Suppressing What Hurts: The Role of Prefrontal-Thalamic Pathways in Inhibitory Control of Intrusive Thoughts and Conditioned Fear

Mahek Kirpalani

St Edmund's College

UNIVERSITY OF
CAMBRIDGE

August 2025

This thesis is submitted for the degree of Doctor of Philosophy

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted, or is being concurrently submitted, for any degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Preface

The Mega-TNT data set analysed in **Chapter 3** which consists of data from 10 previous imaging Think/No-Think studies was compiled by Dr. Dace Apšvalka.

The segmentation protocol to delineate the Nucleus Reuniens (NRe) in the human brain, described in **Chapter 4** was recommended by primate anatomist Dr. Zikopolous. The MATLAB-based application which was used to delineate the NRe ROIs was developed in collaboration with Dr Maité Crespo-García.

Keren Hijmensen, a placement student contributed to data collection for the experiments written up in **Chapters 5** and **6**. Three figures as part of **Chapters 5** and **6** were developed in collaboration with Dr Molly Rowlands.

The research presented in **Chapters 3–5** was disseminated at several national and international conferences:

- **Kirpalani, M., Nardo, D., Zikopoulos, V., & Anderson, M.C.** *The role of the Nucleus Reuniens in inhibitory control.* Poster presented at the **Society for Neuroscience (SfN)** Annual Meeting, Chicago, 2024.
  (Chapters 3–5)
- **Kirpalani, M., Nardo, D., Zikopoulos, V., & Anderson, M.C.** *Prefrontal–thalamic pathways in retrieval suppression: The role of the Nucleus Reuniens.* Poster presented at the **BNA Festival of Neuroscience**, Brighton, 2023.
  (Chapters 3–5)
- **Kirpalani, M., Nardo, D., Zikopoulos, V., & Anderson, M.C.** *Fronto-hippocampal inhibition and the Nucleus Reuniens.* Oral presentation at the **British Neuropsychological Society Meeting**, London, 2022.
  (Chapter 3)
- **Kirpalani, M., Nardo, D., Zikopoulos, V., & Anderson, M.C.** *Prefrontal–thalamic pathways mediating inhibitory control over the hippocampus.* Poster presented at **New Perspectives in Declarative Memory**, University of East Anglia (UEA), Norwich, 2022.
  (Chapter 3)

Abstract

Humans possess cognitive mechanisms which enable them to control the retrieval of unwanted thoughts. Such inhibitory control becomes extremely important when an individual is confronted with unwanted thoughts and intrusive memories. Prior research has implicated prefrontal-hippocampal interactions mediating suppression in humans; however, the precise neural pathway(s) remain unknown. Converging anatomical and functional evidence in rodents and more recently in primates suggests that the Nucleus Reuniens (NRe), a small but vital nucleus situated in the ventral midline of the thalamus plays an important role in mediating such top-down control. Further, fear extinction which forms the basis of exposure therapy has been hypothesized to employ retrieval stopping processes. If this holds true, humans may be capable of extinguishing fear by suppressing its retrieval. In rodents, the NRe has been widely implicated in fear extinction. However, to this date, there have been very few attempts to study this nucleus in the human brain, possibly due to its tiny size and lack of delineation.

This thesis primarily investigates the role of the NRe in memory suppression in the human brain along with studying whether memory suppression and fear extinction share common neural pathways. Chapter 1 and Chapter 2 lay the foundation for this thesis by reviewing evidence across species demonstrating that memory suppression and fear extinction may both rely on shared prefrontal-thalamic-hippocampal mechanisms with the NRe being the thalamic hub which mediates this control. Chapter 3 consists of a meta-analysis of fMRI data (n=330) using the Think/No-Think (TNT) task which revealed that the NRe is functionally active during retrieval suppression, showing task-specific activation and connectivity with the right dorsolateral prefrontal cortex (rDLPFC) and hippocampus (HpC). Building on these findings, Chapter 4 delves into a new segmentation protocol developed for localizing the NRe in individual human brains, informed by anatomy in rodents and primates. Chapter 5 and 6 apply this segmentation to a within-subject fMRI design to study whether memory suppression and fear extinction share common neural pathways and whether the NRe plays a role in these mechanisms. Specifically, chapter 5 provides subject specific evidence that the NRe is involved during retrieval suppression, especially during reactive control as tested via the TNT task. Chapter 6 investigates fear extinction and evaluates if the NRe along with regions involved in the broader inhibitory control network are engaged during extinction. Finally,

chapter 7 summarises findings from all chapters as part of this thesis and makes recommendations about future research avenues.

Taken together, this thesis offers the first anatomically grounded evidence that the NRe is involved in memory suppression in humans and provides a within-subject test of the hypothesis that memory suppression and fear extinction share common neural pathways. The findings from these studies lay groundwork for future research into thalamic contributions to the regulation of maladaptive memory and emotion thus opening translational avenues for treating disorders characterised by intrusive thoughts.

Acknowledgements

As I set out to write this section, I am in disbelief, but I feel deeply grateful. The only way in which I have managed to survive and at times even thrive through this PhD is because of the unwavering support of my people. To each of you, I will always be thankful.

I am most thankful to Bhoomika ma'am, my master's thesis supervisor back in India who first envisioned me pursuing my PhD abroad. Her belief in me pushed me to write to Mike about joining his lab, and that truly changed my life.

To my supervisor, Mike, I am tremendously grateful to you for enabling me to pursue my PhD at the CBU under your mentorship. I have thoroughly enjoyed and learned from our discussions, and your enthusiasm and excitement in ideating has been infectious. You have taught me so much about creativity, belief in thought and self-belief inside and outside the lab. You went out of your way to support me through funding applications, hone my presentation skills and gave me the trust and confidence to bring my ambitious ideas to life.

When I started my PhD, I was less experienced than most of my peers and I was drowning in impostor syndrome. It was Davide, my advisor who believed in me when I could not believe in myself, helped me identify my strengths, and consistently stood by me from the ground up. I have tested his patience and added tremendously to his workload, but he has never once complained and instead only encouraged me to do my best. Much of my computational skill, I owe to him.

When I requested Rik to come on board as my advisor for my PhD project, I was nervous. I was on a tight timeframe, and I had no independent fMRI experience. Rik gave me immense confidence. He met me where I was, without any judgement, without attachment to results and only focused in teaching me, encouraging my questions and honing my skills. He has been extremely generous with his time and little does he know that each time I left his office, I was not only scribbling down his fMRI advice but also making mental notes about his way of life.

I am also grateful to Duncan for his advice and for always giving me a listening ear without judgment, many times over the course of this PhD. I will not forget your kindness. Thank you also to Fionnuala, Camilla, and Tim for your time, support, and advice.

I've had the absolute privilege of working with and learning from Maite, one of the kindest, most resilient, and hardworking people I know. I am grateful to her for her time.

It has also been an absolute pleasure to spend some time working with Dace, whose dedication and work ethic I deeply admire. I've enjoyed working with Akul, who over time has become a close friend. I respect the way he does science, and I am lucky to have learned from him. I am very thankful to Keren Hijmensen, who joined as an intern in our lab. She assisted me with data collection, but more importantly, her questions and enthusiasm about the project kept me going even on the most difficult days. I am also thankful to my wonderful colleagues in the Memory Control Lab — Molly, Subbu, Fred, Zahira, Zuly, Golan, Julia, and Mohith. You have all inspired me in more ways than one. Marie, thank you for uplifting me when I was feeling low and for making the effort to get me out of my thesis every few days, I am grateful for you.

I would like to thank the radiographers, Marius, Steve, and Karen for your imaging advice and for enabling a serious yet light-hearted environment during the long, arduous hours of scanning. I enjoyed your company (for the most part!). Thank you also to Marta for helping me choose the MRI sequence. I am very thankful to Mark and Gary for being extremely patient and going out of their way to enable the complex setup of my study, despite multiple glitches. I am also grateful to Joe for helping me navigate the logistics of grant applications, and to Tony for managing financial logistics throughout my time at the CBU. Thank you to Matthew, Luke, Victoria, Kevin and Andy for your time and support. Importantly, I am very thankful to the IT team, Russell, Anthony, Howard, and Jeff for patiently helping me solve the countless issues I created through my ignorance. I am also thankful to my college tutor, Fernando and college pastor Fr. Ed for their encouragement.

When I was starting out my PhD, I had the privilege and luck of sharing this fairytale town, Cambridge, with two of my closest friends from back home. Meera and Aarushi, you both gave me a home away from home. In the beginning, when I did not have complete funding and was drowning in self-doubt, you gave me stability, love, and cheer. You genuinely enabled me to thrive, and I will always cherish the home we built and the time we spent together. Looking back, I cannot imagine how it would have been without your unconditional love.

Meera, you have been so deeply present, especially in the last few months that I know when I finally submit this thesis, you will feel a palpable sense of relief yourself. That is how much you have held me up, carried me, and made this journey bearable.

Shraddha and Ishita have been my oldest friends since school. They have always cheered me on and had immense faith in my abilities, even when I've had none. When I received

my PhD offer, Shraddha gave me courage, she took it upon herself to help me find funding and crushed every seed of apprehension I had, head-on. She has consistently done whatever it takes to push me to do what is best for me, whether it was taking on this offer or dating my then boyfriend and now fiancé. Ishita has given me constancy, showing up for me for as long as I can remember. When the going has been rough, she has been my unwavering support, especially in the final stages of this PhD. I am also very thankful to Aunty Shammi, who has consistently checked in on me and been there for me throughout this journey.

Rohini and I met during my MSc, and she has been instrumental in helping me keep my sanity during difficult moments. She has always been optimistic about my abilities and has been just a phone call away. I have vented to her more times than I remember, and I've always felt lighter after. Abhilasha, my MSc roommate, has also been an important part of my life. She has always known when to check in, even after long silences, and has reassured me at the right times. Santosh, her partner, has also been lovely, and I am grateful for his advice: "You need to be selfish for now and think about your own needs alone!"

To Sahithyan, thank you for believing in me and telling me things I didn't particularly want to hear, but needed to. To Aswini and Christelle, thank you for cheering me on and having my back. To Shruthi, Sakshi, and Yagika, who entered my life when we all started applying for PhDs, thank you for walking this parallel journey. It has been wonderful to share this path with you all. Shruthi has put in enormous effort to keep in touch, and I've had most of my PhD meltdowns with her. She has always made me feel validated and less alone.

Lamiya, also finishing her PhD back home, has been rooting for me, often giving me optimism even while struggling with her own. After talking to her, I often felt like everything will be alright. Ria has always been curious about my work and has listened to me ramble on endlessly. Hele, Alisha, and Sawla have always cheered me along. Yashesh has been my consistent cheerleader, always pushing me to climb higher. I am thrilled to have him in my corner. Alexia, my school friend, a fierce career woman and a busy mum, has somehow always made time to check in and cheer me on. Baby Skylar's videos have brought me more joy than I can describe.

Manish and I met in Cambridge, and I'm so glad our friendship has endured (largely thanks to his effort). After every conversation with him, I've fretted less and remembered there is so much more to life. I'm usually in a good mood after annoying him for a while.

I am also grateful to Chaitra, my college neighbour for being there for me and who along with Manish, taught me how to cook.

Molly and Elena, you have been my pillars within the CBU, both academically and personally. Life in Cambridge would have been much harder without you both, and I feel very lucky to have you in my life.

I am thoroughly thankful to Kalu and Aaji for their infinite wisdom, their check-ins, and their courage that has strengthened me.

Most importantly, I am indebted to my family, Mum, Dad, and Krish. Your love, patience, and sacrifices have been the foundation of everything I have achieved. This journey has not been easy for you, having me so far away, but you never let me feel the distance. You tirelessly called, cared for me across time zones, and made sure I never felt alone. You gave me courage when I faltered, strength when I was exhausted, and perspective when I was lost. You all have been my safe space, never letting me forget who I am even when I doubted myself. Krish, you have particularly shown me the lighter side of life, making me laugh when I have felt crushed under the weight of it all. I delved into memory research after the Alzheimer's diagnosis of my beloved grandmother. So much of my motivation, my determination, and my will to keep going has come from her memory and from my family's faith in me. I truly believe that everything good that has come to me in life is because of you. I hope I can always make you proud. This thesis is dedicated to you.

I am also grateful to Bharti, for her support and belief in me.

And to Ali, against the magical backdrop of Cambridge, I found not just a partner, but my person, my soulmate. It feels like a rare and precious gift to have found love in the very midst of a PhD, a time when I was at my most vulnerable, overwhelmed, and often doubting myself. Somehow, in the middle of stress and chaos, I found you and with you, I found strength, comfort, and joy. Without you, this thesis would not have been possible. You gave me the courage to "do it scared" and the strength to keep showing up, even when I wanted to give up. You stood by me through innumerable lows, sleepless nights, rejections, doubts, tears and through it all, you have always had my back and lifted me higher. Your patience, encouragement, and unwavering faith in me have carried me more than you will ever know. To your parents as well, I am deeply grateful, their warmth and encouragement have made me feel embraced as family. Ali, this PhD was my dream, but you made it bearable and joyful. For all of that, and for the rare and precious gift of finding love when I least expected it, thank you.

In the midst of this PhD, a very close friend was also diagnosed with a severe mental health condition, and I became their primary caregiver. It was heartbreaking to witness their suffering, but I am grateful for the strength I found and for the help I received from Anna Bevan at the CBU. This experience transformed how I view my work. Though the aftermath was complex, I am beyond grateful to all my people who did what it took to help me through. I am also thankful to my remote co-caregiver, those days would have been harder without your support. I dedicate this thesis also to all caregivers out there, who, while making mistakes, are ready to learn and to show up for their loved ones.

Importantly, I owe deep gratitude to my funders namely, the Cambridge Trust, the Medical Research Council and the Cambridge Philosophical Society. This journey would have been impossible without their generous support.

