.

Our pattern reinstatement analysis demonstrated that activity patterns during perception were reinstated during memory retrieval in the visual processing areas, parietal lobe, and hippocampus lob after 24 hours. Then, we investigated whether different modulation (i.e. retrieval and suppression) could modulate the memory reactivation process, indexed by different decoding outcomes for different memory associations. We used two decoding outcomes as the neural evidence for memory reactivation: [1] decoding accuracy. We analyzed predicted labels generated by the Support Vector Classification (SVC) classifier for different kinds of associations separately and calculated the average decoding accuracies of *RETRIEVAL, SUPPRESSION*, and *CONTROL* associations for each participant. The higher decoding accuracy for *RETRIEVAL* associations compared to *CONTROL* associations may reflect stronger direct memory reactivations induced by repeated retrieval. [2] decision value (d value)(Linde-Domingo et al., 2019). During trial-by-trial classification, together with predicted labels, we also generated the distance to the hyper-plane using the "decision_function" implemented in the SVC function. This distance measure indicates how confident the classifier was about generated predicted labels at the single-trial level. Raw d values could be either positive or negative. To use d values for further analysis, we transformed each raw d value to its absolute value, resulting in d value that larger value always reflects more confident classification. Then we calculated the averaged values for each kind of association for each participant. Finally, we compared the evidence for memory reactivation (i.e., accuracy and d value) between different associations. As shown in **Table S4.1-S4.2**, for all ROIs investigated, we did not find the effect of modulation on decoding outcomes.

3. Robustness of changes in activity pattern similarity for *RETRIEVAL ASSOCIATIONS*.

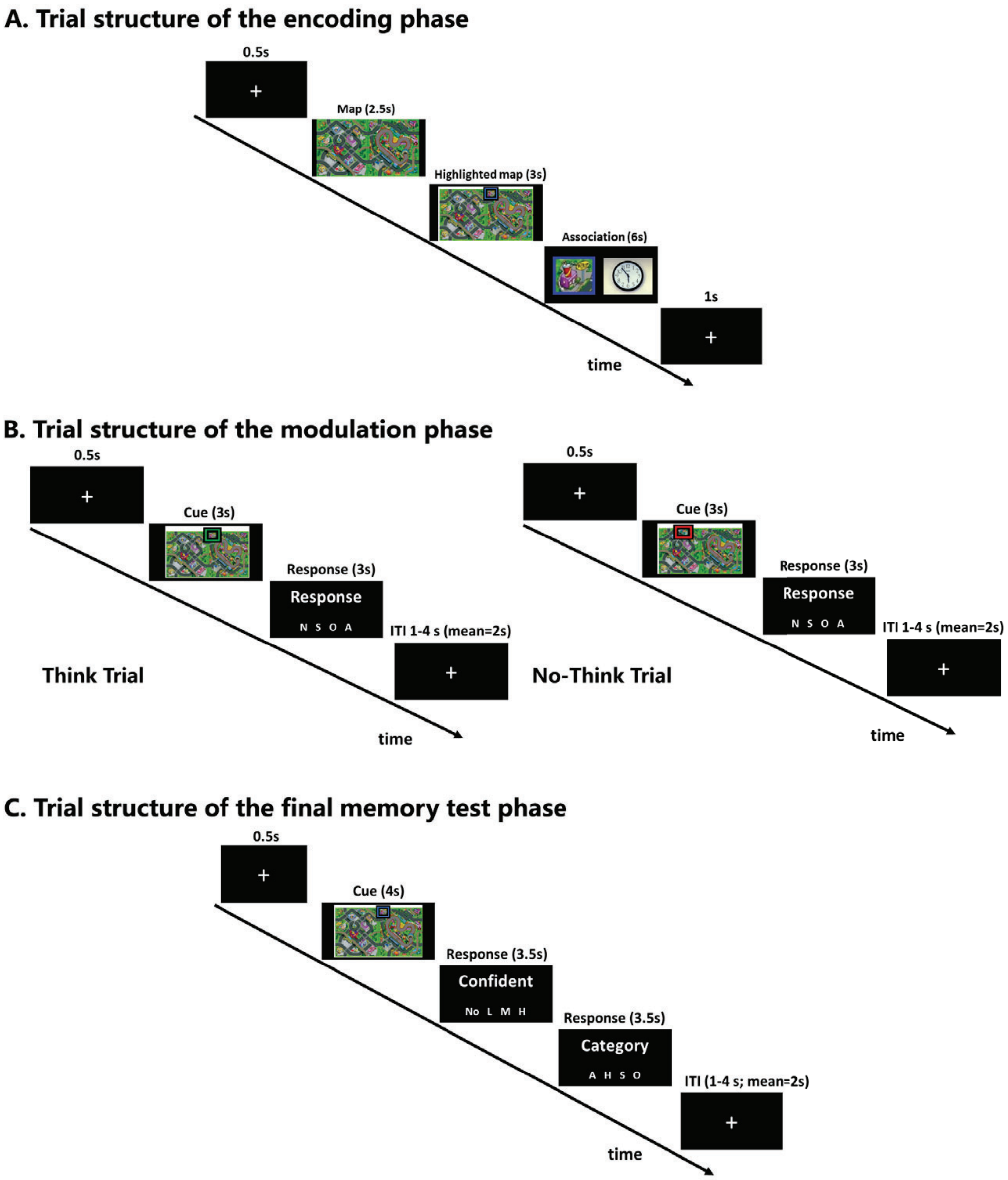
We performed three control analyses to assess the robustness of retrieval-induced increase in VVC's activity pattern variability for *RETRIEVAL ASSOCIATIONS.* Firstly, we investigated if the observed change in activity pattern variability only exists for remembered associations. We reanalyzed the activity pattern variability for all associations without considering the individual differences in the objective memory performance, and also found the higher activation pattern variability in the VVC for *RETRIEVAL ASSOCIATIONS* compared to *CONTROL ASSOCIATIONS* ($t=3.34$, $p=0.002$; **Figure S4.5B**). Then, we examined if the observed variability change depends on the arbitrary threshold (uncorrected $p_{voxel}<0.001$) used for picture-sensitive voxels selection. Results showed that increased activity pattern variability for *RETRIEVAL ASSOCIATIONS* could be also detected under two different thresholds ($p_{voxel}<0.01$, $t=2.4$, $p=0.023$; $p_{voxel}<0.05$, $t=2.41$, $p=0.022$; **Figure S4.6**). Finally, we further localized category-sensitive voxels within the VVC (**Figure S4.7B**) and calculated the activity pattern variability for these voxels separately for *RETRIEVAL ASSOCIATIONS* and *CONTROL ASSOCIATIONS.* The same contrast also revealed higher activation pattern variability (lower pattern similarity) for *RETRIEVAL ASSOCIATIONS* compared to the *CONTROL ASSOCIATIONS* ($t=2.5$, $p=0.018$).

4. Test the univariate and pattern similarity analysis in a control region

In the main text, we already showed that repeated retrieval reduced activity amplitude, but enhanced the distinctiveness of activity patterns (measured by lower pattern similarity) in the ventral visual cortex and parietal areas. However, all of the investigated ROIs locate within the brain network where memory contents are reinstated during memory retrieval.

Here, we performed a series of similar analyses to a control region that also showed a comparable univariate reduction but would not be expected to represent memory contents. We defined the control region (i.e., left precentral gyrus) based on the following steps: (1) we made a mask of left precentral gyrus based on the AAL atlas; (2) we performed the whole-brain contrast between retrieval associations and baseline associations during the final memory test, and saved voxels with $Z>2$ as a map for relative liberal univariate effects; (3) we generated the final mask below by overlapping AAL-based left precentral gyrus and the map of univariate effects (**Figure S4.6A**). Then, we performed the pattern reinstatement analysis to examine whether activation patterns of this area represent memory content during retrieval. Even for the category-based decoding, which requires less information than the item-based decoding, activation patterns of this area cannot be used to classify memory contents (group mean accuracy=27%) (**Figure S4.6B**). Next, we performed the univariate contrast and pattern similarity contrast of the control region and found the univariate activity reduction ($t=-3.56$, $p=0.001$) and the de-correlation effect ($t=-3.38$, $p=0.002$) for the retrieval associations compared to baseline associations. Finally, because we found the de-correlation effect during the final memory test, we further examined the potential gradual change of the de-correlation effect during the modulation. Unlike the VVC, the control region did not show the modulation-specific reduction in pattern similarity during the modulation phase (**Figure S4.6C**).

In sum, our control analyses cannot fully rule out the possibility that "noise correlations" could play a role in our pattern similarity analysis, but these results highlighted the advantages of our methods and study design. We showed a series of evidence from both modulation and final memory test phase to support the conclusion of dynamic adaptive visual memory representations in the VVC, but not control region.



A. Trial structure of the encoding phase

B. Trial structure of the modulation phase

C. Trial structure of the final memory test phase

Figure S4.1 Trial structure with timing for different experimental phases. (A) Trial structure of the encoding phase. After the fixation (0.5s), the entire map was presented (2.5s), and then one specific location was highlighted (BLUE) for 3s. Finally, the enlarged location was presented together with one picture for 6s. A constant ITI of 1 s was used at the end of each encoding trial. **(B)** Trial structure of the modulation phase. There are two kinds of trials during this phase (i.e., Think trial and No-Think trial) with the same trial structure. After the fixation (0.5s), one specific location was highlighted for 3s as the memory cue. According to the color of the frame (GREEN: Think trial; RED: No-Think trial) around the location, participants were instructed to actively retrieve associated pictures in mind ("think") or suppress the tendency to recall them ("no-think"). Then, participants had 3s to report their experience during the cue presentation using one of the four options (i.e., Never (N), Sometimes (S), Often (O), and Always (A)). After the rating, a fixation appeared for 1-4s (mean=2s) as the Inter-Trial Interval (ITI). **(C)** Trial structure of the final memory test phase. After the fixation (0.5s), one specific location was highlighted (BLUE) for 3s, and participants were instructed to retrieve the associated picture in their mind. Then, they rated the confidence of that memory using one of the four options (i.e., No memory (No), low confidence (L), middle confidence (M), and high confidence (H), and the category of that picture (Animal (A), Human (H), Scene (S), and Object (O)). For each rating, a maximum response window of 3.5s was given. In the end, a fixation appeared for 1-4s (mean=2s) as the Inter-Trial Interval (ITI).

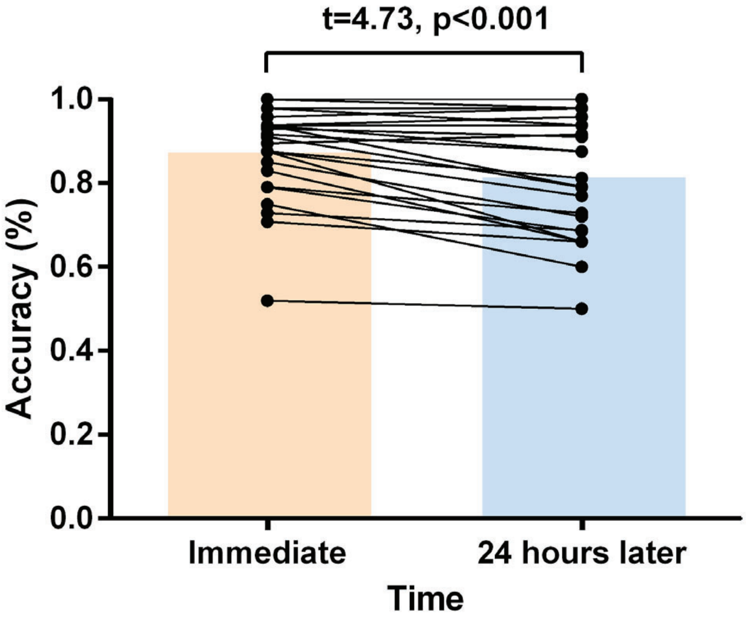


Figure S4.2 Memory performance during typing test immediately after study and 24 hours later. During the immediate typing test (day1), 88.01% of the associated pictures were described correctly (SD= 10.87%; range from 52% to 100%). Twenty-four hours later, participants could recall 82.15% of all associations (SD = 13.87%; range from 50% to 100%). Although we observed less accurate memory 24 hours later ($t(26) = 4.73$, $p < 0.001$, Cohen's $d = 0.912$), participants could still remember most location-picture associations well.

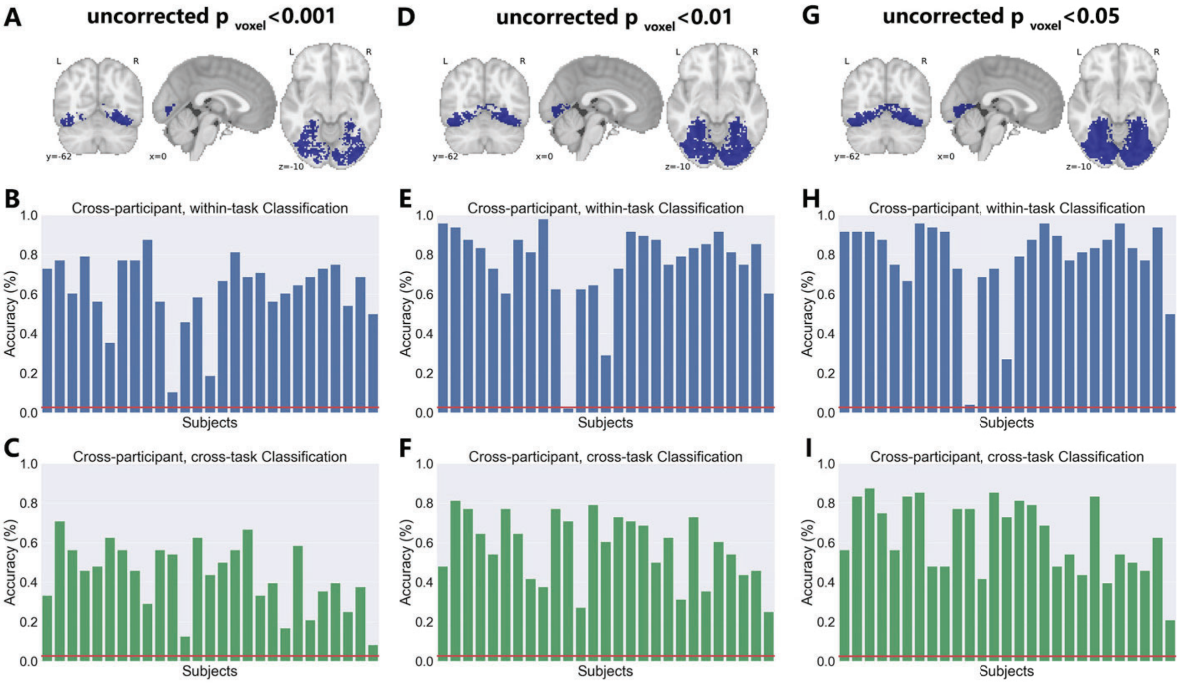


Figure S4.3 Effect of different thresholds during cluster formation on the subsequent classifications. **(A)** *Picture-sensitive voxels* within the ventral visual cortex identified by the searchlight method (uncorrected $p_{voxel}<0.001$). **(B)** *Picture-sensitive voxels* (uncorrected $p_{voxel}<0.001$) could enable the cross-participant picture classification during perception (mean accuracy=61.88%, SD=17.71%). **(C)** The same classifier can decode the memory contents based on activity patterns of *Picture-sensitive voxels* (uncorrected $p_{voxel}<0.001$) during retrieval (mean accuracy=43.13%, SD=16.52%). **(D)** *Picture-sensitive voxels* within the ventral visual cortex identified by the searchlight method (uncorrected $p_{voxel}<0.01$). **(E)** *Picture-sensitive voxels* (uncorrected $p_{voxel}<0.01$) could enable the cross-participant picture classification during perception (mean accuracy=75.54%, SD=20.42%). **(F)** The same classifier can decode the memory contents based on activity patterns of *Picture-sensitive voxels* (uncorrected $p_{voxel}<0.01$) during retrieval (mean accuracy=57.56%, SD=16.76%). **(G)** *Picture-sensitive voxels* within the ventral visual cortex identified by the searchlight method (uncorrected $p_{voxel}<0.05$). **(H)** *Picture-sensitive voxels* (uncorrected $p_{voxel}<0.05$) could enable the cross-participant picture classification during perception (mean accuracy=78.24%, SD=20.88%). **(I)** The same classifier can decode the memory contents based on activity patterns of *Picture-sensitive voxels* (uncorrected $p_{voxel}<0.05$) during retrieval (mean accuracy=63.27%, SD=17.90%). Red line represents the 95th percentile of the accuracy within 1000 randomly label-shuffled null distribution.

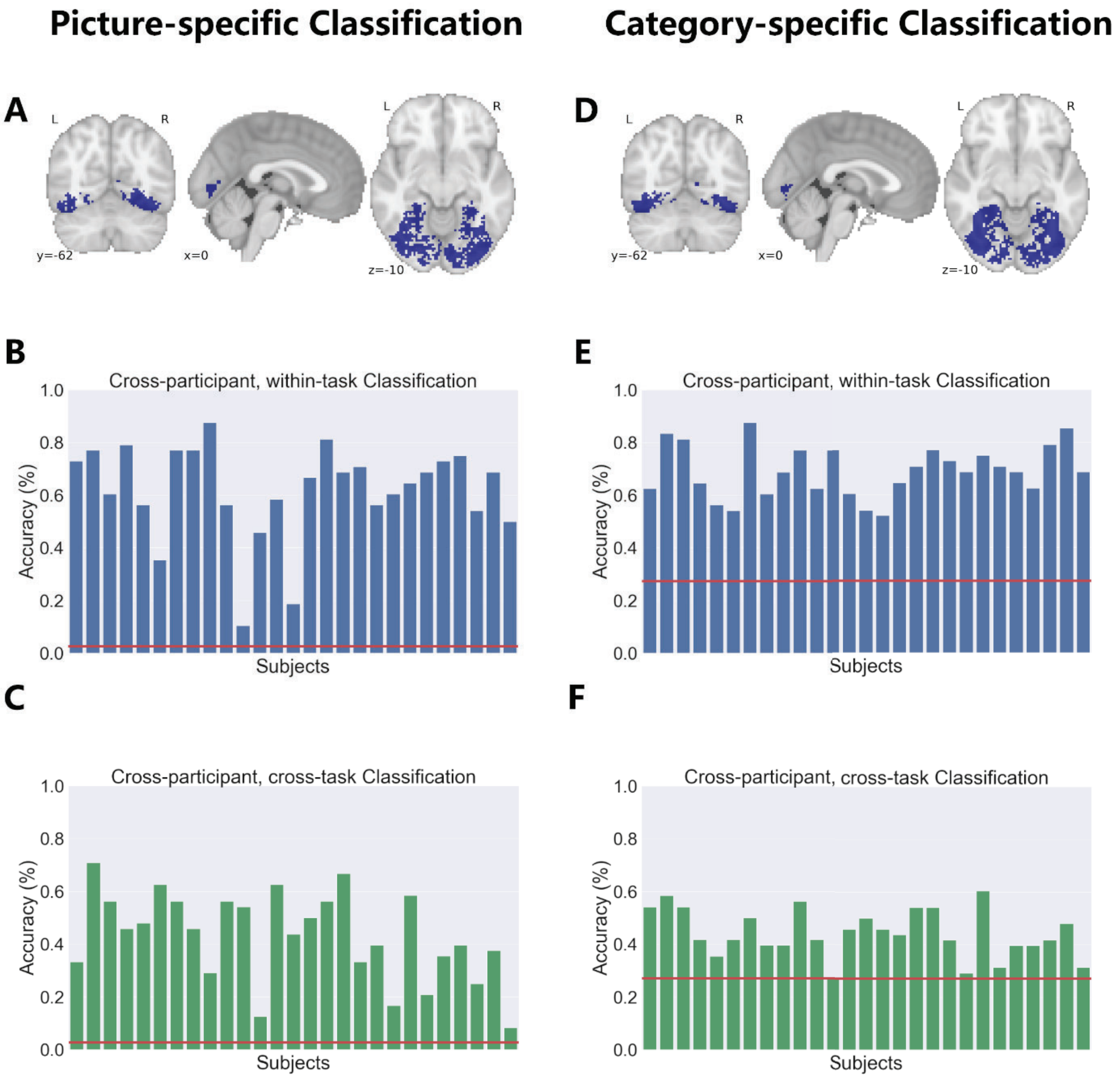


Figure S4.4 Comparison between picture-specific classification and category-specific classification. **(A)** *Picture-sensitive voxels* within the ventral visual cortex identified by the searchlight method. **(B)** *Picture-sensitive voxels* could enable the cross-participant picture classification during perception. **(C)** Activation patterns of *picture-sensitive voxels* during memory retrieval, together with the classifier for different pictures, could enable cross-participant, cross-task classification of memory contents. **(D)** *Category-sensitive voxels* within the ventral visual cortex identified by the searchlight method. **(E)** *Category-sensitive voxels* could enable the cross-participant picture classification during perception (mean accuracy=69.13%, SD=9.6%). **(F)** Activation patterns of *category-sensitive voxels* during memory retrieval, together with the classifier for different categories, could enable cross-participant, cross-task classification of memory contents (mean accuracy=44.29%, SD=8.9%). Redline represents the 95th percentile of the accuracy within 1000 randomly label-shuffled null distribution.

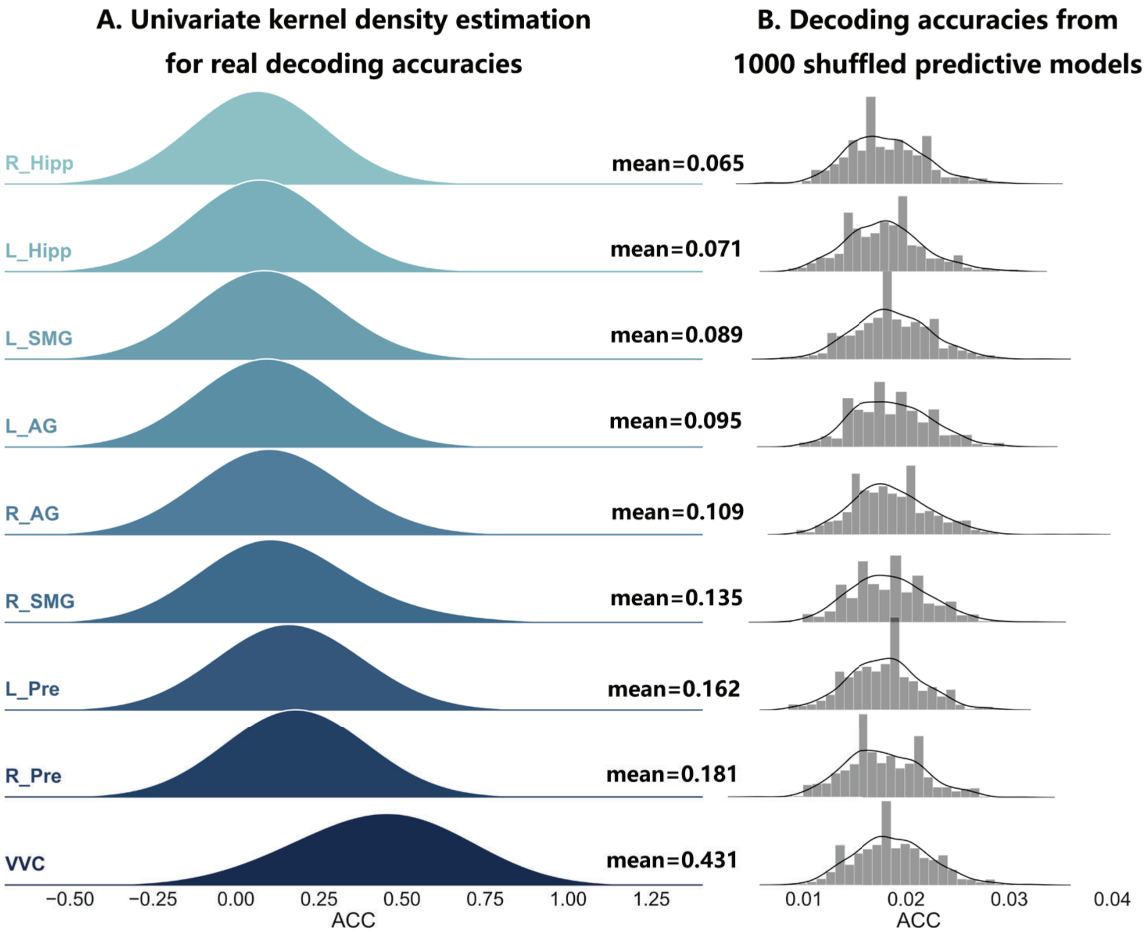


Figure S4.5 Decoding accuracies estimated by the 3-fold cross-validation across all ROIs. (A) Density estimation of decoding accuracies for each ROI. ROIs are ranked based on mean accuracies from less accurate to more accurate classification. **(B)** Distributions of decoding accuracies from 1000 shuffled predictive models. Hipp=Hippocampus; SMG=supramarginal gyrus; AG=angular gyrus; Pre= Precuneus; VVC= Ventral Visual Cortex.

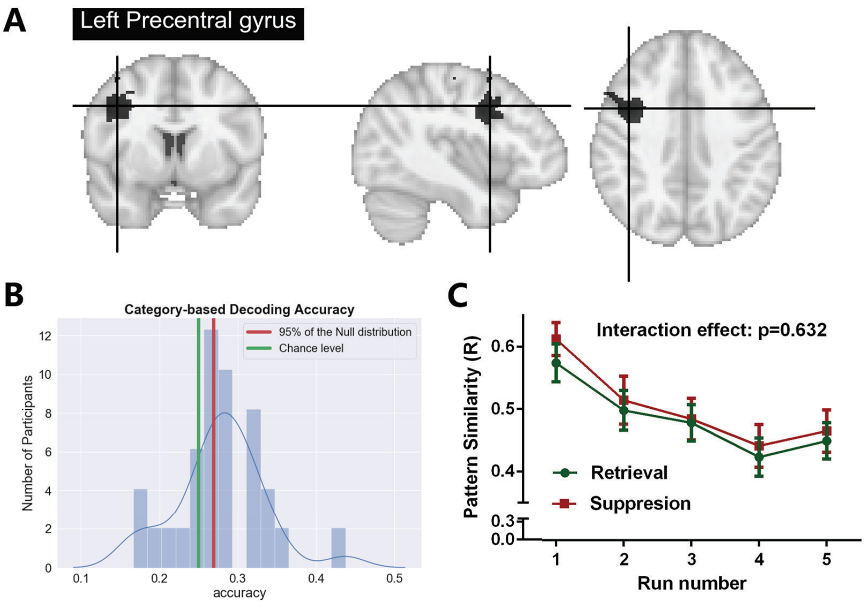


Figure S4.6 **Results from the control region (i.e., left precentral gyrus). (A)** Visualization of the mask of the control region. **(B)** Comparison between individual category decoding accuracies with theoretical chance level and 95th of the null distribution. **(C)** Retrieval and suppression modulate pattern similarity of the control region during the modulation phase in a similar way.

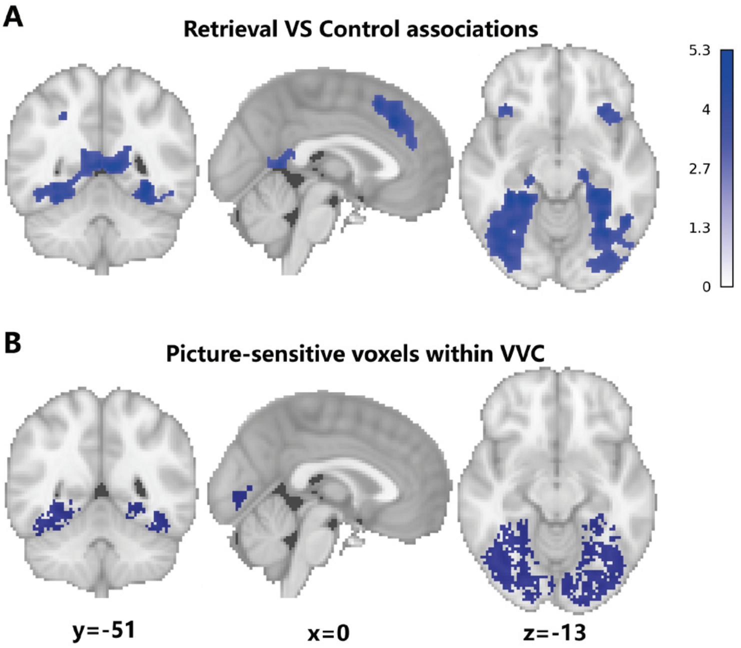

A
Retrieval VS Control associations
5.3
4
2.7
1.3
0
B
Picture-sensitive voxels within VVC
y=-51
x=0
z=-13

Figure S4.7 **Similar spatial pattern between areas showed reduced activity amplitude and picture-sensitive voxels within the ventral visual cortex. (A)** Brain regions showed less activation when retrieved *RETRIEVAL ASSOCIATIONS* compared to *CONTROL ASSOCIATIONS*. Compared to *CONTROL ASSOCIATIONS*, retrieval of *RETRIEVAL ASSOCIATIONS* was associated with less activation in the medial occipital cortex, fusiform gyrus, supplementary motor area (SMA), anterior/medial cingulate cortex (MCC), precuneus, bilateral insula, and bilateral inferior frontal gyrus (IFG) (voxelwise $_{\text{uncorrected}}\ p<0.001$, $p_{\text{FWE-cluster}}<0.05$). **(B)** Searchlight analysis identified picture-sensitive voxels within the ventral visual cortex.

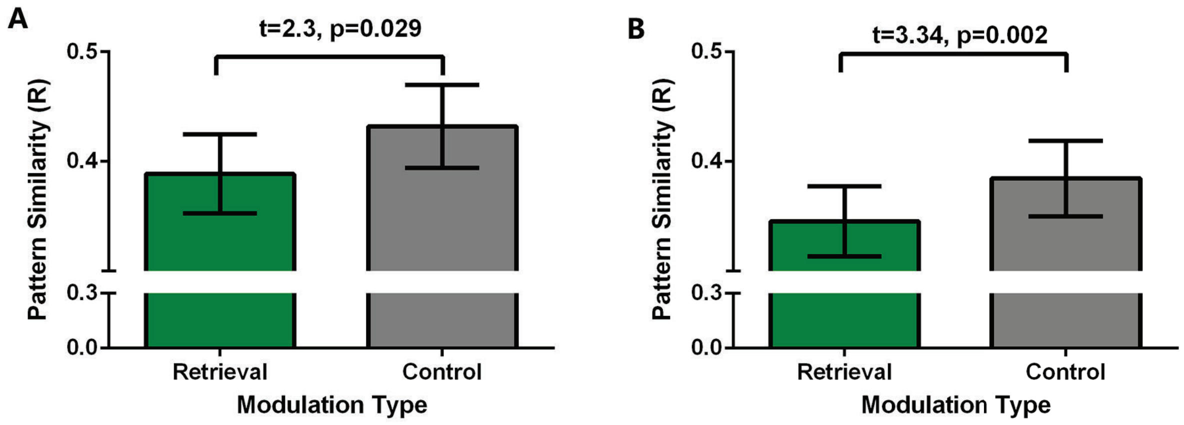


Figure S4.8 **Activity pattern variability analyses using all associations or only remembered associations.** **(A)** Using only the activation patterns of remembered associations, we found higher activation pattern variability (lower pattern similarity) in the VVC for *RETRIEVAL ASSOCIATIONS* compared to the *CONTROL ASSOCIATIONS* (t=2.3, p=0.029). **(B)** Using activation patterns of all associations, we also found the higher activation pattern variability in the VVC for *RETRIEVAL ASSOCIATIONS* (t=3.34, p=0.002).