Lastly. I am deeply grateful to my examiners, Amy and Jarrod, for taking the time to carefully read my thesis and for the thoughtful, stimulating questions they posed during my viva. I genuinely appreciated our engaging conversation, their insightful suggestions for strengthening the thesis, and their encouragement at the end of this long journey.

Table of Contents

Abbreviations

AC–PC	Anterior–Posterior Commissure
AGm	Medial Agranular Cortex
AIC	Anterior Insular Cortex
antOFC	Anterior Orbitofrontal Cortex
BA	Brodmann Areas
BLA	Basolateral Amygdala
CA1	Cornu Ammonis area 1
CA3	Cornu Ammonis area 3
CeA	Central Nucleus of the Amygdala
CS+/CS-	Conditioned Stimulus Positive/Negative
CB	Calbindin
CR	Calretinin
CS	Conditioned stimulus
dACC	Dorsal ACC
DCM	Dynamic Causal Modelling
dmPFC	Dorsomedial Prefrontal Cortex
DTI	Diffusion Tensor Imaging
EEG	Electroencephalography
EPI	Echo-Planar Imaging
ERP	Event-Related Potential
FC	Fear Conditioning
FDR	False Discovery Rate
fMRI	Functional Magnetic Resonance Imaging
FWE	Family-wise Error
GABA	Gamma-aminobutyric Acid
GLM	General Linear Model
HBREC	Human Biology Research Ethics Committee
HpC	Hippocampus
HRF	Hemodynamic Response Function
ICV	Intracranial volume
IL	Infralimbic
ITI	Inter-trial Interval

MNI	Montreal Neurological Institute
mPFC	Medial Prefrontal Cortex
MRS	Magnetic Resonance Spectroscopy
MTL	Medial Temporal Lobe
MVPA	Multivoxel Pattern Analysis
NRe	Nucleus Reuniens
OCD	Obsessive Compulsive Disorder
PAG	Periaqueductal gray
PL	Prelimbic
PPI	Psychophysiological interaction
PTSD	Post-Traumatic Stress Disorder
PV	Parvalbumin
rDLPFC	Right Dorsolateral Prefrontal Cortex
ROI(s)	Region(s) of interest
RSA	Representational Similarity Analysis
rVLPFC	Right Ventrolateral Prefrontal Cortex
SAM	Self-Assessment Manikin
SCR	Skin Conductance Response
SIF	Suppression-Induced Forgetting
SLM	Stratum Lacunosum-Moleculare
SPM	Statistical Parametric Mapping
SVD	Singular Value Decomposition
TCAQ	Thought Control Ability Questionnaire
TE	Echo time
tES	Transcranial Electrical Stimulation
TMS	Transcranial Magnetic Stimulation
TNT	Think/No-Think
TR	Repetition time
US	Unconditioned Stimulus
vmPFC	Ventromedial Prefrontal Cortex
VR	Virtual Reality

Chapter 1

Prefrontal–Thalamic Circuits for Memory Suppression: The Role of the Nucleus Reuniens

1.1 The Nature and Function of Memory Suppression

Memory is fundamental to our sense of being. It anchors us to who we are by enabling us to draw upon past experiences to guide present actions and make future decisions. Without memory, we lose continuity of self and the ability to make sense of the past or navigate the present (Tulving, 1985). In recent decades, cognitive science has demonstrated that memory is not just a passive storage system, but a dynamic process that we can regulate to a certain extent. This ability is increasingly seen as part of a domain-general inhibitory control system, in which prefrontal regions-such as the right dorsolateral cortex (rDLPFC) and right ventrolateral prefrontal cortex (rVLPFC) can flexibly target different brain systems to suppress a range of mental contents, including motor actions, emotional responses, and internal thoughts like memories (Apšvalka et al., 2022).

In line with this, research suggests that the same kind of executive control we use to stop physical actions can also be applied to memory, thereby letting us either retrieve or intentionally suppress it (Anderson & Green, 2001). This becomes especially important when dealing with intrusive memories which can be defined as persistent, unwanted thoughts that can derail attention and emotional stability (Clark, 2005). Retrieval suppression, an active, voluntary attempt to stop a memory from coming to mind, is a core example of memory control, allowing individuals to keep distressing content out of awareness (Anderson & Hulbert, 2021). Loss of this ability has been linked to conditions like Post-Traumatic Stress Disorder (PTSD) and anxiety (Catarino et al., 2015; Mary et al., 2020; Stramaccia et al., 2015).

Brain imaging studies have shown that the prefrontal cortex, especially the rDLPFC plays a key role in this process by reducing activity in the hippocampus (HpC), a region central to memory retrieval (Anderson et al., 2004; Benoit & Anderson, 2012). When suppression is repeated, it can lead to adaptive forgetting, making the memory less likely to come back (Levy & Anderson, 2002). This kind of forgetting can be helpful,

allowing us to realign memory with our current needs, lighten emotional load, and redirect attention.

1.2 Cross-Species Evidence for Memory Control

1.2.1 Rodents

Rodents don't seem to possess a direct anatomical equivalent of the primate DLPFC, which in humans plays a major role in executive functions like working memory and inhibition (Preuss, 1995; Uylings et al., 2003) but they do have medial frontal areas that seem to serve similar functions. In particular, the prelimbic (PL) and infralimbic (IL) cortices, both part of the medial prefrontal cortex (mPFC) are thought to map, functionally if not anatomically, onto parts of the primate PFC. The PL region is involved in cognitive tasks like decision-making and emotional flexibility, like the lateral PFC in primates, while the IL region is more tied to autonomic/visceral regulation, echoing functions of the orbitomedial PFC (Vertes, 2004). Thus, the rodent PL, IL, and anterior cingulate cortex (ACC) regions contribute to memory and behavioural regulation in ways that mirror higher mammals.

More recently, Bekinschtein et al. (2018) demonstrated that pharmacological inactivation of the mPFC with muscimol abolished retrieval-induced forgetting in rats. Consistent with an active control role, c-Fos expression in the mPFC was elevated during the early stages of retrieval practice, when interference was high, and declined as interference decreased. These findings provide evidence that the rodent mPFC contributes to suppressing competing memories. Further, in learning contexts that require forgetting-like extinction, rodents recruit vmPFC pathways to dial down both emotional responses and memory expression. The vmPFC can modulate fear expression through descending projections to the amygdala (Quirk & Mueller, 2008). The IL cortex has been implicated in suppressing memory traces, electrical stimulation of the IL is found to reduce conditioned fear and strengthen the extinction memory (Milad & Quirk, 2002; Milad et al., 2004). Work by Frankland and Bontempi, (2005) demonstrates that memory retrieval becomes increasingly dependent on mPFC circuits as memories age, suggesting that the mPFC gradually acquires a more central role in guiding memory expression. Neurophysiological work has also shown that mPFC inactivation disrupts selective hippocampal ensemble firing which led to the nonselective retrieval of competing memories (Navawongse & Eichenbaum, 2013). In this study, hippocampal neurons encoded the context appropriate odour only when the mPFC was functional; when the

mPFC was inactivated, these neurons were found to fire indiscriminately across contexts. This suggests that the mPFC can suppress hippocampal representations which are irrelevant to the current situation.

These findings suggest that the rodent mPFC plays a broader role in regulating memory and emotion that closely resemble prefrontal control in primates, as discussed next. While it remains unclear whether rodents possess a domain-general control system like that proposed in humans, the presence of prefrontal-hippocampal pathways involved in emotional and memory regulation could indicate a shared evolutionary basis.

1.2.2 Non-Human Primates

Non-human primates are closer to the human brain in terms of prefrontal organization and connectivity. Rhesus monkeys, for example, have well-differentiated PFC regions, including lateral areas like Brodmann areas (BA) 9 and 46, and medial areas like BA 24 and 32 in the ACC, as well as area 25, together resembling the cytoarchitecture of the human PFC (Barbas & Pandya, 1989; Petrides & Pandya, 1999).

These frontal regions have direct connections to the medial temporal lobe (MTL) (Barbas & Blatt, 1995). The monkey mPFC, including the ACC, has robust bidirectional communication with the HpC and surrounding areas. While direct PFC-to-HpC projections are sparse, strong HpC-to-PFC pathways support ongoing functional connectivity (Barbas & Blatt, 1995; Wang, John, & Barbas, 2021). In primates, communication between the MTL and PFC appears to rely on intermediate hubs such as the entorhinal and perirhinal cortices (Suzuki & Amaral, 1994). While the nucleus reuniens (NRe) has been recognized as a key relay in rodents, recent anatomical studies in macaques demonstrate that it also has bidirectional connections with the HpC and medial prefrontal cortex, suggesting a similar integrative role in primates (Joyce et al., 2022). BA 32 (dorsal ACC) and area 25 (ventromedial PFC) project to MTL regions, including entorhinal, perirhinal, and parahippocampal cortices, which then communicate with the HpC (Barbas & Blatt, 1995; Carmichael & Price, 1995). Altogether, this forms a network linking PFC, thalamus, and HpC forming a network of anatomical substrates theoretically capable of supporting memory control.

Joyce et al. (2020) demonstrated that subregions of the primate ACC and the ventromedial prefrontal cortex (vmPFC) are anatomically embedded within circuits that could support both emotional and memory functions. These regions are interconnected with the amygdala, hypothalamus, and brainstem structures involved in autonomic and

affective regulation, as well as with MTL areas such as the entorhinal, perirhinal, and parahippocampal cortices that relay to the HpC.

While subjective memory intrusions (as discussed later, in section 1.3.2.2) cannot be tested for in monkeys, their flexible, memory-guided behaviour suggests evidence for control over retrieval. For instance, if one memory leads to no reward and another does, monkeys learn to recall the useful memory, suggesting selective retrieval (Basile & Hampton, 2017). These kinds of strategic retrieval behaviours could rely on the same PFC–MTL circuits that humans engage during memory suppression.

1.2.3 Humans

A widely used method is the Think/No-Think (TNT) task, designed to examine how people intentionally avoid retrieving unwanted memories (Anderson & Green, 2001). Studies using this paradigm have shown that people can limit access to specific memories, often reducing their later accessibility.

Brain imaging studies, using Functional Magnetic Resonance Imaging (fMRI) and Electroencephalography (EEG), have shown that memory suppression is linked to increased activity in frontal control areas like the rDLPFC and ACC, and reduced activity in the MTL, including the HpC. This pattern reflects a top-down process where prefrontal regions inhibit memory-related activity (Anderson et al., 2004). Participants can make subjective ratings, use different strategies like direct suppression or thought substitution, and their failure and success can be evaluated. At the same time, because neural circuits in humans cannot be manipulated at the current level as in preclinical animal models, it is harder to pinpoint the underlying mechanisms with precision. This limitation has fuelled an interest in combining findings from human neuroscience with detailed circuit-level studies in animals, to build a more complete understanding of mechanisms underlying memory control.

1.3 Laboratory Studies of Memory Control in Humans

1.3.1 The Think/No-Think Paradigm

The TNT task typically consists of three phases (Figure 1). First, during the training phase, participants learn pairs of stimuli (e.g., word pairs like ordeal–roach; object–scene pairs) to a specific accuracy threshold, ensuring the associations are well encoded in long-term memory.

Next comes the Think/No-Think phase, where participants see only the cue items from the learned pairs. Each cue is presented multiple times (e.g., 8–12 repetitions per

item), and for each, participants are instructed either to recall the associated target (Think trials) or to suppress it (No-Think trials). Cues appear in distinct colours, green for Think trials and red for No-Think trials to indicate whether a memory should be retrieved or suppressed. Typically, a Think cue is always a Think cue, and a No-Think cue remains a No-Think cue throughout the task. This consistent assignment allows participants to build up retrieval or suppression abilities for specific memories over repeated trials.

On No-Think trials, cues tend to trigger automatic retrieval of the target due to the earlier learning. When this happens, participants are asked to actively stop the retrieval process, either preventing the memory from rising to awareness or pushing it out of mind once it begins to surface. In the standard direct suppression variant of the task, participants are instructed not to distract themselves with substitute thoughts but instead to push away the unwanted memory itself.

Think trials, by contrast, involve intentionally recalling the target and holding it in mind for several seconds. Repeating these cues across multiple trials provides participants with extended practice in either retrieving or suppressing specific memories allowing the effects of repeated suppression on later memory recall to be tested.

After this Think/No-Think phase, participants receive a final surprise memory test for all the targets to assess how the prior instructions affected later recall. Crucially, this test includes not only the Think and No-Think items but also some baseline items (pairs studied during the training phase but not presented during the TNT phase). By comparing performance on these item types, we can measure the impact of retrieval and suppression practice. The typical finding is that memories that were repeatedly retrieved (Think items) show enhanced recall relative to Baseline (a kind of practice or reminder benefit), whereas memories that were repeatedly suppressed (No-Think items) show impaired recall relative to Baseline. In other words, preventing retrieval multiple times in response to a cue makes it harder to recall that memory later. This effect is called suppression-induced forgetting (SIF) (Anderson & Hanslmayr, 2014; Anderson & Hulbert, 2021). This occurs even though the No-Think items had equal or more exposure to the cues compared to Baseline, indicating that it's not mere lack of practice, instead, some active process during No-Think trials hindered memory retention. The TNT procedure, therefore, provides a robust behavioural measure of voluntary memory suppression.

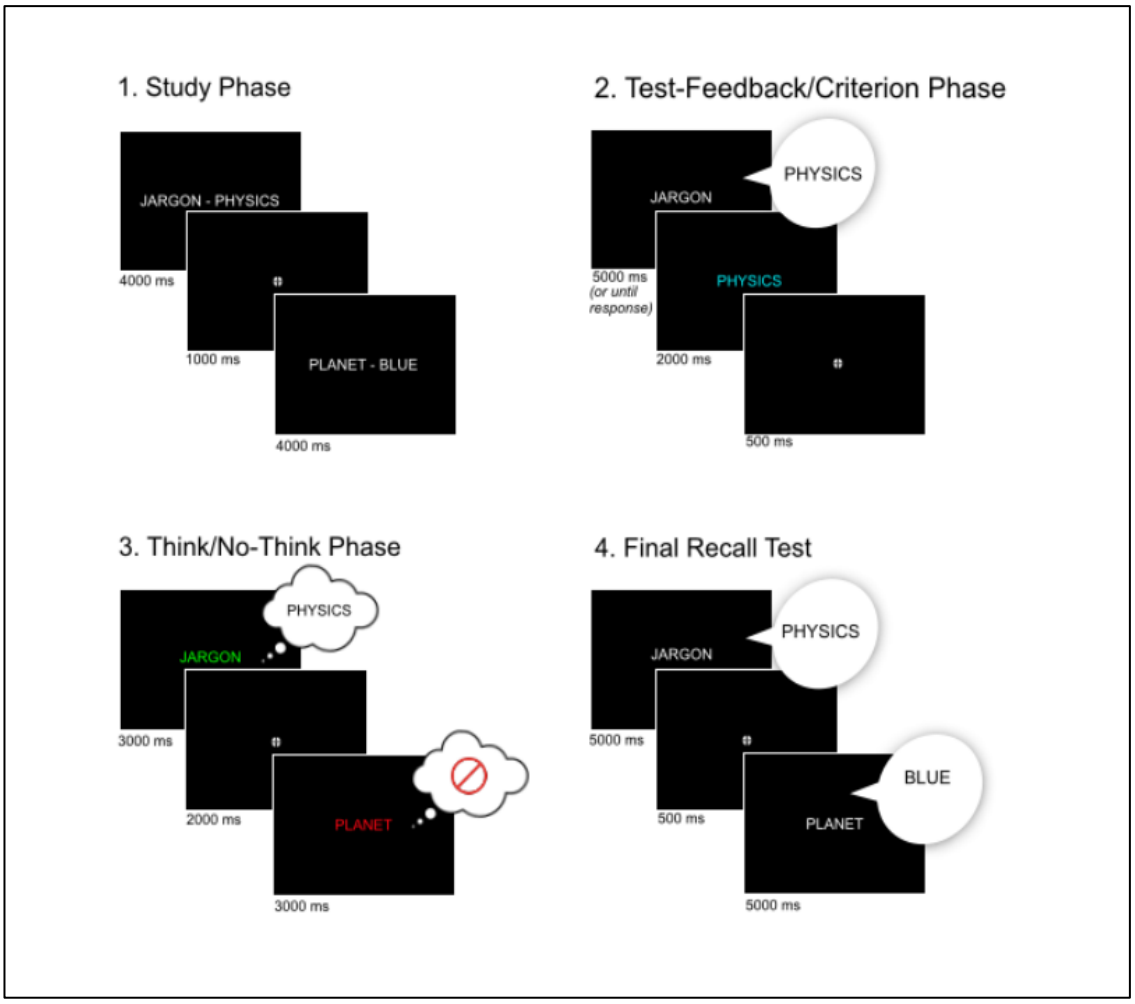

Figure 1. Think/No-Think (TNT) Experimental Overview.

Participants learn cue-target pairs (word–word, word–picture or picture-picture). If participants meet the learning criterion, they enter the main TNT phase: green cues (Think cues) prompt retrieval, red cues (No-Think cues) require suppression of the associated target. A final surprise recall test assesses memory for targets, associated with Think and No-Think cues. It also includes baseline targets not presented in the TNT phase. Suppression-induced forgetting (SIF) is calculated as the difference in recall between baseline and No-Think items, and this is used as an index of successful memory suppression.

1.3.2 Behavioural Indices of Memory Control

Two key behavioural indices have emerged as signatures of successful memory control in the TNT paradigm: SIF and the frequency of memory intrusions during suppression. These measures are related, and reflect cognitive control mechanisms at play i.e., forgetting in this case and the difficulty of exerting control i.e., memory intrusions.

1.3.2.1 Suppression-Induced Forgetting (SIF)

SIF is the phenomenon where memories that an individual repeatedly tries to suppress become harder to recall later, even compared to baseline memories that had no additional practice. SIF provides a behavioural quantification of how effective retrieval suppression

was in weakening the memory. The robust finding that No-Think items are recalled less often than Baseline, indicates that the act of suppressing retrieval causes forgetting. This effect has been demonstrated across many studies and with various types of stimuli. For instance, SIF has been observed not only with simple word pairs (Anderson & Green, 2001), but also with picture–picture associations (Depue et al., 2006) and even autobiographical memories (Noreen & MacLeod, 2013). Moreover, SIF occurs for both, content of neutral and negative valence (Depue et al., 2006; Gagnepain et al., 2017).

SIF is often assessed not only with the originally learned cue, but also with novel test cues to probe the same memory i.e., the independent probe test. For example, after suppressing ordeal–roach, one might be probed for, "Insect r____?" (a category-plus-stem cue that could evoke roach) during the final test. It has been found that SIF usually generalizes to these novel cues, meaning the memory is less accessible even when prompted in a new way. This cue-independence is a critical property: it suggests the memory trace itself was weakened or inhibited, rather than just the specific cue-target association being unlearned. This generalized impairment is consistent with the idea that inhibitory control disrupted the target memory's representation. It was rendered less retrievable even when cued with an independent probe. (Anderson & Green, 2001).

Another property is the interference dependence of SIF (Levy & Anderson, 2012). That is, SIF tends to be larger when the to-be-suppressed memory strongly intrudes or competes during suppression attempts. If there is little to no interference i.e., if the cue fails to remind the individual of the target at all, then they might not need to engage much control, and correspondingly the memory remains relatively intact. This aligns with the idea that inhibitory control is recruited in response to conflict between an unwanted memory and the goal to suppress. Studies have found that SIF is often correlated with how much difficulty individuals experienced on suppressing the item. When an unwanted memory intruded frequently and forced one to push it out of mind multiple times, the later forgetting effect is found to be larger because those items received the most sustained inhibitory suppression (Levy & Anderson, 2012; Hellerstedt et al., 2016). However, if a No-Think item rarely came to mind (low interference), the individual might go through those trials without engaging suppression, and then little or no forgetting is observed for that item. Thus, this pattern contradicts a pure passive decay or context-shift account and highlights the role of effortful motivated control in inducing voluntary forgetting.

1.3.2.2 Memory Intrusions

An intrusion is recorded when the participant reports that the to-be-suppressed memory involuntarily popped into mind on a No-Think trial, even though they were trying not to think of it. Typically, studies using intrusion reports ask participants to indicate, after each No-Think trial, whether the associated memory came to mind-using a three-point scale: Never, Briefly, or Often, reflecting whether the memory did not intrude, was quickly dismissed, or persisted in awareness. In the beginning of suppression training, intrusions are more common. However, with repeated practice, participants can get better at suppression, and the intrusion rate can drop from around 60% to around 30% on later repetitions (Levy & Anderson, 2012).

Importantly, the degree of intrusion reduction is often predictive of later forgetting. Participants or items that show larger decreases in intrusions across the task tend to exhibit more SIF on the final test (Hellerstedt et al., 2016; Levy & Anderson, 2012). Thus, in this manner, intrusion frequency during training serves as a proxy for how much control was needed. Memories which interfere to a greater degree provoke more suppression effort and end up being suppressed to a greater extent if the exertion of inhibitory control is successful.

Intrusions also reveal a proactive vs. reactive control distinction. Ideally, one would like to stop the memory from intruding at all (proactive control), but if it intrudes, one can still react by pushing it out (reactive control). Some No-Think trials involve no intrusion (proactive control), whereas others involve an intrusion that is then purged (reactive control). Levy and Anderson (2012) found that when proactive control succeeds (no intrusion), hippocampal activity is moderately reduced but when only reactive control is available (an intrusion occurs), hippocampal activity shows a much larger down-regulation to push the memory out of mind. Thus, not all suppression attempts are equal, the response may differ in intensity depending on whether proactive control is applied or if the intrusion is purged out via reactive control.

Thus, SIF and intrusion metrics together paint a picture of memory control in action. When people attempt to suppress memories: (1) People often partially fail (intrusions occur), especially initially, but improve with practice. (2) Their successful efforts lead to forgetting of the suppressed content, measured as reduced recall relative to baseline. (3) The degree of forgetting is linked to the level of internal conflict as indexed by intrusions: the more a memory intrudes, the more it could be pushed down by strong inhibitory control, producing greater SIF (Levy & Anderson, 2012).

1.4 Neural Mechanisms of Memory Suppression

A central insight from the past two decades of memory control research is that suppressing unwanted memories is an active process which recruits a dynamic, distributed network of control systems in the brain. A core component of this network is the PFC, which mediates top-down inhibition of memory-related activity, particularly in the HpC, the MTL structure crucial for episodic retrieval. As this body of work has grown, it has revealed a sophisticated network that detects the emergence of intrusive content, initiates suppression, and adapts in real time to the demands of inhibitory control.

The first compelling evidence for top-down memory suppression came from Anderson et al. (2004). Using fMRI, they showed that when participants actively attempted to suppress retrieval (No-Think trials), activation increased in regions including the bilateral DLPFC, VLPFC and the ACC while activation in the HpC reduced. Greater DLPFC engagement and stronger hippocampal downregulation predicted more subsequent forgetting of the successfully suppressed items as compared to the items which were remembered. These findings provided a neural correlate for voluntary memory suppression, and the DLPFC was established as a plausible source of inhibitory control over hippocampal retrieval processes.

This framework was further expanded, (Depue et al., 2007, 2016) incorporating additional regions into the suppression network, including the medial superior frontal gyrus, the ACC, and basal ganglia. The right inferior frontal gyrus and the right medial frontal gyrus were found to suppress sensory components and emotional components of the memory representations. Beyond these core regions, other brain areas contribute to suppression. The bilateral VLPFC is found to be consistently active during No-Think trials and may be involved in maintaining suppression goals or deploying cognitive strategies such as thought substitution. The rVLPFC is found to be implicated during memory suppression and the left VLPFC is thought to be implicated during thought substitution (Benoit & Anderson, 2012). When the suppressed content is emotionally charged, the amygdala also becomes part of this network. Studies have shown that suppressing unpleasant memories involves modulation of the amygdala by the right lateral prefrontal cortex, leading to altered long-term affective responses. (Depue et al., 2007; Gagnepain et al., 2017).

Functional connectivity analyses offered a more fine-grained understanding of these interactions. Dynamic causal modelling (DCM) analyses have shown that activity

in the HpC can be modulated by top-down control via the anterior rDLPFC (Apšvalka et al., 2022; Benoit & Anderson, 2012; Benoit et al., 2016; Gagnepain et al., 2017; Schmitz et al., 2017). Further, when emotionally negative content was employed, activity in the amygdala was reduced which was not merely an indirect consequence of hippocampal suppression but both, the HpC and amygdala were targeted, in parallel, by top-down inhibitory control signals which originated from the right aDLPFC (Gagnepain et al., 2017).

Known for its function in conflict monitoring and error detection, the ACC appears to serve as an internal alarm system for intrusive thoughts. Crespo-García et al. (2022) found that the dACC exhibited distinct activity pattern, depending on the stage of control, revealed through simultaneous EEG-fMRI. Early activation (300-450 ms), marked by elevated mid-frontal theta and N2 amplitudes was associated with reduced hippocampal activity and reflected proactive control, potentially stopping the memory before pattern completion started. In comparison, later activation (500-700 ms) occurred at a latency consistent with memory retrieval. This was interpreted as reflecting a response to a likely intrusion, when a memory had already intruded. Crucially, this late dACC signal predicted increased effective connectivity to the right aDLPFC, which subsequently exerted top-down hippocampal downregulation. High-intrusion trials also showed an initial burst of hippocampal theta consistent with a short-lived retrieval followed by rapid attenuation once suppression mechanisms engaged. These results reveal a hierarchical process in which the ACC monitors for mnemonic conflict and, when necessary, calls upon lateral PFC to implement suppression.

While connectivity analyses clarify where inhibitory control originates, neurochemical evidence explains how hippocampal activity is silenced. One hypothesis is that local inhibitory interneurons in the HpC, which release gamma-aminobutyric acid (GABA), are responsible for suppressing pyramidal cell firing during retrieval. In order to test this, Schmitz et al. (2017) employed magnetic resonance spectroscopy (MRS) to measure resting state GABA concentration in the HpC. It was found that individuals with higher hippocampal GABA concentrations exhibited stronger HpC deactivation. This resulted in suppression and greater forgetting of No-Think items. DCM analyses further confirmed that top-down inhibitory control was observed in participants with high hippocampal GABA. Importantly, GABA levels in the rDLPFC and visual cortex were unrelated to suppression success which suggests that suppression depended on the availability of GABA in the HpC alone.

Secondly, hippocampal GABA was not predictive of performance on a motor inhibition task, demonstrating a dissociation between thought suppression and motor stopping. While both depend on frontal control, only memory suppression required local inhibitory capacity within the HpC (Anderson & Hulbert, 2021; Schmitz et al., 2017). This distinction lends further support to the idea that memory control recruits a network which is at interplay between domain-general control systems and domain-specific inhibitory mechanisms.

Together, this body of work paints a picture of memory suppression as a multi-stage process. It begins with early detection (via the ACC), engages targeted control (via rDLPFC), and implements suppression through local inhibition (via hippocampal GABAergic circuits).