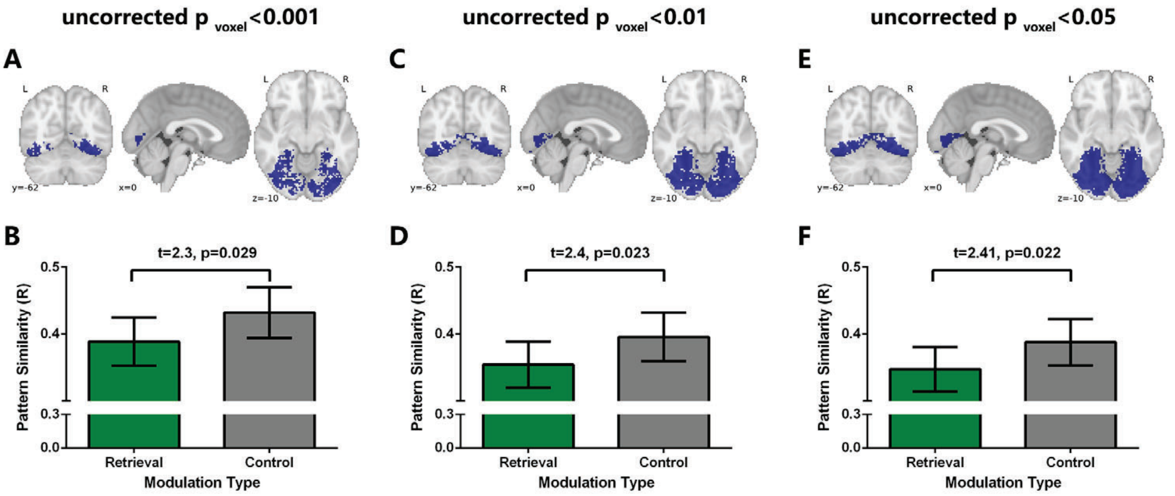


Figure S4.9 **Activity pattern variability analyses based on different cluster-formation thresholds. (A)** *Picture-sensitive voxels* within the ventral visual cortex identified by the searchlight method (uncorrected $p_{voxel}<0.001$). **(B)** Based on the VVC region-of-interest formed by threshold p<0.001, we found higher activation pattern variability (lower pattern similarity) in the VVC for *RETRIEVAL ASSOCIATIONS* compared to the *CONTROL ASSOCIATIONS* (t=2.3, p=0.029). **(C)** *Picture-sensitive voxels* within the ventral visual cortex identified by the searchlight method (uncorrected $p_{voxel}<0.01$). **(D)** Based on the VVC region-of-interest formed by threshold p<0.01, we also found higher activation pattern variability (lower pattern similarity) in the VVC for *RETRIEVAL ASSOCIATIONS* compared to the *CONTROL ASSOCIATIONS* (t=2.4, p=0.023). **(E)** *Picture-sensitive voxels* within the ventral visual cortex identified by the searchlight method (uncorrected $p_{voxel}<0.05$). **(F)** Based on the VVC region-of-interest formed by threshold p<0.01, we also found higher activation pattern variability (lower pattern similarity) in the VVC for *RETRIEVAL ASSOCIATIONS* compared to the *CONTROL ASSOCIATIONS* (t=2.41, p=0.022).

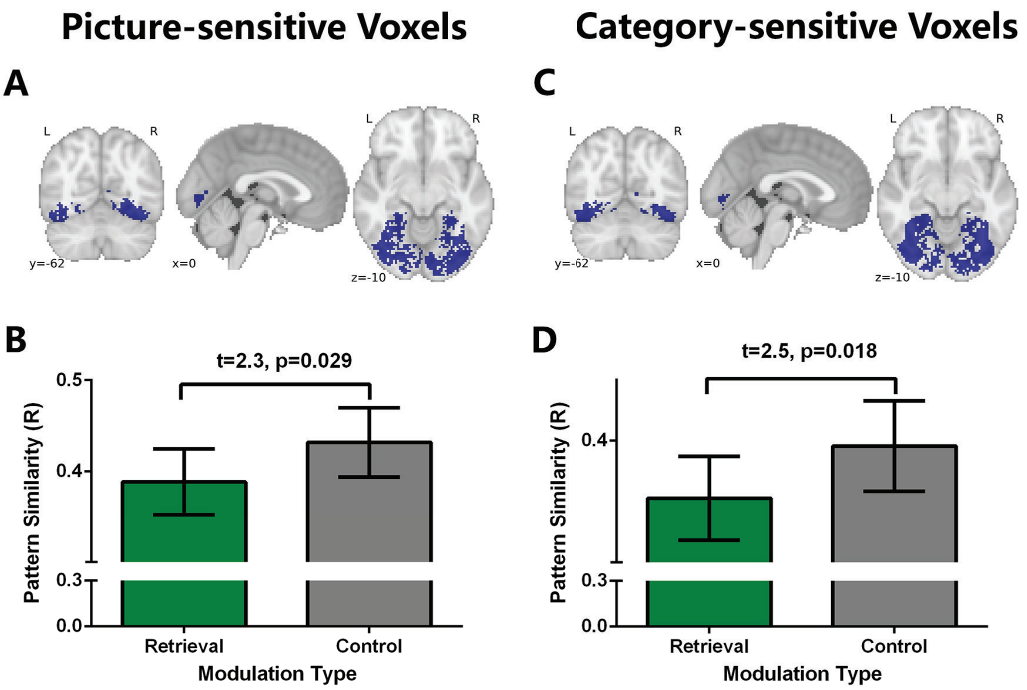


Figure S4.10 **Activity pattern variability analyses based on picture-sensitive voxels or category-sensitive voxels. (A)** *Picture-sensitive voxels* within the ventral visual cortex identified by the searchlight method. **(B)** Based on activation patterns of *Picture-sensitive voxels* within the ventral visual cortex, we found higher activation pattern variability (lower pattern similarity) in the VVC for *RETRIEVAL ASSOCIATIONS* compared to the *CONTROL ASSOCIATIONS* (t=2.3, p=0.029). **(C)** *Category-sensitive voxels* within the ventral visual cortex identified by the searchlight method. **(D)** Based on activation patterns of *Category-sensitive voxels* within the ventral visual cortex, we also found higher activation pattern variability (lower pattern similarity) in the VVC for *RETRIEVAL ASSOCIATIONS* compared to the *CONTROL ASSOCIATIONS* (t=2.5, p=0.018).

Table S1 **Comparisons of cross-participant[a] decoding accuracies between different modulations**

Region of Interest (ROI)	**Retrieval association (%)** (mean(SD))	**Suppression association (%)** (mean(SD))	**Control association (%)** (mean(SD))	**Chi-square[b]**	**p[b]**
Ventral Visual Cortex	43.0(22.2)	43.3(22.1)	45.0(21.7)	0.05	0.97
Left hippocampus	7.9(8.8)	6.5(8.7)	6.6(7.3)	0.72	0.69
Right hippocampus	7.0(6.9)	6.8(8.0)	5.9(7.3)	0.08	0.95
Left AG	8.4(8.0)	9.6(10.2)	10.5(12.4)	2.11	0.34
Right AG	10.6(11.2)	10.9(11.7)	11.3(14.9)	0.44	0.80
Left SMG	8.5(10.3)	7.6(8.2)	7.7(9.5)	0.09	0.95
Right SMG	16.0(17.2)	10.4(12.9)	15.6(16.2)	2.67	0.26
Left precuneus	17.0(13.1)	17.5(14.1)	13.7(8.5)	0.63	0.72
Right precuneus	17.8(12.1)	16.3(13.0)	22.3(9.7)	5.36	0.06

a. Decoding outputs presented in this table were calculated using the cross-participant approach: we performed three-fold cross-validation. Pattern localization data of the 2/3 participants (i.e., training sample) were used to train classifiers and final memory test data of the remaining 1/3 participants (i.e., testing sample) were used to probe memory reactivation; b. Statistical values (i.e., Chi-square values) and p values were generated by the Friedman Test. The Friedman Test is the non-parametric equivalent to the one-way ANOVA with repeated measures used to test the main effect of modulation condition on decoding outputs. SMG=supramarginal gyrus; AG=angular gyrus.

Table S2 Comparisons of cross-participant[a] decoding decision value (d) between different modulations

Region of Interest (ROI)	Retrieval association (d) (mean(SD))	Suppression association (d) (mean(SD))	Control association (d) (mean(SD))	Chi-square[b]	p[b]
Ventral Visual Cortex	38.89(5.11)	39.83(5.17)	39.49(6.51)	1.18	0.55
Left hippocampus	27.87(4.17)	27.01(3.78)	26.86(4.11)	0.96	0.61
Right hippocampus	27.96(4.34)	26.36(3.93)	26.99(3.95)	0.96	0.61
Left AG	26.53(7.49)	27.33(7.29)	26.54(7.12)	0.96	0.61
Right AG	28.55(7.84)	27.23(6.88)	27.37(7.75)	0.22	0.89
Left SMG	27.18(6.19)	27.56(7.18)	27.78(6.36)	0.22	0.89
Right SMG	32.21(7.45)	30.86(6.50)	31.94(6.84)	1.55	0.45
Left precuneus	30.20(6.55)	30.55(7.19)	31.60(4.99)	2.29	0.31
Right precuneus	30.80(4.97)	30.06(7.08)	31.56(4.15)	3.18	0.20

a. Decoding outputs presented in this table were calculated using the cross-participant approach: we performed three-fold cross-validation. Pattern localization data of the 2/3 participants (i.e., training sample) were used to train classifiers and final memory test data of the remaining 1/3 participants (i.e., testing sample) were used to probe memory reactivation; b. Statistical values (i.e., Chi-square values) and p values were generated by the Friedman Test. The Friedman Test is the non-parametric equivalent to the one-way ANOVA with repeated measures used to test the main effect of modulation condition on decoding outputs. SMG=supramarginal gyrus; AG=angular gyrus.

Table S3 Comparisons of within-participant[a] decoding accuracies between different modulations

Region of Interest (ROI)	Retrieval association (%) (mean(SD))	Suppression association (%) (mean(SD))	Control association (%) (mean(SD))	Chi-square[b]	p[b]
Ventral Visual Cortex	55.9(28.6)	53.0(28.4)	53.0(26.7)	0.84	0.65
Left hippocampus	3.0(4.7)	3.8(6.3)	2.7(4.0)	1.21	0.54
Right hippocampus	1.9(3.3)	2.7(4.0)	3.1(5.8)	0.64	0.72
Left AG	5.6(6.0)	4.8(6.8)	5.0(8.2)	0.30	0.86
Right AG	10.5(9.1)	10.5(12.6)	10.6(11.6)	0.15	0.92
Left SMG	7.8(11.9)	5.0(8.7)	7.2(9.0)	2.49	0.28
Right SMG	4.4(6.4)	3.6(5.7)	3.9(5.2)	0.25	0.87
Left precuneus	11.6(13.4)	13.9(16.2)	13.1(12.5)	0.62	0.73
Right precuneus	9.2(9.8)	13.0(13.8)	8.7(11.0)	1.37	0.50

a. Decoding outputs presented in this table were calculated using the within-participant approach: classifiers were trained on pattern localizer data of one particular participant and tested on his/her retrieval data; b. Statistical values (i.e., Chi-square values) and p values were generated by the Friedman Test. The Friedman Test is the non-parametric equivalent to the one-way ANOVA with repeated measures used to test the main effect of modulation condition on decoding outputs. SMG=supramarginal gyrus; AG=angular gyrus.

Table S4 Comparisons of within-participant[a] decoding decision value (d) between different modulations

Region of Interest (ROI)	**Retrieval association (d)** (mean(SD))	**Suppression association (d)** (mean(SD))	**Control association (d)** (mean(SD))	**Chi-square**[b]	**p**[b]
Ventral Visual Cortex	40.64(7.21)	40.67(6.20)	40.52(7.10)	0.07	0.96
Left hippocampus	25.30(4.10)	26.96(4.69)	24.42(3.80)	1.40	0.49
Right hippocampus	23.91(3.97)	23.67(3.55)	26.94(5.02)	2.74	0.25
Left AG	26.99(5.20)	29.04(5.05)	28.42(6.23)	3.18	0.20
Right AG	31.24(6.25)	30.82(7.45)	31.57(5.25)	0.22	0.89
Left SMG	28.83(6.19)	27.68(6.39)	28.57(4.85)	0.96	0.61
Right SMG	30.23(4.72)	30.54(4.19)	30.51(6.00)	0.29	0.86
Left precuneus	32.58(7.08)	32.18(6.38)	31.40(7.26)	1.40	0.49
Right precuneus	31.11(6.81)	30.65(7.64)	30.68(5.94)	0.51	0.77

a. Decoding outputs presented in this table were calculated using the within-participant approach: classifiers were trained on pattern localizer data of one particular participant and tested on his/her retrieval data; b. Statistical values (i.e., Chi-square values) and p values were generated by the Friedman Test. The Friedman Test is the non-parametric equivalent to the one-way ANOVA with repeated measures used to test the main effect of modulation condition on decoding outputs. SMG=supramarginal gyrus; AG=angular gyrus.

Table S5 **Brain regions showed less activation when retrieved *RETRIEVAL ASSOCIATIONS* compared to *CONTROL ASSOCIATIONS.***

Brain region	Hemisphere	Peak MNI coordinates	Cluster size (mm³)	Cluster mean
IFG	L	-36 42 8	184	-3.23
Left DLPFC	L	-46 46 6	224	-3.20
Precuneus	L	-8 -80 46	360	-3.23
Precentral gyrus	L	-38 2 38	3024	-3.56
Insula	R	32 26 -6	9528	-3.61
Insula	L	-30 26 -2	10936	-3.69
ACC/MCC/SMA	R/L	2 28 48	11712	-3.68
Medial occipital cortex/ Fusiform gyrus	R/L	32 -46 -8	71664	-3.66

IFG= Inferior Frontal Gyrus; DLPFC= Dorsolateral Prefrontal Cortex; ACC= Anterior Cingulate Cortex; MCC= Middle Cingulate Cortex; SMA= Supplementary Motor Area

Table S6 Comparison of pattern variability change between participants who showed strong and weak suppression

Region of Interest (ROI)	**Pattern Variability change (Strong)** (mean (SD))	**Pattern Variability change (Weak)** (mean (SD))	**t**	**p**	**d**
Ventral Visual Cortex	-0.02(0.06)	-0.006(0.08)	-0.644	0.526	-0.253
Left hippocampus	-0.005(0.02)	0.009(0.01)	-0.612	0.546	-0.246
Right hippocampus	-0.0005(0.03)	-0.0004(0.02)	-0.01	0.99	-0.005
Left AG	-0.015(0.04)	0.013(0.06)	-1.288	0.21	-0.505
Right AG	-0.03(0.09)	0.02(0.08)	-1.57	0.129	-0.616
Left SMG	-0.007(0.04)	-0.015(0.05)	0.4	0.693	0.157
Right SMG	-0.012(0.06)	-0.011(0.04)	-0.03	0.973	-0.013
Left precuneus	-0.02(0.04)	0.01(0.05)	-1.82	0.073	-0.734
Right precuneus	-0.028(0.06)	0.002(0.04)	-1.35	0.189	-0.53

Pattern Variability change=pattern variability of suppression associations minus variability of control associations; Strong=the group of participants who showed stronger suppression (i.e. more negative suppression slope); weak= the group of participants who showed weaker suppression (i.e. less negative suppression slope); SMG=supramarginal gyrus; AG=angular gyrus.

Chapter 5

Common neural and transcriptional correlates of inhibitory control underlie emotion regulation and memory control

This chapter is based on:
Liu. W., Peeters, N., Fernandez, G., & Kohn, N. (2020). Common neural and transcriptional correlates of inhibitory control across emotion, memory, and response inhibition. **Soc. Cogn. Affect. Neurosci.**

Abstract

Inhibitory control is crucial for regulating emotions, and may also enable memory control. However, evidence for their shared neurobiological correlates is limited. Here, we report meta-analyses of neuroimaging studies on emotion regulation, or memory control, and link neural commonalities to transcriptional commonalities using the Allen Human Brain Atlas (AHBA). Based on 95 fMRI studies, we reveal a role of the right inferior parietal lobule embedded in a frontal-parietal-insular network during emotion and memory control, which is similarly recruited during response inhibition. These co-activation patterns also overlap with the networks associated with "inhibition", "cognitive control", and "working memory" when consulting the Neurosynth. Using the AHBA, we demonstrate that emotion and memory control-related brain activity patterns are associated with transcriptional profiles of a specific set of "inhibition-related" genes. Gene ontology enrichment analysis of these "inhibition-related" genes reveal associations with the neuronal transmission and risk for major psychiatric disorders as well as seizures and alcoholic dependence. In summary, this study identified a neural network and a set of genes associated with inhibitory control across emotion regulation and memory control. These findings facilitate our understanding of the neurobiological correlates of inhibitory control and may contribute to the development of brain stimulation and pharmacological interventions.

Key Words: emotion regulation; memory control; inhibitory control; gene expression; transcriptional network

Introduction

One of the most influential cognitive theories of emotion regulation proposes that it is primarily supported by inhibitory control (Gross & Thompson, 2007; Ochsner & Gross, 2005). A potentially related cognitive process, memory control, is defined as an ability to actively reduce the accessibility of memories. Memory control is also thought to be supported by inhibitory control (Michael C. Anderson & Hanslmayr, 2014; Michael C Anderson, 2004; Michael C Anderson & Green, 2001). Recently, Engen and Anderson summarized the conceptual link between emotion regulation and memory control and raised the question of whether there is a common neurobiological mechanism supporting these two processes (Engen & Anderson, 2018). Here, we set out to empirically demonstrate the neurobiological commonalities between the two processes by first analyzing fMRI data and then linking neuroimaging findings with post mortem gene expression data.

The idea that inhibitory control, as the fundamental cognitive process, supports other higher-level processes such as emotion regulation and memory control is supported by behavioral and neural evidence. Behaviorally, a positive correlation between emotion regulation and memory control performances has been found (B. E. Depue et al., 2016). Human neuroimaging results showed overlapping recruitment of right superior medial frontal gyrus (rSFG) and right inferior frontal gyrus (rIFG) in both emotion regulation (Buhle et al., 2014; Kohn et al., 2014; Ochsner et al., 2002) and memory control (Michael C Anderson, 2004; Guo et al., 2018). Beyond these frontal regions, the parietal cortex is another candidate region that may play a critical role in both emotion regulation and memory control. Starting from the idea of the multiple-demand system in the brain (Duncan, 2010), both the evidence from task-fMRI (Fedorenko et al., 2013) and resting-state fMRI (Dosenbach et al., 2007; Power et al., 2011) suggest that a fronto-parietal network supports the initiation of top-down control and control adjustment in response to task-goals and feedbacks.

No previous formal meta-analysis of human neuroimaging studies has investigated the neural commonalities between emotion regulation and memory control to pinpoint the overlap in inhibitory control, although common neural correlates between emotion regulation or memory control with response inhibition tasks (e.g., Stop-signal and Go/Nogo task) have been investigated in two recent meta-analyses (Guo et al., 2018; Langner et al., 2018). Using activation likelihood estimation (ALE), we aim to demonstrate that emotion regulation and memory control evoke activation of similar brain regions with a spatial pattern that is similar to the activation of typical response inhibition paradigms, including Stop-signal and Go/Nogo tasks. Moreover, beyond overlapping regional brain activity, we also are interested in overlapping co-activation patterns of associated brain regions. We used meta-analytic connectivity modeling (MACM) to test whether these brain regions act as an interconnected functional network. We expected to find a similar set of co-activated brain regions that are associated with both emotion regulation and memory control, defining a core “inhibition-related” network.

Identifying an underlying neural network is relevant for potentially brain stimulation interventions including but not limited to Transcranial Magnetic Stimulation (TMS) and Transcranial Direct Current Stimulation (tDCS). However, further dissection of molecular underpinnings is necessary to expand our understanding of inhibition control, which in turn may pave the way for pharmacological interventions. Therefore, we reasoned whether emotion regulation and memory control are not only related to a common neural network defined by activity and connectivity, but also by associated similarities in spatial transcriptional profiles. Conventional imaging-genetic methods including candidate gene methods and genome-wide association approaches cannot reveal the relationship between localized gene transcription and task-related brain activity, even though they already showed the association between multiple common gene variants and neuroimaging measures (Elliott et al., 2018; Hariri et al., 2002). Thus, to explore the relationship between spatial transcriptional profiles and emotion regulation or memory control-related brain activity, we adopted a recently developed approach to associate spatial maps of gene expression in post-mortem brain tissues with brain activation measures (K. Gorgolewski et al., 2014; X. Kong et al., 2017). Spatial pattern analysis of gene expression maps of the Allen Human Brain Atlas (AHBA) together with neuroimaging revealed fundamental features of transcriptional regulation (See review by Fornito, Arnatkevičiūtė, & Fulcher, 2018) and related disruptions in brain disorders (Grothe et al., 2018; McColgan et al., 2018; Romero-Garcia et al., 2019; Romme et al., 2017). However, the correspondence between spatial transcriptional profiles and neural functionality has not yet provided full insight into how genetic correlates are linked to core cognitive abilities (e.g., inhibitory control in this study). Based on the assumption that spatial transcriptional profiles not only co-vary with the connectional architecture but also support the task-evoked, synchronous brain activity (Berto et al., 2018; K. Gorgolewski et al., 2014; X. Kong et al., 2017; Shine et al., 2019) we expected to find similar spatial transcriptional profiles between emotion regulation, memory control, and response inhibition.

To investigate the neural and transcriptional commonality of emotion regulation and memory control, we combined task fMRI data, neuroimaging meta-analytic approaches, and postmortem gene-expression data. First, we used the ALE method to generate the task activation map for each paradigm of interest (e.g., emotion regulation or memory control) based on 95 published task fMRI studies with in total 1995 healthy participants (**Figure 5.1A**). Then, we used the brain regions identified by the ALE as seed regions and estimated the meta-analytic connectivity map (or co-activation map), using information from BrainMap (**Figure 5.1B**). And to explore the behavioral relevance of these findings, we examined the associations of these co-activation patterns in a data-driven way via Neurosynth with behavioral domains (**Figure 5.1C**). Next, we calculated the spatial association between task activation maps and human gene expression maps to identify their common spatial transcriptional profiles (**Figure 5.1D**). Finally, we performed a systematic and integrative analysis of the resulting gene list to gain an in-depth understanding of putative biological functions and disease associations of the identified "inhibition-related "gene set (**Figure 5.1E**).

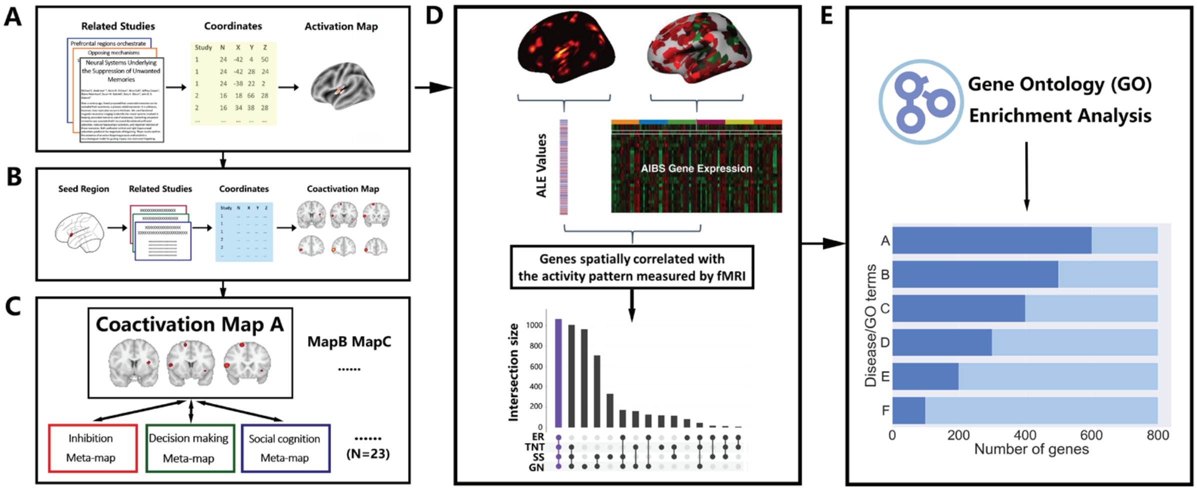


Figure 5.1 Schematic of the pipeline for the investigation of neural and transcriptional commonality. **(A)** Activation likelihood estimation (ALE) meta-analyses of functional MRI studies. MRI coordinates of reported brain activations from relevant studies were used to generate the activation map for each task of interest. **(B)** Meta-analytic connectivity modeling of seed regions resulting from the ALE analyses was conducted to search co-activated regions over all studies in BrainMap. Coactivation maps were estimated via ALE. **(C)** To generate behavioral profiles of co-activation maps, we used Neurosynth to acquire a list of meta-maps, including the behavioral terms from low-level to high-level cognitive processes (total Number of terms=23). Next, each co-activation map was compared to all meta-maps (e.g., inhibition, decision making, and social cognition) to compute the similarity index for each meta-map. **(D)** The activation patterns were associated with gene expression maps from the Allen Human Brain Atlas (AHBA). Each gene expression map was vectorized based on the expression measures within the three-dimension maps. The ALE values at the corresponding brain regions on the activation map were extracted and vectorized. To identify the genes whose spatial patterns are similar to a specific activation map, the similarity between all vector pairs were quantified. "Inhibitory-related" genes were defined by identifying genes of which the expression patterns are correlated with the activation maps of four paradigms (ER, Emotion Regulation; TNT, Think/No-Think; SS, Stop-signal; GN, Go/No-go). **(E)** "Inhibition-related" genes were associated with biological functions (GO terms) and disease terms for the interpretation using Gene Ontology Enrichment Analysis (GOEA).

Materials and Methods

Literature searches, selection, and coordinates extraction

In total, we performed literature searches for four task paradigms (think/no-think, emotion regulation, stop-signal, and go/no-go) and one network ("default mode network"). To avoid biases, we used the following inclusion criteria during the search.

(1) We only included data from studies on healthy adults with no prior report of neurological, medical, or psychiatric disorders in the current meta-analysis, while results of patients or specific sub-group effects (e.g., gender differences) were not included. Articles including patients were only selected if they reported results for a control group separately, and only the control group was included here.

(2) Only neuroimaging studies, which used whole-brain fMRI and reported coordinates for brain activation or deactivation in standard anatomical reference space (Talairach/Tournoux; Montreal Neurological Institute (MNI)) were considered. Coordinates originally published in Talairach space were converted to MNI space using the algorithm implemented in GingerALE 3.0.2 (http://www.brainmap.org/ale/)

(3) Only studies reporting results of General Linear Model (GLM) contrasts were included, while studies focusing on functional connectivity, structural, resting-state, or brain-behavior correlations were excluded.

(4) Only studies reporting whole-brain analyses were included, while studies based on partial coverage or employing only region-of-interest analyses were excluded.

Detailed search and extraction procedures were as following for each paradigm:

Think/No-Think studies

A step-wise procedure was used to search articles, published before February 2020, using functional MRI to investigate brain activity during Think/No-Think paradigm. First, we used standard search in PubMed and ISI Web of Science to perform the search. More specifically, we used the combination of following keywords during the search: “memory regulation”, “memory control”, “memory suppression regulation”, “memory inhibition”, “think/no-think”, “fMRI”, “neuroimaging”, “functional magnetic resonance imaging”, or “functional MRI”. At the same time, we carefully exclude studies using the “directed forgetting” paradigm, which targets the memory control during the encoding. Next, two lab members compared the search results with a recent review article (Michael C. Anderson & Hanslmayr, 2014) to find additional relevant studies. The same two lab members independently extracted the coordinates and other essential information (e.g., sample size, type of stimulus) extraction based on the identified think/no-think literature and then cross-validated the coordinates. In summary, this search and inclusion/ exclusion criteria led to 15 think/no-think studies (491 subjects and 256 foci).

Emotional Regulation Studies

We used the databases of previously published meta-analyses on emotion regulation (Kohn et al., 2014; Morawetz et al., 2017). We used the keywords: “emotion regulation”, “affective regulation”, “implicit emotion regulation”, “explicit emotion regulation”, “interpersonal emotion regulation”, “extrinsic emotion regulation, “intrinsic emotion regulation”, “reappraisal”, “suppression”, “distraction”, “detachment”, “labelling”, “affective labelling”, “reinterpretation”, “rumination”, “fMRI”, “neuroimaging”, “functional magnetic resonance imaging”, or “functional MRI”. In the case that a study did not report the contrast of interest for this meta-analysis, the corresponding authors were contacted and asked to provide more information on their data. The term “experiment” refers to any single contrast analysis, while the term “study” refers to a scientific publication, usually reporting several contrasts, i.e. experiments. This search and the employed inclusion/exclusion criteria led to a total inclusion of 107 studies from peer-reviewed journals by July 31st, 2017 (385 experiments, 3204 participants).

Each experiment was manually coded by the authors of the previous meta-analysis (C.M. and N.K) with terms that described the experimental design with respect to contrast, stimulus type utilized, emotion regulation strategy, goal of the strategy, valence of the stimuli, tactics of the strategy and the task nature. To achieve a more appropriate comparison between think/no-think and emotion regulation, we restricted inclusion to ER studies that used the “suppression” or “distraction” strategy. This led to a similar amount of studies compared to memory control and most crucially suppression or distraction of emotions is conceptually also closer to the process of suppressing memories. These criteria led to the inclusion of 15 emotion regulation studies (387 subjects and 165 foci).

Go/No-go and Stop-signal studies

A similar procedure was used to search published whole-brain functional MRI studies using the “Go/Nogo” and “Stop-signal” paradigm. To confirm the completeness of our search, we compared our results with used studies in a recent meta-analysis of motor inhibitory and memory control (Guo et al., 2018). Again, two lab members (W.L and N.P) performed the coordinates and study information extraction for the Go/No-go and Stop-signal studies.

Default Mode Network (DMN) (task-negative network) identification

We also performed a coordinated-based meta-analysis to identify the DMN. Instead of manually searching related studies and extracting coordinates, we used the BrainMap database (Fox & Lancaster, 2002; Angela R Laird et al., 2005) to find peak coordinates of task-independent deactivation reported in neuroimaging literature. This method was used before by Laird and colleagues (A. R. Laird et al., 2009) to identify the core regions in DMN. More specifically, we searched the BrainMap for all contrasts that were labeled as “deactivation” and “low-level control” during submission. “Deactivation” refers to contrasts in which stronger signal was observed during a baseline condition than during task condition (e.g., Control-Task); “low-level control" are conditions in which either fixation or resting was defined as the baseline. Our search was further limited to "normal mapping", which means that participants who were diagnosed with disease or disorders were excluded. In total, 105 studies (1588 foci) matched our search criteria and were used in the following analyses.

Activation Likelihood Estimation (ALE) analyses

The ALE analyses were based on the revised ALE algorithm (Eickhoff et al., 2009) in GingerALE 2.3. Firstly, two separate meta-analyses were conducted for the Think/No-Think (contrast: No-Think vs. Think) and emotion regulation (contrast: Regulation vs. View) tasks using cluster-level inference (FWE cluster-level correction $p<0.05$, uncorrected cluster-forming threshold $p<0.001$, threshold permutations=1000). Secondly, contrast analyses (Eickhoff et al., 2011) were conducted between the Think/No-Think and emotion regulation tasks. In these contrast analyses, the thresholded activation maps from the two separate analyses, as well as the pooled results from both tasks, were used as inputs. Conjunction and contrast maps between the conditions were given as output. For the output images, the same cluster-level threshold correction was used (FWE cluster-level correction $p<0.05$, uncorrected cluster-forming threshold $p<0.001$, threshold permutations=1000).

Additionally, meta-analyses of published fMRI studies using the Stop-signal and Go/No-go paradigms were also performed. The same software (GingerALE 2.3) and threshold (cluster-level $p<0.05$) was used to perform the analysis. Due to the unbalanced number of studies and subjects included, no conjunction or contrast analyses were performed for four tasks to identify the overlap.

Co-activation analyses using BrainMap

We conducted the Meta-Analytic Connectivity Modeling (MACM) analyses on the regions from the ALE meta-analysis. More specifically, for each ROI, we used the BrainMap database (A R Laird et al., 2009; Angela R Laird et al., 2011) to search for experiments that also activated the particular ROI. Next, we retrieved all foci reported in the identified experiments. Finally, ALE analyses were performed over these foci to identify regions of significant convergence. Sequentially, raw co-activation maps were corrected for multiple comparisons (Voxel-wise False Discovery Rate(FDR)<0.05. All clusters sizes$>$ 200mm^3)

Functional profiles of the co-activation maps

To assess associated functional terms of the co-activation maps generated by MACM, we used the NeuroSynth meta-analytic database (www.neurosynth.org) (Yarkoni et al., 2011). We followed the methodology (See Meta-analytic Functional Gradients section) used in a previous study to assess topic terms associated with the principal connectivity gradient in the human brain (Margulies et al., 2016). More specifically, we conducted a term-based meta-analysis for the same list of NeuroSynth topic terms as Margulies and colleagues did. This list covered well-studied functional terms from low-level cognition (e.g., visual perception, and auditory) to high-level cognition (e.g., language, and rewarding). Sequentially, we examined the association between these term-based activation maps with our four sets of co-activation maps (emotion regulation, think/no-think, stop-signal, go/no-go). For each co-activation map, a spatial similarity index (r statistic) between each co-activation map and each meta map of the functional term was provided. The terms were then ordered based on the average correlation for interpretation and visualization.

Activation-gene expression association analysis

We used a recently developed activation-gene expression association analysis to link the task-related brain activity to gene expression in postmortem human brains. This analysis can identify a list of associated genes based on the MRI-space statistical map(s). This analysis presupposes that if certain gene(s) are associated with the cognitive task of interest, then spatial distributions of their expression values and task-related activation pattern measured by functional MRI should be similar.

The Allen Human Brain Atlas (http://www.brainmap.org) was used in the gene expression decoding analysis. The atlas provided genome-wide microarray-based gene expression data based on six postmortem human brains (Gene expression level of over 62000 gene probes for around 1000 sampling sites across the whole brain were recorded) (Hawrylycz, Lein, Guillozet-Bongaarts, Shen, Jones, et al., 2012). Additionally, structural brain imaging data of each donor was collected and provided, which enable users to visualize gene expression in its naive space and perform the registration to the standard MRI MNI space.

Previous studies used slightly different statistical methods to associate task-independent MRI-based brain measures (e.g., cortical thickness, functional or structural networks) with the gene expression data (Richiardi et al., 2015; Seidlitz et al., 2018; Wang et al., 2015). We used the method developed by Gorgolewski and colleagues implemented in the alleninfo tool (K. Gorgolewski et al., 2014) (https://github.com/chrisfilo/alleninf). This method was originally designed for the association analysis between voxel-vise statistical maps and gene expression maps. This method was also by default implemented in Neurovault (K. J. Gorgolewski et al., 2015) (https://neurovault.org/). The method has two important features: (1) nonlinear coregistration of the donor's brain with MNI space was allowed and (2) the ability to use a random-effects model makes it possible to generalize the results to the whole population. The activation-gene expression association analysis works as following: (1) data from gene probes were aggregated for each gene, resulting in 20787 gene expression maps. (2) For each gene expression map, MNI coordinates of each sampling location (the locations in which brain tissues were analyzed for the gene expression data) were extracted to draw a spherical ROI (r=4mm). We used these ROIs to extract the average values of the ALE statistical map within each ROI. Next, the gene expression and meta-analysis vectors were correlated. (3) This extraction and correlation procedure was repeated for each gene expression map to quantify the spatial pattern similarity between the statistical maps and gene expression map. (4) Different threshold (i.e., 500, 1000, 1500, 2000) was implemented to generate the significantly associated gene list(s) among the 20787 genes. We mainly presented results under the threshold of 2000 most significant genes, but results under other thresholds can be found in the *Supplemental Material*. Since the negative correlation between brain measures and gene expression is difficult to explain, we only considered the positively correlated genes. Additionally, because of fundamental differences in gene expression between cortical and subcortical regions, we only performed our analyses within the cortical regions.

The described association analysis takes MRI statistical map(s) as input(s) and will output a list of significantly associated genes. To investigate the common transcriptional signatures (associated genes) of emotion regulation and memory control, we used the unthresholded statistical maps from the ALE analyses to identify the task-associated gene list via the association algorithm. Next, the common gene list was generated by overlapping the emotion regulation-associated gene list and memory control-associated gene list. To further investigate whether the two gene lists significantly overlapped with each other, we generated the null distribution of the number of overlapping genes by creating two gene lists (with the same size as the real gene lists) from the 20787 genes. To control for autocorrelation, we generated 5000 lists in this way and for each doner, we estimated the overall spatial similarity between identified genes (i.e., emotion regulation-related and memory control-related genes) for each donor (**Table S5.3**). Identified genes demonstrated modest spatial correlations (mean=0.063, standard deviation=0.27) across different thresholds and donors. Based on these correlation values, 1718 out of 5000 pairs of randomly generated gene lists with a similarity ranging from 0.05 to 0.07, and the standard deviation ranging from 0.25 to 0.29 were further selected. The significance of overlapping genes was estimated by comparing the real number of overlapping genes with the number of overlapping genes within these 1718 pairs.

An alternative method (i.e., “permutated” statistical maps) was also used to evaluate the significance of the overlap further. We permutated the spatial distribution of the fMRI statistical maps used in the activation-gene expression association analysis for 100 times. Voxel-wise statistical values that were stored at the real maps (i.e., emotion regulation or memory control map) were relocated to other locations, testing for spatial specificity of the original task-related spatial patterns, but containing all voxel-wise values. It is notable that this relocation procedure was restricted by the functional-defined brain parcellation (Schaefer et al., 2018). In this way, intrinsic connectivity associations that might underlie functional networks and brain parcels are retained. These "permutated" statistical maps were used for the activation-gene expression association analysis to generate 100 pairs of gene lists. Before the overlapping calculation, we also calculated overall spatial similarity with each pair. Seventy-two out of 100 pairs of these gene lists had a similarity ranging from 0.05 to 0.07 (mean=0.06) with a standard deviation ranging from 0.25 to 0.29 (standard deviation=0.27). These results were used as the null distribution to estimate the p-value of the overlap. To identify the "inhibition-related" genes, we first used the same activation-gene expression association analysis to identify the stop-signal and go/no-go-associated gene list and then defined the overlap between four gene lists (emotion regulation, think/no-think, stop-signal, and go/no-go) as "inhibitory-related" genes.

We aim to improve the specificity of our activation-gene expression association analysis and safeguard the possibility that these identified genes may only support the general functional brain network architecture instead of particular cognitive functions. To rule out this possibility, we performed a dedicated control analysis: (1) An unthreholded DMN ALE map was used as a comparison with four inhibitory control tasks because control processes hardly take place in the DMN and the DMN has been associated with homeostasis and undirected thought or mind-wandering. (2) A pair of "permutated" statistical maps with comparable spatial similarity level (mean ranging from 0.05 to 0.07, standard deviation ranging from 0.25 to 0.29) was also used to calculate the transcriptional overlap with four inhibitory control tasks.

We used a range of different numbers of genes (***x*** from 1 to 2000, with step size 10) as thresholds to identify the top ***x*** most similar genes. We calculated three kinds of overlap under different thresholds: the first one is "within control" overlap, which is the overlap in gene expression association between two inhibitory tasks (e.g., think/no-think and emotion regulation, or stop-signal and go/no-go); the second one is "inhibitory & DMN" overlap, which is the transcriptional overlap between one of the control-related tasks and DMN (e.g., think/no-think and DMN, or emotion regulation and DMN). The third one is the "inhibition & permutation" overlap, which is the overlap of associated gene lists between one of the actual statistical maps (i.e., emotion regulation or memory control) and one of the permutated statistical maps. Finally, we averaged the percentage of each kind of overlap for a certain threshold x and compared the average percentages across all thresholds.

Gene Ontology Enrichment Analysis

The Gene Ontology(GO) is a widely used bioinformatics tool to interpret the complex gene list based on the knowledge regarding functions of genes and gene products (Ashburner et al., 2000; D. W. Huang et al., 2009). To systematically investigate the biological meaning of the "inhibitory- related" genes, we use GO to perform a binary version of the overrepresentation test for Biological Processes (BP), Molecular Function (MF), and Cellular Component (CC). We did not use additional parameters to restrict the selection of Go categories. Currently, experimental findings from over 140 000 published papers are represented as over 600 000 experimentally-supported GO annotations in the GO knowledgebase. For the input gene list, "Gene Ontology enrichment analysis" can identify relevant groups of genes that function together, and associate and statistically test the relationship between these groups and Go annotations. In this way, it can reduce a big list of genes to a much smaller number of biological functions (BP, MF, or CC), and make the input list more easily comprehendible. This method has been applied successfully to understand the output of many gene expression studies, including the study that also used the Allen Human Brain Atlas (Richiardi et al., 2015). More specifically, we used GOATOOLS (Ashburner et al., 2000) to perform the GO analyses based on Ontologies and Associations downloaded on 15th November 2018. All of the significant GO items were corrected by FDR correction ($p<0.05$). We further identified the frequently seen words within all the significant GO items by counting the frequency of the words after removal of not meaningful words (e.g. "of" and "the").

Disease-related gene set enrichment

Although GO analysis can provide insights into the biological functions (BP, MF, or CC) of the overlapped gene list, the approach does not provide sufficient information to identify disease associations of the gene list. We leveraged *ToppGene* (Jing Chen et al., 2009) to explore the gene-diseases associations. The ToppGene platform can cluster groups of genes together according to their disease associations, and perform a statistical test as well as the multiple testing error corrections. The latest version of DisGeNET included associations between 17,549 genes and 24,166 diseases, disorders, or abnormal human phenotypes. In this study, our analysis was based on one of the sub-database of DisGeNET, the DisGeNET Curated (expert-curated associations obtained from UniProt and CTD human datasets). We did not use additional parameters to restrict the selection of disease items and performed a binary version of the over-representation analysis. All of the significant diseases items were FDR-corrected ($p<0.05$).

Data and code availability.

All the data (excluding neuroimaging data) is stored in the Open Science Framework (OSF). Open access data includes study summary, extracted coordinates for ALE, significantly associated genes for each task paradigm, and overlapped gene list(s)(OSF link: https://doi.org/10.17605/OSF.IO/6WZ2J). Neurovault was used to store all the neuroimaging data (e.g., results of ALE from Figure2, MACM from Figure3, and "Top Genes" expression maps from Figure 4) and provide 3D visualization of all the statistical maps (Neurovault link: https://neurovault.org/collections/4845/).

Functional-defined brain parcellation can be found in (https://github.com/ThomasYeoLab/CBIG/tree/master/stable_projects/brain_parcellation/Schaefer2018_LocalGlobal) Other research data supporting reported findings are available from the authors upon request.

For the neuroimaging meta-analysis, two software (GingerALE and Sleuth) (http://www.brainmap.org/software.html) were used. The Neurosynth Tool (https://github.com/neurosynth/neurosynth), a Python package that covered most of the functions on the Neurosynth (http://www.neurosynth.org/) website, was used in functional profile analyses and co-activation analyses. Activation-gene expression association analyses were based on alleninfo tool (https://github.com/chrisfilo/alleninf) and web application of gene decoder within the Neurovault (https://neurovault.org/). Nilearn (https://nilearn.github.io/) was used to load, manipulate, and visualize MRI statistical maps.

GOATOOLS (https://github.com/tanghaibao/goatools), a Python library, was used for Gene Ontology analyses and related visualization. The Ontology data was downloaded from the Gene Ontology website (http://geneontology.org/ontology/), and the Association data was downloaded from the National Center for Biotechnology Information (ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/). ToppGene Suite (https://toppgene.cchmc.org/) was used for disease-related gene set enrichment. The gene-disease association database can be downloaded from http://www.disgenet.org/.

Anaconda (https://www.anaconda.com/) Python 3.6 version for Win10 was used as the platform for all the programming and statistical analyses. Custom python scripts were written to perform all analyses described based on the mentioned Python packages and are released via the OSF.

Results

Regional brain activity associated with emotion regulation and memory control

Taking data from 95 published studies including in total 1995 subjects, we used 15 Emotion Regulation studies to represent emotion regulation, and 15 Think/No-think studies to represent memory control. It is noteworthy that we explicitly only focus on emotion regulation studies using the “suppression” or “distraction” as the regulation techniques rather than the more common “cognitive reappraisal” because of the conceptual relation with memory control, and equal statistical power between contrasted conditions. Furthermore, 27 Go/No-go studies and 38 Stop-Signal studies were used to represent response inhibition. (A list of studies and coordinates are available via the Open Science Framework; Search and inclusion criteria in *Materials and Methods*).

Regional brain activity of emotion regulation

The meta-analysis of the emotion regulation studies revealed six brain regions that are active during "regulation" compared to a "passive viewing" condition (FWE-cluster level corrected $p<.05$, uncorrected $p<.001$, threshold permutations=1000). Emotion regulation task consistently led to activation in right insula/inferior frontal gyrus (IFG), left IFG, left insula, middle cingulate gyrus, right inferior parietal lobule (IPL) and left supplementary motor area (SMA) (**Table 5.1**, **Figure 5.2A**).

Regional brain activity of memory control

Memory control studies revealed five brain regions during the "No-Think" condition compared to the "Think" condition (FWE-cluster level corrected $p<.05$, uncorrected $p<.001$, threshold permutations=1000): left insula/IFG, right dorsolateral prefrontal cortex (DLPFC), right middle frontal gyrus, bilateral IPL/supramarginal gyrus, bilateral precuneus, and SMA (**Table 5.1**, **Figure 5.2B**).

Regional brain activity of response inhibition

We used the same meta-analytical approach for all Go/No-go and Stop-signal studies and found similar significant clusters during the "control" condition compared to "baseline" condition (No-go vs go; Stop vs Go) in the insula, IFG, middle cingulate gyrus, SMA, IPL, and DLPFC (**Figure 5.2C** and **Figure 5.2D**, **Table S5.1** and **Table S5.2**).

Analyses of convergence and divergence

To examine the spatial convergence and divergence between emotion regulation and memory control activations, we performed formal conjunction and contrast analysis (FWE-cluster level corrected $p<.05$, threshold permutations=1000). The conjunction and contrast analysis did not find any reliable clusters after correction. Informal overlap analysis of two thresholded maps (i.e., emotion regulation and memory control) revealed a shared cluster of right IPL (MNI:60/-42/42; BA40).

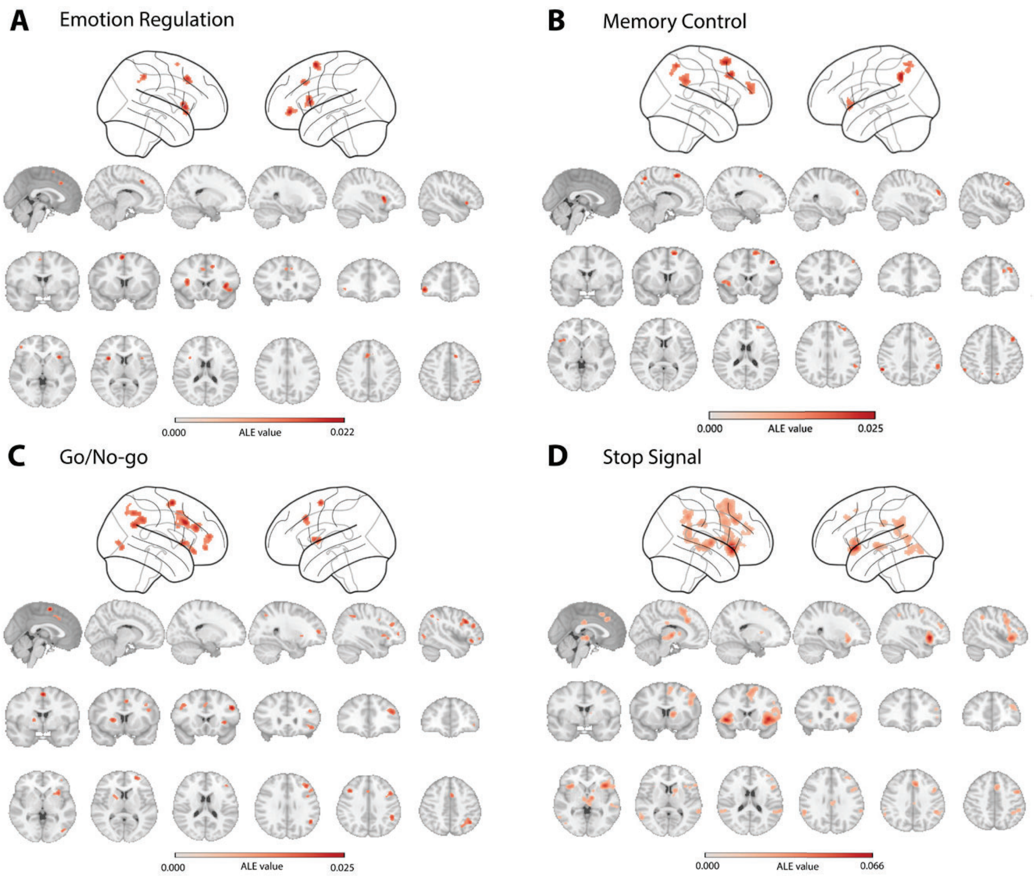


Figure 5.2 Brain activity underlying four inhibitory control tasks revealed by Activation likelihood estimation (ALE) meta-analyses. (A) Brain regions significantly more activated in the “Regulation” compared to the “Baseline” condition during the Emotion Regulation task. (B) Brain regions significantly more activated in the “Control” compared to the “non-control” condition during the memory control task. (C) Brain regions significantly more activated in the “No-Go” compared to the “Go” condition during the Go/No-go task. (D) Brain regions significantly more activated in the “Stop” compared to the “Go” condition during the Stop-signal task.

Table 5.1 Significant activated clusters during emotion regulation and think/no-think task

Task	Brain region	Hemisphere	MNI coordinates	Cluster size	Peak voxel Value
Emotion Regulation	Insula/IFG	R	36 16 6	209	0.02056
	IFG	L	-46 44 -4	138	0.02051
	Insula	L	-34 16 12	156	0.02089
	Middle Cingulate Gyrus	L/R	10 22 44	215	0.0189
	Inferior parietal lobule	R	60 -42 46	90	0.01688
	SMA	L	-4 6 62	110	0.0217

Memory Control	IFG/Insula	L	-36 20 -4	115	0.017
	DLPFC	R	36 48 24	221	0.018
	Supramarginal Gyrus/IPL	R	56 -44 32	167	0.02
	Supramarginal Gyrus/IPL	L	-58 -52 38	125	0.025
	Middle Frontal Gyrus	R	42 20 42	126	0.022
	Precuneus	R	8 -56 54	140	0.021
	Precuneus/SPL	L	-16 -64 56	114	0.017
	SMA	R	12 12 60	190	0.023

IFG: inferior frontal gyrus; SMA: supplementary motor area; DLPFC: dorsolateral prefrontal cortex; IPL: Inferior parietal lobule; SPL: Superior parietal lobule

Co-activation maps of regions associated with emotion regulation and memory control

To gain deeper insight into the co-activation profiles of brain regions associated with emotion regulation and memory control, we used a large database of fMRI studies (*BrainMap*: http://www.brainmap.org/) and applied meta-analytic connectivity modeling (MACM). This approach reveals brain regions that are consistently activated together with the seed regions resulting from the ALE meta-analyses. To control for specific modeling methods and sizes of seed regions, we also performed similar co-activation analyses using *Neurosynth* (http://neurosynth.org/) based on peak voxel coordinates of each seed region instead of the entire clusters. The two approaches yielded similar co-activation maps for each ROI. In the main test, we only report results from the *BrainMap* analysis (*Results from the Neurosynth can be found in* ***Figure S5.1 and S5.2****)*

In total, we analyzed six ROIs from the meta-analysis of emotion regulation studies. MACM analyses for the left IFG and right IPL revealed co-activation patterns in the direct vicinity of the seed regions. Co-activation analyses of other seed regions, including the bilateral insula, SMA, MACC revealed co-activation between these ROIs and with the cerebellum, IFG, IPL, thalamus, and DLPFC (**Figure 5.3A**). Similarly, we analyzed eight seed regions from the meta-analysis of memory control studies. ROIs tended to be co-activated with each other. Only the MACM analysis for the left IPL ROI identified modest co-activation patterns (**Figure 5.3B**). In addition, we investigated the co-activation profiles of response inhibition tasks. The co-activation patterns of 11 ROIs from Go/No-go meta-analysis and ten ROIs from Stop-signal meta-analysis were estimated using the same method. We found a brain network including bilateral DLPFC, insular, IPL, thalamus, SMA, and MACC across the co-activation patterns of ROIs from response inhibition.

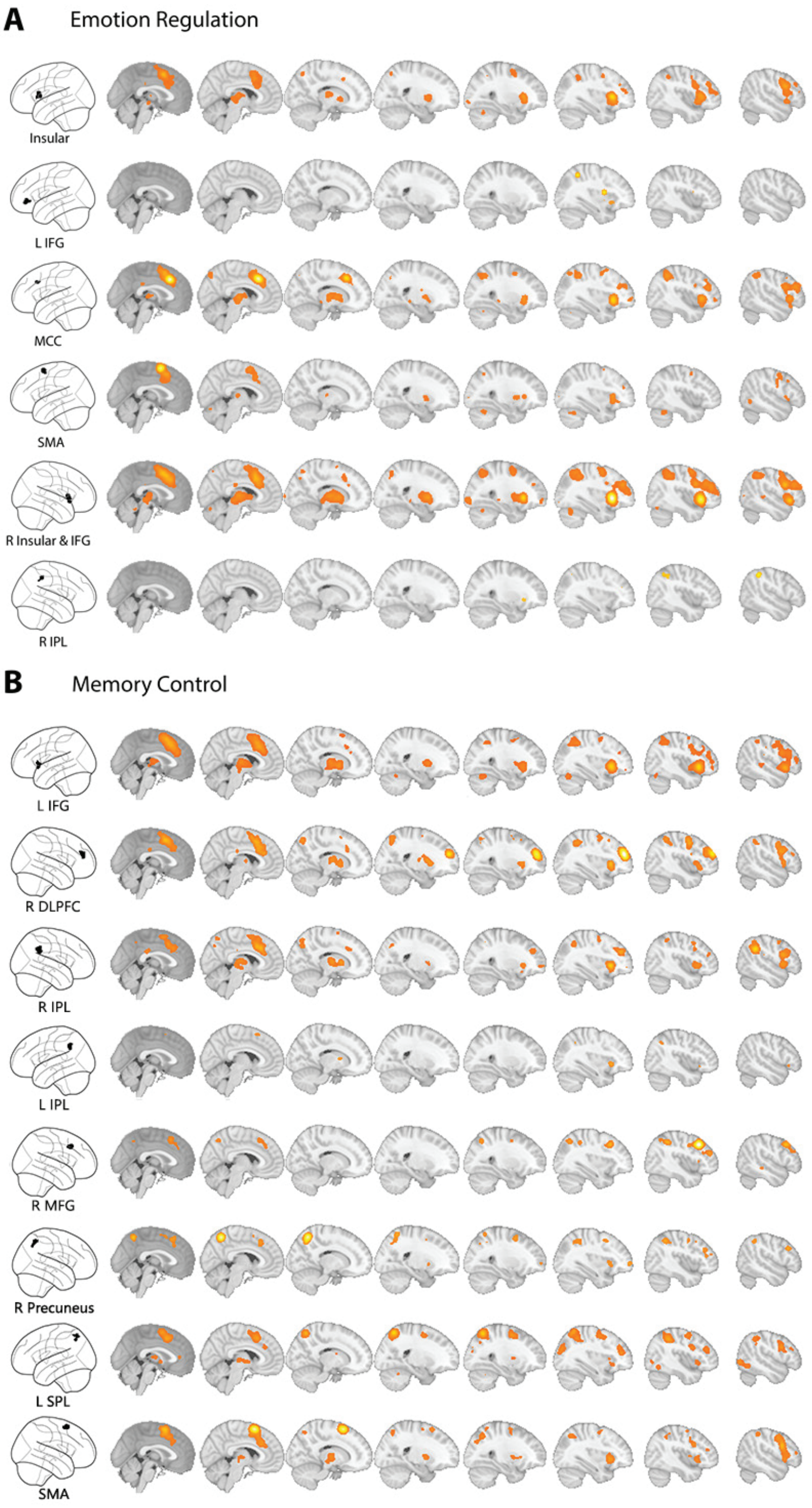


Figure 5.3 Co-activation maps of Emotion Regulation and Memory Control revealed by Meta-analytic connectivity modeling (MACM). (A) Co-activation maps for regions of interest (ROIs) resulting from ALE analysis of the emotion regulation. (B) Co-activation maps for ROIs resulting from ALE analysis of the memory control. ALE-based ROIs are projected onto glass brains (left column), and co-activation patterns are rendered on an MNI template second to ninth column (right columns).

Behavioral profiles of the co-activation maps

To identify the cognition domains most strongly associated with these co-activation maps, we created the behavioral profiles of each co-activation maps using the *Neurosynth* database. First, we quantified the behavioral profiles of co-activation maps of emotion regulation and memory control. As expected, these two sets of co-activation maps were both strongly characterized by terms such as “inhibition”, “cognitive control”, and “working memory”. Then, the behavioral profiles of Stop-signal and Go/Nogo co-activation maps were estimated in the same way. Finally, we investigated the commonalities of behavioral characteristics of all co-activation maps across four paradigms and revealed that these co-activation maps have the highest correlations to terms including “inhibition”, “cognitive control”, and “working memory”, compared to other items (**Figure 5.4**).

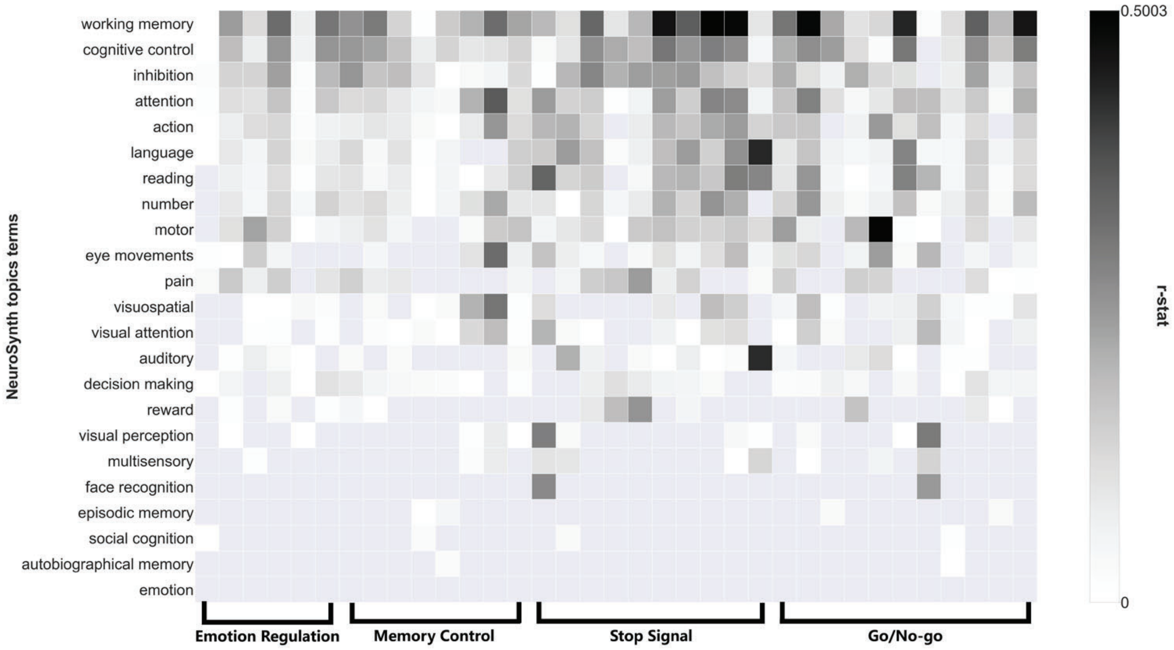


Figure 5.4 Behavioral profiles of co-activation maps across four inhibitory control tasks. Spatial similarity between Neurosynth meta-maps and co-activation maps across 23 topic terms. Terms are ordered by the mean r-values across the row(s) (all co-activation maps). “Working memory”, “cognitive control”, and “inhibition” are located at the top, suggesting the common stronger association. Domain-specific cognitive functions (e.g. autobiographical memory, emotion) are located at the bottom, suggesting a limited association.

Transcriptional signatures underlying the Emotion Regulation and Memory control

Next, to understand the transcriptional correlates that may be associated with brain activity elicited by each task, we combined the Allen Human Brain Atlas (AHBA), a brain-wide atlas of gene expression of postmortem brains and spatial pattern correlation. More specifically, we aimed to identify the genes for which the spatial transcriptional patterns are similar to the spatial pattern of brain activity during a given task.

Common transcriptional correlates of brain activity

We used the activation-gene expression association analysis to identify two lists of genes whose spatial patterns correlate with emotion regulation-related brain activity patterns or memory control-related brain activity patterns separately. The top ten genes with the highest spatial similarities are represented in Figure 5A, and their expression patterns in the brain are depicted in **Figure 5.5B** (*Complete gene lists in the OSF repository*). We found a substantial overlap (**Figure 5.5A**) between the two identified gene lists for the emotion regulation and memory control activation networks: there are in total 1212 genes (60.6% of all correlated genes) whose expression pattern correlated with the brain activity pattern of both tasks. We evaluated the significance of the overlap by generating a set of null distributions of the overlapping genes under the restriction of spatial similarity. Specifically, two gene lists with identical sizes (list1=2531; list2=1529) were randomly selected from all the genes (N=20787), and then overlapping genes between the two random lists were identified. This procedure was repeated 5000 times. 1718 out of 5000 pairs of randomly sampled genes demonstrated a similar level of spatial similarity as our gene lists of interests. We found that the amount of overlapping genes between the memory control and the emotion regulation was significantly larger than the number of overlapping genes within these 1718 randomly sampled overlapping genes lists across different thresholds (threshold=2000, real overlapping=1212, random overlapping=193.28 ± 12.3, $p<0.001$) (Results at different thresholds presented in **Table S5.4**). Furthermore, the significance of the overlap was estimated by generating 100 pairs of “permutated” statistical maps. 72 out of 100 pairs of genes that were associated with these “permutated” maps were used for the estimation because they showed a comparable level of spatial similarity across genes. The amount of real overlapping genes was significantly higher than the number of overlapping genes within these 72 pairs of genes across different thresholds (threshold=2000, real overlapping=1212, permutated overlapping=667.3 ± 147.39, $p<0.01$). (Full results in **Table S5.5**).

Specificity of “inhibition-related” genes

To test the specificity of the “inhibition-related” genes versus genes related to neuronal activity and brain function in general, we first generated, also using an ALE meta-analysis, the activation map of “Default Mode Network” (DMN), or “Task-Negative Network” (See Materials and Method; **Figure S5.3**). Next, we identified a list of genes associated with the DMN network and “permutated” maps emotion regulation and memory control via the AHBA and the same activation-gene expression association analysis. Given the fact that the DMN has been associated with homeostasis and undirected thought or mind-wandering, we expected that DMN-associated genes are different from inhibition-association genes. To test this, we calculated the number of overlapping genes between all possible combinations of two out of the seven reported gene lists (memory control, emotion regulation, stop-signal, and go/no-go, DMN, “permutated” memory control, “permutated” emotion regulation) across different thresholds. As depicted in **Figure**

5.5C, genes associated with inhibitory-control tasks demonstrate more overlap between each other than the associated genes between the inhibitory-control tasks and the DMN (mean Within Inhibition =41.24%, mean Inhibition & DMN =35.18%, t=3.44, p=0.0006). They also showed more overlap compared to associated genes between inhibitory-control tasks and "permutated" inhibitory control maps (mean Within Inhibition =41.24%, mean Inhibition & Permutation =29.53%, t=6.97, $p=1.26\times10^{-11}$).

Biological functions and disease associations of the "inhibition-related" genes

Using gene ontology (GO) (Ashburner et al., 2000), a widely-used literature-based gene-to-function annotation analysis, we generated a list of biological functions related to the "inhibition-related" genes. Furthermore, due to the common finding of impairment in inhibitory control in neurological and psychiatric disorders, we explored the human disease association of the identified gene list.

Enrichment for biological functions

Gene Ontology Enrichment Analysis (GOEA) of 779 "inhibition-related" genes identified 5 related Biological Processes (BP), which are associated with information communications between neurons (***Table S5.6***). More specifically, 19 genes within the list are associated with the neuropeptide signaling pathway, 30 genes with the chemical synaptic transmission, and 17 genes with the potassium ion transmembrane transport, 38 with the cell adhesion, and 60 with signal transduction.

Enrichment for human diseases

Additionally, we used *ToppGene* (https://toppgene.cchmc.org/) to perform the gene list enrichment analysis of 708 "inhibition-related" genes for human diseases. Using *DisGeNET* (www.disgenet.org), a comprehensive database on the relationship between human diseases and genes, we found significant associations between our gene list and 36 disease terms (***Table S5.7***). Among these diseases, seizures/epilepsy, chronic alcoholic intoxication, depression, bipolar disorder, autism spectrum disorders, and schizophrenia were ranked top 10 of the list (i.e., most significant associations) (**Figure 5.5D**).