1.5 Pathways for Memory Suppression: The Dual Pathway Hypothesis

Although there is a clear understanding of the major players in memory suppression such as the rDLPFC exerting top-down control over the HpC, the ACC detecting intrusions and initiating regulation, and hippocampal GABA enabling local inhibition-the precise anatomical route(s) through which this control is implemented remains unknown. Identifying the specific neural pathways involved is essential not just for understanding the basic science of memory control, but also for developing more targeted interventions for mental health conditions such as PTSD and depression which are characterised by a high frequency of intrusive thoughts. To address this gap, Anderson, Bunce, and Barbas (2016) proposed the Dual Pathway Hypothesis, which outlines two complementary anatomical routes by which the PFC could suppress hippocampal activity.

1.5.1 Cortical Gating via the Entorhinal Cortex

The first proposed route for PFC control over the HpC involves modulation of the entorhinal cortex, the primary interface through which cortical inputs reach the HpC. Most afferents to the HpC, particularly to the dentate gyrus and Cornu Ammonis area 3 (CA3) are routed through superficial layers (II and III) of the medial and lateral entorhinal cortex. However, these layers are themselves regulated by deeper layers (V/VI), which receive projections from the ACC and broader medial prefrontal regions. Anderson, Bunce, and Barbas (2016) propose that the ACC exerts control by activating GABAergic parvalbumin-positive interneurons in these deep layers, thereby inhibiting signal relay to superficial layers. This inhibition serves two complementary functions. First, it blocks incoming sensory or cue-based inputs from triggering hippocampal pattern completion,

effectively gating off the HpC and preventing retrieval. Second, it reduces feedback re-entry from hippocampal output (Cornu Ammonis area 1 (CA1) and subiculum) that would normally re-enter via the entorhinal cortex, thus dampening reverberatory loops that sustain memory activation. Together, this pathway provides a mechanism for rapid, proactive control over retrieval via filtering cues before they can elicit an unwanted memory.

1.5.2 Thalamic Modulation via the Nucleus Reuniens

The second route proposed by Anderson, Bunce, and Barbas (2016) involves the NRe of the midline thalamus as an intermediary between the mPFC, particularly the ACC (area 32) and the HpC. The NRe is densely interconnected with both the PFC and HpC, receiving input from medial frontal regions and projecting excitatory signals to the HpC and back to the PFC. While the projections from the NRe are excitatory, and they exert a net inhibitory effect on hippocampal output by strongly activating local GABAergic interneurons. These interneurons suppress pyramidal cell activity, thereby dampening memory retrieval through an indirect, network-mediated mechanism.

In this model, strong activation of the NRe could recruit local inhibitory interneurons in the HpC to suppress neural ensemble firing, thus silencing memory traces already in the process of retrieval. This pathway is particularly relevant to reactive control-situations in which an intrusive memory has already been partially activated. If entorhinal gating fails to prevent retrieval at the input stage, the NRe-mediated pathway may serve to interrupt the hippocampal response midstream, halting recollection and weakening the trace through disrupted neural replay.

Because the NRe also projects to entorhinal and perirhinal cortices, its engagement may further suppress MTL activity beyond the HpC, thereby limiting the propagation of mnemonic content into conscious awareness. Anderson et al. (2016) suggest that this route may underlie the particularly strong hippocampal deactivations observed during reported memory intrusions, and that it operates on a slightly slower, reactive timescale compared to the faster, proactive entorhinal gating mechanism.

While the dual-pathway hypothesis proposes that entorhinal gating and thalamic suppression operate at distinct temporal stages, proactive and reactive, respectively, it is also possible that both pathways are recruited in parallel. Their co-activation could allow for more robust modulation of hippocampal dynamics, especially under high cognitive load. When a cue is rapidly identified as needing suppression, the PFC may engage the

entorhinal cortex to block input into the HpC, thereby preventing retrieval altogether, a means of proactive control. If this mechanism fails and a memory intrudes, a second, reactive pathway may come into play: the PFC signals via the NRe to globally inhibit hippocampal activity and suppress the emerging recollection. This division could explain empirical findings such as the distinction between modest hippocampal deactivation on most No-Think trials and the more dramatic suppression seen during trials involving intrusions (Levy & Anderson, 2012). While the dual-pathway framework remains a theoretical model, it has prompted increasing interest in the specific role of the NRe in memory control. The next section focuses on this structure, which is uniquely positioned to mediate top-down and bottom-up communication between the PFC and HpC.

As discussed earlier, studies have established a link between memory control and psychological wellbeing. However, the neural pathways that underlie this emotional regulation remain unknown. Understanding the precise role of the NRe could not only shed light on how adaptive forgetting occurs but also how its breakdown could contribute to mental health conditions which are characterised by intrusive thoughts.

1.6 Clinical Relevance of Memory Suppression

The ability to suppress unwanted memories is not only an adaptive function, but it also has meaningful consequences for psychological health. Across studies, repeated suppression of aversive memories has been shown to make them less intrusive and less emotionally charged over a period of time (Depue et al., 2007; Gagnepain et al., 2017).

Individuals who are unable to suppress unwanted memories often experience more persistent intrusions, a hallmark of conditions such as anxiety, depression, and OCD (Catarino et al., 2015; Stramaccia et al., 2021). Further, patients suffering from PTSD exhibited impaired memory suppression whereas, trauma exposed individuals showed suppression abilities comparable in healthy controls (Mary et al., 2020). Thus, inhibitory control could serve as a protective factor. This view was further supported by a behavioural study by Mamat & Anderson, 2023, wherein they found that training participants to suppress their future fears reduced negative affect, anxiety and depression symptoms and these benefits lasted for at least three months.

Such findings position memory suppression as a function with clinical utility. Strengthening the ability to suppress retrieval to unwanted and intrusive thoughts could be an important therapeutic goal. Emerging evidence in rodents points to the NRe as being

an important region in this circuitry underlying inhibitory control which could potentially offer new insights into how the brain achieves emotional control over unwanted memories.

1.7 Evidence for the Nucleus Reuniens in Memory Control

To understand how the nucleus reuniens might support such control, we examined its anatomical connectivity and physiological function across species.

1.7.1 Anatomical Connectivity across Species

Anatomically, the NRe of the midline thalamus occupies a strategic position within the known circuitry that underlies memory suppression. In rodents, medial prefrontal regions have been shown to interact with hippocampal dynamics (Jayachandran et al., 2019; Shapiro et al., 2017). However, these cortical areas do not project directly to the HpC (Vertes et al.,2007), raising the question of how top-down control is implemented. The NRe addresses this gap by serving as a key thalamic relay: it could enable the prefrontal cortex to exert indirect influence over hippocampal activity. Specifically, the NRe receives dense glutamatergic input from mPFC and sends robust projections to hippocampal subfields such as CA1 and the subiculum, positioning it as a functional intermediary (Dolleman-Van der Weel et al., 2019; Vertes et al., 2006). A population of NRe cells were also found to project to both the CA1 and mPFC (Vienna et al., 2020) and these dual-projecting cells also adds to the growing evidence that the NRe supports bidirectional communication and could regulate the flow of information from the mPFC to the HpC.

This bidirectional connectivity could enable the NRe to mediate both feedforward control by shaping how incoming cues are interpreted and modulating feedback via expressing or gating hippocampal output. Importantly, the NRe is not a passive relay. It integrates inputs from both executive and limbic structures, including the ACC and the amygdala (Vertes et al., 2006). This ability could enable the NRe to weigh task relevance against emotional salience in regulating memory expression. Such integration is important in emotionally charged contexts, such as fear learning or trauma, where NRe dysfunction impairs both memory specificity and the suppression of conditioned fear responses (Ramanathan et al. 2018; Xu & Südhof, 2013).

While rodent studies provide foundational insight, recent anatomical tracing work in non-human primates reveals species-specific adaptations that suggest a broader and more nuanced role for the NRe in memory regulation. In macaques, the core anatomical architecture of the NRe is conserved but incorporates important differences. Joyce et al.

(2022) demonstrated that the macaque NRe receives prominent input from area 25 (subgenual ACC/vmPFC), a region centrally involved in mood regulation, valuation, and introspective control. Thus, in primates, the NRe could support affect modulation in comparison to basic gating alone.

Unlike rodents, the primate NRe contains local parvalbumin-positive GABAergic interneurons, which could contribute to local inhibitory modulation of the HpC. Furthermore, anatomical tracing studies indicate that the NRe is not a homogeneous structure but contains functionally specialized subdomains. Some subdivisions preferentially connect with the dorsal HpC, involved in spatial and contextual memory, while others project to the ventral HpC, which supports emotional memory (Varela et al., 2014). Similarly, inputs from distinct mPFC regions, such as the ACC and orbitomedial PFC, map onto discrete territories within the NRe (Dolleman-Van der Weel & Witter, 2019).

Taken together, both rodent and primate findings converge on the view that the NRe is more than a passive conduit between the PFC and HpC. Through its integrative inputs, specialized inhibitory architecture, and topographic organization, the NRe emerges as a crucial potential modulator of memory accessibility by filtering, transforming, and distributing prefrontal control signals to enable adaptive, goal-directed modulation of memory.

1.7.2 Electrophysiological Dynamics and Oscillatory Coordination

Electrophysiological studies converge on the idea that the NRe plays a dynamic role in modulating oscillatory coherence between the PFC and the HpC, with distinct frequency bands emerging depending on cognitive demands.

During working memory tasks, specifically, during delay periods, the dHpC theta frequency oscillations are thought to send information about the previous trial to the mPFC to construct a behaviourally relevant plan resulting in theta-gamma coupling. When the NRe was inactivated, this coherence between the HpC and mPFC was disrupted, impairing performance, which suggests that NRe is necessary for sustaining ongoing network states during delay periods and exploratory behaviours (Hallock et al., 2016).

As discussed above, theta rhythms have been associated with HpC→PFC communication; the mechanisms underlying PFC→HpC communication are less understood due to the absence of direct anatomical projections. Roy et al., 2017, identified a distinct 2–5 Hz oscillation in the PFC that synchronizes with hippocampal activity via

the NRe. Simultaneous recordings in urethane-anesthetized rats revealed that this low-frequency coupling was present across PFC, NRe, and HpC, and was selectively disrupted when NRe was inactivated. Importantly, this effect was observed under urethane anesthesia, in the absence of overt behaviour, suggesting a circuit mechanism by which the NRe could support PFC→HpC communication through a slow oscillatory dynamic distinct from canonical theta-band synchrony.

Further, Jayachandran et al. (2023) provided direct evidence that NRe activity is also involved in enabling beta-frequency synchrony (~20 Hz) between PFC and HpC during goal-directed sequence memory retrieval. In a non-spatial sequence memory task in rodents, beta oscillations in PFC and HpC reliably followed NRe activation during retrieval epochs. Optogenetic stimulation of NRe at beta frequency enhanced PFC-HpC synchrony in the beta band and improved task performance, indicating that NRe can shift the network into a memory-focused state. During non-memory-related behaviours such as locomotion, theta oscillations predominated, and NRe activity was relatively low.

These findings suggest that NRe serves as a frequency-flexible coordinator of inter-regional synchrony. It may facilitate theta coupling during maintenance and exploration, and beta synchrony during controlled retrieval, enabling communication between the mPFC and the HpC to support different memory operations.

1.7.3 Behavioural and Lesion Evidence

The NRe has also emerged as a crucial node for temporal sequence memory, a core component of episodic memory that involves the ordered recall of events (Jayachandran et al., 2019). In an odor sequence task in rodents, item recognition was preserved, the animals could not recall the correct temporal order, indicating a disruption in sequential integration rather than content representation. Evidence suggests that the mPFC makes contributions to sequence memory and specifically, silencing mPFC-NRe synapses could repeatedly abolish memory sequence. This result determines the NRe's importance for maintaining the temporal continuity of memory episodes potentially by sustaining top-down inputs necessary for encoding or retrieving the sequence structure.

Rodent studies reveal that disruption of the NRe impairs tasks that require integration of executive and mnemonic information. Lesions to NRe result in deficits in spatial working memory and interference control. For example, rodents with NRe inactivation fail in dual alternation tasks requiring mPFC-HpC coordination (Hallock et

al., 2016), suggesting the PFC cannot influence hippocampal contextual signals without NRe mediation.

Along with spatial working memory, lesions or inactivation of the NRe impair performance in tasks requiring cognitive flexibility. Affected animals struggle to resolve competition between overlapping memory traces and exhibit spatial perseverative behaviours suggesting that the NRe supports PFC-mediated inhibition of irrelevant hippocampal content (Viena et al., 2018). Supporting this idea, Ito et al. (2015) showed that trajectory dependent firing was observed in the rat mPFC, NRe and CA1. Disrupting communication between the rat mPFC and NRe by lesioning the NRe led to reduced firing in the CA1 which impaired the HpC's ability to represent future goal-directed paths during spatial navigation.

Further, Xu et al. (2013) demonstrated that activation of mPFC projections to the NRe during learning enhanced the precision of contextual fear memories, whereas optogenetic inhibition of this pathway promoted fear generalization, suggesting a critical role for mPFC–NRe communication in context discrimination.

Beyond sequence memory and spatial working memory, evidence from the Warburton laboratory demonstrates that the NRe is essential for long-term associative recognition memory. Using both permanent lesions and temporary inactivation, Barker and Warburton (2018) showed that disrupting the NRe selectively impaired long-term object-in-place recognition memory, while sparing short-term performance as well as single-item and object-location memory. Complementing these findings, electrophysiological work has shown that coordinated activation of the NRe and HpC, specifically HpC activation followed by NRe activation induces NDMA-dependent synaptic plasticity in the mPFC (Banks, Warburton & Bashir, 2021), supporting the idea that the NRe contributes to hippocampal–prefrontal information integration.

Collectively, these findings illustrate that NRe is necessary not only for selecting relevant memories but also for potentially suppressing irrelevant or maladaptive ones, functioning as a cognitive arbitrator of PFC-HpC communication.