We performed a preliminary investigation of the unique diseases associations for inhibition, but not DMN. First, we calculated the differences between "inhibition-related" genes and DMN-related genes under four different thresholds and included the results into the ToppGene. Only under the threshold of 1500 genes, the algorithm identified a significant association between 103 "inhibition-unique" genes and the risk for Schizophrenia. In other words, variance in these genes, which are associated with inhibition, but not DMN, are reported to be associated with an increased risk for Schizophrenia. We also performed the disease association analysis for "inhibition-related" and DMN-related genes separately. These two sets of genes were associated with similar brain disorders (e.g., Schizophrenia, Autistic Disorder, Unipolar Depression, Alcoholic Intoxication) (See OSF folder). One important difference between the two sets of genes is that “DMN-related” genes are uniquely associated with Alzheimer’s Disease and Age-Related Memory Disorders (**Figure S5.4**), which is consistent with our knowledge regarding the relationship between DMN, aging, and memory.

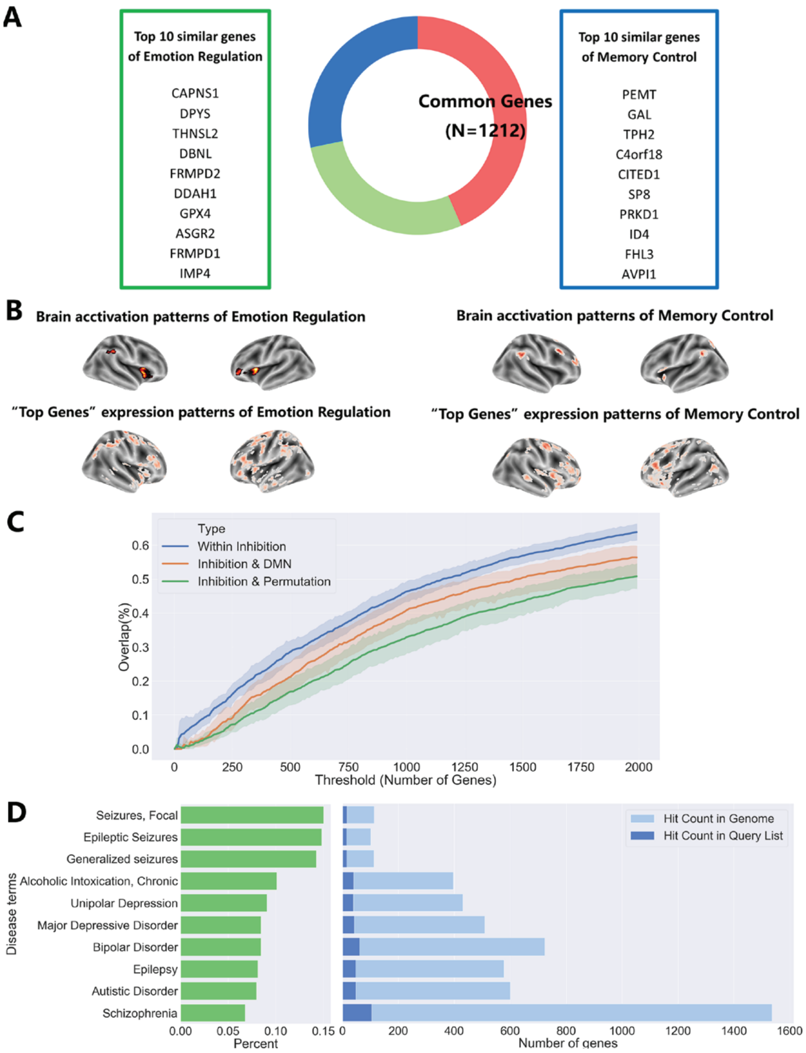


Figure 5.5 Common transcriptional profiles revealed by activation-gene expression association analysis. (A) Transcriptional patterns of 1212 genes correlated with both the emotion regulation and memory control-related brain activity patterns. (B) Visualization and comparison of brain activation patterns and expression patterns of “Top Genes”. “Top Genes” are 10 genes with the most similar expression patterns as the brain activation. (C) The gene lists of two inhibitory control tasks were more overlap with each other (Within Inhibitory) compared to one of the inhibitory control and DMN (Inhibitory & DMN) or one of the real inhibitory control and one of permutated inhibition inhibitory control (Inhibition & Permutation). (D) Disease associations of “inhibitory-related” genes. “Inhibition-related” genes are associated with genetic risks for depression, bipolar disorder, autism spectrum disorders, schizophrenia, seizures/epilepsy, and chronic alcoholic intoxication. Terms are ordered by the percentage (Left Green) of hit counts in the query list to hit count in the genome. The hit count in the query list is the number of genes belongings to “inhibitory-related” genes (Right Dark Blue). The hit count in the genome is the total number of genes associated with the risk for certain disease terms based on the DisGeNET (Right Light Blue).

Discussion

Inhibitory control is a fundamental cognitive function supporting other processes like emotion regulation and response inhibition. Here, we provided neurobiological evidence from human neuroimaging and transcriptional mapping to support the concept that there is one generic neural network of inhibitory control with a set of "inhibition-related" genes modulating not only response inhibition but also emotion regulation and memory control. Our meta-analysis of 95 neuroimaging studies revealed a common role of the right inferior parietal lobule and related regions in a frontal-parietal-insular network during emotion regulation and memory control. These co-activation patterns were also similar to the meta-analysis results of response inhibition tasks, and "inhibition-related" network reported in the literature. Additionally, we used the Allen Human Brain Atlas as an avenue to link this neural network to common transcriptional profiles and identified "inhibition-related" genes, which are associated with the neuronal transmission, and risk for major psychiatric disorders as well as seizures and alcoholic dependency.

The idea that inhibitory control is the underlying core cognitive function in emotion regulation was already suggested before (McRae et al., 2012; Ochsner et al., 2002; Ochsner & Gross, 2005; Schmeichel et al., 2008; Wager et al., 2008). Similarly, it was also suggested that inhibitory control plays a fundamental role in memory control (Benjamin J Levy & Anderson, 2002). A unified theory proposed a central role of inhibitory control in various psychological domains (e.g., motor inhibition, emotional response, and memory retrieval) depending on external task requirements (Aron et al., 2004, 2014; B. E. Depue et al., 2016). Engen and Anderson recently reviewed behavioral and neuroimaging studies in this field and proposed the conceptual link between emotion regulation and memory control (Engen & Anderson, 2018). However, there has been little empirical support for this link. Our multimodal analysis provides rich evidence beyond neuroimaging supporting a conceptual link and suggests that inhibitory control, as well as its underlying neural and transcriptional correlates, modulate both emotion regulation and memory control.

Brain activation patterns of emotion regulation and memory control found here are in accordance with previous meta-analyses (emotion regulation: e.g., Buhle et al., 2014; Kohn et al., 2014; memory control: e.g., Guo et al., 2018). We found one overlapping region between emotion regulation and memory control, the right inferior parietal lobule. Due to the central role of inhibitory control in both tasks, only one overlapping brain region seems surprising at first glance. However, meta-analytic connectivity modeling revealed that other regions (including IFG, insula, preSMA/MACC, IPL) that lacked significantly overlapping activations across emotion regulation and memory control, form a tightly integrated network. Our results suggest that although these regions do not overlap strictly, they belong to the same functional network. Our behavioral profile analysis corroborated this interpretation: both co-activation maps of emotion regulation and memory control, as well as stop-signal and go/no-go paradigms, have comparable behavioral profiles and were characterized by terms like “inhibition,” “cognitive control,” and “working memory.”

We have demonstrated the neural commonality of emotion regulation and memory control. Next, we proceeded to investigate if transcriptional profiles overlap with activity patterns in a similar way. Critically, this study adopted an imaging-genetic approach to investigate the common transcriptional signatures across neural networks of emotion regulation and memory control. Activation-gene expression association analysis revealed a largely overlapping gene list whose expression patterns were similar to the activation patterns. Furthermore, we identified a list of “inhibition-related” genes and characterized their biological function and disease associations. “Inhibition-related” genes were primarily associated with the neuropeptide signaling pathway, chemical synaptic transmission, the potassium ion transmembrane transport, and cell adhesion. One common feature of these biological functions is that they are critical for information communications between neurons. Together with our neuroimaging finding of frontal-parietal-insular network, these genes may act as molecular correlates underlying synchronous brain activity during inhibitory control. Inhibition-related” genes were also associated with risks for several psychiatric disorders (e.g., depression, bipolar disorder, schizophrenia, and autism), seizures, and alcoholic dependence.

Recently, neuroimaging (Goodkind et al., 2015; Sha et al., 2019) and genetic studies (Anttila et al., 2018; Cross-Disorder Group of the Psychiatric Genomics Consortium., 2013; Schork et al., 2019) collectively demonstrated the potential common biological roots across psychiatric disorders. However, common phenotypes across disorders are less well understood. Thus, the National Institute of Mental Health’s Research Domain Criteria (NIMH’s RDoC) (https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/index.shtml) summarized several domains of phenotypes, where inhibitory control is a central aspect. Here, our results highlighted the critical role of inhibitory control and its biological underpinning across psychiatric disorders. Firstly, dysfunctional inhibitory control (e.g., impaired response inhibition, lack of emotion regulation, and compromised memory control) is evident across different psychiatric disorders (Amstadter, 2008; Catarino et al., 2015; Ehring & Quack, 2010; Erk et al., 2010; Falconer et al., 2008; Joormann & Gotlib, 2010; Lipszyc & Schachar, 2010; Magee & Zinbarg, 2007; Price & Mohlman, 2007; Sacchet et al., 2017; Tull et al., 2007; Yang et al., 2016). Also, inhibitory control deficits can be further linked to some disorder-specific symptoms (e.g., lack of inhibition of negative thoughts (or rumination) in depression, lack of fear control in anxiety, and failure to avoid retrieval of traumatic memories in PTSD). Secondly, our results suggest an overlap between brain regions (frontal-parietal-insular network) involved in inhibitory control and regions whose structural abnormalities were observed consistently in a variety of psychiatric diagnoses (Goodkind et al., 2015). Thirdly, “inhibition-related” genes, which we identified by spatial transcriptional profiles that overlap with activation patterns of inhibitory control tasks, were also associated with the risks for a variety of psychiatric disorders. Furthermore, although brain disorders such as epilepsy and alcoholic dependence may involve different neurobiological underlies compared to psychiatric disorders, impairment of inhibitory control seems to also be a critical behavioral aspect of the epilepsy (Elger et al., 2004; Helmstaedter, 2001) and alcoholic dependence (Courtney et al., 2013; Lawrence et al., 2009; Papachristou et al., 2013).

Our study has two limitations that should be mentioned. First, our neural commonality analyses were based on fMRI studies only. However, overlap in fMRI activation or co-activation patterns lacks temporal information of the underlying cognitive processes. Electroencephalography (EEG) or Magnetoencephalography (MEG) in humans could provide further confirmatory evidence for the idea of a common cognitive process. Recently, Castiglione and colleagues reported that memory control elicited an electrophysiological signature, increased right frontal beta, which was seen in the Stop Signal task (Castiglione et al., 2019). Follow-up electrophysiological studies or even a meta-analysis of them might confirm the idea of a common electrophysiological signature. Second, the current activation-gene expression association analysis is still preliminary (e.g., low sample size and spatial resolution of the post-mortem data), and without the possibility of testing the specific relationship between expression maps and cognitive function. For example, although our preliminary results demonstrated that “inhibition-specific” genes are associated with the risk for Schizophrenia, while DMN-related genes are more closely linked to Alzheimer’s Disease, we also identified considerable transcriptional overlaps between “inhibition-related” genes and DMN-related genes. The latter may suggest that “inhibition-related” genes, as defined in our study may include both “inhibition-specific” genes and other genes supporting general brain function. However, it is challenging to separate them using methods currently available. Nevertheless, the method already showed great potential when helping to understand basic molecular principles of both the structural and functional connectome (*See review by Fornito et al., 2018*) and it identified molecular mechanisms underlying changes in brain structure or function associated with brain disorders (e.g., autism spectrum disorder (Romero-Garcia et al., 2019), Huntington’s disease (McColgan et al., 2018), schizophrenia (Romme et al., 2017), and Alzheimer’s disease (Grothe et al., 2018)). Results from these studies are consistent with the genetics of neuropsychiatric disorders using conventional methods like genome-wide association studies or animal models. Taken together, although preliminary, neuroimaging-gene expression association analysis has demonstrated its potential to bridge brain structure-function associations and to reveal its underlying molecular processes. To detect more specific associations between spatial transcriptional profiles and neuroimaging data, large sets of postmortem gene-expression data with higher spatial resolution need to be collected. Also, more dedicated analytical methods need to be developed and validated (Arnatkevičiūtė et al., 2019) with new methods that may better control for the effects of domain-general genes that are supporting brain function in general and the bias in gene-set enrichment analyses of brain-wide gene expression data, probably induced by the gene-gene co-expression or autocorrelation (Fulcher et al., 2020).

In summary, our multimodal analysis identified a frontal-parietal-insular neural network and a set of genes associated with inhibitory control across emotion regulation, memory control, and response inhibition. The integrative approach established here bridges between cognitive, neural, and molecular correlates of inhibitory control and can be used to study other higher-level cognitive processes. Our findings may deepen our understanding of emotion regulation and memory control in health and pave the way for better emotion regulation and memory control by targeting the core inhibitory-related network or related molecular targets in patients with such deficit at issue.

Acknowledgments: We thank Prof. Hans van Bokhoven for comments on the early results of this study, Dr. Yuhua Guo for providing brain activation coordinates of unpublished studies, Dr. Dace Apšvalka for providing valuable feedback on our preprint, and Dr. Xiangzhen Kong for suggestions on our revision. **Funding:** W.L. is funded by the Ph.D. scholarship (201606990020) of the Chinese Scholarship Council. **Author contributions:** The experiment was designed by W.L., G.F., and N.K. Data were collected by W.L., N.P., and N.K. Analyses were performed by W.L. and W.L., N.P. wrote the manuscript. W.L, G.F., and N.K. revised and edited the manuscript. All authors provided essential feedback on the manuscript at different stages, contributed substantially to, and approved the final version of the manuscript. **Competing interests:** All authors declare that they have no competing interests.

Supplementary Material for

Common neural and transcriptional correlates of inhibitory control underlie emotion regulation and memory control

This file includes:
Supplementary Text
Table S5.1 to S5.7
Figure S5.1 to S5.4

1. Comparison between BrainMap and Neurosynth-based co-activation analysis

Co-activation analysis can be performed based on large-scale databases of fMRI studies such as BrainMap or Neurosynth. During the initial data analysis, we took advantage of the relative strengths and weaknesses of the two methods. We augmented the coordinate-based meta-analysis on the Neurosynth with meta-analytic connectivity modelling (MACM) based on the BrainMap to explore co-activation patterns. The idea behind Neurosynth is similar to MACM (which is described in the main text). The Neurosynth algorithm searches for brain regions which co-activate with input coordinates within the same functional contrast, and summarize the results as a co-activation map. Neurosynth has several advantages over BrainMap: (1) Neurosynth is based on automated text mining, therefore includes a higher number of studies and at the same time decreases the potential selection bias of the users. (2) ROI-based MACM is largely dependent on the size (number of voxels) and the shape of the input ROIs. It could be a potential problem because we included far more studies in the ALE analysis of Stop-signal (SS) paradigm and Go/No-go (GN) paradigm compared to the emotion regulation and memory control, leading to on average larger ROIs for stop signal and go/no-go paradigm. On the contrary, Neurosynth-based co-activation analysis does not depend on the ROIs, and it can also base on the MRI coordinates, providing a control for the effect of different size of ROIs. The disadvantage of Neurosynth compared to BrainMap is that automated text mining does not separate different contrasts or experiments within one article. Even though these differences, both methods yielded highly similar co-activation patterns for each ROIs (**Figure S5.1** and **Figure S5.2**), suggesting that the effect of methodology and database on the neural network analysis in our study is negligible. We nevertheless chose to present the results of MACM in the main text.

2. Data sharing of non-imaging data via the Open Science Framework (OSF)

We used the OSF database (https://osf.io/6wz2j/) as a venue to share non-imaging data generated within this study.

2.1 Studies and coordinates used in the meta-analyses

For each task (e.g. ER: emotion regulation; TNT: think/no-think; GN: go/no-go; SS: stop-signal), an excel file with all coordinates used in the meta-analyses is uploaded to the folder (ALE_coordinates_data) within the OSF.

2.2 Code

Custom python scripts used in this study can be found.

2.3 Details of MACM results

Within the MACM analyses, we estimated in total co-activation patterns of all ROIs from 4 different tasks. Because of the limitation of the number of figures presented in the supplemental materials, we cannot provide all the detailed results for these analyses here. Instead, we used a python package (atlasreader: https://github.com/miykael/atlasreader) to generate coordinate tables and region labels from all co-activation images. All results can be found in an OSF folder (folder name: MACM_results_details, filename: MACM_results_altasreader.zip)

2.4 Lists of genes

Complete gene lists, together with the statistical results, is uploaded in one folder (complete_gene_list) within the OSF database. All files started with "gene_decoding" are the genes whose expression patterns correlated with the brain activity pattern elicited by each task of interest. All files started with "overlapped_gene_list" are genes whose expression patterns correlated with two (ER and TNT) or more (ER, TNT, GN, and SS) task-related activity patterns simultaneously. Numbers (e.g. 500, 1000) were thresholds used in the analysis to select the genes with most similar spatial patterns. All files started with "difference" are genes whose spatial patterns correlated with inhibition tasks, but not DMN.

2.5 Tables generated by gene ontology enrichment analyses

Tables for enrichment analyses of biological functions (GOEA.zip) or diseases items (disease.zip). Results from different thresholds were presented.

3. Data sharing and 3D visualization of statistical maps via the Neurovault

We uploaded all the statistical maps generated within this study to our Neurovault database (https://neurovault.org/collections/4845/) for data sharing purpose and 3D, interactive visualization. There are in total of 46 maps within the images collection.

- Images names starting with the task names (e.g. ER) and contrast names (e.g. Regulation vs Baseline) are maps resulting from the ALE meta-analysis.
- Images names ending with threshold methods are corrected ALE images (e.g. ALE C05 1K stands for $p<.05$, uncorrected $p<.001$, threshold permutations=1000).
- Images names starting with "MACM" are co-activation maps from the MACM analyses.
- No images from the coordinate-based co-activation analysis using Neurosynth were uploaded.

Table S5.1 Significant activations resulting from the meta-analysis of Go/No-go paradigm

Brain region	Hemisphere	MNI coordinates	Cluster size	Peak voxel Value
Insular/IFG	R	32 16 2	212	0.020847
Inferior temporal gyrus	R	48 -72 -4	98	0.017735
Superior frontal gyrus	R	24 54 10	182	0.018687
Putamen	L	-22 8 6	127	0.019816
DLPFC	R	40 34 24	180	0.021783
IFG, opercular part	R	46 18 32	266	0.024922
Angular	R	50 -54 28	238	0.021796
DLPFC	L	-40 22 38	114	0.021692
Middle cingulate gyrus	R/L	4 10 46	98	0.017317
Inferior Parietal Lobule	R	40 -56 44	200	0.019583
SMA	R/L	0 0 60	95	0.024901

IFG: inferior frontal gyrus; SMA: supplementary motor area; DLPFC: dorsolateral prefrontal cortex

Table S5.2 Significant activations resulting from the meta-analysis of Stop-signal paradigm

Brain region	Hemisphere	MNI coordinates	Cluster size	Peak voxel Value
Insular/IFG	R	36 20 -4	2050	0.066141
Fusiform	L	-40 -60 -12	252	0.026161
Insular/IFG	L	-38 18 -6	690	0.062398
Inferior parietal lobule	R	48 -44 40	1011	0.038367
Thalamus	R	10 -10 2	812	0.045329
Supramarginal gyrus/ Inferior parietal lobule	L	-58 -46 28	601	0.032832
DLPFC	R	36 46 20	200	0.024995
Middle cingulate gyrus	R/L	2 -24 30	181	0.031727
SMA	R	6 24 34	861	0.042154

IFG: inferior frontal gyrus; SMA: supplementary motor area; DLPFC: dorsolateral prefrontal cortex

Table S5.3 **Estimation of spatial similarity between gene expression maps.**

Threshold\Donor	ID9861	ID10021	ID12876	ID14380	ID15496	ID15697
500	0.07(0.24)	0.03(0.26)	0.08(0.26)	0.02(0.28)	0.06(0.29)	0.07(0.29)
1000	0.07(0.24)	0.03(0.26)	0.09(0.26)	0.03(0.28)	0.07(0.29)	0.08(0.29)
1500	0.07(0.24)	0.03(0.26)	0.09(0.26)	0.03(0.28)	0.07(0.29)	0.08(0.29)
2000	0.08(0.24)	0.03(0.26)	0.10(0.26)	0.03(0.28)	0.07(0.29)	0.08(0.29)

Mean (standard deviation)

Table S5.4 **Estimation of the significance of overlap genes under the restriction of spatial similarity.**

Threshold	Real overlap	Significance	Random overlap (mean)	Random overlap (standard deviation)
500	145	P<0.001	12.01	3.32
1000	445	P<0.001	48.35	6.62
1500	808	P<0.001	108.89	9.7
2000	1212	P<0.001	193.28	12.3

Table S5.5 **Estimation of the significance of overlap genes based on permutated statistical maps.**

Threshold	Real overlap	Significance	Permutated overlap (mean)	Permutated overlap (standard deviation)
500	145	P<0.01	48.62	25.14
1000	445	P<0.01	190.41	70.43
1500	808	P<0.01	405.04	113.70
2000	1212	P<0.01	667.30	147.39

Table S5.6 Gene Ontology Enrichment Analysis results of the "inhibition-related" genes

GO	NS	enrichment	name	ratio_in_study	ratio_in_pop	p_uncorrected	depth	study_count	p_fdr_bh
GO:0007218	BP	e	neuropeptide signaling pathway	19/779	101/20913	5,3E-09	6	19	3,7E-05
GO:0007268	BP	e	chemical synaptic transmission	30/779	239/20913	6,2E-09	7	30	3,7E-05
GO:0071805	BP	e	potassium ion transmembrane transport	17/779	114/20913	1,1E-06	8	17	0,00434
GO:0007155	BP	e	cell adhesion	38/779	463/20913	5,3E-06	2	38	0,01603
GO:0007165	BP	e	signal transduction	60/779	898/20913	1,2E-05	4	60	0,02976

BP: Biological Processes

Table S5.7 Disease associations of the "inhibition-related" genes

Name	pValue	FDR B&H	FDR B&Y	Bonferroni	Genes from Input	Genes in Annotation
Schizophrenia	1,92E-07	5,15E-04	4,37E-03	5,15E-04	104	1537
Bipolar Disorder	5,85E-07	7,86E-04	6,66E-03	1,57E-03	61	723
Alcoholic Intoxication, Chronic	4,58E-06	4,10E-03	3,48E-02	1,23E-02	40	396
Unipolar Depression	1,33E-04	8,94E-02	7,58E-01	3,58E-01	39	430
Epilepsy	1,85E-04	8,95E-02	7,59E-01	4,98E-01	47	578
Major Depressive Disorder	2,22E-04	8,95E-02	7,59E-01	5,96E-01	43	509
Autistic Disorder	2,33E-04	8,95E-02	7,59E-01	6,27E-01	48	601
Seizures, Focal	6,19E-04	2,08E-01	1,76E+00	1,66E+00	17	113
Generalized seizures	2,67E-03	7,96E-01	6,75E+00	7,16E+00	16	112
Epileptic Seizures	3,33E-03	8,95E-01	7,58E+00	8,95E+00	15	101
Mental disorders	5,58E-03	1,16E+00	9,84E+00	1,50E+01	29	320

Mood Disorders	8,09E-03	1,16E+00	9,84E+00	2,18E+01	28	309
Withdrawal Symptoms	9,48E-03	1,16E+00	9,84E+00	2,55E+01	12	72
Neuroblastoma	1,15E-02	1,16E+00	9,84E+00	3,09E+01	94	1683
Gustatory seizure	1,26E-02	1,16E+00	9,84E+00	3,38E+01	14	99
Atonic Absence Seizures	1,26E-02	1,16E+00	9,84E+00	3,38E+01	14	99
Olfactory seizure	1,26E-02	1,16E+00	9,84E+00	3,38E+01	14	99
Visual seizure	1,26E-02	1,16E+00	9,84E+00	3,38E+01	14	99
Seizures, Sensory	1,26E-02	1,16E+00	9,84E+00	3,38E+01	14	99
Generalized Absence Seizures	1,26E-02	1,16E+00	9,84E+00	3,38E+01	14	99
Seizures, Clonic	1,26E-02	1,16E+00	9,84E+00	3,38E+01	14	99
Single Seizure	1,26E-02	1,16E+00	9,84E+00	3,38E+01	14	99
Jacksonian Seizure	1,26E-02	1,16E+00	9,84E+00	3,38E+01	14	99
Nonepileptic Seizures	1,26E-02	1,16E+00	9,84E+00	3,38E+01	14	99
Non-epileptic convulsion	1,26E-02	1,16E+00	9,84E+00	3,38E+01	14	99
Vertiginous seizure	1,26E-02	1,16E+00	9,84E+00	3,38E+01	14	99
Absence Seizures	1,26E-02	1,16E+00	9,84E+00	3,38E+01	14	99
Seizures, Somatosensory	1,26E-02	1,16E+00	9,84E+00	3,38E+01	14	99
Seizures, Auditory	1,26E-02	1,16E+00	9,84E+00	3,38E+01	14	99
Drug Dependence	1,30E-02	1,16E+00	9,84E+00	3,49E+01	19	170
Myoclonic Seizures	1,41E-02	1,18E+00	1,00E+01	3,80E+01	14	100
Epileptic drop attack	1,41E-02	1,18E+00	1,00E+01	3,80E+01	14	100
Mental Depression	1,45E-02	1,18E+00	1,00E+01	3,91E+01	41	559
Drug Withdrawal Symptoms	1,70E-02	1,34E+00	1,14E+01	4,57E+01	10	53
Tonic Seizures	1,78E-02	1,36E+00	1,15E+01	4,79E+01	14	102
Depressive disorder	1,86E-02	1,36E+00	1,15E+01	4,99E+01	43	604

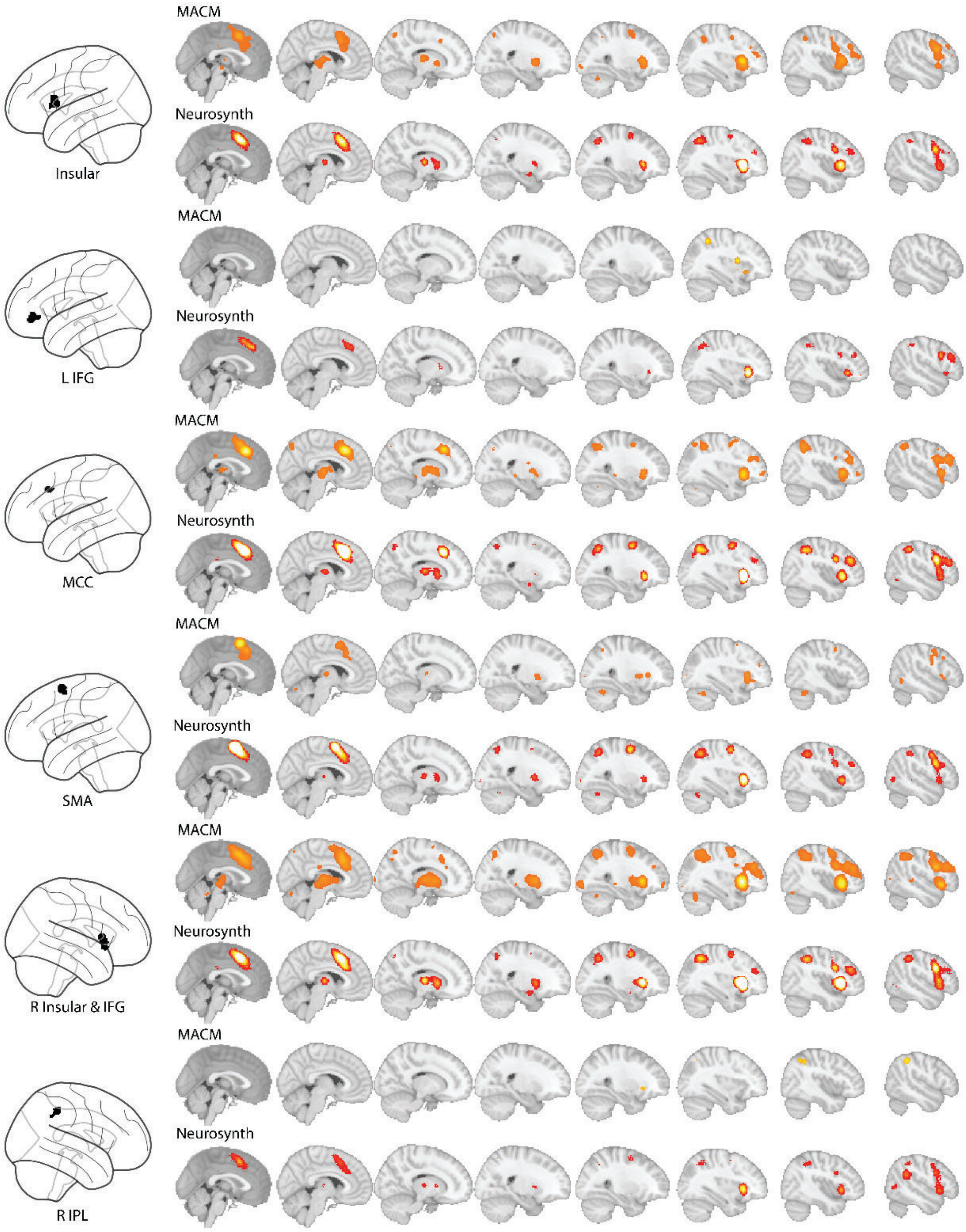


Figure S5.1 Comparisons between BrainMap-based and Neurosynth-based co-activation map for the emotion regulation studies MACM: meta-analytic connectivity modelling; IFG: inferior frontal gyrus; MCC: middle cingulate gyrus; IPL: inferior parietal lobule; SMA: supplementary motor area

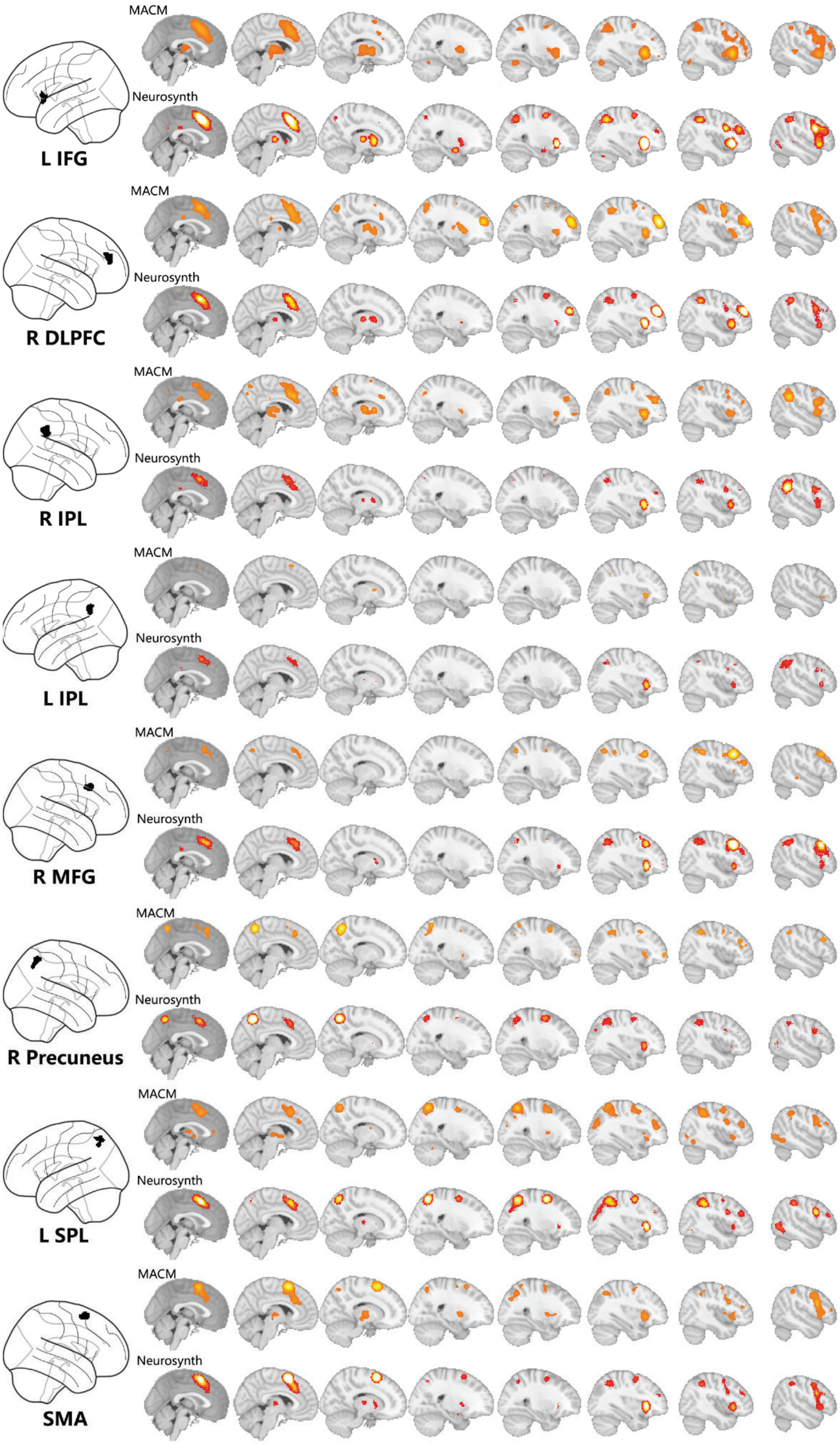


Figure S5.2 Comparisons between BrainMap-based and Neurosynth-based co-activation map for the memory control studies MACM: meta-analytic connectivity modelling; DLPFC: dorsolateral prefrontal cortex; IPL: inferior parietal lobule; SMA: supplementary motor area

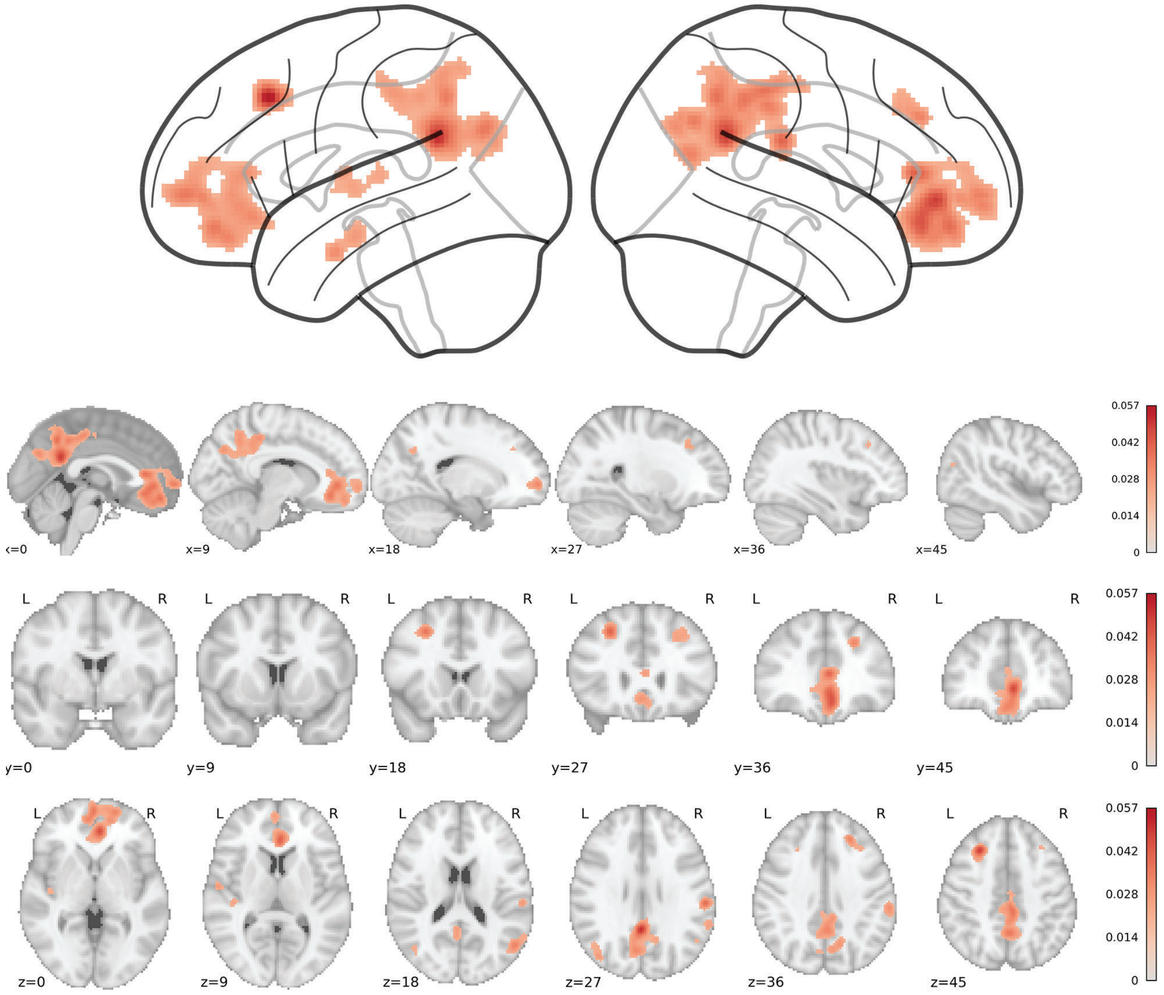
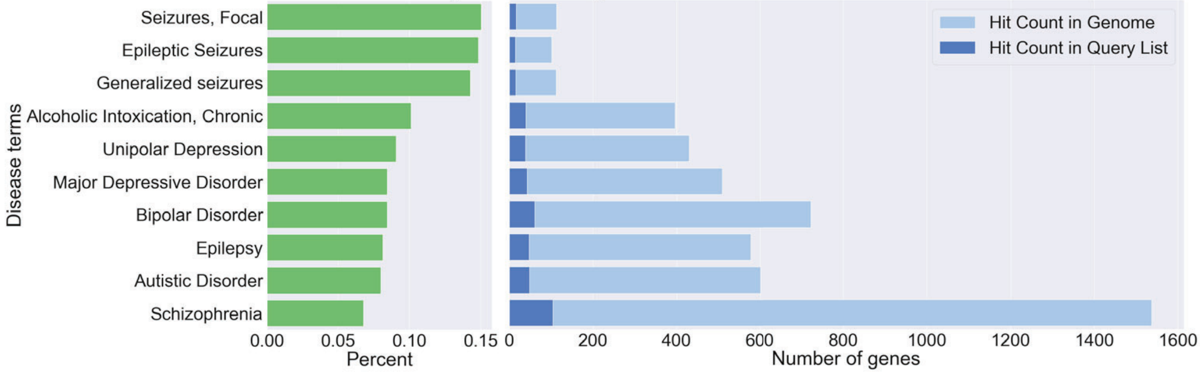


Figure S5.3 ALE meta-analysis of the Default-Mode Network (Task-negative network)

A. Disease associations of “inhibitory-related” genes.



B. Disease associations of “DMN” genes.

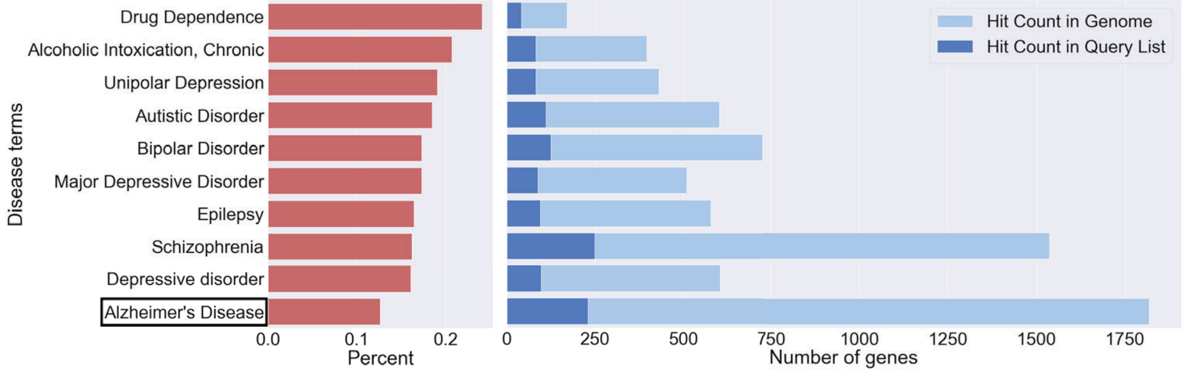


Figure S5.4 Comparison of diseases association between “inhibition-related” genes and “DMN-related” genes.

Chapter 6

General Discussion

At the beginning of this thesis, I discussed the ideas of *process* and *strength* dynamics of memory. These dynamics can be measured non-invasively in the encoding-retrieval network based on the principle of pattern reinstatement. They also have close relationships with other core cognitive operations, such as emotion perception and cognitive control. Here, I answered specific questions regarding the neural dynamics of memories:

1. **How do we transform continuous experience into discrete memories?**
2. **How does the brain flexibly switch between memory retrieval and memory control?**
3. **How does memory modulation re-organize memory traces and change their memory strength after overnight consolidation?**
4. **Why are changes in memory strength also accompanied by alterations of emotional intensity?**

I will start with a summary of our experimental findings in the order of the above questions. Then, I will seek to deliberate the relationship between our findings and existing literature. Next, I will attempt to discuss how future studies could be conducted to address new questions raised by our current results. Lastly, a general conclusion will be presented.

Summary of findings

I presented four experimental chapters on the topic of *process* and *strength* dynamics of memories. In brief, we used fMRI in healthy human subjects to elicit brain activity during different memory tasks, and combined approaches from genetics and machine learning. Here I will give a summary of our main findings.

1. How do we transform continuous experience into discrete memories?

How do we encode continuous information is critical for subsequent retrieval. Theories of event memories proposed that different neural states are used to represent discrete events. We demonstrated that successful encoding of continuous information was dependent on events being represented with dissimilar activity patterns in a network centered on the hippocampus and medial prefrontal cortex. At the same time, we found the potential neural correlates for event integration: similar connectivity patterns of these regions linked events and preserved the narrative order in which they were encoded (**Chapter 2**).

2. How does the brain flexibly switch between memory retrieval and memory control?

During a continuous task, sometimes we need to switch between two or more task demands. This process is particularly challenging for the brain when the switch is between memory retrieval and memory control, which requires the need to coordinate two opposite neural states with partly overlapping neural networks. As the classical task-switch studies, we found the effect that preceding mental processing has an impairing effect on the current processing. We also found that the switch between retrieval and control involves large-scale adaptations between memory retrieval and inhibitory control networks. This adaptation is less flexible immediately after the task switching and associated with behavioral switch costs. Thus, we reasoned that the timely reconfiguration between memory and control networks is the key to flexible memory processing (**Chapter 3**).

3. How does the memory modulation re-organize memory traces and change their memory strength after overnight consolidation?

Memory modulation immediately after their formation can modify neural representations of memory traces and change their strength. We found that similar neural effects can be observed after initial consolidation (i.e., 24 hours after encoding). Specifically, repeated retrieval reduced overall activity amplitude, but seems to promote episode-unique mnemonic representations in visual processing and parietal regions. In contrast, repeated memory control was associated with the reduced lateral prefrontal activity, but relative intact mnemonic representations (**Chapter 4**).

4. Why are changes in memory strength also accompanied by alterations of emotional intensity?

Memory modulations such as memory control can alter not only memory accessibility and neural representations, but also the valence of these memory traces. We found that memory control and emotion regulation are supported by the same frontal-parietal-insular network, which is involved in inhibitory control. Moreover, neuroimaging-gene expression analysis identified the association between task-induced brain activations and a set of "inhibition-related" genes. These genes were reported to be associated with neuronal transmission and risk for major psychiatric disorders, as well as epilepsy and alcohol dependence (**Chapter 5**).

Memory dynamics: integration of literature and our findings

Dynamics of segmentation and integration contribute to memory formation

Successful memory formation in humans was studied intensively by the subsequent memory paradigm (Kim, 2011). It is well established that increased activation in the hippocampus, MTL, and prefrontal cortex is the neural signature of successful encoding (Brewer et al., 1998; Fernández et al., 1999; Wagner et al., 1998). Because unrelated to-be-remembered materials are often presented in isolation, *temporal dynamics* during the continuous experience are largely ignored in previous investigations.

In **Chapter2**, combining naturalistic stimuli and MVPA, we revealed two complementary neural dynamics (i.e., event segmentation and integration). Related neural processing in the hippocampus and mPFC could predict subsequent retrieval success and order. It is noticeable that this is in line with one recent report which investigated the neural coding of continuous experience in rodents (Sun et al., 2020). They reported "event-specific rate remapping" (ESR) cells in CA1 simultaneously tracked subdivisions of a continuous experience (i.e., events) and their sequential relationship. Results from **Chapter2** suggests that humans may use a similar code for everyday-like memory, and this neural code exists not only in the hippocampus but also in several prefrontal regions, mainly in the mPFC.

While previous neuroimaging studies demonstrated the role of human hippocampus in event segmentation (Baldassano et al., 2017; Ben-Yakov et al., 2013; Ben-Yakov & Dudai, 2011; Ben-Yakov & Henson, 2018; DuBrow & Davachi, 2016; Williams et al., 2019), and in establishing links between different events (DuBrow & Davachi, 2016; Benjamin J Griffiths & Fuentemilla, 2020; Silva et al., 2019; Sols et al., 2017), the relationship these neural representational processing and subsequent memory performance remain unclear.

We showed that neural correlations of event segmentation and integration in the hippocampus and neocortical regions (mainly mPFC) during encoding contribute to successful episodic memory formation. Here, the hippocampus may use pattern separation-like neural processes to represent distinct episodic events during the continuous experience (Bakker et al., 2008; Yassa & Stark, 2011). The role of mPFC in the continuous event processing may relate to its function of establishing links between elements across time and space: for instance, memory integration (Preston & Eichenbaum, 2013; Schlichting et al., 2014; Schlichting & Preston, 2015; Zeithamova et al., 2012), knowledge accumulation (Berkers et al., 2018; Kumaran et al., 2009), and schema learning (van Kesteren et al., 2013, 2014; Van Kesteren et al., 2010).

To summarize, we highlighted the role of hippocampal and medial prefrontal event segmentation and integration during the *temporal dynamics* of memories and how they contribute to naturalistic memory formation. Our experimental results showed that subsequent memory retrieval could be predicted by the interaction between "units" of encoding: separate neural representation of these "units" are the neural substrates of segmentation, and consistent context encoding across "units" supports integration.

Task switching-related process dynamics

Based on the experimental results from **Chapter 3**, we proposed that the task-switching paradigm is another window to observe *process dynamics* and their related neural reconfiguration. During task switching, suboptimal neural coordinations across large-scale brain networks could limit the executive control resource that can be deployed for certain cognitive processing, in our case, memory retrieval, and control.

We found the dynamic interaction between memory retrieval and memory control when participants were required to switch between these two opposite demands. Specifically, we found that it is more challenging for participants to control unwanted memories when the task demand just switched from retrieval to control. This result can be integrated with the broader literature of task switching (Jersild, 1927; Meiran, 2010; Monsell, 2003; Spector & Biederman, 1976) as a specific kind of switch cost between memory retrieval and control. Also, it suggests that different task demands, as processing units, interact with each other. These interactions have behavioral relevance and could be the source of switch costs.

Our fMRI results provided a new perspective of neural state transitions to understand the *process dynamics* during task switching. First, we used the multivariate decoding method to show that distributed fMRI signals across the inhibitory control and memory retrieval network can be used to differentiate two opposite task demands (i.e., memory retrieval and control), although all trials remained highly consistent throughout the experiment at the perceptual level. It is well established that human brain could demonstrate diverse brain states during different cognitive tasks (Cocuzza et al., 2019; Gonzalez-Castillo et al., 2015; Sadaghiani et al., 2015; Shine et al., 2016; Shine & Poldrack, 2018; Westphal et al., 2017). Our decoding results added additional evidence for this line of research. Second, initial evidence suggested that transitions between task-specific neural states have behavioral relevance: participants who showed less distinct neural states across tasks are more likely to perform worse in these tasks (Gonzalez-Castillo et al., 2015). However, that study design contained only a limited number of task "units" and transitions, limiting the ability to probe deeply into the state transitions between cognitive tasks. We compared the neural representations of task demands between the switch and non-switch periods. By applying the decoding method to the switch and non-switch periods separately, we showed that the decoder performed less accurately when participants were just instructed to just switch to another task demand. In other words, immediately after the shift in task demand, the decoder was more likely to mistakenly classify the neural state as the previous state instead of the current new state. Because our task only contained two opposite task demands, this delayed neural transition could be even more detrimental for optimal behavioral performance. Our following analyses confirmed it by associating delayed transitions to worse performance on the trial-by-trial basis. Our results, together with the previous study (Gonzalez-Castillo et al., 2015), demonstrated that if the underlying neural state is not well configurated for the current task, participants' task performance is compromised.

In summary, **Chapter 3** of this thesis investigated the *process dynamics* of neural states transitions during task-switching. We revealed how the interaction between opposite processing "units" (i.e., memory retrieval and memory control) generates behavioral switch costs. If the brain cannot timely reconfigure in time its neural states between these "units," behavioral performances are compromised.

Retrieval practice and memory control induces different strength dynamics after consolidation

After memory formation, memory traces could be modulated in different ways (Phelps & Hofmann, 2019). Among them, behavioral modulations such as retrieval practice (Karpicke & Blunt, 2011; Roediger III & Butler, 2011) can enhance memory traces while memory suppression can disrupt them (Michael C Anderson & Green, 2001; Michael C Anderson & Hanslmayr, 2014). However, after initial consolidation, what happens to the neural representations of individual memory traces during and after different kinds of modulations is still under ongoing investigation. In **Chapter 4**, we investigated this question with specific interests in how established memories (i.e., 24 hours after initial formation) are dynamically modulated by retrieval and suppression using a two-day fMRI design.

A theoretical framework of how retrieval practice can help to create long-lasting memories is provided by a fast memory consolidation hypothesis of memory retrieval (Antony et al., 2017). This hypothesis suggests that online retrieval triggers the fast reactivation of associative information, supporting the creation of hippocampal-neocortical representation (Antony et al., 2017). Ferreira and co-workers reported rapid increases mPFC activity, and slow hippocampal disengagement across retrieval attempts (Ferreira et al., 2019), which is partly consistent with neural changes during memory consolidation (Frankland & Bontempi, 2005; Takashima et al., 2006, 2009). Critically, using multivariate fMRI methods, they demonstrated that episodic-unique information was enhanced in the parietal cortex after retrieval practice. Building on this work (Ferreira et al., 2019) and previous literature of retrieval-mediate learning (Eriksson et al., 2011; Kuhl et al., 2010; Nelson, Arnold, Gilmore, & McDermott, 2013; G. van den Broek et al., 2016; G. S. E. van den Broek et al., 2013; Maria Wimber et al., 2008, 2011; Wing et al., 2013; Wirebring et al., 2015), we demonstrated that, for initially consolidated memories, reported neural effects of retrieval practice on memory representations could also be observed, and we found a similar decreasing hippocampal activity during and after the retrieval practice. Furthermore, our data showed the enhanced episodic-unique neural representations in the visual processing areas and precuneus, which in turn associated with decreased retrieval-related univariate activity amplitude of the same regions. This association may reflect a process during the retrieval practice that suppresses redundant neural activity, and thus, only the fine-grained neural patterns are reinstated, enabling more distinctive memory representations with higher fidelity.

Understanding how memory suppression can cause motivated forgetting of existing memory traces is not only relevant for understanding the flexible nature of memories during modulation (Kroes & Fernández, 2012) but also has its clinical values for memory-related psychotherapy and psychopathology (Dillon & Pizzagalli, 2018; Mary et al., 2020). In most memory suppression literature, memory traces were suppressed immediately after their initial formation (Michael C Anderson & Green, 2001; Michael C Anderson & Hanslmayr, 2014). However, unwanted memories (usually traumatic memories) which are needed to be suppressed are usually remote, and therefore it is more challenging to modify them. Indeed, a recent report showed that initially consolidated emotional memories are more resistant to suppression and retain larger emotional reactivity compared to newly-formed ones. At the same time, Yunzhe and colleagues demonstrated a shift of hippocampally centered representations to distributed neocortical memory representations after the memories have been consolidated (Y. Liu et al., 2016). Similarly, we found the limited behavioral memory suppression effect. Our design allowed us to examine the neural processes further when participants try to retrieve those memories which were suppressed during the TNT. We revealed that retrieving suppressed memories involves lower prefrontal engagement, but rather intact item-specific neural representations during the subsequent memory retrieval. The observed prefrontal effect during subsequent retrieval might be a long-lasting effect of reduced prefrontal activity during suppression (Brendan E Depue et al., 2007) across repeated attempts and suggests that limited cognitive control resources hamper retrieval.

In summary, **Chapter 4** of this thesis showed how different memory modulations techniques induce strength dynamics and their different interactions with initial consolidation. After consolidation, active retrieval can further promote the episodic-unique memory representations in neocortical regions, while memory suppression demonstrates a modest effect on memory performance and underlying neural representations. I conclude that consolidation and retrieval practice seems to have an additive effect on creating long-lasting memories, while consolidation partly prevents the modification induced by suppression.

Dual modulation of memory and emotion: from strength dynamics to valence dynamics

Recent evidence suggested that memory suppression, controlling the tendency to retrieve during retrieval, maybe act as another routine for emotion regulation. Suppression can disrupt existing memory representations by the inhibitory processes, causing the failures in explicit memory retrieval (Michael C Anderson & Green, 2001; Michael C Anderson & Hanslmayr, 2014) and subconscious processing (Gagnepain et al., 2014). The dynamical change of memory strength may also be associated with the reductions in the emotional reactivity of the memory contents (De Vito & Fenske, 2017). This effect does not simply result from the repeated suppression, but more likely from implicit emotion regulation (Engen & Anderson, 2018). The dual regulation of memory content and its valence was hypothesized to be supported by similar neural networks (Engen & Anderson, 2018), and an experiment showed that both hippocampal and amygdalar activity is suppressed in parallel by the DLPFC (Gagnepain et al., 2017). The role of dorsomedial PFC in the conjunction between memory retrieval and emotion regulation has been highlighted separately from the perspective of aging (Ford & Kensinger, 2017, 2018).

The interplay between memory and emotion is one of the central topics of memory research, while the number of conducted research seems asymmetric (Kensinger & Ford, 2020). Specifically, much of prior research has emphasized the effect of emotion on memory (e.g., emotional memories are more frequently and more vividly remembered than neutral memories (Dewhurst & Parry, 2000; Ochsner, 2000)), but not the other way around. Depending on task instructions, retrieval of emotional memories can also affect the subsequent emotion perception. For example, emotional memory retrieval can be used as a strategy of emotion regulation (Gross, 1998) at the time of retrieval (Pasupathi, 2003; Pillemer, 2009).

Our results from **Chapter 5** provided a comprehensive investigation of common neurobiological underpinnings of memory control and emotion regulation using methods from neuroimaging and imaging-genetics. I propose that the *strength dynamics* of memories can be considered together with *valence dynamics* within a two-dimensional space. The movement of one particular memory trace within the space is supported by the common neurobiological correlations of inhibitory control. Therefore, improvement of inhibitory control may be beneficial for specifical cognitive processing such as memory control and emotion regulation.

Summary of memory dynamics

In the first two experimental chapters (***Chapters 2-3***) of the thesis, I investigated how the brain uses non-overlapping neural representations for different processes within a continuous experience (i.e., *process dynamics*). In ***Chapter 2***, different episodic events were represented by separate neural states. At the same time, these separate neural states were further integrated across time forming a coherent narrative (**Figure 6.1.A**). In ***Chapter 3***, different memory-related task demands are represented by distinct neural states. During the switch from one to the other task, the brain adapts dynamically to the corresponding state (**Figure 6.1.B**). These results together call for focusing on interactions between different mnemonic processes that are executed close in time and how these interactions are related to memory performance.

With studies in ***Chapters 4 and 5***, we aimed at revealing the neural changes underlying the *strength dynamics*. In ***Chapter 4***, we demonstrated how two memory modulation techniques either increase or decrease memory strength and how they interact with memory consolidation (**Figure 6.1.C**). In ***Chapter 5***, we showed that the dual modulation of memory strength and emotional intensity is supported by similar neurobiological correlates (**Figure 6.1.D**). In sum, these results suggest that memory "engrams" should be viewed as states that could be susceptible to change, and contain multiple related elements of one memory (e.g., emotion).

Although not investigated in this thesis, process- and strength dynamics may interact with each other. For example, people may have a daily routine from getting up, washing his/her face, drinking coffee, taking the train before finally sitting at the office desk finally. And then one day, he/she witnesses a tragic car accident while commuting. In the following days or even months, he/she will keep retrieving the memory of the accident while taking the train. In this way, the memory representation gets increasingly stronger. At the same time, because taking the train belongs to a continuous experience, the strengthening memory will also affect the strengths of other adjacent memories, possibly enhancing all neighboring (e.g., biking from home to the train station). Interactions between strength dynamics and process dynamics together generate the rich collection of episodic memories with corresponding temporal and strength information.

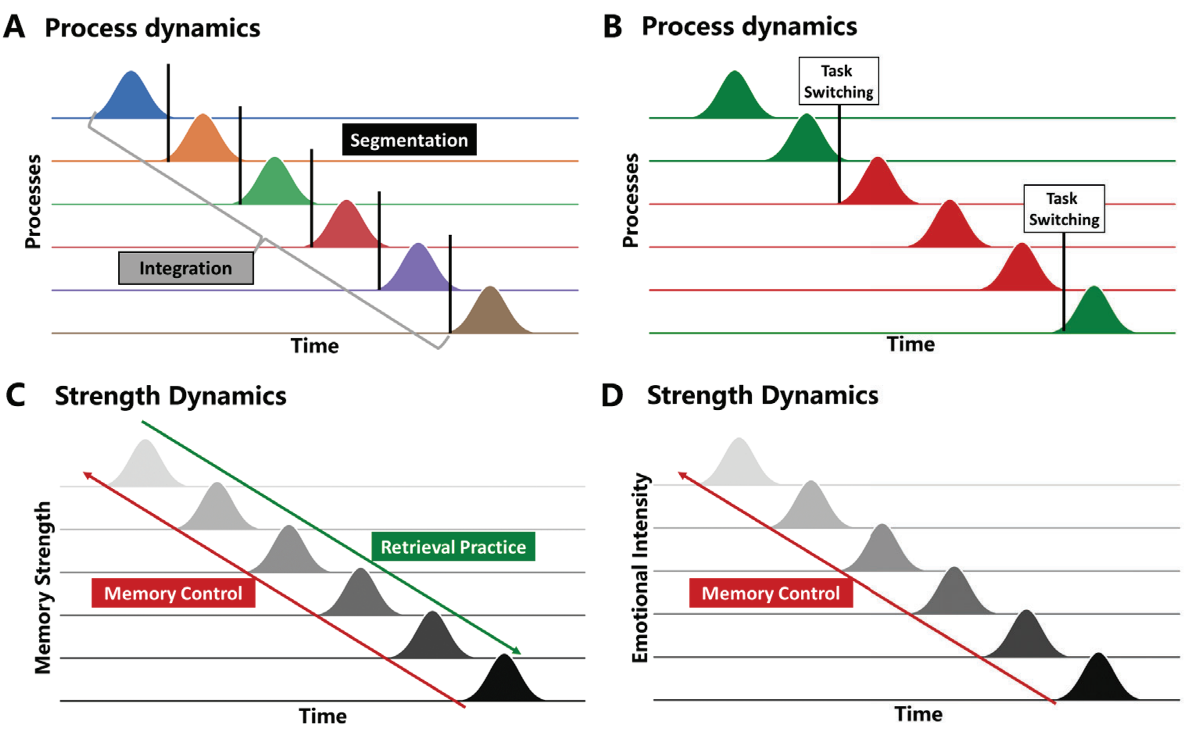


Figure 6.1 Role of two memory dynamics in specific memory processes. (A) During encoding of the continuous experience, processing units are events that to be encoded by non-overlapping neural states. The brain simultaneously performs the segmentation and integration for episodic events (*Chapter2*). **(B)** When the brain performs continuous task-switching, processing units are "task demand" for each time window. The brain remains in the same neural state for the same task demand but needs to reconfigure its neural state for a different task demand during switching (*Chapter3*). **(C)** For a specific memory trace, external modulations can lead to strength dynamics: retrieval practice can increase its memory strength, while memory control can decrease its strength (*Chapter4*). **(D)** After memory control, the dynamics of memory strength of a memory trace (from high to low) could also associate with the reduction in its emotional intensity (*Chapter5*).

The work presented in this thesis has answered several research questions but also raised new questions. I will discuss them in the following sections and suggest alternative methods or future studies with a new experimental design.

Future directions

1. Towards the individualized nested event memory processing in the brain

In **Chapter 2**, we investigated the relationship between neural similarities across event boundaries and episodic memory formation. However, the event boundary data was determined by a separate group of participants whose fMRI data is not available. Although the approach is widely used in the literature and demonstrated optimal correspondence with neural responses (Baldassano et al., 2017; Ben-Yakov & Henson, 2018; Janice Chen et al., 2017), current analyses were based on the assumption that individual differences in event perception can be ignored. This assumption was challenged by behavioral evidence: individual differences in behavioral event segmentation have been observed and were associated with memory retrieval performance (Sargent et al., 2013). To reveal individual differences in event processing, fMRI data, and event boundary data should be collected from one large group of participants with deep phenotyping of relevant variables.

Also, these boundaries were determined at the coarsest time scale, with a duration of about 1 minute per event. Current coarse boundaries can only allow us to evaluate the subsequent memory performance for entire events, but not sub-elements within them. Ideally, the continuous experience would be segmented at a different level. For example, the highest conceptual level with the complete episodes as events or lowest perceptual level with the continuous presentation of one visual element (e.g., a dog) as an event. These boundaries could form a nested event structure with one high-level event is the combination of several low-level events. Aside from events labeled by human participants, methods from computer vision and natural language processing can be adapted to analyze the content of a movie (Rohrbach et al., 2013; Töreyin et al., 2006), and event boundaries can be determined in a data-driven way based on different models (e.g., item model, context model, character model, and so on). Modeling the nested event structure with fMRI data can give us more insights into the region-specific timescales of changes in cortical patterns changes during the continuous perception of a narrative (Baldassano et al., 2017).

2. Memory-related pattern reinstatements in the human brain

In **Chapter 4**, we used RSA-based MVPA to detect episode-unique pattern reinstatements during memory retrieval. Even though MVPA is a powerful method to probe the reinstatement of episodic memories (Xue, 2018), we need to evaluate each result critically and be fully aware of the limitation of fMRI-based MVPA, compared to other invasive methods (Dubois et al., 2015).

First, comprehensive investigations should be conducted to examine the relationship between different fMRI-based MVPA measures and memory retrieval outcomes (e.g., speed, accuracy, subjective confidence, temporal order). Future research could try to apply both RSA-based MVPA and classification-based MVPA to the same dataset and compare their results. It is possible that different MVPA indexes, which were thought to all measure pattern reinstatements, may track different aspects of memory. Moreover, because MVPA can be influenced by many factors, such as experimental setup (Coutanche & Thompson-Schill, 2012; Mumford et al., 2014), preprocessing-and analysis choices (de Beeck, 2010; Haynes, 2015), choice of regions-of-interests (Kriegeskorte et al., 2006), and signal-to-noise ratio (A. T. Smith et al., 2011) (for an in-depth discussion, see **Chapter 4**). Future studies can try different analytic choices, to see how they might affect the results, and aim for developing a relative optimal pipeline for measuring pattern reinstatement signals in the brain.