1.7.4 The role of the NRe in the Acquisition and Extinction of Conditioned Fear

In addition to regulating episodic and sequential memory, the NRe is also crucial for the modulating affective responses. Sierra et al. (2017) found that the formation of both, recent and remote fear memories became inaccessible on inhibiting the ACC. The

formation of these memories could become possible only when both, the ACC and NRe were reactivated. This finding suggests that the NRe could play a role in enabling memory access under specific regulatory conditions. Thus, the NRe could potentially support flexible modulation of retrieval, either inhibiting or facilitating access depending on task demands and prefrontal input.

Ramanathan et al. (2018, 2019) demonstrated that fear conditioning and extinction of contextual fear memories are both found to rely on the NRe. Inactivating the NRe or silencing mPFC projections to the NRe prevents both, the encoding and retrieval of the extinction memory. Further, they found that the NRe contributes hippocampal-dependent encoding of fear memories and is crucial to facilitating the discrimination of safe versus threatening contexts on fear retrieval. Pharmacological inactivation of the NRe induced fear relapse after extinction and was also found to impair c-Fos early gene expression in both, the mPFC and the HpC along with abolishing the HpC-mPFC theta synchrony, as discussed earlier (Totty et al., 2023).

Taken together, anatomical, electrophysiological, and behavioural findings converge to position the NRe as a central conduit for prefrontal regulation of hippocampal and limbic systems. Its capacity to modulate both memory specificity and affective salience supports the idea that the NRe participates in a domain-general control mechanism. This anticipates the retrieval stopping model of extinction, which reconceptualises extinction learning as a form of active memory suppression.

Chapter 2

Reconceptualising Fear Extinction: From Associative Learning to Memory Control

Advances in cognitive neuroscience have shown that intentionally suppressing unwanted memories engages prefrontal inhibitory control networks that modulate hippocampal retrieval. Building on this foundation, the retrieval stopping model of extinction (Anderson & Floresco, 2022) proposes that similar inhibitory mechanisms underlie fear extinction reframing it not just as new associative learning, but as a process of suppressing the originally learnt aversive memory traces. This chapter explores the convergence between memory suppression and extinction, highlighting shared neural pathways along with evaluating this reconceptualised narrative.

2.1. Introduction to Fear Extinction

Research concerning fear conditioning can be traced back to Ivan Pavlov (1927) whose studies established the basic principles of classical conditioning though his main experiments involved appetitive learning rather than fear. He demonstrated that a neutral cue, like a bell, could trigger salivation in dogs after the bell and the food were repeatedly paired with each other. Although Pavlov did not directly study fear, the foundational principles from his experiments informed the development of modern Pavlovian fear-conditioning paradigms. In these paradigms, a neutral stimulus (the conditioned stimulus or CS) becomes associated with a biologically significant neutral or negative stimulus (unconditioned stimulus or US) and results in a learned conditioned response (CR). Pavlov's experiments also demonstrated fear extinction of conditioned responding wherein the conditioned response could be diminished when the CS was repeatedly presented without the aversive US.

During extinction, the original memory is not erased, rather, a new inhibitory memory is formed which then competes with the original memory (Bouton et al., 2021). This model suggests that extinction involves the creation of a CS–no US association, which counteracts the original CS–US link. Thus, the original memory remains intact and can re-emerge under certain conditions for example, a change in context known as fear renewal, the mere passage of time known as spontaneous recovery or re-exposure to the unconditioned stimulus known as reinstatement. This retention of the original fear

association challenges older models like the Rescorla-Wagner theory (1972) which framed extinction as a form of unlearning.

There is neurobiological evidence supporting this CS–no US association model. Extinction engages brain regions which are typically involved in both, emotion and memory regulation. The ventromedial prefrontal cortex (vmPFC) plays a role in both, the learning and retrieval of extinction memories (Phelps et al., 2004) where it modulates the amygdala (Quirk & Mueller, 2008) which underlies emotion. The vmPFC might project to inhibitory interneurons within the basolateral amygdala (BLA). The hippocampus (HpC) provides contextual information (Ji & Maren, 2007) which helps to determine situations in which extinction learning should be applicable. Importantly, these extinction memories are fragile and can be susceptible to relapse.

However, a more recent complementary view has emerged which states that extinction may also involve executive control mechanisms like those used in suppressing unwanted memories (as discussed in the previous chapter). Just as individuals can suppress intrusive thoughts via the right dorsolateral prefrontal cortex (rDLPFC), extinction may similarly depend on the brain's capacity to actively suppress reactivation of fear memories. This perspective forms the basis of the retrieval stopping model (Anderson & Floresco, 2022) which proposes that the brain actively blocks retrieval of the original fear trace through top-down inhibitory mechanisms. Extinction can be seen as a process which does not merely involve new learning, but as a dynamic control process over memory retrieval itself through active suppression of pre-existing threat memories.

Gaining an understanding of whether extinction employs inhibitory learning becomes important in advancing emotional regulation strategies influencing the theoretical basis of exposure-based therapies which are commonly used to treat post-traumatic stress disorder (PTSD), phobias and obsessive-compulsive disorder (OCD). This chapter aims to explore this view by examining the neural, behavioural and theoretical overlap between memory suppression and fear extinction.

2.2 Experimental Paradigms

Experimental paradigms for investigating fear extinction have been developed across species. These have been widely employed and provide a translational bridge between basic neuroscience and clinical research on anxiety-related disorders.

2.2.1 Rodent and Non-human Primate Paradigms

In rodent models, extinction is typically examined using classical Pavlovian conditioning frameworks followed by extinction protocols. Conditioned fear responses are quantified using behavioural indices such as freezing (absence of voluntary movement), conditioned place avoidance, etc. (Fanselow, 1980; Meehan et al., 1994). Extinction is studied by presenting the CS+ repeatedly in the absence of the US, which tends to result in a graded reduction in the conditioned response. In rodents, pharmacological inactivation, excitotoxic lesions and optogenetic modulation have enabled precise temporal and spatial targeting of brain regions involved in conditioning and extinction. These regions are the infralimbic (IL) subdivision of the medial prefrontal cortex, the basolateral and central nuclei of the amygdala, the HpC (Quirk & Mueller, 2008) and midline thalamic structures such as the nucleus reuniens (NRe) (Moscarello, 2020; Ramanathan et al., 2018; Ratigan et al., 2023 Troyner et al., 2018; Totty et al., 2023). These approaches have shed light on the contributions of these regions in distinct phases of extinction learning and recall.

While extinction research has mostly been carried out in rodent models, fear conditioning has also been adapted for study in non-human primates. In primate models, conditioned fear is inferred via eye tracking, autonomic responses and/or facial expressions (Kalin et al., 2004; Prather et al., 2001). A marmoset study demonstrated that, like in rodents, the PFC, specifically the ventrolateral prefrontal cortex (vLPFC) and anterior orbitofrontal cortex (antOFC) were involved in fear learning and extinction (Agustín-Pavón et al., 2012). However, the use of primates in experimental paradigms is constrained by ethical considerations and the complexity of behavioural training, which limit the feasibility of such studies in these species.

2.2.2 Human Paradigms

In humans, fear extinction is studied using conditioning paradigms that are conceptually aligned with animal models but adapted for subjective evaluation. Participants are exposed to a CS+ paired with an aversive outcome (e.g., mild electric shock, aversive tones) and a CS- that is never paired. During extinction training, the CS+ is presented with partial reinforcement or without any reinforcement, to observe reductions in conditioned responses over time (Phelps et al., 2004).

These studies commonly employ psychophysiological and subjective measures to assess conditioning and extinction. The metrics used consist of skin conductance response (SCR), pupil dilation and startle reflex potentiation along with subjective fear ratings,

arousal ratings and expectancy ratings. A commonly tested protocol is the two-day extinction recall paradigm in which participants undergo fear acquisition and extinction on day 1 and delayed extinction recall on day 2. This design enables the study of consolidation and retrieval of extinction memory and its susceptibility to relapse effects such as spontaneous recovery or reinstatement (Phelps et al., 2004).

2.3 Neural Correlates of Fear Conditioning, Extinction, and Reinstatement

To evaluate whether extinction involves active memory suppression, we first outline the core neural circuitry supporting fear learning, extinction, and relapse. This section highlights the key brain systems implicated in these processes, exploring how top-down inhibitory mechanisms could shape the dynamics of extinction.

2.3.1 Fear Conditioning

During fear conditioning in humans, sensory and contextual information about the CS and US is distributed within a network of brain regions. Fullana et al (2016) performed a meta-analysis concerning human studies wherein they found that sensory information about CSs such as visual and auditory information is processed through primary sensory cortices and is relayed via thalamic nuclei particularly the mediodorsal, centromedial, and ventrolateral thalamic nuclei before engaging regions like the anterior insular cortex (AIC) and dorsal anterior cingulate cortex (dACC). Information about the US, typically aversive stimuli like electric shocks, is processed through parallel somatosensory and interoceptive pathways, with prominent involvement of hypothalamus, and brainstem nuclei including the periaqueductal gray (PAG).

This contextual information, encoded through these multimodal inputs, is processed by the HpC and vmPFC, and are consistently deactivated in response to the CS+ compared to the CS−. Fullana et al., suggest that these regions play a role in encoding safety signals or modulating anticipatory responses to threat.

According to Bouton et al. (2021) these CS and US inputs converge functionally within the basolateral amygdala (BLA). The BLA is thought to mediate associative learning between CS and US and project to the central nucleus of the amygdala (CeA), which drives conditioned fear responses via efferent projections to the PAG (mediating freezing behaviour) and to autonomic control regions such as the ventrolateral medulla, which modulate cardiovascular arousal.

2.3.2 Fear Extinction

Extinction learning involves the presentation of the CS in the absence of the US. Over time, this leads to a decrease in the conditioned response and a competing safety memory is formed (as thought of traditionally). The vmPFC plays a central role in this process. vmPFC activity during extinction recall is predictive of successful retention of the safety memory and has been linked to the attenuation of amygdala responses (Milad & Quirk, 2002). This pattern was demonstrated directly by Milad et al. (2007) who showed that during human extinction recall the vmPFC responds strongly to the extinguished CS+ and that this vmPFC activity correlates with behavioural extinction retention. They also observed hippocampal engagement reflecting contextual gating, along with strong vmPFC–hippocampal coupling, indicating that successful extinction retrieval depends on coordinated hippocampal support for vmPFC-mediated inhibition of amygdala outputs. Anatomical tracing studies have identified direct projections from the vmPFC to the amygdala (Hurley et al., 1991) with specific targeting of inhibitory regions such as the intercalated cell masses, which are thought to gate information flow within the amygdala (Vertes, 2004).

The HpC continues to contribute to extinction by signaling the contextual boundary conditions under which the CS should be interpreted as safe rather than threatening (Ji & Maren, 2007). Contextual modulation of extinction memory is believed to rely on hippocampal projections to both the vmPFC and amygdala; during extinction recall, the HpC may signal contextual match with the extinction environment, enabling the vmPFC to suppress amygdala-driven fear responses via feed-forward inhibition and thereby prevent context-inappropriate retrieval of fear memories (Marek et al., 2018; Milad & Quirk, 2012;). This contextual modulation is key to explaining phenomena such as renewal, wherein fear returns when the CS is encountered outside the original extinction context suggesting that the extinction memory is context specific.

Beyond the canonical HpC-vmPFC-amygdala circuit, increasing evidence points to the involvement of lateral prefrontal regions especially the rDLPFC and VLPFC during extinction. As discussed earlier, these regions are associated with higher-order executive functions, including inhibitory control and voluntary memory suppression. In a meta-analysis of human fear extinction studies conducted by Fullana et al. (2018), they found consistent activation of the right DLPFC, implicating it in top-down modulation. More recently, Rowlands et al. (2024) demonstrated a meta-analytic overlap between fear

extinction and memory suppression networks. They reported activity in the right aDLPFC, VLPFC and ACC and suggested that extinction and memory suppression engage domain-general inhibitory control mechanisms.

2.3.3 Fear Reinstatement

Despite successful extinction, conditioned fear can re-emerge, a phenomenon broadly referred to as relapse. As discussed earlier, three well-documented forms include renewal, reinstatement and spontaneous recovery. These forms of relapse demonstrate that extinction does not erase the original fear memory but instead overlays it with an inhibitory trace that under certain conditions, could be disrupted.

Fear reinstatement is associated with the reactivation of the dACC and amygdala, regions which are implicated in threat detection and fear expression. A multivariate neuroimaging study (Hennings et al., 2022) demonstrated that the dACC selectively reinstates neural patterns associated with the original fear memory, especially in the absence of vmPFC-mediated extinction signals. Along with the dACC, the anterior insula was also found to be implicated both being hubs of the salience network. This suggests that reinstatement may occur when the brain fails to access or apply the extinction memory, allowing the original fear trace to dominate behaviour.

Reinstatement is also modulated by the HpC, which helps determine whether the current context matches the extinction environment. In mismatched contexts, the HpC may fail to support retrieval of the extinction memory, leading to renewed fear expression. The posterior HpC has been specifically linked to reinstatement of fear memories, whereas the anterior HpC appears more involved in extinction recall (Hennings et al., 2022). This functional dissociation along the hippocampal long axis could be the reason why fear can return in certain contexts despite prior extinction learning.

While these canonical models offer a compelling framework in the understanding of the extinction of conditioned fear, they have certain limitations. The next section discusses these limitations and introduces the emerging perspective that extinction could be a form of cognitive control rather than the canonical model which assumes associative updating.

2.4 Limitations of Canonical Models

If extinction simply reflects a new CS-noUS association, why is it so easily overridden by changes in context, time, or emotional state? The idea that extinction is merely “new learning” competing with the original fear memory explains relapse to some extent, but

it does not fully capture the active, regulatory nature of extinction observed in behavioural and neural data.