Second, fMRI-based measures of pattern reinstatements are not suitable to detect the fine-grained temporal information during memory retrieval. Memory-related pattern reinstatements are not only about reactivating the same neural patterns that were present when the memory was first experienced in the same region (i.e., spatial patterns) (Janice Chen et al., 2017; Kent & Lamberts, 2008; S.-H. Lee et al., 2019; Polyn et al., 2005; Tulving, 1984; Xue, 2018), but also how these patterns unfold over time (i.e., temporal patterns). Although the role of sequential replay in memory retrieval has been proposed (Carr et al., 2011) and supported by animal data from spatial memory paradigms (Pfeiffer, 2020; Takahashi, 2015), it is until recently that evidence from human electrophysiology link it with episodic memory retrieval (Michelmann et al., 2016, 2019; Vaz et al., 2020). There is an exciting avenue for future research: how the retrieval of the same memory trace could induce distinct pattern reinstatements in the spatial and temporal domain? What is the relationship between the spatial and temporal domains? Simultaneous EEG-fMRI recording is a powerful method to answer these questions and was already used by one pioneering work (Benjamin James Griffiths et al., 2019).

3. Dynamical brain state transitions: all-or-none or continuous process?

In **Chapter 2,** we used event boundaries to separate neural processing for different episodic events: the time points before and after a certain boundary belong to two states. In **Chapter 3**, we used a time-resolved multivariate binary neural decoder to track the transition between two neural states. Therefore, in both studies, we assumed that these brain state transitions during the task are all-or-none processes, and without the mixture of two or more states at the same time. Let us first assume that brain state transitions indeed follow the all-or-none principle. Then, the critical question is, how can we better detect the exact point of state transitions. There are at least two advanced methods available to detect the exact point of transitions that define the states based on continuous time series. For instance, the Hidden Markov Model (HMM) (Blunsom, 2004) can be used to identify hidden states and their boundaries through continuous data. This method has been used in identifying distinct neural representations of events in participants that were watching a movie (Baldassano et al., 2017). Alternatively, spatial standard deviation (SSD), together with the clustering method can be used to detect the time points transition (Y. B. Lee et al., 2019). This method successfully detected the hidden states within the dynamically changing of fMRI brain network states.

Another more likely possibility is that these transitions are continuous: at a certain time-point, the brain can be in both two (or more) states, but the balance between them changes over time. This idea is similar to the storage of items in the working memory: several items can be held at the same time, but their relative strength may differ based on the task demands. Results from **Chapter 2** and **Chapter 3** already suggested the potential existence of continuous processing. In **Chapter 2**, we found that connectivity patterns persist across event boundaries, potentially serving as the neural correlates of event integration. In **Chapter 3,** we found that brain states usually cannot switch immediately according to the task demand at hand. That is to say, near the switching boundaries, two brain states may compete with each other for mental resources. Future research should try to identify more evidence for continuous processing, and whether there is a boundary condition between the all-or-none and a continuous transition during continuous cognition. Nevertheless, the fMRI signal has its internal limitations when the research question is about precise temporal information. Future studies may also measure the neural activity during similar paradigms using electroencephalography (EEG), magnetoencephalography (MEG), or intracranial EEG.

Conclusions

Our results provide a perspective on the dynamic nature of memory in terms of temporal and strength. For the temporal dynamics, our experiments revealed that during the continuous experience, the brain uses separate neural states to segment information into events and simultaneously binds them into a coherent narrative by context encoding. During fast switches between memory retrieval and memory control, the brain needs to reconfigure its neural states in time. Otherwise, the remaining state of retrieval may cause failures in memory control. For the strength dynamics, after initial consolidation, active retrieval seems to promote episode-unique mnemonic representations, leading to enhanced memory strength. By contrast, memory control disengages prefrontal involvement during retrieval, causing compromised memory strength. These changes in memory strength are associated with changes in the emotional intensity of individual memory traces. This dual modulation phenomenon is supported by the common inhibitory control network and corresponding transcriptional correlates. Future research on memory dynamics might lead to non-pharmacological, cognitive approaches that can enhance the encoding efficiency and persistence of everyday memories or can modify traumatic memories and their emotional impacts. Such methods could potentially provide both fundamental and applied knowledge for memory-related symptoms in memory disorders and affective disorders.

Appendix

References
Research data management
Summary of thesis
List of publications
Academic portfolio DGCN
Curriculum Vitae
Acknowledgments
Donders Graduate School for Cognitive Neuroscience

References

Amstadter, A. (2008). Emotion regulation and anxiety disorders. *Journal of Anxiety Disorders*, *22*(2), 211–221.

Anderson, Michael C., & Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. *Trends in Cognitive Sciences*, *18*(6), 279–292.

Anderson, Michael C. (2004). Neural Systems Underlying the Suppression of Unwanted Memories. *Science*, *303*(5655), 232–235.

Anderson, Michael C, & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, *410*(6826), 366–369.

Anderson, Michael C, & Huddleston, E. (2012). Towards a cognitive and neurobiological model of motivated forgetting. In *True and false recovered memories* (pp. 53–120). Springer.

Andersson, J. L. R., Jenkinson, M., Smith, S., & others. (2007). Non-linear registration aka Spatial normalisation FMRIB Technial Report TR07JA2. *FMRIB Analysis Group of the University of Oxford*.

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

Anttila, V., Bulik-Sullivan, B., Finucane, H. K., Walters, R. K., Bras, J., Duncan, L., Escott-Price, V., Falcone, G. J., Gormley, P., Malik, R., Patsopoulos, N. A., Ripke, S., Wei, Z., Yu, D., Lee, P. H., Turley, P., Grenier-Boley, B., Chouraki, V., Kamatani, Y., … Neale, B. M. (2018). Analysis of shared heritability in common disorders of the brain. *Science*, *360*(6395), eaap8757.

Arnatkevičiūtė, A., Fulcher, B. D., & Fornito, A. (2019). A practical guide to linking brain-wide gene expression and neuroimaging data. *NeuroImage*, *189*, 353–367.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, *8*(4), 170–177.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: one decade on. *Trends in Cognitive Sciences*, *18*(4), 177–185.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29.

Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, *7*(5), 358–366.

Backus, A. R., Schoffelen, J.-M., Szebényi, S., Hanslmayr, S., & Doeller, C. F. (2016). Hippocampal-prefrontal theta oscillations support memory integration. *Current Biology*, *26*(4), 450–457.

Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. L. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science*, *319*(5870), 1640–1642.

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*, *95*(3), 709-721.e5.

Bellmund, J. L., Deuker, L., & Doeller, C. F. (2019). Mapping sequence structure in the human lateral entorhinal cortex. *ELife*, *8*, 458133.

Ben-Yakov, A., & Dudai, Y. (2011). Constructing realistic engrams: poststimulus activity of hippocampus and dorsal striatum predicts subsequent episodic memory. *Journal of Neuroscience*, *31*(24), 9032–9042.

Ben-Yakov, A., Eshel, N., & Dudai, Y. (2013). Hippocampal immediate poststimulus activity in the encoding of consecutive naturalistic episodes. *Journal of Experimental Psychology: General*, *142*(4), 1255–1263.

Ben-Yakov, A., & Henson, R. N. (2018). The Hippocampal Film Editor: Sensitivity and Specificity to Event Boundaries in Continuous Experience. *Journal of Neuroscience*, *38*(47), 10057–10068.

Benoit, R. G., Davies, D. J., & Anderson, M. C. (2016). Reducing future fears by suppressing the brain mechanisms underlying episodic simulation. *Proceedings of the National Academy of Sciences*, *113*(52), E8492--E8501.

Benoit, R. G., Hulbert, J. C., Huddleston, E., & Anderson, M. C. (2015). Adaptive top--down suppression of hippocampal activity and the purging of intrusive memories from consciousness. *Journal of Cognitive Neuroscience*, *27*(1), 96–111.

Bergström, Z. M., Anderson, M. C., Buda, M., Simons, J. S., & Richardson-Klavehn, A. (2013). Intentional retrieval suppression can conceal guilty knowledge in ERP memory detection tests. *Biological Psychology*, *94*(1), 1–11.

Berkers, R. M. W. J., van der Linden, M., Neville, D. A., van Kesteren, M. T. R., Morris, R. G. M., Murre, J. M. J., & Fernandez, G. (2018). Neural dynamics of accumulating and updating linguistic knowledge structures. *BioRxiv*, 495168.

Berto, S., Wang, G.-Z., Germi, J., Lega, B. C., & Konopka, G. (2018). Human Genomic Signatures of Brain Oscillations During Memory Encoding. *Cerebral Cortex*, *28*(5), 1733–1748.

Blunsom, P. (2004). Hidden markov models. *Lecture Notes, August*, *15*(18–19), 48.

Bode, S., & Haynes, J.-D. (2009). Decoding sequential stages of task preparation in the human brain. *Neuroimage*, *45*(2), 606–613.

Borota, D., Murray, E., Keceli, G., Chang, A., Watabe, J. M., Ly, M., Toscano, J. P., & Yassa, M. A. (2014). Post-study caffeine administration enhances memory consolidation in humans. *Nature Neuroscience*, *17*(2), 201–203.

Bosch, S. E., Jehee, J. F. M., Fernández, G., & Doeller, C. F. (2014). Reinstatement of associative memories in early visual cortex is signaled by the hippocampus. *Journal of Neuroscience*, *34*(22), 7493–7500.

Bovy, L., Tendolkar, I., Fernández, G., & Dresler, M. (2019). Sleep, Emotional Memories, and Depression. In *Handbook of Behavioral Neuroscience* (Vol. 30, pp. 519–531). Elsevier.

Braver, T. S., Reynolds, J. R., & Donaldson, D. I. (2003). Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron*, *39*(4), 713–726.

Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1998). Making memories: brain activity that predicts how well visual experience will be remembered. *Science*, *281*(5380), 1185–1187.

Buckner, R. L., Koutstaal, W., Schacter, D. L., Dale, A. M., Rotte, M., & Rosen, B. R. (1998). Functional--anatomic study of episodic retrieval: II. Selective averaging of event-related fMRI trials to test the retrieval success hypothesis. *Neuroimage*, *7*(3), 163–175.

Buckner, R. L., Krienen, F. M., & Yeo, B. T. T. (2013). Opportunities and limitations of intrinsic functional connectivity MRI. *Nature Neuroscience*, *16*(7), 832–837.

Buhle, J. T., Silvers, J. A., Wager, T. D., Lopez, R., Onyemekwu, C., Kober, H., Weber, J., & Ochsner, K. N. (2014). Cognitive reappraisal of emotion: a meta-analysis of human neuroimaging studies. *Cerebral Cortex*, *24*(11), 2981–2990.

Bunge, S. A., Kahn, I., Wallis, J. D., Miller, E. K., & Wagner, A. D. (2003). Neural circuits subserving the retrieval and maintenance of abstract rules. *Journal of Neurophysiology*, *90*(5), 3419–3428.

Buysse, D. J., Reynolds III, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Research*, *28*(2), 193–213.

Carr, M. F., Jadhav, S. P., & Frank, L. M. (2011). Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nature Neuroscience*, *14*(2), 147–153.

Castiglione, A., Wagner, J., Anderson, M., & Aron, A. R. (2019). Preventing a Thought from Coming to Mind Elicits Increased Right Frontal Beta Just as Stopping Action Does. *Cerebral Cortex*, *29*(5), 2160–2172.

Catarino, A., Küpper, C. S., Werner-Seidler, A., Dalgleish, T., & Anderson, M. C. (2015). Failing to forget: Inhibitory-control deficits compromise memory suppression in posttraumatic stress disorder. *Psychological Science*, *26*(5), 604–616.

Chen, Janice, Honey, C. J., Simony, E., Arcaro, M. J., Norman, K. A., & Hasson, U. (2016). Accessing real-life episodic information from minutes versus hours earlier modulates hippocampal and high-order cortical dynamics. *Cerebral Cortex*, *26*(8), 3428–3441.

Chen, Janice, Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, *20*(1), 115–125.

Chen, Jing, Bardes, E. E., Aronow, B. J., & Jegga, A. G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Research*, *37*(suppl_2), W305--W311.

Chien, H.-Y. S., & Honey, C. J. (2020). Constructing and Forgetting Temporal Context in the Human Cerebral Cortex. *Neuron*, *106*(4), 675-686.e11.

Cocuzza, C. V., Ito, T., Schultz, D. H., Bassett, D. S., & Cole, M. W. (2019). Flexible coordinator and switcher hubs for adaptive task control. *BioRxiv*, 822213.

Cohen, J. D., Daw, N., Engelhardt, B., Hasson, U., Li, K., Niv, Y., Norman, K. A., Pillow, J., Ramadge, P. J., Turk-Browne, N. B., & Willke, T. L. (2017). Computational approaches to fMRI analysis. *Nature Neuroscience*, *20*(3), 304–313.

Cohen, M. R., & Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature Neuroscience*, *14*(7), 811–819.

Cole, M. W., Etzel, J. A., Zacks, J. M., Schneider, W., & Braver, T. S. (2011). Rapid transfer of abstract rules to novel contexts in human lateral prefrontal cortex. *Frontiers in Human Neuroscience*, *5*, 142.

Cole, M. W., Ito, T., Schultz, D., Mill, R., Chen, R., & Cocuzza, C. (2019). Task activations produce spurious but systematic inflation of task functional connectivity estimates. *NeuroImage*, *189*, 1–18.

Combrisson, E., & Jerbi, K. (2015). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of Neuroscience*

Methods, *250*, 126–136.

Courtney, K. E., Ghahremani, D. G., & Ray, L. A. (2013). Fronto-striatal functional connectivity during response inhibition in alcohol dependence. *Addiction Biology*, *18*(3), 593–604.

Coutanche, M. N., & Thompson-Schill, S. L. (2012). The advantage of brief fMRI acquisition runs for multi-voxel pattern detection across runs. *Neuroimage*, *61*(4), 1113–1119.

Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet*, *381*(9875), 1371–1379.

de Beeck, H. P. O. (2010). Probing the mysterious underpinnings of multi-voxel fMRI analyses. *Neuroimage*, *50*(2), 567–571.

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, *22*(9), 764–779.

De Vito, D., & Fenske, M. J. (2017). Suppressing memories of words and familiar objects results in their affective devaluation: Evidence from Think/No-think tasks. *Cognition*, *162*, 1–11.

Deisseroth, K. (2011). Optogenetics. *Nature Methods*, *8*(1), 26–29.

Depue, B. E., Orr, J. M., Smolker, H. R., Naaz, F., & Banich, M. T. (2016). The Organization of Right Prefrontal Networks Reveals Common Mechanisms of Inhibitory Regulation Across Cognitive, Emotional, and Motor Processes. *Cerebral Cortex*, *26*(4), 1634–1646.

Depue, Brendan E, Curran, T., & Banich, M. T. (2007). Prefrontal Regions Orchestrate Suppression of Emotional Memories via a Two-Phase Process. *Science*, *317*(5835), 215–219.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., & others. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, *31*(3), 968–980.

Dewhurst, S. A., & Parry, L. A. (2000). Emotionality, distinctiveness, and recollective experience. *European Journal of Cognitive Psychology*, *12*(4), 541–551.

Dillon, D. G., & Pizzagalli, D. A. (2018). Mechanisms of memory disruption in depression. *Trends in Neurosciences*, *41*(3), 137–149.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158–164.

Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., Fox, M. D., Snyder, A. Z., Vincent, J. L., Raichle, M. E., & others. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences*, *104*(26), 11073–11078.

Dove, A., Pollmann, S., Schubert, T., Wiggins, C. J., & Von Cramon, D. Y. (2000). Prefrontal cortex activation in task switching: an event-related fMRI study. *Cognitive Brain Research*, *9*(1), 103–109.

Dubois, J., de Berker, A. O., & Tsao, D. Y. (2015). Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *Journal of Neuroscience*, *35*(6), 2791–2802.

DuBrow, S., & Davachi, L. (2013). The influence of context boundaries on memory for the sequential order of events. *Journal of Experimental Psychology: General*, *142*(4), 1277–1286.

DuBrow, S., & Davachi, L. (2016). Temporal binding within and across events. *Neurobiology of Learning and Memory*, *134*, 107–114.

DuBrow, S., Rouhani, N., Niv, Y., & Norman, K. A. (2017). Does mental context drift or shift? *Current Opinion in Behavioral Sciences*, *17*, 141–146.

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience*, *2*(11), 820–829.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*(4), 172–179.

Dunsmoor, J. E., Murty, V. P., Davachi, L., & Phelps, E. A. (2015). Emotional learning selectively and retroactively strengthens memories for related events. *Nature*, *520*(7547), 345–348.

Ebbinghaus, H. (1885). *Über das gedächtnis: untersuchungen zur experimentellen psychologie*. Duncker & Humblot.

Ehring, T., & Quack, D. (2010). Emotion regulation difficulties in trauma survivors: The role of trauma type and PTSD symptom severity. *Behavior Therapy*, *41*(4), 587–598.

Eickhoff, S. B., Bzdok, D., Laird, A. R., Roski, C., Caspers, S., Zilles, K., & Fox, P. T. (2011). Co-activation patterns distinguish cortical modules, their connectivity and functional differentiation. *Neuroimage*, *57*(3), 938–949.

Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping*, *30*(9), 2907–2926.

Elger, C. E., Helmstaedter, C., & Kurthen, M. (2004). Chronic epilepsy and cognition. *The Lancet Neurology*, *3*(11), 663–672.

Elliott, L. T., Sharp, K., Alfaro-Almagro, F., Shi, S., Miller, K. L., Douaud, G., Marchini, J., & Smith, S. M. (2018). Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature*, *562*(7726), 210–216.

Engen, H. G., & Anderson, M. C. (2018). Memory Control: A Fundamental Mechanism of Emotion Regulation. *Trends in Cognitive Sciences*, *22*(11), 982–995.

Eriksson, J., Kalpouzos, G., & Nyberg, L. (2011). Rewiring the brain with repeated retrieval: a parametric fMRI study of the testing effect. *Neuroscience Letters*, *505*(1), 36–40.

Erk, S., Mikschl, A., Stier, S., Ciaramidaro, A., Gapp, V., Weber, B., & Walter, H. (2010). Acute and sustained effects of cognitive emotion regulation in major depression. *Journal of Neuroscience*, *30*(47), 15726–15734.

Etzel, J. A., Cole, M. W., Zacks, J. M., Kay, K. N., & Braver, T. S. (2016). Reward motivation enhances task coding in frontoparietal cortex. *Cerebral Cortex*, *26*(4), 1647–1659.

Falconer, E., Bryant, R., Felmingham, K. L., Kemp, A. H., Gordon, E., Peduto, A., Olivieri, G., & Williams, L. M. (2008). The neural networks of inhibitory control in posttraumatic stress disorder. *Journal of Psychiatry & Neuroscience: JPN*, *33*(5), 413.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences*, *110*(41), 16616–16621.

Fenno, L., Yizhar, O., & Deisseroth, K. (2011). The Development and Application of Optogenetics. *Annual Review of Neuroscience*, *34*(1), 389–412.

Ferguson, M. A., Lim, C., Cooke, D., Darby, R. R., Wu, O., Rost, N. S., Corbetta, M., Grafman, J., & Fox, M. D. (2019). A human memory circuit derived from brain lesions causing amnesia. *Nature Communications*, *10*(1), 1–9.

Fernández, G., Effern, A., Grunwald, T., Pezer, N., Lehnertz, K., Dümpelmann, M., Van Roost, D., & Elger, C. E. (1999). Real-time tracking of memory formation in the human rhinal cortex and hippocampus. *Science*, *285*(5433), 1582–1585.

Fernández, G., & Morris, R. G. M. (2018). Memory, novelty and prior knowledge. *Trends in Neurosciences*, *41*(10), 654–659.
Ferreira, C. S., Charest, I., & Wimber, M. (2019). Retrieval aids the creation of a generalised memory trace and strengthens episode-unique information. *NeuroImage*, *201*, 115996.

Ford, J. H., & Kensinger, E. A. (2017). Prefrontally-mediated alterations in the retrieval of negative events: Links to memory vividness across the adult lifespan. *Neuropsychologia*, *102*, 82–94.

Ford, J. H., & Kensinger, E. A. (2018). Older adults use a prefrontal regulatory mechanism to reduce negative memory vividness of a highly emotional real-world event. *NeuroReport*, *29*(13), 1129–1134.

Fornito, A., Arnatkevičiūtė, A., & Fulcher, B. D. (2019). Bridging the Gap between Connectome and Transcriptome. *Trends in Cognitive Sciences*, *23*(1), 34–50.

Fox, P. T., & Lancaster, J. L. (2002). Mapping context and content: the BrainMap model. *Nature Reviews Neuroscience*, *3*(4), 319–321.

Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience*, *6*(2), 119–130.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, *2*(4), 189–210.

Frithsen, A., & Miller, M. B. (2014). The posterior parietal cortex: Comparing remember/know and source memory tests of recollection and familiarity. *Neuropsychologia*, *61*, 31–44.

Fulcher, B. D., Arnatkeviciute, A., & Fornito, A. (2020). Overcoming bias in gene-set enrichment analyses of brain-wide transcriptomic data. *BioRxiv*.

Gagnepain, P., Henson, R. N., & Anderson, M. C. (2014). Suppressing unwanted memories reduces their unconscious influence via targeted cortical inhibition. *Proceedings of the National Academy of Sciences*, *111*(13), E1310–E1319.

Gagnepain, P., Hulbert, J., & Anderson, M. C. (2017). Parallel Regulation of Memory and Emotion Supports the Suppression of Intrusive Memories. *The Journal of Neuroscience*, *37*(27), 6423–6441.

Ganesh, G., Minamoto, T., & Haruno, M. (2019). Activity in the dorsal ACC causes deterioration of sequential motor performance due to anxiety. *Nature Communications*, *10*(1), 1–11.

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, *15*(4), 870–878.

Gerlicher, A. M. V, Tüscher, O., & Kalisch, R. (2018). Dopamine-dependent prefrontal reactivations explain long-term benefit of fear extinction. *Nature Communications*, *9*(1), 1–9.

Gilbert, S. J. (2011). Decoding the content of delayed intentions. *Journal of Neuroscience*, *31*(8), 2888–2894.
Gilmore, A. W., Nelson, S. M., & McDermott, K. B. (2015). A parietal memory network revealed by multiple MRI methods. *Trends in Cognitive Sciences*, *19*(9), 534–543.

Gonzalez-Castillo, J., Hoy, C. W., Handwerker, D. A., Robinson, M. E., Buchanan, L. C., Saad, Z. S., & Bandettini, P. A. (2015). Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(28), 8762–8767.

Goodkind, M., Eickhoff, S. B., Oathes, D. J., Jiang, Y., Chang, A., Jones-Hagata, L. B., Ortega, B. N., Zaiko, Y. V, Roach, E. L., Korgaonkar, M. S., & others. (2015). Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry*, *72*(4), 305–315.

Gorgolewski, K., Fox, A., Chang, L., Schäfer, A., Arélin, K., Burmann, I., Sacher, J., & Margulies, D. (2014). Tight fitting genes: Finding relations between statistical maps and gene expression patterns. *Organization for Human Brain Mapping. Hamburg, Germany*.

Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., Sochat, V. V, Nichols, T. E., Poldrack, R. A., Poline, J.-B., & others. (2015). NeuroVault. org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers in Neuroinformatics*, *9*, 8.

Goschke, T. (2000). Intentional reconfiguration and J-TI involuntary persistence in task set switching. *Control of Cognitive Processes: Attention and Performance XVIII*, *18*, 331.

Gottlieb, L. J., Uncapher, M. R., & Rugg, M. D. (2010). Dissociation of the neural correlates of visual and auditory contextual encoding. *Neuropsychologia*, *48*(1), 137–144.

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, *48*(1), 63–72.

Griffiths, Benjamin J, & Fuentemilla, L. (2020). Event conjunction: How the hippocampus integrates episodic memories across event boundaries. *Hippocampus*, *30*(2), 162–171.

Griffiths, Benjamin James, Mayhew, S. D., Mullinger, K. J., Jorge, J., Charest, I., Wimber, M., & Hanslmayr, S. (2019). Alpha/beta power decreases track the fidelity of stimulus-specific information. *ELife*, *8*.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, *2*(3), 271–299.

Gross, J. J. (2013). *Handbook of emotion regulation*. Guilford publications.

Gross, J. J., & Thompson, R. A. (2007). *Emotion regulation: Conceptual foundations.* Guilford Press.

Grothe, M. J., Sepulcre, J., Gonzalez-Escamilla, G., Jelistratova, I., Schöll, M., Hansson, O., Teipel, S. J., & Initiative, A. D. N. (2018). Molecular properties underlying regional vulnerability to Alzheimer's disease pathology. *Brain*, *141*(9), 2755–2771.

Gruber, M. J., Ritchey, M., Wang, S.-F., Doss, M. K., & Ranganath, C. (2016). Post-learning hippocampal dynamics promote preferential retention of rewarding events. *Neuron*, *89*(5), 1110–1120.

Gruber, O., Karch, S., Schlueter, E. K., Falkai, P., & Goschke, T. (2006). Neural mechanisms of advance preparation in task switching. *Neuroimage*, *31*(2), 887–895.

Guo, Y., Schmitz, T. W., Mur, M., Ferreira, C. S., & Anderson, M. C. (2018). A supramodal role of the basal ganglia in memory and motor inhibition: Meta-analytic evidence. *Neuropsychologia*, *108*, 117–134.

Han, J.-H., Kushner, S. A., Yiu, A. P., Cole, C. J., Matynia, A., Brown, R. A., Neve, R. L., Guzowski, J. F., Silva, A. J., & Josselyn, S. A. (2007). Neuronal competition and selection during memory formation. *Science*, *316*(5823), 457–460.

Han, J.-H., Kushner, S. A., Yiu, A. P., Hsiang, H.-L. L., Buch, T., Waisman, A., Bontempi, B., Neve, R. L., Frankland, P. W., & Josselyn, S. A. (2009). Selective erasure of a fear memory. *Science*, *323*(5920), 1492–1496.

Hariri, A. R., Mattay, V. S., Tessitore, A., Kolachana, B., Fera, F., Goldman, D., Egan, M. F., & Weinberger, D. R. (2002). Serotonin transporter genetic variation and the response of the human amygdala. *Science*, *297*(5580), 400–403.

Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: memory as an integral component of information processing. *Trends in Cognitive Sciences*, *19*(6), 304–313.

Hasson, U., Furman, O., Clark, D., Dudai, Y., & Davachi, L. (2008). Enhanced Intersubject Correlations during Movie Viewing Correlate with Successful Episodic Encoding. *Neuron*, *57*(3), 452–462.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, *303*(5664), 1634–1640.

Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., Shen, E. H., Ng, L., Miller, J. A., Van De Lagemaat, L. N., Smith, K. A., Ebbert, A., Riley, Z. L., Abajian, C., Beckmann, C. F., Bernard, A., Bertagnolli, D., Boe, A. F., Cartagena, P. M., Mallar Chakravarty, M., Chapin, M., Chong, J., … Jones, A. R. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*, *489*(7416), 391–399.

Haynes, J.-D. (2015). A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives. *Neuron*, *87*(2), 257–270.

Hellerstedt, R., Johansson, M., & Anderson, M. C. (2016). Tracking the intrusion of unwanted memories into awareness with event-related potentials. *Neuropsychologia*, *89*, 510–523.

Helmstaedter, C. (2001). Behavioral aspects of frontal lobe epilepsy. *Epilepsy & Behavior*, *2*(5), 384–395.
Hermans, E. J., Kanen, J. W., Tambini, A., Fernández, G., Davachi, L., & Phelps, E. A. (2017). Persistence of amygdala--hippocampal connectivity and multi-voxel correlation structures during awake rest after fear learning predicts long-term expression of fear. *Cerebral Cortex*, *27*(5), 3028–3041.

Hermans, E. J., Van Marle, H. J. F., Ossewaarde, L., Henckens, M. J. A. G., Qin, S., Van Kesteren, M. T. R., Schoots, V. C., Cousijn, H., Rijpkema, M., Oostenveld, R., & others. (2011). Stress-related noradrenergic activity prompts large-scale neural network reconfiguration. *Science*, *334*(6059), 1151–1153.

Himmer, L., Schönauer, M., Heib, D. P. J., Schabus, M., & Gais, S. (2019). Rehearsal initiates systems memory consolidation, sleep makes it last. *Science Advances*, *5*(4), eaav1695.

Ho, J., Tumkaya, T., Aryal, S., Choi, H., & Claridge-Chang, A. (2019). Moving beyond P values: data analysis with estimation graphics. *Nature Methods*, *16*(7), 565–566.

Howard, M. W., & Eichenbaum, H. (2013). The hippocampus, time, and memory across scales. *Journal of Experimental Psychology: General*, *142*(4), 1211–1230.

Howard, M. W., Fotedar, M. S., Datey, A. V, & Hasselmo, M. E. (2005). The Temporal Context Model in Spatial Navigation and Relational Learning: Toward a Common Explanation of Medial Temporal Lobe Function

Across Domains. *Psychological Review*, *112*(1), 75–116.

Hu, X., Cheng, L. Y., Chiu, M. H., & Paller, K. A. (2020). Promoting memory consolidation during sleep: A meta-analysis of targeted memory reactivation. *Psychological Bulletin*, *146*(3), 218–244.

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57.

Huang, P., Carlin, J. D., Alink, A., Kriegeskorte, N., Henson, R. N., & Correia, M. M. (2018). Prospective motion correction improves the sensitivity of fMRI pattern decoding. *Human Brain Mapping*, *39*(10), 4018–4031.

Huk, A., Bonnen, K., & He, B. J. (2018). Beyond Trial-Based Paradigms: Continuous Behavior, Ongoing Neural Activity, and Natural Stimuli. *The Journal of Neuroscience*, *38*(35), 7551–7558.

Hulbert, J. C., Henson, R. N., & Anderson, M. C. (2016). Inducing amnesia through systemic suppression. *Nature Communications*, *7*(1), 11003.

Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., Della Penna, S., Duyn, J. H., Glover, G. H., Gonzalez-Castillo, J., & others. (2013). Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, *80*, 360–378.

Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458.

Jacques, P. L. S., Olm, C., & Schacter, D. L. (2013). Neural mechanisms of reactivation-induced updating that enhance and distort memory. *Proceedings of the National Academy of Sciences*, *110*(49), 19671–19678.

Jafarpour, A., Griffin, S., Lin, J. J., & Knight, R. T. (2019). Medial orbitofrontal cortex, dorsolateral prefrontal cortex, and hippocampus differentially represent the event saliency. *Journal of Cognitive Neuroscience*, *31*(6), 874–884.

Jamalabadi, H., Alizadeh, S., Schönauer, M., Leibold, C., & Gais, S. (2016). Classification based hypothesis testing in neuroscience: Below-chance level classification rates and overlooked statistical properties of linear parametric classifiers. *Human Brain Mapping*, *37*(5), 1842–1855.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*(2), 825–841.

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). Fsl. *Neuroimage*, *62*(2), 782–790.

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156.

Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology*, *14*(89), 81–86.

Jimura, K., Cazalis, F., Stover, E. R. S., & Poldrack, R. A. (2014). The neural basis of task switching changes with skill acquisition. *Frontiers in Human Neuroscience*, *8*, 339.

Jonker, T. R., Dimsdale-Zucker, H., Ritchey, M., Clarke, A., & Ranganath, C. (2018). Neural reactivation in parietal cortex enhances memory for episodically linked information. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(43), 11084–11089.

Joormann, J., & Gotlib, I. H. (2010). Emotion regulation in depression: relation to cognitive inhibition.

Cognition and Emotion, *24*(2), 281–298.

Josselyn, S. A., Köhler, S., & Frankland, P. W. (2015). Finding the engram. *Nature Reviews Neuroscience*, *16*(9), 521–534.

Josselyn, S. A., & Tonegawa, S. (2020). Memory engrams: Recalling the past and imagining the future. *Science*, *367*(6473), eaaw4325.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*(6018), 772–775.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*(5865), 966–968.

Karpicke, J. D., & Roediger III, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*(2), 151–162.

Kensinger, E. A., & Ford, J. H. (2020). Retrieval of emotional events from memory. *Annual Review of Psychology*, *71*, 251–272.

Kent, C., & Lamberts, K. (2008). The encoding--retrieval relationship: retrieval as mental simulation. *Trends in Cognitive Sciences*, *12*(3), 92–98.

Kerrén, C., Linde-Domingo, J., & Hanslmayr, S. (2018). An optimal oscillatory phase for pattern reactivation during memory retrieval. *Current Biology*, *28*(21), 3383–3392.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin*, *136*(5), 849–874.

Kim, H. (2010). Dissociating the roles of the default-mode, dorsal, and ventral networks in episodic memory retrieval. *Neuroimage*, *50*(4), 1648–1657.

Kim, H. (2011). Neural activity that predicts subsequent memory and forgetting: a meta-analysis of 74 fMRI studies. *Neuroimage*, *54*(3), 2446–2461.

Kim, H. (2013). Differential neural activity in the recognition of old versus new events: An Activation Likelihood Estimation Meta-Analysis. *Human Brain Mapping*, *34*(4), 814–836.

Kohn, N., Eickhoff, S. B., Scheller, M., Laird, A. R., Fox, P. T., & Habel, U. (2014). Neural network of cognitive emotion regulation—an ALE meta-analysis and MACM analysis. *Neuroimage*, *87*, 345–355.

Kok, P., Jehee, J. F. M., & De Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*, *75*(2), 265–270.

Kong, X.-Z., Tzourio-Mazoyer, N., Joliot, M., Fedorenko, E., Liu, J., Fisher, S. E., & Francks, C. (2020). Gene Expression Correlates of the Cortical Network Underlying Sentence Processing. *Neurobiology of Language*, *1*(1), 77–103.

Kong, X., Song, Y., Zhen, Z., & Liu, J. (2017). Genetic Variation in S100B Modulates Neural Processing of Visual Scenes in Han Chinese. *Cerebral Cortex*, *27*, 1326–1336.

Kosslyn, S. M., Thompson, W. L., & Alpert, N. M. (1997). Neural systems shared by visual imagery and visual perception: A positron emission tomography study. *Neuroimage*, *6*(4), 320–334.

Kowalczyk, A., & Chapelle, O. (2005). An analysis of the anti-learning phenomenon for the class symmetric polyhedron. *International Conference on Algorithmic Learning Theory*, 78–91.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, *103*(10), 3863–3868.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4.

Kring, A. M., & Sloan, D. M. (2009). *Emotion regulation and psychopathology: A transdiagnostic approach to etiology and treatment*. Guilford Press.

Kroes, M. C. W., & Fernández, G. (2012). Dynamic neural systems enable adaptive, flexible memories. *Neuroscience & Biobehavioral Reviews*, *36*(7), 1646–1666.

Kudrimoti, H. S., Barnes, C. A., & McNaughton, B. L. (1999). Reactivation of hippocampal cell assemblies: effects of behavioral state, experience, and EEG dynamics. *Journal of Neuroscience*, *19*(10), 4090–4101.

Kuhl, B. A., Shah, A. T., DuBrow, S., & Wagner, A. D. (2010). Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nature Neuroscience*, *13*(4), 501–506.

Kumaran, D., Summerfield, J. J., Hassabis, D., & Maguire, E. A. (2009). Tracking the emergence of conceptual knowledge during human decision making. *Neuron*, *63*(6), 889–901.

Laird, A. R., Eickhoff, S. B., Li, K., Robin, D. A., Glahn, D. C., & Fox, P. T. (2009). Investigating the Functional Heterogeneity of the Default Mode Network Using Coordinate-Based Meta-Analytic Modeling. *Journal of Neuroscience*, *29*(46), 14496–14505.

Laird, A R, Eickhoff, S. B., Kurth, F., Fox, P. M., Uecker, A. M., Turner, J. A., Robinson, J. L., Lancaster, J. L., & Fox, P. T. (2009). ALE meta-analysis workflows via the BrainMap database: Progress towards a probabilistic functional brain atlas. *Frontiers in Neuroinformatics*, *3*, 23.

Laird, Angela R, Eickhoff, S. B., Fox, P. M., Uecker, A. M., Ray, K. L., Saenz, J. J., McKay, D. R., Bzdok, D., Laird, R. W., Robinson, J. L., Turner, J. A., Turkeltaub, P. E., Lancaster, J. L., & Fox, P. T. (2011). The BrainMap strategy for standardization, sharing, and meta-analysis of neuroimaging data. *BMC Research Notes*, *4*(1), 349.

Laird, Angela R, Lancaster, J. J., & Fox, P. T. (2005). Brainmap. *Neuroinformatics*, *3*(1), 65–77.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, *1*, 39–58.

Langner, R., Leiberg, S., Hoffstaedter, F., & Eickhoff, S. B. (2018). Towards a human self-regulation system: Common and distinct neural signatures of emotional and behavioural control. *Neuroscience & Biobehavioral Reviews*, *90*, 400–410.

Lansink, C. S., Goltstein, P. M., Lankelma, J. V, Joosten, R. N., McNaughton, B. L., & Pennartz, C. M. A. (2008). Preferential reactivation of motivationally relevant information in the ventral striatum. *Journal of Neuroscience*, *28*(25), 6372–6382.

Lashley, K. S. (1950). In search of the engram. *In Society for Experimental Biology, Physiological Mechanisms in Animal Behavior.*, 454–482.

Lawrence, A. J., Luty, J., Bogdan, N. A., Sahakian, B. J., & Clark, L. (2009). Impulsivity and response inhibition in alcohol dependence and problem gambling. *Psychopharmacology*, *207*(1), 163–172.

Lee, H., Chun, M. M., & Kuhl, B. A. (2017). Lower parietal encoding activation is associated with sharper information and better memory. *Cerebral Cortex*, *27*(4), 2486–2499.

Lee, S.-H., Kravitz, D. J., & Baker, C. I. (2019). Differential Representations of Perceived and Retrieved Visual Information in Hippocampus and Cortex. *Cerebral Cortex*, *29*(10), 4452–4461.

Lee, Y. B., Yoo, K., Roh, J. H., Moon, W. J., & Jeong, Y. (2019). Brain-State Extraction Algorithm Based on the State Transition (BEST): A Dynamic Functional Brain Network Analysis in fMRI Study. *Brain Topography*, *32*(5), 897–913.

Levy, B. J., & Anderson, M. C. (2012). Purging of Memories from Conscious Awareness Tracked in the Human Brain. *Journal of Neuroscience*, *32*(47), 16785–16794.

Levy, Benjamin J, & Anderson, M. C. (2002). Inhibitory processes and the control of memory retrieval. *Trends in Cognitive Sciences*, *6*(7), 299–305.

Liégeois, R., Li, J., Kong, R., Orban, C., Van De Ville, D., Ge, T., Sabuncu, M. R., & Yeo, B. T. T. (2019). Resting brain dynamics at different timescales capture distinct aspects of human behavior. *Nature Communications*, *10*(1), 1–9.

Linde-Domingo, J., Treder, M. S., Kerrén, C., & Wimber, M. (2019). Evidence that neural information flow is reversed between object perception and object reconstruction from memory. *Nature Communications*, *10*(1), 179.

Lindquist, M. A., Geuter, S., Wager, T. D., & Caffo, B. S. (2019). Modular preprocessing pipelines can reintroduce artifacts into fMRI data. *Human Brain Mapping*, *40*(8), 2358–2376.

Lipszyc, J., & Schachar, R. (2010). Inhibitory control and psychopathology: a meta-analysis of studies using the stop signal task. *Journal of the International Neuropsychological Society*, *16*(6), 1064–1076.

Liu, W., Kohn, N., & Fernández, G. (2020). Probing the neural dynamics of mnemonic representations after the initial consolidation. *BioRxiv*.

Liu, W., Peeters, N., Fernandez, G., & Kohn, N. (2020). Common neural and transcriptional correlates of inhibitory control across emotion, memory, and response inhibition. *BioRxiv*.

Liu, X., Ramirez, S., Pang, P. T., Puryear, C. B., Govindarajan, A., Deisseroth, K., & Tonegawa, S. (2012). Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature*, *484*(7394), 381–385.

Liu, Y., Lin, W., Liu, C., Luo, Y., Wu, J., Bayley, P. J., & Qin, S. (2016). Memory consolidation reconfigures neural pathways involved in the suppression of emotional memories. *Nature Communications*, *7*, 13375.

Loose, L. S., Wisniewski, D., Rusconi, M., Goschke, T., & Haynes, J. D. (2017). Switch-independent task representations in frontal and parietal cortex. *Journal of Neuroscience*, *37*(33), 8033–8042.

Lositsky, O., Chen, J., Toker, D., Honey, C. J., Shvartsman, M., Poppenk, J. L., Hasson, U., & Norman, K. A. (2016). Neural pattern change during encoding of a narrative predicts retrospective duration estimates. *ELife*, *5*, 1–40.

MacDonald, C. J., Lepage, K. Q., Eden, U. T., & Eichenbaum, H. (2011). Hippocampal “time cells” bridge the gap in memory for discontiguous events. *Neuron*, *71*(4), 737–749.

Magee, J. C., & Zinbarg, R. E. (2007). Suppressing and focusing on a negative memory in social anxiety: Effects on unwanted thoughts and mood. *Behaviour Research and Therapy*, *45*(12), 2836–2849.

Manns, J. R., Howard, M. W., & Eichenbaum, H. (2007). Gradual changes in hippocampal activity support remembering the order of events. *Neuron*, *56*(3), 530–540.

Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., Bezgin, G., Eickhoff, S. B., Castellanos, F. X., Petrides, M., Jefferies, E., & Smallwood, J. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences*, *113*(44), 12574–12579.

Mary, A., Dayan, J., Leone, G., Postel, C., Fraisse, F., Malle, C., Vallée, T., Klein-Peschanski, C., Viader, F., de la Sayette, V., Peschanski, D., Eustache, F., & Gagnepain, P. (2020). Resilience after trauma: The role of memory suppression. *Science*, *367*(6479), eaay8477.

McColgan, P., Gregory, S., Seunarine, K. K., Razi, A., Papoutsi, M., Johnson, E., Durr, A., Roos, R. A. C., Leavitt, B. R., Holmans, P., & others. (2018). Brain regions showing white matter loss in Huntington's disease are enriched for synaptic and metabolic genes. *Biological Psychiatry*, *83*(5), 456–465.

McDermott, K. B., Szpunar, K. K., & Christ, S. E. (2009). Laboratory-based and autobiographical retrieval tasks differ substantially in their neural substrates. *Neuropsychologia*, *47*(11), 2290–2298.

McGaugh, J. L., & Roozendaal, B. (2002). Role of adrenal stress hormones in forming lasting memories in the brain. *Current Opinion in Neurobiology*, *12*(2), 205–210.

McRae, K., Jacobs, S. E., Ray, R. D., John, O. P., & Gross, J. J. (2012). Individual differences in reappraisal ability: Links to reappraisal frequency, well-being, and cognitive control. *Journal of Research in Personality*, *46*(1), 2–7.

Meiran, N. (2010). Task Switching: Mechanisms Underlying Rigid vs. Flexible Self-Control. In *Self Control in Society, Mind, and Brain* (pp. 202–220). Oxford University Press.

Michelmann, S., Bowman, H., & Hanslmayr, S. (2016). The temporal signature of memories: identification of a general mechanism for dynamic memory replay in humans. *PLoS Biology*, *14*(8).

Michelmann, S., Staresina, B. P., Bowman, H., & Hanslmayr, S. (2019). Speed of time-compressed forward replay flexibly changes in human episodic memory. *Nature Human Behaviour*, *3*(2), 143–154.

Momennejad, I., & Haynes, J.-D. (2013). Encoding of prospective tasks in the human prefrontal cortex under varying task loads. *Journal of Neuroscience*, *33*(44), 17342–17349.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140.

Montchal, M. E., Reagh, Z. M., & Yassa, M. A. (2019). Precise temporal memories are supported by the lateral entorhinal cortex in humans. *Nature Neuroscience*, *22*(2), 284–288.

Morawetz, C., Bode, S., Derntl, B., & Heekeren, H. R. (2017). The effect of strategies, goals and stimulus material on the neural mechanisms of emotion regulation: A meta-analysis of fMRI studies. *Neuroscience & Biobehavioral Reviews*, *72*, 111–128.

Morgan, S. E., Seidlitz, J., Whitaker, K. J., Romero-Garcia, R., Clifton, N. E., Scarpazza, C., van Amelsvoort, T., Marcelis, M., van Os, J., Donohoe, G., & others. (2019). Cortical patterning of abnormal morphometric similarity in psychosis is associated with brain expression of schizophrenia-related genes. *Proceedings of the National Academy of Sciences*, *116*(19), 9604–9609.

Müller, G. E., & Pilzecker, A. (1900). *Experimentelle Beiträge zur Lehre vom Gedächtniss* (Vol. 1). JA Barth.
Mumford, J. A., Davis, T., & Poldrack, R. A. (2014). The impact of study design on pattern estimation for single-trial multivariate pattern analysis. *Neuroimage*, *103*, 130–138.

Nader, K., Schafe, G. E., & Le Doux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, *406*(6797), 722–726.

Nastase, S. A., Gazzola, V., Hasson, U., & Keysers, C. (2019). Measuring shared responses across subjects using intersubject correlation. *Social Cognitive and Affective Neuroscience*, *14*(6), 667–685.

Nelson, S. M., Arnold, K. M., Gilmore, A. W., & Mcdermott, K. B. (2013). Neural signatures of test-potentiated learning in parietal cortex. *Journal of Neuroscience*, *33*(29), 11754–11762.

Noreen, S., & MacLeod, M. D. (2013). It's all in the detail: Intentional forgetting of autobiographical memories using the autobiographical think/no-think task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 375–393.

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*(6), 1013–1023.

Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology: General*, *129*(2), 242–261.

Ochsner, K. N., Bunge, S. A., Gross, J. J., & Gabrieli, J. D. E. (2002). Rethinking feelings: an FMRI study of the cognitive regulation of emotion. *Journal of Cognitive Neuroscience*, *14*(8), 1215–1229.

Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, *9*(5), 242–249.

Otten, L. J. (2007). Fragments of a larger whole: retrieval cues constrain observed neural correlates of memory encoding. *Cerebral Cortex*, *17*(9), 2030–2038.

Oudiette, D., & Paller, K. A. (2013). Upgrading the sleeping brain with targeted memory reactivation. *Trends in Cognitive Sciences*, *17*(3), 142–149.

Papachristou, H., Nederkoorn, C., Havermans, R., Bongers, P., Beunen, S., & Jansen, A. (2013). Higher levels of trait impulsiveness and a less effective response inhibition are linked to more intense cue-elicited craving for alcohol in alcohol-dependent patients. *Psychopharmacology*, *228*(4), 641–649.

Pastalkova, E., Itskov, V., Amarasingham, A., & Buzsáki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science*, *321*(5894), 1322–1327.

Pasupathi, M. (2003). Emotion regulation during social remembering: Differences between emotions elicited during an event and emotions elicited when talking about it. *Memory*, *11*(2), 151–163.

Patil, A., Murty, V. P., Dunsmoor, J. E., Phelps, E. A., & Davachi, L. (2017). Reward retroactively enhances memory consolidation for related items. *Learning & Memory*, *24*(1), 65–69.

Paz, R., Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R., & Fried, I. (2010). A neural substrate in the human hippocampus for linking successive events. *Proceedings of the National Academy of Sciences*, *107*(13), 6046–6051.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., & others. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine*

Learning Research, *12*, 2825–2830.

Pfeiffer, B. E. (2020). The content of hippocampal “replay.” *Hippocampus*, *30*(1), 6–18.

Phelps, E. A., & Hofmann, S. G. (2019). Memory editing from science fiction to clinical practice. *Nature*, *572*(7767), 43–50.

Pillemer, D. B. (2009). Twenty years after Baddeley (1988): Is the study of autobiographical memory fully functional? *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *23*(8), 1193–1208.

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science*, *310*(5756), 1963–1966.

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage*, *59*(3), 2142–2154.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., Vogel, A. C., Laumann, T. O., Miezin, F. M., Schlaggar, B. L., & others. (2011). Functional network organization of the human brain. *Neuron*, *72*(4), 665–678.

Preston, A. R., & Eichenbaum, H. (2013). Interplay of Hippocampus and Prefrontal Cortex in Memory. *Current Biology*, *23*(17), R764–R773.

Price, R. B., & Mohlman, J. (2007). Inhibitory control and symptom severity in late life generalized anxiety disorder. *Behaviour Research and Therapy*, *45*(11), 2628–2639.

Pruim, R. H. R., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J. K., & Beckmann, C. F. (2015). ICA-AROMA: a robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage*, *112*, 267–277.

Qin, Y.-L., McNaughton, B. L., Skaggs, W. E., & Barnes, C. A. (1997). Memory reprocessing in corticocortical and hippocampocortical neuronal ensembles. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *352*(1360), 1525–1533.

Ramirez, S., Liu, X., Lin, P.-A., Suh, J., Pignatelli, M., Redondo, R. L., Ryan, T. J., & Tonegawa, S. (2013). Creating a false memory in the hippocampus. *Science*, *341*(6144), 387–391.

Ramirez, S., Liu, X., MacDonald, C. J., Moffa, A., Zhou, J., Redondo, R. L., & Tonegawa, S. (2015). Activating positive memory engrams suppresses depression-like behaviour. *Nature*, *522*(7556), 335–339.

Ranganath, C., & Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Nature Reviews Neuroscience*, *13*(10), 713–726.

Reverberi, C., Görgen, K., & Haynes, J.-D. (2012). Compositionality of rule representations in human prefrontal cortex. *Cerebral Cortex*, *22*(6), 1237–1246.

Richiardi, J., Altmann, A., Milazzo, A.-C., Chang, C., Chakravarty, M. M., Banaschewski, T., Barker, G. J., Bokde, A. L. W., Bromberg, U., Buchel, C., Conrod, P., Fauth-Buhler, M., Flor, H., Frouin, V., Gallinat, J., Garavan, H., Gowland, P., Heinz, A., Lemaitre, H., … Tahmasebi, A. (2015). Correlated gene expression supports synchronous activity in brain networks. *Science*, *348*(6240), 1241–1244.

Richter, F. R., & Yeung, N. (2014). Neuroimaging studies of task switching. In *Task switching and cognitive control*. Oxford University Press Oxford, England.

Roediger III, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27.

Roelofs, J., van Breukelen, G., de Graaf, L. E., Beck, A. T., Arntz, A., & Huibers, M. J. H. (2013). Norms for the Beck Depression Inventory (BDI-II) in a large Dutch community sample. *Journal of Psychopathology and Behavioral Assessment*, *35*(1), 93–98.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*(2), 207–231.

Rohrbach, M., Qiu, W., Titov, I., Thater, S., Pinkal, M., & Schiele, B. (2013). Translating video content to natural language descriptions. *Proceedings of the IEEE International Conference on Computer Vision*, 433–440.

Romero-Garcia, R., Warrier, V., Bullmore, E. T., Baron-Cohen, S., & Bethlehem, R. A. I. (2019). Synaptic and transcriptionally downregulated genes are associated with cortical thickness differences in autism. *Molecular Psychiatry*, *24*(7), 1053–1064.

Romme, I. A. C., de Reus, M. A., Ophoff, R. A., Kahn, R. S., & van den Heuvel, M. P. (2017). Connectome Disconnectivity and Cortical Gene Expression in Patients With Schizophrenia. *Biological Psychiatry*, *81*(6), 495–502.

Rothbaum, B. O., & Schwartz, A. C. (2002). Exposure therapy for posttraumatic stress disorder. *American Journal of Psychotherapy*, *56*(1), 59–75.

Roy, D. S., Park, Y.-G., Ogawa, S. K., Cho, J. H., Choi, H., Kamensky, L., Martin, J., Chung, K., & Tonegawa, S. (2019). Brain-wide mapping of contextual fear memory engram ensembles supports the dispersed engram complex hypothesis. *BioRxiv*, 668483.

Ruge, H., Jamadar, S., Zimmermann, U., & Karayanidis, F. (2013). The many faces of preparatory control in task switching: reviewing a decade of fMRI research. *Human Brain Mapping*, *34*(1), 12–35.

Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Current Opinion in Neurobiology*, *23*(2), 255–260.

Sacchet, M. D., Levy, B. J., Hamilton, J. P., Maksimovskiy, A., Hertel, P. T., Joormann, J., Anderson, M. C., Wagner, A. D., & Gotlib, I. H. (2017). Cognitive and neural consequences of memory suppression in major depressive disorder. *Cognitive, Affective, & Behavioral Neuroscience*, *17*(1), 77–93.

Sadaghiani, S., Poline, J.-B., Kleinschmidt, A., & D'Esposito, M. (2015). Ongoing dynamics in large-scale functional connectivity predict perception. *Proceedings of the National Academy of Sciences*, *112*(27), 8463–8468.

Sakai, K., & Passingham, R. E. (2003). Prefrontal interactions reflect future task operations. *Nature Neuroscience*, *6*(1), 75–81.

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., Eisenberg, M. L., & Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, *129*(2), 241–255.
Schacter, D. L. (2012). *Forgotten ideas, neglected pioneers: Richard Semon and the story of memory*. Psychology Press.

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex*, *28*(9), 3095–3114.

Schiller, D., Monfils, M.-H., Raio, C. M., Johnson, D. C., LeDoux, J. E., & Phelps, E. A. (2010). Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature*, *463*(7277), 49–53.

Schlichting, M. L., & Preston, A. R. (2015). Memory integration: neural mechanisms and implications for behavior. *Current Opinion in Behavioral Sciences*, *1*, 1–8.

Schlichting, M. L., Zeithamova, D., & Preston, A. R. (2014). CA1 subfield contributions to memory integration and inference. *Hippocampus*, *24*(10), 1248–1260.

Schmeichel, B. J., Volokhov, R. N., & Demaree, H. A. (2008). Working memory capacity and the self-regulation of emotional expression and experience. *Journal of Personality and Social Psychology*, *95*(6), 1526.

Schork, A. J., Won, H., Appadurai, V., Nudel, R., Gandal, M., Delaneau, O., Revsbech Christiansen, M., Hougaard, D. M., Bækved-Hansen, M., Bybjerg-Grauholm, J., Giørtz Pedersen, M., Agerbo, E., Bøcker Pedersen, C., Neale, B. M., Daly, M. J., Wray, N. R., Nordentoft, M., Mors, O., Børglum, A. D., … Werge, T. (2019). A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment. *Nature Neuroscience*, *22*(3), 353–361.

Schuck, N. W., & Niv, Y. (2019). Sequential replay of nonspatial task states in the human hippocampus. *Science*, *364*(6447), eaaw5181.

Schultz, D. H., & Cole, M. W. (2016). Higher Intelligence Is Associated with Less Task-Related Brain Network Reconfiguration. *The Journal of Neuroscience*, *36*(33), 8551–8561.

Schwabe, L., Nader, K., & Pruessner, J. C. (2014). Reconsolidation of human memory: Brain mechanisms and clinical relevance. *Biological Psychiatry*, *76*(4), 274–280.

Seidlitz, J., Váša, F., Shinn, M., Romero-Garcia, R., Whitaker, K. J., Vértes, P. E., Wagstyl, K., Kirkpatrick Reardon, P., Clasen, L., Liu, S., Messinger, A., Leopold, D. A., Fonagy, P., Dolan, R. J., Jones, P. B., Goodyer, I. M., Raznahan, A., & Bullmore, E. T. (2018). Morphometric Similarity Networks Detect Microscale Cortical Organization and Predict Inter-Individual Cognitive Variation. *Neuron*, *97*(1), 231-247.e7.

Semon, R. W. (1923). *Mnemic psychology*. G. Allen & Unwin, Limited.

Sha, Z., Wager, T. D., Mechelli, A., & He, Y. (2019). Common Dysfunction of Large-Scale Neurocognitive Networks Across Psychiatric Disorders. *Biological Psychiatry*, *85*(5), 379–388.

Shanahan, L. K., Gjorgieva, E., Paller, K. A., Kahnt, T., & Gottfried, J. A. (2018). Odor-evoked category reactivation in human ventromedial prefrontal cortex during sleep promotes memory consolidation. *Elife*, *7*, e39681.

Shen, E. H., Overly, C. C., & Jones, A. R. (2012). The Allen Human Brain Atlas: comprehensive gene expression mapping of the human brain. *Trends in Neurosciences*, *35*(12), 711–714.

Shine, J. M., Bissett, P. G., Bell, P. T., Koyejo, O., Balsters, J. H., Gorgolewski, K. J., Moodie, C. A., & Poldrack, R. A. (2016). The dynamics of functional brain networks: integrated network states during cognitive task performance. *Neuron*, *92*(2), 544–554.

Shine, J. M., Breakspear, M., Bell, P. T., Ehgoetz Martens, K. A., Shine, R., Koyejo, O., Sporns, O., & Poldrack, R. A. (2019). Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. *Nature Neuroscience*, *22*(2), 289–296.

Shine, J. M., & Poldrack, R. A. (2018). Principles of dynamic network reconfiguration across diverse brain states. *NeuroImage*, *180*, 396–405.

Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V., & Greicius, M. D. (2012). Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cerebral Cortex*, *22*(1), 158–165.

Shohamy, D., & Wagner, A. D. (2008). Integrating Memories in the Human Brain: Hippocampal-Midbrain Encoding of Overlapping Events. *Neuron*, *60*(2), 378–389.

Siegel, J. S., Mitra, A., Laumann, T. O., Seitzman, B. A., Raichle, M., Corbetta, M., & Snyder, A. Z. (2017). Data quality influences observed links between functional connectivity and behavior. *Cerebral Cortex*, *27*(9), 4492–4502.

Silva, M., Baldassano, C., & Fuentemilla, L. (2019). Rapid Memory Reactivation at Movie Event Boundaries Promotes Episodic Encoding. *The Journal of Neuroscience*, *39*(43), 8538–8548.

Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., & Hasson, U. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension. *Nature Communications*, *7*, 12141.

Smith, A. M., Floerke, V. A., & Thomas, A. K. (2016). Retrieval practice protects memory against acute stress. *Science*, *354*(6315), 1046–1048.

Smith, A. T., Kosillo, P., & Williams, A. L. (2011). The confounding effect of response amplitude on MVPA performance measures. *Neuroimage*, *56*(2), 525–530.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155.

Sols, I., DuBrow, S., Davachi, L., & Fuentemilla, L. (2017). Event Boundaries Trigger Rapid Memory Reinstatement of the Prior Events to Promote Their Representation in Long-Term Memory. *Current Biology*, *27*(22), 3499-3504.e4.

Sonkusare, S., Breakspear, M., & Guo, C. (2019). Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends in Cognitive Sciences*, 1–16.

Spalding, K. N., Schlichting, M. L., Zeithamova, D., Preston, A. R., Tranel, D., Duff, M. C., & Warren, D. E. (2018). Ventromedial prefrontal cortex is necessary for normal associative inference and memory integration. *Journal of Neuroscience*, *38*(15), 3767–3775.

Spector, A., & Biederman, I. (1976). Mental set and mental shift revisited. *The American Journal of Psychology*, 669–679.

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745–756.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–1006.

Sun, C., Yang, W., Martin, J., & Tonegawa, S. (2020). Hippocampal neurons represent events as transferable units of experience. *Nature Neuroscience*, *23*(5), 651–663.

Takahashi, S. (2015). Episodic-like memory trace in awake replay of hippocampal place cell activity sequences. *Elife*, *4*, e08105.

Takashima, A., Nieuwenhuis, I. L. C., Jensen, O., Talamini, L. M., Rijpkema, M., & Fernández, G. (2009). Shift from hippocampal to neocortical centered retrieval network with consolidation. *Journal of Neuroscience*, *29*(32), 10087–10093.

Takashima, A., Petersson, K. M., Rutters, F., Tendolkar, I., Jensen, O., Zwarts, M. J., McNaughton, B. L., & Fernandez, G. (2006). Declarative memory consolidation in humans: a prospective functional magnetic resonance imaging study. *Proceedings of the National Academy of Sciences*, *103*(3), 756–761.

Tambini, A., & Davachi, L. (2013). Persistence of hippocampal multivoxel patterns into postencoding rest is related to memory. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(48), 19591–19596.

Tambini, A., & Davachi, L. (2019). Awake Reactivation of Prior Experiences Consolidates Memories and Biases Cognition. *Trends in Cognitive Sciences*, 1–15.

Teng, X., Ma, M., Yang, J., Blohm, S., Cai, Q., & Tian, X. (2020). Constrained Structure of Ancient Chinese Poetry Facilitates Speech Content Grouping. *Current Biology*, *30*(7), 1299-1305.e7.

Thavabalasingam, S., O'Neil, E. B., Tay, J., Nestor, A., & Lee, A. C. H. (2019). Evidence for the incorporation of temporal duration information in human hippocampal long-term memory sequence representations. *Proceedings of the National Academy of Sciences*, *116*(13), 6407–6414.

Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics*, *27*(1), 77–83.

Tonegawa, S., Liu, X., Ramirez, S., & Redondo, R. (2015). Memory engram cells have come of age. *Neuron*, *87*(5), 918–931.

Töreyin, B. U., Dedeoğlu, Y., Güdükbay, U., & Çetin, A. E. (2006). Computer vision based method for real-time fire and flame detection. *Pattern Recognition Letters*, *27*(1), 49–58.

Tsao, A., Sugar, J., Lu, L., Wang, C., Knierim, J. J., Moser, M.-B., & Moser, E. I. (2018). Integrating time from experience in the lateral entorhinal cortex. *Nature*, *561*(7721), 57–62.

Tull, M. T., Barrett, H. M., McMillan, E. S., & Roemer, L. (2007). A preliminary investigation of the relationship between emotion regulation difficulties and posttraumatic stress symptoms. *Behavior Therapy*, *38*(3), 303–313.

Tulving, E. (1984). Precis of elements of episodic memory. *Behavioral and Brain Sciences*, *7*(2), 223–238.

van den Broek, G. S. E., Takashima, A., Segers, E., Fernández, G., & Verhoeven, L. (2013). Neural correlates of testing effects in vocabulary learning. *NeuroImage*, *78*, 94–102.

van den Broek, G., Takashima, A., Wiklund-Hörnqvist, C., Karlsson Wirebring, L., Segers, E., Verhoeven, L., & Nyberg, L. (2016). Neurocognitive mechanisms of the "testing effect": A review. *Trends in Neuroscience and Education*, *5*(2), 52–66.

van den Broek, G., Takashima, A., Wiklund-Hörnqvist, C., Karlsson Wirebring, L., Segers, E., Verhoeven, L., & Nyberg, L. (2016). Neurocognitive mechanisms of the "testing effect": A review. *Trends in Neuroscience and Education*, *5*(2), 52–66.

Van Den Heuvel, M. P., & Pol, H. E. H. (2010). Exploring the brain network: a review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, *20*(8), 519–534.

van der Bij, A. K., de Weerd, S., Cikot, R. J. L. M., Steegers, E. A. P., & Braspenning, J. C. C. (2003). Validation of the dutch short form of the state scale of the Spielberger State-Trait Anxiety Inventory: considerations for usage in screening outcomes. *Public Health Genomics*, *6*(2), 84–87.

van Kesteren, M. T. R., Beul, S. F., Takashima, A., Henson, R. N., Ruiter, D. J., & Fernández, G. (2013). Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: from congruent to incongruent. *Neuropsychologia*, *51*(12), 2352–2359.

Van Kesteren, M. T. R., Fernández, 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*, *107*(16), 7550–7555.

van Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J., & Fernández, G. (2010). Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. *Journal of Neuroscience*, *30*(47), 15888–15894.

van Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J., Morris, R. G. M., & Fernández, G. (2014). Building on prior knowledge: schema-dependent encoding processes relate to academic performance. *Journal of Cognitive Neuroscience*, *26*(10), 2250–2261.

van Schie, K., & Anderson, M. C. (2017). Successfully controlling intrusive memories is harder when control must be sustained. *Memory*, *25*(9), 1201–1216.

Vaz, A. P., Wittig, J. H., Inati, S. K., & Zaghloul, K. A. (2020). Replay of cortical spiking sequences during human memory retrieval. *Science*, *367*(6482), 1131–1134.

Vetere, G., Tran, L. M., Moberg, S., Steadman, P. E., Restivo, L., Morrison, F. G., Ressler, K. J., Josselyn, S. A., & Frankland, P. W. (2019). Memory formation in the absence of experience. *Nature Neuroscience*, *22*(6), 933–940.

Wager, T. D., Davidson, M. L., Hughes, B. L., Lindquist, M. A., & Ochsner, K. N. (2008). Prefrontal-subcortical pathways mediating successful emotion regulation. *Neuron*, *59*(6), 1037–1050.

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., Rosen, B. R., & Buckner, R. L. (1998). Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, *281*(5380), 1188–1191.

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences*, *9*(9), 445–453.

Wang, G.-Z., Belgard, T. G., Mao, D., Chen, L., Berto, S., Preuss, T. M., Lu, H., Geschwind, D. H., & Konopka, G. (2015). Correspondence between Resting-State Activity and Brain Gene Expression. *Neuron*, *88*(4), 659–666.