Firstly, these models do not adequately account for the variability in extinction success across individuals. Some individuals, particularly those with anxiety or PTSD struggle to inhibit fear even after repeated extinction training (Milad & Quirk, 2012). These deficits are not due to impaired learning per se but seem to reflect difficulties in retrieving and expressing the extinction memory at the right moment. This pattern could resonate more with failures in cognitive control than a failure in associative updating.

Secondly, extinction has increasingly been linked to executive processes, including conflict detection, goal maintenance, and inhibitory control-functions traditionally studied in the context of memory regulation rather than associative learning. For instance, extinction recall has been associated with increased activity in the rDLPFC (Fullana et al., 2018), a region not typically implicated in associative learning models. As discussed previously, the rDLPFC is known to support top-down regulation of internal content, including the suppression of unwanted memories, thoughts, and emotions. Its recruitment during extinction suggests that extinction may engage inhibitory control mechanisms to actively suppress the fear memory, when the CS triggers retrieval of the aversive experience.

Thirdly, emerging theoretical perspectives suggest that extinction learning may be hindered by memory intrusions, the involuntary reactivation of the original fear memory despite attempts to encode a new, non-threatening association (Anderson & Floresco, 2022). While intrusions have been well-characterized in the context of memory suppression (Benoit et al., 2015; Levy & Anderson, 2012) their specific role during fear extinction remains underexplored. However, findings from reactivation–based reconsolidation paradigms indicate that the timing of memory reactivation can determine whether the original threat memory intrudes upon, and potentially interferes with, new learning. Extinction delivered within vs. outside the reconsolidation window produced dramatically different outcomes, with timely reactivation reducing the return of fear (Schiller et al., 2010). Together, these findings suggest that extinction can fail when the original threat memory intrudes during learning, compromising attempts to form a new safety association.

Lastly, classical extinction paradigms do not consider the timing and dynamics of neural mechanisms underlying these processes. Emerging electrophysiological evidence in rodents has found that prefrontal regions, particularly, the IL cortex can initiate

inhibitory responses within hundreds of milliseconds after the CS has been presented (Likhtik et al., 2014). This rapid engagement of the IL cortex suggests that extinction may involve both, rapid inhibitory control processes along with slower cumulative learning processes which are cumulative across repeated exposure. While classical associative models conceptualize extinction as an error-driven updating of outcome expectations (Rescorla & Wagner, 1972) recent findings from Likhtik et al., (2014) show that rapid prefrontal activity may flexibly modulate fear expression. Moreover, the effectiveness of extinction learning appears to depend not only on what is learned, but when it is learned, with studies showing that the timing of extinction relative to memory reactivation can significantly shape outcomes (Oyarzún et al., 2012). Although speculative, this opens the possibility that neural circuits may operate in both pre-emptive and reactive modes of fear regulation, a distinction yet to be fully explored in empirical research

Together, these limitations motivate a shift in perspective, captured by the retrieval stopping model of extinction, which reconceptualizes extinction as a form of active memory suppression. Complementing these theoretical concerns, empirical findings from human neuroimaging studies have further challenged the reliability of the canonical extinction model. Meta-analyses by Fullana et al. (2016, 2018) failed to identify consistent vmPFC or amygdala activation during extinction or its recall, raising questions about whether these regions operate as reliably in humans as they do in rodent models. Instead, these analyses reported reliable activation in regions associated with autonomic arousal and conflict monitoring, specifically in the dACC and anterior insula, even when the CS was not paired with the US during extinction. This could suggest that some commonly observed activity during extinction may reflect residual threat processing rather than successful regulation (Fullana et al., 2018; Visser et al., 2021). These inconsistencies further promote the need for a more flexible framework which can not only explain the neural markers of successful extinction, but also its variability. Such a framework may be better captured by models emphasizing dynamic inhibitory control, such as retrieval suppression.

Although Pavlovian fear conditioning paradigms in humans are modelled on those employed with rodents, several key methodological and psychological differences limit the extent to which findings can be directly compared across species. As emphasised by Lonsdorf et al. (2017) several factors complicate the translational interpretations of fear learning and extinction. Firstly, ethical constraints in human research restrict the intensity of aversive stimuli that can be employed. Whereas rodents typically receive footshocks

which are known to robustly engage defensive systems, human studies rely on milder and more variable unconditioned stimuli such as brief electrical stimulation, aversive sounds, or airpuffs. These weaker stimuli produce less reliable fear responses and greater inter-individual variability thus, reducing the robustness of conditioned responding compared to what is observed in animal studies. Second, humans engage explicit cognitive processes during fear learning that rodents might not have the ability to perform. Participants form conscious expectations about CS–US contingencies, and their fear responses are modulated by contingency awareness, instructions, and appraisal. As a result, human conditioning reflects an interaction of associative learning along with higher-order cognition, whereas rodent conditioning more closely entails a pure associative process. Third, the expression of conditioned fear is variable across species. Rodents display clear, quantifiable defensive behaviours such as freezing, whereas human fear responses are typically assessed using subjective ratings along with physiological indices such as skin conductance, pupil dilation, and/or startle potentiation. These measures index autonomic arousal or anticipation rather than overt defensive behaviour, and they are found to show greater intra-individual variability. Fourth, contextual processing is more tightly controlled in rodent studies than in human experiments. Rodent conditioning chambers differ markedly in lighting, odour, and general structure thus producing robust contextual modulation of fear. In contrast, human participants experience conditioning within a richer environment and more abstract cognitive context i.e., they know they are safe in the lab, which weakens contextual influences and makes renewal effects less reliable. Finally, humans often apply emotion-regulation strategies, intentionally or unintentionally during conditioning and extinction. Processes such as reappraisal, distraction, or anticipatory reasoning can dampen conditioned responses or mimic extinction-like reductions in fear, thereby complicating the interpretation of learning-based mechanisms. Together, these differences highlight that human fear conditioning cannot be assumed to operate identically to rodent models. Recognising these constraints is essential when drawing translational inferences about extinction, inhibitory learning, and their underlying neural mechanisms.

2.5 Clinical Relevance of Fear Extinction

Fear extinction is the core mechanism which is thought to underlie exposure-based therapies commonly employed as treatments for anxiety disorders including PTSD and phobias. These therapies are found to be effective but are prone to relapse under stress, in

novel contexts or after time has passed (Craske et al., 2014). This fragility of the extinction memory highlights a crucial limitation of the standard associative model.

Relatedly, impairments in extinction are a symptom of PTSD. Individuals suffering from PTSD were found to exhibit reduced vmPFC activity along with heightened amygdala responses during extinction which reflect persistent fear expression. During extinction learning and especially extinction recall, PTSD patients exhibit vmPFC hypoactivity and exaggerated amygdala activation, neural patterns that correspond to their difficulty recalling that previously threatening cues are now safe (Milad et al., 2009). These findings imply dysfunctional prefrontal control over fear expression. This could be the reason why trauma reminders evoke intense fear despite the therapeutic effort of promoting safety learning.

Currently, pharmacological agents like D-cycloserine along with neuromodulation techniques like transcranial magnetic stimulation (TMS) are commonly employed to enhance extinction (Fonzo et al., 2017; Norberg et al., 2008; Raij et al., 2018) and demonstrate mixed efficacy. These approaches are thought to conceptualize extinction as the inability to acquire and/or consolidate safety memories. However, if the retrieval stopping model of extinction holds true, this failure could in fact stem from impaired top-down inhibitory control over the retrieval of the fear memory. The reframing of extinction as active suppression may offer a deeper mechanistic account of treatment resistance and may open doors for alternative interventions.

2.6 The Retrieval Stopping Model of Fear Extinction: Theoretical Foundations

The retrieval stopping model of extinction (Anderson & Floresco, 2022) offers a novel reinterpretation of fear extinction. Rather than merely forming a competing inhibitory association (CS-noUS), this model suggests that extinction engages the same prefrontal inhibitory systems implicated in the suppression of unwanted episodic memories. Specifically, exposure to a CS+ may involuntarily trigger the retrieval of the aversive US representation. This retrieval acts as a mnemonic intrusion that provokes a fear response. To regulate the resulting aversive effect, the brain recruits top-down control processes that inhibit the reactivated memory trace, analogous to the suppression mechanisms engaged in TNT paradigms. Just as the rDLPFC suppresses hippocampal retrieval in TNT tasks (Anderson et al., 2004; Gagnepain et al., 2017) extinction may involve rDLPFC-mediated suppression of intrusive fear memories.

Supporting this framework, Fullana et al.'s meta-analysis of over 1,000 participants found consistent rDLPFC activity during extinction learning. This region is not typically highlighted in classical models. Likewise, Fonzo et al., (2017) demonstrated that higher activation in the rDLPFC during emotional reactivity tasks (e.g., viewing fearful faces) pre-treatment, was associated with better treatment outcomes following prolonged exposure therapy in individuals suffering from PTSD. Further, in a subset of their sample, they employed TMS and observed that stronger TMS-induced inhibition of the left amygdala by rDLPFC stimulation predicted greater symptom improvement post-treatment.

The model further suggests that hippocampal and amygdala deactivations observed during extinction reflect the successful suppression of reactivated fear memories. The retrieval stopping framework thus offers a mechanistic account of extinction that links memory control and emotional regulation, reframing it not merely as associative updating but as a dynamic interplay between memory retrieval, affective processing, and inhibition. The following sections will examine the evidence for this model across species and explore how it reshapes our understanding of extinction as a neurocognitive control process.

2.6.1 Empirical Evidence for Extinction as Memory Suppression

In this section, we examine the empirical support for the retrieval stopping model from neuroimaging studies in humans, circuit-level analyses in rodents, and behavioural paradigms that incorporate suppression-like instructions. Together, these findings yield some support for the view that extinction is not merely new learning, but an active, inhibitory process, potentially grounded in retrieval stopping mechanisms.

2.6.1.1 Lateral prefrontal involvement in extinction and suppression tasks

While extinction research has traditionally emphasized the role of the vmPFC, recent findings suggest that lateral prefrontal regions, particularly the rDLPFC contributes by supporting inhibitory control (Fullana et al., 2018; Rowlands, 2024). These areas are well-known from memory suppression paradigms such as TNT and yet have received little attention in extinction research. In an fMRI study, Delgado et al. (2008) demonstrated that emotional regulation trials were associated with increased activation in the DLPFC and vmPFC activation along with decreased amygdala activity. Based on these findings and correlational analysis, they concluded that the DLPFC could influence fear responses via influencing vmPFC modulation of the amygdala. As discussed earlier, Fullana et al.

(2018) reported the engagement of the DLPFC during extinction learning. Supporting this finding, Rowlands et al. (2024) provided further evidence based on their meta-analysis that extinction and retrieval suppression both recruit overlapping prefrontal control regions, particularly, the right aDLPFC and ACC. These regions are not typically considered central to the canonical extinction network, indicating the presence of higher-order executive mechanisms. Similarly, Jovanovic et al. (2013) found that reduced activation in the vmPFC during an inhibitory control Go/No-Go task was associated with impaired fear inhibition, as measured by physiological fear-potentiated startle responses, suggesting a shared inhibitory function across both, behavioural inhibition and fear extinction.

Further, amygdala downregulation when observed during extinction is well characterised, the interpretation of hippocampal suppression is not as straightforward. This stems from the dual role of the HpC; not only does it contribute to fear memory retrieval, but it is also known to encode contextual safety. Nevertheless, emerging evidence of overlapping activation patterns across prefrontal-limbic-hippocampal circuits suggests that extinction could recruit broad system-level inhibitory control mechanisms.

2.6.1.2 Circuit-level parallels between extinction and memory suppression

Rodent studies provide foundational evidence that extinction engages neural mechanisms consistent with retrieval suppression. Extinction learning in rodents depends critically on PFC regulation of hippocampal and amygdala circuits regions which mirror the neural substrates of memory control in humans. In particular, the IL cortex, considered a rodent homolog of the human vmPFC, has been shown to exert inhibitory control over the amygdala during extinction recall (Milad & Quirk, 2002) which is thought to be achieved by intercalated amygdala neurons which interact with the CeA and BLA, which mediate feed-forward inhibition of fear output pathways (Likhtik et al., 2008). IL activity during extinction training is associated with the consolidation of the extinction memory. Optogenetic activation of IL neurons projecting to the amygdala was found to enhance extinction retention, whereas silencing impairs it and led to increased freezing (Bukalo et al., 2015). These effects are thought to stem from IL-driven activation of BLA neurons which in turn recruit inhibitory intercalated cells, suppressing CeA output. Thus, extinction success is found to be tied to prefrontal-amygdala inhibitory control.

As discussed earlier, more recent work has also identified the NRe as a critical intermediary in this circuit, coordinating pre-frontal influence over the HpC via CA1-targeting projections. It is uniquely positioned to relay excitatory signals from the mPFC

to HpC projecting neurons in the NRe, creating an indirect pathway through which the mPFC can influence and potentially gate hippocampal processing and output (Vertes et al., 2007). Lesting et al. (2011) demonstrated that theta synchrony between the mPFC and hippocampal CA1 increased during fear memory retrieval, but it significantly decreased during extinction recall. This decoupling could reflect prefrontal suppression of the original fear memory. Further, Guise and Shapiro (2017) found that mPFC activity reduces interference by modifying encoding patterns in CA1 which increases memory specificity. Thus, the mPFC could potentially bias the HpC to gate certain memory traces. Ramanathan et al. (2018) offered further evidence by showing that optogenetic silencing of mPFC projections to the NRe disrupted extinction learning, while pharmacological inactivation of NRe impaired extinction memory retrieval, implicating a mPFC→NRe→HpC pathway. Malik et al. (2022) identified a direct long-range GABAergic projection from the mPFC to the dorsal HpC and proposed that this could inhibit local excitatory activity via direct prefrontal suppression. Totty et al. (2023) further demonstrated that extinction recall depends on synchronized theta oscillations between mPFC and HpC; inactivation of NRe impaired this synchrony and retrieval, while artificial theta stimulation of NRe enhanced extinction and reduced relapse.