Waskom, M. L., Kumaran, D., Gordon, A. M., Rissman, J., & Wagner, A. D. (2014). Frontoparietal representations of task context support the flexible control of goal-directed cognition. *Journal of Neuroscience*, *34*(32), 10743–10755.

Westphal, A. J., Wang, S., & Rissman, J. (2017). Episodic memory retrieval benefits from a less modular brain network organization. *Journal of Neuroscience*, *37*(13), 3523–3531.

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences*, *97*(20), 11125–11129.

Williams, A. N., Postans, M., & Hodgetts, C. J. (2019). How the Human Brain Segments Continuous Experience. *The Journal of Neuroscience*, *39*(17), 3172–3174.

Wimber, M., Schott, B. H., Wendler, F., Seidenbecher, C. I., Behnisch, G., Macharadze, T., Bäuml, K. H. T., &

Richardson-Klavehn, A. (2011). Prefrontal dopamine and the dynamic control of human long-term memory. *Translational Psychiatry*, *1*(June), 1–7.

Wimber, Maria, Alink, A., Charest, I., Kriegeskorte, N., & Anderson, M. C. (2015). Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nature Neuroscience*, *18*(4), 582–589.

Wimber, Maria, Bäuml, K.-H., Bergström, Z., Markopoulos, G., Heinze, H.-J., & Richardson-Klavehn, A. (2008). Neural markers of inhibition in human memory retrieval. *Journal of Neuroscience*, *28*(50), 13419–13427.

Wimber, Maria, Schott, B. H., Wendler, F., Seidenbecher, C. I., Behnisch, G., Macharadze, T., Bäuml, K. H. T., & Richardson-Klavehn, A. (2011). Prefrontal dopamine and the dynamic control of human long-term memory. *Translational Psychiatry*, *1*(7), e15--e15.

Wing, E. A., Marsh, E. J., & Cabeza, R. (2013). Neural correlates of retrieval-based memory enhancement: an fMRI study of the testing effect. *Neuropsychologia*, *51*(12), 2360–2370.

Wirebring, L. K., Wiklund-Hörnqvist, C., Eriksson, J., Andersson, M., Jonsson, B., & Nyberg, L. (2015). Lesser neural pattern similarity across repeated tests is associated with better long-term memory retention. *Journal of Neuroscience*, *35*(26), 9595–9602.

Wisniewski, D., Goschke, T., & Haynes, J.-D. (2016). Similar coding of freely chosen and externally cued intentions in a fronto-parietal network. *Neuroimage*, *134*, 450–458.

Wisniewski, D., Reverberi, C., Tusche, A., & Haynes, J.-D. (2015). The neural representation of voluntary task-set selection in dynamic environments. *Cerebral Cortex*, *25*(12), 4715–4726.

Woolgar, A., Afshar, S., Williams, M. A., & Rich, A. N. (2015). Flexible coding of task rules in frontoparietal cortex: an adaptive system for flexible cognitive control. *Journal of Cognitive Neuroscience*, *27*(10), 1895–1911.

Woolgar, A., Thompson, R., Bor, D., & Duncan, J. (2011). Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. *Neuroimage*, *56*(2), 744–752.

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *Neuroimage*, *14*(6), 1370–1386.

Xiao, X., Dong, Q., Gao, J., Men, W., Poldrack, R. A., & Xue, G. (2017). Transformed neural pattern reinstatement during episodic memory retrieval. *Journal of Neuroscience*, *37*(11), 2986–2998.
Xue, G. (2018). The neural representations underlying human episodic memory. *Trends in Cognitive Sciences*, *22*(6), 544–561.

Yang, W., Chen, Q., Liu, P., Cheng, H., Cui, Q., Wei, D., Zhang, Q., & Qiu, J. (2016). Abnormal brain activation during directed forgetting of negative memory in depressed patients. *Journal of Affective Disorders*, *190*, 880–888.

Yang, W., Liu, P., Zhuang, K., Wei, D., Anderson, M. C., & Qiu, J. (2020). Behavioral and neural correlates of memory suppression in subthreshold depression. *Psychiatry Research: Neuroimaging*, *297*, 111030.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*(8), 665.

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515–525.

Ye, Z., Shi, L., Li, A., Chen, C., & Xue, G. (2020). Retrieval practice facilitates memory updating by enhancing and differentiating medial prefrontal cortex representations. *ELife*, *9*, 1–51.

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., & others. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125.

Yonelinas, A. P., Otten, L. J., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *Journal of Neuroscience*, *25*(11), 3002–3008.

Zacks, J. M. (2020). Event Perception and Memory. *Annual Review of Psychology*, *71*(1), 165–191.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychological Bulletin*, *133*(2), 273.

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*(1), 29–58.

Zeithamova, D., Dominick, A. L., & Preston, A. R. (2012). Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron*, *75*(1), 168–179.

Zhang, J., Kriegeskorte, N., Carlin, J. D., & Rowe, J. B. (2013). Choosing the rules: distinct and overlapping frontoparietal representations of task rules for perceptual decisions. *Journal of Neuroscience*, *33*(29), 11852–11862.

Research data management

Ethics

This thesis is based on results from (healthy) human experiments, which were conducted in accordance with the principles of the Declaration of Helsinki and followed the applicable laws and ethical guidelines. Research Data Management was conducted according to the FAIR principles. The paragraphs below specify in detail how this was achieved. Local data acquisition (*Chapter3 and Chapter4*) was approved by the medical and ethical review board Committee on Research involving Human Subjects Region Arnhem Nijmegen, Nijmegen, the Netherland. For the open-access dataset used (*Chapter2)*, the experimental procedures were approved by the Princeton University Institutional Review Board.

Findable Accessible

All collected electronic data was archived at the central project archive of the Donders Centre for Cognitive Neuroimaging, and further shared publicly using the Donders Repository (https://data.donders.ru.nl/). The project was named as *Tracking the involuntary retrieval of unwanted memory in the human brain with functional MRI* in the Repository (https://doi.org/10.34973/5afg-7r41).

Chapter	DAC	RDC	DSC	DSC License	Note
2	NA	NA	http://arks.princeton.edu/ark:/88435/dsp01nz8062179	CC-BY-NC	Open Dataset
3	DAC_3013067.01_584	RDC_3013067.01_003	https://doi.org/10.34973/5afg-7r41	RU-DI-HD-1.0	Local Data Acquistion
4	DAC_3013067.01_584	RDC_3013067.01_003	https://doi.org/10.34973/5afg-7r41	RU-DI-HD-1.0	Local Data Acquistion
5	NA	NA	https://doi.org/10.17605/OSF.IO/6WZ2J	CC-BY-NC	Meta-analysis

DAC=Data acquistion Collection; RDC=Research Document Collection; DSC=Data Sharing Collection

Within each experimental chapter, there is a dedicated section for the *Data and Code Availability*. There, we provided detailed information and links for data and code sharing. Furthermore, an Open Science Framework (OSF) folder was created for each chapter to share intermediate results and in-house scripts of critical data analyses. Technical supports to use these scripts are available from the corresponding author on reasonable request. No paper materials were collected from participants during experiments. Data beyond the neuroimaging and behavioral log files (e.g., questionnaires) were collected using the online electronic data management website called Castor EDC. Data were exported from Castor EDC to comma-separated value (.csv) files and archived together with other data at the central project archive.

For chapters 2, 3, 4 and 5 research data have also been stored on the project/networkdrive (respectively…). These data were accessible to all members involved in the project. After finalization of the project data were removed from the project/networkdrives. The maniscript of chapter 1 and 2 is still under revision and review respectively. Informed consent was obtained on paper following the Centre procedure. The forms are archived in the central archive of the Centre for 10 years after termination of the studies.

Privacy

The following measures were used to protect the privacy of the participants: (1) unique individual subject code of the participant recruitment system was used as participant identity. (2) This code corresponded with subject ID within specific research projects. (3) The correspondences between subject IDs and individual subject codes were kept separate from any research data. (4) Sensitive personal data collected during the experiments (e.g., e-mail address, the exact date of birth, address of family doctors (i.e., GP)), which are not relevant for research questions, were removed from the Castor EDC half a year after the end of data acquisition.

Summary of thesis

[English]

I presented four experimental chapters on the topic of *process* and *strength* dynamics of memories. In brief, we used fMRI in healthy human subjects to elicit brain activity during different memory tasks, and combined approaches from genetics and machine learning. Here I will give a summary of our main findings.

1. How do we transform continuous experience into discrete memories?

How do we encode continuous information is critical for subsequent retrieval. Theories of event memories proposed that different neural states are used to represent discrete events. We demonstrated that successful encoding of continuous information was dependent on events being represented with dissimilar activity patterns in a network centered on the hippocampus and medial prefrontal cortex. At the same time, we found the potential neural correlates for event integration: similar connectivity patterns of these regions linked events and preserved the narrative order in which they were encoded (**Chapter 2**).

2. How does the brain flexibly switch between memory retrieval and memory control?

During a continuous task, sometimes we need to switch between two or more task demands. This process is particularly challenging for the brain when the switch is between memory retrieval and memory control, which requires the need to coordinate two opposite neural states with partly overlapping neural networks. As the classical task-switch studies, we found the effect that preceding mental processing has an impairing effect on the current processing. We also found that the switch between retrieval and control involves large-scale adaptations between memory retrieval and inhibitory control networks. This adaptation is less flexible immediately after the task switching and associated with behavioral switch costs. Thus, we reasoned that the timely reconfiguration between memory and control networks is the key to flexible memory processing (**Chapter 3**).

3. How does the memory modulation re-organize memory traces and change their memory strength after overnight consolidation?

Memory modulation immediately after their formation can modify neural representations of memory traces and change their strength. We found that similar neural effects can be observed after initial consolidation (i.e., 24 hours after encoding). Specifically, repeated retrieval reduced overall activity amplitude, but seems to promote episode-unique mnemonic representations in visual processing and parietal regions. In contrast, repeated memory control was associated with reduced lateral prefrontal activity, but relative intact mnemonic representations (**Chapter 4**).

4. Why are changes in memory strength also accompanied by alterations of emotional intensity?

Memory modulations such as memory control can alter not only memory accessibility and neural representations, but also the valence of these memory traces. We found that memory control and emotion regulation are supported by the same frontal-parietal-insular network, which is involved in inhibitory control. Moreover, neuroimaging-gene expression analysis identified the association between task-induced brain activations and a set of "inhibition-related" genes. These genes were reported to be associated with neuronal transmission and risk for major psychiatric disorders, as well as epilepsy and alcohol dependence (**Chapter 5**).

Conclusions

Our results provide a perspective on the dynamic nature of memory in terms of temporal and strength. For the temporal dynamics, our experiments revealed that during the continuous experience, the brain uses separate neural states to segment information into events and simultaneously binds them into a coherent narrative by context encoding. During fast switches between memory retrieval and memory control, the brain needs to reconfigure its neural states in time. Otherwise, the remaining state of retrieval may cause failures in memory control. For the strength dynamics, after initial consolidation, active retrieval seems to promote episode-unique mnemonic representations, leading to enhanced memory strength. By contrast, memory control disengages prefrontal involvement during retrieval, causing compromised memory strength. These changes in memory strength are associated with changes in the emotional intensity of individual memory traces. This dual modulation phenomenon is supported by the common inhibitory control network and corresponding transcriptional correlates. Future research on memory dynamics might lead to non-pharmacological, cognitive approaches that can enhance the encoding efficiency and persistence of everyday memories or can modify traumatic memories and their emotional impacts. Such methods could potentially provide both fundamental and applied knowledge for memory-related symptoms in memory disorders and affective disorders.

[Nederlands]

Er worden vier experimentele hoofdstukken over de onderwerpen *proces*- en *sterktedynamiek* van herinneringen gepresenteerd. Hierin werd gebruik gemaakt van fMRI in gezonde proefpersonen om hersenactiviteit tijdens verschillende geheugentaken aan het licht te brengen en combineerde we benaderingen uit de genetica en het 'machine learning'. Onderstaand geef ik een samenvatting van onze belangrijkste bevindingen geven.

1. Hoe transformeren we continue ervaringen in discrete herinneringen?

Hoe we continue informatie coderen is cruciaal voor het later terughalen van de informatie. Theorieën over geheugen van bepaalde gebeurtenissen stelden voor dat verschillende neurale toestanden worden gebruikt om discrete gebeurtenissen weer te geven. Wij toonde aan dat een succesvolle codering van continue informatie afhangt van verschillende activiteitspatronen in een netwerk dat gecentreerd is op de hippocampus en de mediale prefrontale cortex, die elk een gebeurtenis representeren. Tegelijkertijd vonden we de potentiële neurale correlaten voor de integratie van gebeurtenissen: vergelijkbare connectiviteitspatronen van regio's verbinden gebeurtenissen en behouden de narratieve volgorde waarin ze werden gecodeerd **(hoofdstuk 2).**

2. Hoe schakelt het brein flexibel tussen het ophalen van het geheugen en de controle van het geheugen?

Tijdens een continue taak moeten we soms switchen tussen twee of meer taakvereisten. Dit proces is voor het brein vooral een uitdaging wanneer er een switch gemaakt moet worden tussen het ophalen van het geheugen en het controleren van geheugen. Dit is het geval omdat dit van het brein vereist om twee tegenovergestelde neurale staten met gedeeltelijk overlappende neurale netwerken te coördineren. Net zoals in de klassieke 'task-switch' studies vonden we dat voorafgaande mentale verwerking een nadelig effect had op huidige verwerking. Daarnaast vonden we dat de switch tussen het ophalen en het controleren van geheugens samengaat met grootschalige aanpassingen tussen netwerken verantwoordelijk voor het ophalen van geheugen ophalen en remmende controle. Deze aanpassing is minder flexibel direct na de taakwisseling en is geassocieerd met gedragsveranderingskosten. Daarom beredeneerde we dat de tijdige her configuratie tussen geheugen- en controlenetwerken belangrijk is voor een flexibele geheugenverwerking (**Hoofdstuk 3**).

3. Hoe reorganiseert geheugenmodulatie geheugensporen en verandert het geheugensterkte na nachtelijke consolidatie?

Geheugenmodulatie direct na de vorming van geheugens kan de neurale representaties van geheugensporen wijzigen en de sterkte ervan veranderen. Wij vonden dat na de initiële consolidatie (d.w.z. 24 uur na het coderen) soortgelijke neurale effecten kunnen worden waargenomen. Meer specifiek, herhaalde retrieval verminderde de algehele activiteit amplitude maar lijkt episode-unieke mnemonische representaties in de visuele verwerking en partiële regio's te bevorderen. Daarentegen werd herhaalde geheugen controle geassocieerd met verminderde laterale prefrontale activiteit, maar relatief intacte mnemonische representaties **(Hoofdstuk 4).**

4. Waarom gaan veranderingen in geheugensterkte ook gepaard met veranderingen in emotionele intensiteit?

Geheugenmodulaties zoals geheugencontrole kunnen niet alleen de toegankelijkheid en de neurale representaties van geheugen veranderen, maar ook de waarde van deze geheugenprocessen. Wij vonden dat geheugen controle en emotieregulatie werden ondersteund door hetzelfde frontaal-pariëtaal-insulaire netwerk, welke ook betrokken is bij remmende controle. Daarnaast identificeerde we middels een neuro imaging-genexpressie analyse de associatie tussen taak-geïnduceerde hersenactivatie en een set "inhibitie-gerelateerde" genen. Geïdentificeerde genen werden geassocieerd met neuronale transmissie en risico's voor grote psychiatrische stoornissen, evenals epilepsie en alcoholvergiftiging (**Hoofdstuk 5**).

Conclusie

Onze resultaten bieden een perspectief op het dynamische karakter van geheugen voor zowel de temporele als de sterkte kant. Wat betreft de temporele kant lieten onze experts zien dat de hersenen tijdens een continue ervaring gebruik maken van afzonderlijke neurale toestanden om informatie te segmenteren in gebeurtenissen en deze tegelijkertijd te binden tot een samenhangend verhaal door middel van contextcodering. Om snel te schakelen tussen het ophalen van het geheugen en de controle van geheugen moeten de hersenen hun neurale toestanden her configureren in de tijd. Indien dit niet gebeurd kan de resterende status van het ophalen van het geheugen storingen in geheugencontrole veroorzaken. Wat betreft de sterkte van geheugen lijkt het actief ophalen van geheugen na initiële consolidatie episode-unieke mnemonische representaties te promoten, wat leidt tot een toename in geheugen sterkte. Daarentegen schakelt de geheugencontrole de prefrontale betrokkenheid bij het ophalen van geheugens uit, waardoor de geheugensterkte aangetast wordt. Deze veranderingen in geheugensterkte zijn geassocieerd met veranderingen in de emotionele intensiteit van de individuele geheugensporen. Dit dubbele modulatieverschijnsel wordt ondersteund door het gemeenschappelijke remmende controle netwerk en corresponderende transcriptionele correlaten. Toekomstig onderzoek naar geheugendynamiek zou kunnen leiden tot niet-farmacologische, cognitieve benaderingen die de coderingsefficiëntie en persistentie van alledaagse herinneringen kunnen verbeteren of traumatische herinneringen en hun emotionele impact kunnen wijzigen. Dergelijke methoden zouden zowel fundamentele als toegepaste kennis kunnen opleveren voor geheugen gerelateerde symptomen bij geheugenstoornissen en affectieve stoornissen.

List of publications

Journal Publications:

[1] **WEI LIU.,** Kohn, N., & Fernández, G. (2020). Probing the neural dynamics of mnemonic representations after the initial consolidation. ***NeuroImage***, 221, 117213.

[2] **WEI LIU.**, Peeters, N., Fernandez, G., & Kohn, N. (*2020*). Common neural and transcriptional correlates of inhibitory control across emotion, memory, and response inhibition. ***Soc. Cogn. Affect. Neurosci***.

[3] **WEI LIU.,** Kohn, N., & Fernández, G. (2019). Intersubject similarity of personality is associated with intersubject similarity of brain connectivity patterns. ***NeuroImage***, 186, 56-69.

[4] **WEI LIU**., Wei, D., Chen, Q., Yang, W., Meng, J., Wu, G., ... & Qiu, J. (2017). Longitudinal test-retest neuroimaging data from healthy young adults in southwest China. ***Scientific data***, 4, 170017.

[5] Wei Cheng *, Edmund T Rolls *, Jiang Qiu *, **WEI**, **LIU** *, Yanqing Tang *, Chu-Chung Huang *, XinFa Wang *., ... & Jianfeng Feng. (2016). Medial reward and lateral non-reward orbitofrontal cortex circuits change in opposite directions in depression. ***Brain***, 139(12), 3296-3309. (*These authors contributed equally to this work.)

[6] **WEI**, **LIU.**, Mao, Y., Wei, D., Yang, J., Du, X., Xie, P., & Qiu, J. (2016). Structural Asymmetry of Dorsolateral Prefrontal Cortex Correlates with Depressive Symptoms: Evidence from Healthy Individuals and Patients with Major Depressive Disorder. ***Neuroscience Bulletin***, 1-10.

[7] **WEI**, **LIU.**, Liu, H., Wei, D., Sun, J., Yang, J., Meng, J., ... & Qiu, J. (2015). Abnormal degree centrality of functional hubs associated with negative coping in older Chinese adults who lost their only child. ***Biological psychology***, *112*, 46-55.

Journal articles in revision:

[1] **WEI LIU**, Yingjie Shi, James N.Cousins, Nils Kohn, Guillén Fernandez. (*in revision*). Hippocampal event segmentation and integration contribute to episodic memory formation.

Preprint under review:

[1] **WEI LIU.**, Kohn, N., & Fernández, G. (*under review*). The dynamic transition between neural states is associated with the flexible use of memory

Academic portfolio DGCN

Academic courses

- Donders Toolkit: Advanced analysis and source modeling of EEG and MEG data, 2020
- Cognitive Neuroscience of Memory (6 ECTs), 2020
- FENS-Hertie Winter School: Genetic and epigenetic mechanisms underlying brain disorders (3 ECTs), 2019
- Communication in Cognitive Neuroscience (6 ECTs), 2018
- Academic Writing (3 ECTs), 2017
- Quantitative Brain Networks (6 ECTs), 2017

Ad-Hoc Reviewer

- Neuroimage
- Neuropsychologia
- Psychiatry Research: Neuroimaging
- Brain Connectivity

Conference presentations

- **WEI LIU**, Nancy, P., Kohn, N., & Fernández, G (2019). *Common neural and transcriptional correlates of inhibitory control modulating emotion regulation and memory control. Poster Presentation*. Organization for Human Brain Mapping, June 2019, Italy
- **WEI LIU**, Kohn, N., & Fernández, G (2018). *Intersubject similarity of personality is associated with intersubject similarity of brain connectivity patterns*. Poster Presentation. Organization for Human Brain Mapping, June 2018, Singapore.

Invited presentations

- **WEI LIU**. *Open neuroimaging: from a data provider to a data user*. Meeting for functional neuroimaging analysis. Donders Center for cognitive neuroimaging, Nijmegen, Netherland, March 2020
- **WEI LIU**. *Open neuroimaging: from a data provider to a data user*. Symposium: Medical research in the era of Big data. Donders Center for Medical Neuroscience, Nijmegen, Netherland, February 2020
- **WEI LIU**. *Event segmentation and integration in the human hippocampus*. Chinese Association for Psychological & Brain science Meeting, Utrecht, Netherland, November 2019.
- **WEI LIU**. *Integrating neuroimaging and gene expression data*. Symposium: to understand brain organization in health and disease using imaging genetics. Donders Discussion 2019, Nijmegen, Netherland, November 2019.
- **WEI LIU**. *Neuro-computational principles of event memories*. School of Psychology,

Southwest University, Chongqing, China, May 2019.

- **WEI LIU**. *Common neural and transcriptional correlates of inhibitory control modulating emotion regulation and memory control*. Symposium: Integrating neuroimaging and gene expression data. Donders Discussion 2018, Nijmegen, October 2018.

Extra-curricular training & certification

- Certified user 3T MRI-scanner (Donders Center for cognitive neuroimaging) (2017)
- Super-certifier user 3T MRI-scanner (Donders Center for cognitive neuroimaging) (2020)

Teaching & mentoring experience

- **Supervisor of lab rotation students:**
  Yingjie Shi

- **Co-supervisor of master internship**
  Tom Roovers, Nancy Peeters, Joost Verchuren

- **Co-supervisor of research assistant**
  Joyce van Arendonk, Jette de Vos

- **Lecturer for Donders Toolkits:**
  Introduction to Neuroimaging (2017) (2018) (2019)
  Brain Stimulation (2017) (2018) (2019)

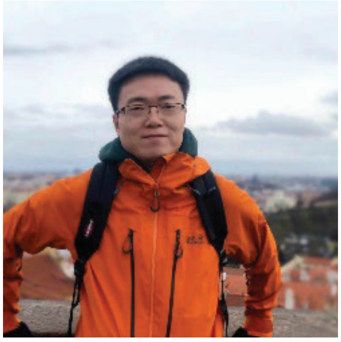
Grants & awards

- Centre for Population Neuroscience and Precision Medicine (PONS)-Early Career Award (2020)
- Donders-Monash exchange travel grant (2500 Euros) (2018)
- 4-year PhD fellowship from Chinse Scholarship Council (2016)

Other academic activities

- Organizer of Donders Discussion 2019, Nijmegen, Netherland, November 2019
- Organizer of Chinese Association for Psychological & Brain science Meeting, Nijmegen, Netherland, November 2018
- Organizer of symposium “Integrating neuroimaging and gene expression data”. Donders Discussion 2018, Nijmegen, October 2018.

Curriculum vitae



Wei Liu was born on 11th of May in 1991 in Changsha, Human Province, China. There, he graduated (2009) from Zhounan High School of Changsha, which was founded in 1905. After graduation, he went on study Psychology at Hunan University of Traditional Medicine and graduated *cum laude* in 2013. He then enrolled in the research master program Cognitive Neuroscience & Psychology, at the Southwest University, Chongqing, China, where he joined the Creative & Affective Neuroscience lab of Prof. Jiang Qiu. Wei used fMRI and network science to study the effects of traumatic effects (e.g., losing the only child) on brain networks of healthy individuals and patients with major depression disorder. During his time as a master student, he also worked together with other colleagues within the lab to develop the standard neuroimaging data preprocessing, management, and sharing protocol, resulting in three datasets public (total N>3000) which are publicly available for and used by researchers worldwide.

In September 2016, Wei started as a Ph.D. student at the Donders Institute for Brain, Cognition, and Behaviour, under the supervision of Prof. dr. Guillén Fernández and Dr. Nils Kohn with the support of his own Ph.D. fellowship. During his Ph.D., he investigated dynamical memory representations in the human brain using functional MRI in combination with methods from genetics and machine learning. In addition to the main Ph.D. project, he collaborated on a rodent-human translational project of fear memory with researchers from École polytechnique fédérale de Lausanne (EPFL), Switzerland. He led the human neuroimaging section of the project and gained new research methods such as skin conductance response and eye-tracking.

From 2021, he will join the lab led by Prof. Gunter Schumann at the Centre for Population Neuroscience and Stratified Medicine (PONS) with the support of the PONS Early Career Award. He will apply and develop novel (multimodal) genomic and computational neuroimaging methodology to identify and characterize brain structural and functional networks for mechanistic characterization and stratification of behavior and mental disorders.

Acknowledgments

To each of you who are reading this, thank you! Thanks for being here with me at the end of my PhD-journey and hopefully also my future scientific career and life!

First of all, I would like to thank my promotor **Guillén** for taking me in as a Ph.D. student, and for giving me suggestions for all scientific and non-scientific problems and struggles. I can still remember the day I decided to choose you as my Ph.D. supervisor, but never imagine that my PhD-journey will be so smooth and fruitful. Your passion, wisdom, humor, and kindness will always inspire me no matter where I can, and I will always be proud of being one of your Ph.D. students.

To my co-promotor **Nils K**, we first met each other around four years ago at the canteen with “Donders pizza” and I hope we can say goodbye to each other with the same kind of pizza. We worked together as a great team. Beyond all the scientific guidance from you, you also helped me a lot with how to navigate the administrational system of the Donders and University. You cannot imagine how important it is for a foreigner who knows nothing about the system here. You are such a great teacher that can explain almost everything so clear. You always gave me the feeling that I could come to you with even the most stupid question.

During my Ph.D., I am honored to be part of the Memory & Emotion Group. **Guillen, Nils K, Geeralien, Martin, Boris, Lisa, Nils M, Anne, Clara, Klara, James, Jasper, David, Hongxia, Yan-nan, and all interns, rotation students, and visitors**, I felt always home around you guys. I enjoyed the lab/cake meetings, pizza meetings, OHBM-trips, group lunches, cooking evenings, picnic, and group lunches. Among our lab members, I would like to thank my paranymphs. **Yan-nan**, I was so happy when you started a fellow Ph.D. student in our group. You have a special sense of humor and brought a lot of fun to the group although your jokes cannot really be translated into other languages. I wish you a bright and fruitful Ph.D. time here and our collaborative project can lead to successful publications. Yingjie, you contributed a lot to my favorite piece of scientific work so far (i.e., our “aging-sherlock” study). I felt a little sad when you decided not to continue with the memory path because I thought you can become the next female memory rockstar. But I am very happy for you that you found such a nice job in the Genetic department. I wish you can enjoy your forthcoming Ph.D. program and make the best out of it.

Thanks for **all PIs** I have encountered along my scientific journey. Although I did not directly work with you during my Ph.D. time, all of the conversations we had enable me to further think about my Ph.D. projects and neuroscience in general from many different perspectives. **Jiang**, thanks for always having my back during my scientific career and taking time with me to discuss the next steps in terms of personal development. **Shaozheng**, all of my adventures in Donders will not happen without you. You are one of my scientific idols since I was a master student. At the same time, you are always so kind and patient about my questions and uncertainties. **Josef**, thanks for offering me the chance to join your lab after my Ph.D. However, I had to turn you down and take another job after all of the crazy things in 2020. I learned a lot about intracranial recordings from our personal meetings and group meetings. I wish our paths will cross again in some ways when the world comes back to normal. **Gunter**, thanks for offering me the PONS-early career award and the chance to start as the post-doc fellow with you. I hope we can solve some fascinating research questions together in the following years.

I would like to thank all **students, interns, and RAs** who worked with me before. You helped me a lot and contributed a lot to the progress of our projects. **Nancy**, you are the first master student I supervised from the beginning to the end. I am still missing our weekly progress meeting because I also learned a lot from you and our interactions. **Jette**, you definitely contributed a lot to our fear memory project because I had almost zero experience with this topic before. I wish you good luck with your Ph.D. journal and I believe you will have a bright scientific career. **Tom**, you are such a special intern for us because you did your bachelor's and master's internship within our lab! I can still remember how quickly you learn about Python program (almost) by yourself. Very sadly, our collaboration has to be stopped by the COVID-19, otherwise, we can achieve more together scientifically.

I would like to thank all administrational staff at the DCCN and CNS. **Tildie, Sandra, Nicole, Ayse, Sabine, Renee, Geeralien, Arthur, Berend, Marek, Mike, Erik, Paul, and Lucia**. Without your contributions, everything cannot run so smoothly at the Donders Institute, and not a single research project is possible.

I would like to thank all my **participants** for trying to remember all those pictures and maps and I am very sorry that some of them must be emotional according to our study design. Moreover, I would like to thank my **reading committee** for my Ph.D. thesis and **reviewers** for my journal submissions. Your critical reading and suggestions helped us rethink my studies and improved my research dramatically.

A big thank you to my **Chinese Neuroscience/Psychology Friends**. It makes me feel good to stay connected with all of the research progress and "gossip" of the Chinese Neuroscience/Psychology community when I am aboard. **Xiangzheng**, thanks for all of the inspirations and suggestions for my SCAN paper in terms of research methods. **Zaixu**, thanks for your suggestions which helped me improved my Neuroimage paper before the submission. **Wei**, thanks for helping me get the job opportunity to join the Centre for Population Neuroscience and Precision Medicine (PONS) and I am looking forward to more collaborations between us in Shanghai. **Jie, Yu, Dongtao, Wenjing, Lei, Jiangzhou, Qunlin, and all my former lab members**, thanks for all suggestions and help during my Ph.D., especially my job hunting. I am happy to move back to the field of population neuroimaging and we can work together on something again.

Haixu, you came into my life and changed my world. You give me courage, "weakness", and importantly a softer heart. Excited to see where life takes us next.

Last but not least, I would like to thank all **scientists, doctors, and nurses** worldwide who were/are risking their own lives to fight COVID-19. This thesis was finished during the pandemic. Without all of your hard-working and even sacrifice, normal people like me cannot get their work done during this global health crisis.

Donders Graduate School for Cognitive Neuroscience

For a successful research Institute, it is vital to train the next generation of young scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School for Cognitive Neuroscience (DGCN), which was officially recognised as a national graduate school in 2009. The Graduate School covers training at both Master's and PhD level and provides an excellent educational context fully aligned with the research programme of the Donders Institute.

The school successfully attracts highly talented national and international students in biology, physics, psycholinguistics, psychology, behavioral science, medicine and related disciplines. Selective admission and assessment centers guarantee the enrolment of the best and most motivated students.

The DGCN tracks the career of PhD graduates carefully. More than 50% of PhD alumni show a continuation in academia with postdoc positions at top institutes worldwide, e.g. Stanford University, University of Oxford, University of Cambridge, UCL London, MPI Leipzig, Hanyang University in South Korea, NTNU Norway, University of Illinois, North Western University, Northeastern University in Boston, ETH Zürich, University of Vienna etc.. Positions outside academia spread among the following sectors: specialists in a medical environment, mainly in genetics, geriatrics, psychiatry and neurology. Specialists in a psychological environment, e.g. as specialist in neuropsychology, psychological diagnostics or therapy. Positions in higher education as coordinators or lecturers. A smaller percentage enters business as research consultants, analysts or head of research and development. Fewer graduates stay in a research environment as lab coordinators, technical support or policy advisors. Upcoming possibilities are positions in the IT sector and management position in pharmaceutical industry. In general, the PhDs graduates almost invariably continue with high-quality positions that play an important role in our knowledge economy.

For more information on the DGCN as well as past and upcoming defenses please visit:
http://www.ru.nl/donders/graduate-school/phd/

DONDERS
INSTITUTE

Max Planck Institute
for Psycholinguistics

ISBN 978-94-6284-232-8

Radboud University

Radboudumc