These converging findings indicate that the mPFC–NRe–Amygdala–HpC circuit serves as an extinction-relevant suppression pathway in rodents, offering a plausible analogue to human retrieval stopping mechanisms.

2.6.1.3 Behavioural paradigms linking suppression to extinction outcomes

While neuroimaging and rodent studies have provided compelling correlational and mechanistic data, behavioural studies directly testing suppression in extinction remain limited. However, recent work has begun to fill this gap. Wang et al. (2021) adapted the TNT task to a fear extinction context. Participants were instructed to either actively suppress thoughts of the aversive US during presentations of the CS+, or to engage in mental diversion by imagining calming natural scenes. Notably, only the suppression condition led to a significant reduction of fear reinstatement. Additionally, this effect was pronounced in individuals with high scores on the Thought Control Ability Questionnaire (TCAQ), a measure of an individual's capacity for thought suppression. Moreover, the reduction in these fear responses also influenced related memory traces that had not undergone suppression which can be compared to the property of the cue-independence effect observed in TNT research.

Conversely, Hennings et al. (2021) reported that suppression instructions impaired extinction generalization, potentially by introducing conditioned inhibition effects. However, their design instructed participants to suppress the CS+ (rather than the memory of the US), which may have bypassed the retrieval process necessary to engage hippocampal inhibition. Thus, rather than reflecting a failure of suppression per se, the counterproductive results in this study could be due to task design, specifically, the type of suppression directed towards the CS+ along with timing of suppression instruction i.e., during extinction instead of during conditioning.

In a related paradigm, Chalkia et al. (2023) showed that directed forgetting, a procedure thought to involve inhibitory control, reduced associative memory and SCRs to fear-conditioned stimuli which were followed by the "forget" condition. These findings suggest that directed forgetting can attenuate physiological fear responses, consistent with the retrieval stopping hypothesis which involves inhibitory control.

Further support comes from a recent behavioural study by Rowlands (2024) which adapted the amnesic shadow paradigm to a fear extinction context. Participants were part of an extinction training protocol wherein bystander items were presented between CS+ and CS- trials. Memory for these bystander items was tested later on wherein, no overall amnesic shadow was found but individuals with greater forgetting for bystanders flanked by CS+ trials exhibited stronger extinction of affective responses, particularly in valence. This finding thus supports the retrieval stopping model by offering a link between the engagement of spontaneous suppression and extinction success.

Another comprehensive behavioural demonstration of suppression within an extinction context comes from Quaedflieg et al. (2025). Using a hybrid paradigm that embedded a modified TNT task into a fear extinction protocol, participants learned aversive object-scene associations paired with a CS+, followed by either suppression (No-Think) or passive extinction (View) of these associations. Suppression led to a reduction in the emotional intensity of aversive scenes, but it did not outperform standard extinction protocols in reducing US expectancy or subjective fear. This experiment lends some support to the retrieval stopping model by showing that suppression can influence conditioned fear.

These studies all vary in their aims, methods, and interpretations, they collectively suggest that inhibitory control can influence extinction outcomes. However, most behavioural studies have focused on instructed suppression. Whether spontaneous

memory suppression occurs during extinction trials, as the retrieval stopping model predicts, remains an open and testable hypothesis.

2.7 The Role of the Nucleus Reuniens in Fear Conditioning and Extinction

Traditionally studied in the context of spatial memory and executive function, the NRe is increasingly recognized as a critical node enabling communication between the mPFC and the HpC, a pathway essential for integrating contextual and emotional information in fear learning and fear extinction.

2.7.1 NRe in Fear Conditioning

Rodent studies have demonstrated that NRe is necessary for the contextual aspects of fear learning. In a 2013 study, Xu and Südhof used viral tracing and demonstrated that the mPFC sends direct projections to the NRe which in turn sends projections to the HpC. The HpC was found to project back to the mPFC. Importantly, inactivation of the NRe, particularly the mPFC→NRe pathway during conditioning led to the overgeneralization of contextual fear memories. Thus, the NRe is found to be responsible for maintaining the precision of contextual fear memories during acquisition.

Sierra at al.'s, 2017 study proposed that the mPFC communicates to the HpC via the ACC which has connections to the NRe which in turn projects to the CA1 region of the HpC via the CA3. They found that inactivating the ACC during fear acquisition impaired fear learning of both, recent and remote memories but this could be rescued during reconsolidation, which required the ACC and NRe to be re-activated together. Thus, the NRe was necessary for fear encoding. Ramanathan et al. (2018) inactivated the NRe with muscimol before conditioning and found impaired contextual freezing in rats. They suggested that the NRe supports hippocampal-dependent encoding of contextual fear memories which is necessary to discriminate safe from unsafe contexts.

Troyner et al.'s study in 2018 found that temporarily inactivating the NRe, immediately after conditioning, led to more generalized memories. This effect was observed at both, recent and remote time points. These memories were also found to be more resistant to extinction and disruption via reconsolidation. Immunohistochemistry enabled visualisation of Arc protein, a marker of synaptic activity which was found to show abnormal activation patterns in the NRe, HpC and mPFC. This further reinforces the mPFC-NRe-HpC pathway's role in fear conditioning and reconsolidation.

2.7.2 NRe in Fear Extinction

Along with encoding, the NRe is also found to play a role in extinction learning, a process which requires the modulation of fear responses when the CS is no longer predictive of threat. A study by Ramanathan et al. (2018) demonstrated that inactivating the NRe prior to extinction training was found impair both, the acquisition as well as the retrieval of extinction memories. Disrupting the mPFC→NRe projection led to contextual freezing which indicated a failure to encode extinction memories. A second study by this group found that the NRe is necessary not only for encoding these memories but also for retrieving precise, hippocampal-dependent extinction memories.

Moscarello (2020) trained rats in a two-way active avoidance paradigm, wherein they reduced freezing as they learnt how to avoid an aversive outcome. The study found that optogenetic inhibition of mPFC projections to the NRe reinstated freezing behaviour. It was concluded that this circuit is essential for suppressing conditioned defensive responses once avoidance behaviour has been learnt. This is different from classical conditioning, but it provides evidence that the NRe can gate fear expression.

Totty et al. (2023) showed that extinction is accompanied by increased theta-band synchrony between the mPFC and HpC which is mediated by the NRe. Inactivation of the NRe reduced c-Fos activation in both regions, disrupted theta coherence, and impaired the retrieval of extinction memories. Conversely, optogenetic stimulation of the NRe enhanced extinction and reduced relapse.

Further, Ratigan et al. (2023) used calcium imaging and identified a role for the NRe→CA1 pathway in contextual fear regulation. NRe axons projecting to the CA1 became tuned to freezing behaviour after conditioning. Chemogenetic inhibition of this pathway led to increased fear generalization, prolonged freezing, and delayed extinction. These findings suggest that the NRe can regulate fear responses during both, extinction learning and extinction retrieval.

2.7.3 A Circuit for Fear Memory Regulation via the NRe

Together, all the above findings suggest that NRe serves a dual function in fear-related learning: (1) during conditioning, it helps bind contextual features to fear associations via mPFC-HpC co-ordination, and (2) during extinction, it enables the mPFC to potentially inhibit hippocampal memory reactivation and fear expression. Its ability to modulate hippocampal activity via GABAergic interneurons (Dolleman-Van der Weel & Witter,

2019) makes it a likely neural substrate for retrieval stopping: a process potentially essential for both memory suppression and fear extinction.

In summary, the NRe seems to act as an integratory hub of cognitive signals, dynamically routing information between brain regions to flexibly control fear memories depending on task demands. Its role in extinction might not just be to enable learning of safety signals, but to suppress maladaptive retrieval of conditioned fear, making it an important target for both mechanistic research and potential therapeutic interventions.

2.8 Clinical and Translational Implications

If memory suppression mechanisms are indeed recruited during fear extinction, this reconceptualization could have far-reaching clinical consequences. Techniques that enhance inhibitory control such as high frequency rTMS over the rDLPFC, pharmacological modulation of GABAergic tone, or cognitive interventions that explicitly train suppression strategies may improve extinction outcomes and reduce relapse.

Variability in the suppression capacity of individuals could account for why exposure therapy is more effective for some patients than others. It could be that patients suffering from PTSD who exhibit impaired hippocampal downregulation or greater activity in the amygdala may not be able to engage in effective extinction. Identifying neuroimaging markers of impaired prefrontal control and training individuals in suppression-based techniques could potentially lead to more efficient and individually tailored treatments.

These insights tap on the translational value of identifying whether the retrieval stopping model of fear extinction is at play along with then identifying the precise neural pathyways, potentially involving the NRe. Further, transdiagnostic research could determine whether deficits in retrieval suppression represent a shared mechanism characterised by intrusive cognition across disorders like PTSD, OCD, anxiety and depression.

2.9 Critical Appraisal of the Retrieval Stopping Model and Future Directions

The retrieval stopping model of fear extinction offers a compelling bridge between memory control and fear extinction. However, this model has some limitations which are discussed in this section.

The biggest challenge lies in distinguishing between active suppression and classical inhibitory learning. Traditional models are based on competition between co-existing associative traces whereas suppression relies on a more direct inhibition of mnemonic representations. Disentangling these mechanisms is difficult in standard conventional extinction paradigms since both mechanisms lead to a reduction in fear expression.

To rigorously validate the retrieval stopping model, future research must test its mechanistic predictions. For instance, causal interventions targeting rDLPFC-NRe-HpC circuitry via non-invasive brain stimulation, pharmacological modulation, or DCM could reveal whether suppression-specific pathways are engaged during extinction. Additionally, modifying behavioural paradigms to experimentally dissociate proactive and reactive control such as through time-locked suppression cues, intrusion monitoring, and physiological timing measures will be essential for demonstrating that extinction recruits dynamic inhibitory processes, rather than merely encoding new associations.

Another open question concerns the intentionality of suppression. Is fear attenuation during extinction always dependent on deliberate control, or can suppression mechanisms operate incidentally? In TNT literature, suppression is typically cued explicitly, whereas extinction training often unfolds without conscious effort to suppress. If spontaneous suppression contributes to extinction, this would mean that the retrieval stopping model could generalize to more naturalistic, non-instructed forms of regulation which will be a critical factor for translational relevance.

However, these challenges also represent opportunities for empirical progress. As delineated in the following chapters, our two-day fMRI study addresses many of these gaps by investigating whether fear extinction and memory suppression engage common inhibitory pathways, including the mPFC–NRe–HpC circuit, within the same subjects. This approach leverages carefully validated cross-species anatomical constraints to identify the human NRe, allowing us to test the retrieval stopping model within the limits of conventional-resolution fMRI. By directly measuring both behavioural suppression effects and underlying neural dynamics, the study aims to provide the first human evidence for a unified, mechanistically specific model of inhibitory regulation across domains.

2.10 Thesis Overview and Structure

This thesis investigates the neural mechanisms of memory suppression, with a particular focus on the exact mechanism via which the brain suppresses the retrieval of unwanted or aversive memories. Key to this investigation is to test the role of the nucleus reuniens (NRe) of the thalamus hypothesized to facilitate communication between the medial prefrontal cortex (mPFC) and the hippocampus (HpC).

Drawing on recent reconceptualizations of extinction and memory suppression as interrelated inhibitory processes, it explores the hypothesis that extinction is not merely a form of new associative learning but also an act of cognitive control over memory.

This thesis addresses **two overarching aims**, through four complementary empirical chapters:

Aim 1: To test whether the NRe plays role in during retrieval suppression

Chapter 3 addresses this aim using a mega-analytic fMRI dataset from the Think/No-Think (TNT) paradigm. This chapter tests whether the NRe is activated during suppression attempts and whether it participates in functionally specific circuits with the dorsolateral prefrontal cortex (rDLPFC) and the HpC. It distinguishes between direct retrieval suppression and reactive control (to counteract intrusions) and provides the first large-scale human evidence implicating the NRe in memory suppression.

Chapter 4 deepens the investigation of the NRe's role by examining its anatomical and functional properties across species. Drawing on rodent and primate data, it reviews the cytoarchitecture and connectivity the NRe and translates these findings to the human brain using an anatomically informed segmentation pipeline. This chapter establishes the anatomical foundation for interpreting the functional data in Chapters 5 and 6 and to validate the plausibility of NRe-mediated memory control and fear extinction respectively, in humans.

Aim 2: To test whether memory suppression and fear extinction rely on shared neural mechanisms.

Chapter 5 and 6 address this aim through a novel within-subject fMRI study. Participants completed both a TNT task and a fear conditioning and extinction task, allowing for a within-subject comparison of neural activity across tasks. This chapter examines whether the NRe is engaged during both memory suppression and fear extinction and whether its connectivity with rDLPFC and HpC reflects a shared inhibitory control mechanism. By

testing for across tasks within the same individuals, it aims to provide evidence for domain-general suppression processes.

Chapter 3

The Role of the Nucleus Reuniens during Retrieval Suppression: A Mega-Analytic Study

As discussed in the previous chapters, the nucleus reuniens (NRe) of the thalamus has been widely implicated in both, rodent and primate models as a critical structure facilitating communication between the medial prefrontal cortex (mPFC) and the hippocampus (HpC), particularly in relation to memory regulation. Additionally, evidence suggests that the NRe also interacts with the amygdala, facilitating pre-fronto-limbic integration. However, despite being well studied in animal models, the NRe remains poorly understood in the human brain. This is mainly because of its very small size and location in the deep midline of the thalamus along with being closely surrounded by neighbouring thalamic nuclei and white matter structures. In standard human brain atlases, the NRe is typically not represented or grouped with other thalamic nuclei. The NRe is difficult to to segment using conventionally used neuroimaging tools and this is particularly the case in datasets of lower resolution.

To address this gap, this chapter leverages bilateral NRe regions of interest (ROIs) provided by our collaborators in Prof. Zikopoulos' lab, anatomically delineated based on postmortem primate histology and localized within MNI152 space. These ROIs were applied consistently across all participants in a large, multi-site neuroimaging dataset. This dataset, the Mega-TNT sample was compiled by Dr. Dace Apšvalka and integrates data from ten previously published Think/No-Think (TNT) functional MRI studies. This well powered dataset enables the study of the NRe across a diverse range of participants engaged in the voluntary suppression of memory.

The analysis in this chapter aims to determine whether the NRe is recruited during memory control, and whether its activation patterns and connectivity reflect prefrontal mechanisms that support the suppression of unwanted memories. Specifically, it examines whether the NRe is engaged to prevent retrieval in response to suppression cues as well as in reactive control, engaged to counteract intruding memories once retrieval has been initiated. By providing the first large-scale evidence for NRe involvement in human memory suppression, this work bridges rodent and primate preclinical findings

with human neuroimaging evidence, establishes a novel thalamic node in the memory control network, and opens new directions for translational research on intrusive memory.

3.1 Hypotheses

To investigate the neural mechanisms underlying retrieval suppression during the TNT task, we focused on the interaction between the right dorsolateral prefrontal cortex (rDLPFC), the NRe of the thalamus, and the HpC. Building on rodent and primate findings, we hypothesized that the NRe provides a relay pathway through which the prefrontal cortex inhibits hippocampal-mediated retrieval. To test this model, we formulated four hypotheses targeting distinct aspects of this system: local activation of the NRe, its responsiveness to memory intrusions, and its functional connectivity with both the rDLPFC and the HpC. Together, these hypotheses provide a mechanistic test of the proposed rDLPFC-NRe-HpC suppression circuit.

3.1.1 Retrieval Stopping Hypothesis

We hypothesize that the NRe shows increased activation during retrieval suppression. Specifically, we predict significantly greater univariate activation in the NRe during No-Think (NT) trials compared to Think (T) trials. This reflects the idea that the NRe supports inhibitory control by influencing hippocampal activity when memory retrieval must be prevented.

Support for this hypothesis would suggest that the NRe plays a general role in stopping retrieval, independent of whether or not an intrusion occurs on a given trial.

3.1.2 Reactive Control Hypothesis

We hypothesize that the NRe shows increased activation during memory intrusions. Specifically, we predict greater univariate activation during Intrusion (I) trials compared to Non-Intrusion (NI) trials within the NT condition. This hypothesis targets the role of the NRe in reactive control; that is, suppressing memory after an unwanted trace has already entered awareness.

Importantly, this hypothesis is conceptually distinct from the Retrieval Stopping Hypothesis. The NRe may respond primarily to the general demand to stop retrieval without being sensitive to intrusions. Conversely, the NRe might be recruited specifically when intrusions happen, reflecting a more reactive role in resolving memory conflict. Thus, both hypotheses could be supported, only one might be true, or neither might hold; each reflects a different control mode within the broader memory suppression framework.

3.1.3 Prefrontal-Thalamic Communication Hypothesis

We hypothesize that functional connectivity between the rDLPFC and the NRe is enhanced during retrieval suppression. Specifically, we predict that Psychophysiological Interaction (PPI) analyses will reveal greater task-dependent connectivity between these regions during NT trials compared to T trials.

This hypothesis reflects the idea that retrieval suppression requires increased communication between prefrontal executive regions and the thalamic relay system. Strengthened rDLPFC-NRe connectivity during NT trials would suggest that prefrontal areas are dynamically recruiting the NRe to support top-down memory control.

3.1.4 Thalamo-Hippocampal Modulation Hypothesis

We hypothesize that functional connectivity between the NRe and the HpC is reduced during retrieval suppression. Specifically, we expect more negative connectivity values between the NRe and HpC during NT trials relative to T trials, as measured by PPI.

This hypothesis reflects the idea that the NRe supports memory suppression by dampening hippocampal activity potentially via projections to inhibitory interneurons in the CA1. Reduced connectivity between the NRe and HpC would support the idea that the NRe enables the implementation of top-down inhibition by disrupting the HpC.

Together, these four hypotheses provide a mechanistic test of the hypothesized suppression circuit consisting of prefrontal, thalamic, and hippocampal nodes. Each hypothesis isolates a specific functional link within this network, enabling us to evaluate various control processes, as well as their underlying neural pathways.

3.2 Methods

3.2.1 Dataset Description

As mentioned, the Mega-TNT dataset was compiled by Dr. Dace Apšvalka, integrating data from ten previously published TNT functional MRI studies (total n = 330). These studies were conducted across multiple years and laboratories, each with different experimental aims, scanner models, acquisition protocols (TRs, TEs, number of runs, voxel sizes), and stimulus materials (words, images, and both negatively and neutrally valenced stimuli). A subset of five studies within the Mega-TNT dataset included trial-wise intrusion ratings (n = 103), enabling us to test hypotheses related specifically to reactive memory control. A detailed breakdown of the individual datasets and their parameters are provided in Table 1 below.

Dataset #	Sample size (n)	No. of TNT runs	Stimulus content	Stimulus valence	Intrusion ratings	TR	TE
1	18	6	word-word	neutral	Y	2.0	0.025
2	18	4	word-word	neutral	Y	2.0	0.025
3	16	5	word-picture	faces and places	Y	2.0	0.03
4	24	5	face-scenes	negative and neutral	Y	2.0	0.03
5	27	5	word-word	neutral	Y	1.12	0.03
6	18	6	word-word	neutral	N	2.0	0.03
7	24	4	word-object	neutral	N	2.0	0.03
8	24	8	word-word	neutral	N	2.0	0.03
9	24	6	word-word	neutral	N	1.5	0.029
10	137	6	word-word	neutral	N	2.0	0.03

Table 1. Characteristics of the 10 datasets included in the Mega-TNT study, detailing scanner parameters, stimulus content, and inclusion of intrusion ratings. All datasets contributed to the analysis of retrieval suppression; a subset of five datasets (total n = 103) included trial-wise intrusion ratings used for testing reactive control hypotheses.

3.2.2 Region of Interest (ROI) Definitions

3.2.2.1 Left and Right NRe

The NRe ROIs were anatomically delineated by Prof. Zikopoulos' lab using a structural image normalized to Montreal Neurological Institute (MNI) space. These ROIs were defined according to known anatomical landmarks surrounding the midline thalamus and applied consistently across all datasets included in the mega-analysis. These served as the basis for all hypothesis-driven analyses involving the NRe.

Although the NRe is a midline structure, we analyzed the left and right ROIs separately. We had no strong a priori prediction about whether suppression-related activity would differ by hemisphere. While most previous research on the NRe has been conducted in rodents and non-human primates-species that show some evidence of lateralization in memory-related processes, particularly in the HpC and PFC, direct evidence for hemispheric specialization in the NRe remains limited. In contrast, humans often exhibit robust hemispheric specialization across cognitive domains, including

memory control. Notably, retrieval suppression has been consistently associated with greater engagement of the rDLPFC compared to the left (Anderson et al., 2004). Therefore, we considered potential hemispheric differences by examining the left and right NRe separately.

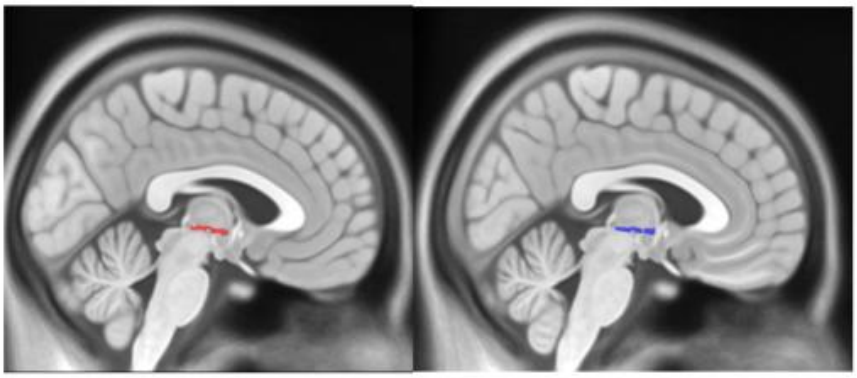

Figure 2. Sagittal sections of the human brain.

Left and right NRe regions of interest (ROIs) are marked in red and blue, respectively, as delineated by Prof. Zikopoulos' lab. ROIs are overlaid on an MNI structural template, with sections spaced 1 mm apart around the thalamus.

3.2.2.2 Right Dorsolateral Prefrontal Cortex (rDLPFC)

The rDLPFC ROI was functionally defined based on a meta-analytic conjunction analysis combining the TNT and Stop-signal tasks as reported in Apšvalka et al. (2022). The rDLPFC mask corresponded to the anterior portion of Brodmann areas 9, 10, and 46, derived from the ALE conjunction map in MNI space. Its role in domain-general inhibitory control was validated through multivoxel pattern analysis (MVPA) and dynamic causal modeling (DCM).

3.2.2.3 Bilateral Hippocampus

The hippocampal ROI was created from the full Mega-TNT dataset (n = 330), which included manually traced hippocampal masks for each participant. These masks were defined in native space and normalized to 1×1×1 mm MNI space. The binarized masks were then summed across participants, divided by the total number of subjects, and multiplied by 100 to generate a voxel-wise percentage overlap map. The final bilateral HpC mask retained only voxels identified as hippocampus in at least 20% of participants (i.e., ≥66 individuals), ensuring both anatomical precision and generalizability across subjects.

3.2.3 Preprocessing

All raw imaging data were pre-processed and analyzed using Statistical Parametric Mapping (SPM12; Wellcome Trust Centre for Neuroimaging, London), implemented in MATLAB vR2018a (The MathWorks, MA, USA). To approximate the orientation of the standard Montreal Neurological Institute (MNI) coordinate space, each participant's functional scans were manually realigned to the anterior-posterior commissure (AC-PC) line, with the origin set at the anterior commissure.

Standard pre-processing procedures were then applied. First, images were realigned to the first functional volume in each series to correct for head motion. Slice timing correction was performed relative to the midpoint of the repetition time (TR/2), to adjust for temporal differences across slice acquisitions. Each participant's mean functional image was co-registered to their anatomical (T1-weighted) scan, which was then segmented into grey matter, white matter, and cerebrospinal fluid.

Segmented tissue maps were submitted to the DARTEL procedure (Ashburner, 2007) to create dataset-specific anatomical templates and participant-specific flow fields. This iterative approach alternates between template creation and individual warping to optimize spatial normalization. The final group template was normalized into MNI space. Spatial smoothing was applied during this stage, using a Gaussian kernel of 4 × 4 × 4 mm.

3.2.4 First-level analysis

The pre-processed functional data were then entered into a first-level general linear model (GLM) for each participant. For each TNT run, regressors were included for T and NT trials, and where available for I and NI trials. Additional regressors were included for residual conditions of no interest, such as filler items, incorrect trials, and, where applicable, intrusion rating periods. Six motion parameters derived from the realignment step were also included as nuisance regressors.

This GLM yielded beta weights for each modelled condition at the individual level. Condition-specific contrasts (NT > T) were computed for each participant and each TNT run. In datasets that included trial-wise intrusion ratings, NT trials were further divided into I and NI conditions, allowing contrasts such as I > NI to be estimated. Because the 10 datasets varied in acquisition parameters (TR, TE, number of runs), analysis scripts were adapted to accommodate each dataset's specific characteristics.

3.2.5 Second-level analysis

Contrast images from the first-level (within-subject) analyses were submitted to second-level (group-level) analyses for each dataset. A repeated-measures ANOVA was conducted, treating subject as a random effect and condition (NT vs. T) as a within-subject factor with two levels. Planned t-contrasts were computed to examine condition-specific effects, including NT > T for all datasets, and I > NI for the subset that included intrusion ratings.

To account for variability across studies in the mega-analysis, a covariate coding dataset membership was included in all models. Analyses were performed in SPM12, constrained by a standard intracranial volume (ICV) brain mask to restrict comparisons to brain tissue.

3.2.6 ROI Analysis of the NRe

To evaluate the role of the NRe in memory suppression, we conducted hypothesis-driven ROI analyses focusing on univariate activation within the anatomically defined NRe masks. These analyses were conducted to test two core hypotheses: the Retrieval Stopping Hypothesis and the Reactive Control Hypothesis.

To test the Retrieval Stopping Hypothesis (Hypothesis 1), we examined the NRe to check if it exhibited an increased activation during NT trials compared to T trials. Greater activation in the NT > T contrast suggests that the NRe is engaged as a broader effort to prevent memory retrieval, irrespective of whether an intrusion occurs.

To test the Reactive Control Hypothesis (Hypothesis 2), we examined whether the NRe was more active during memory intrusions as compared to instances in which no intrusions occurred, and the memory was successfully suppressed. Specifically, we tested for greater activation during I trials relative to NI trials within the NT condition. Evidence for this contrast would support the idea that the NRe is also recruited reactively, to help suppress retrieval after an unwanted memory has entered awareness.

ROI analyses were conducted using the MarsBaR toolbox (version 0.44; https://marsbar-toolbox.github.io/). Mean parameter estimates were extracted from individually normalized data using anatomically defined left and right NRe masks in MNI space. While we had no a priori prediction regarding hemispheric differences, we analyzed left and right NRe separately to accommodate potential lateralization, and we also report results from the combined bilateral NRe mask. Analyses were conducted at the second level using repeated-measures t-tests for each contrast of interest.

3.2.7 Psychophysiological Interaction (PPI) Analysis

To examine task-dependent changes in functional connectivity between key nodes of the proposed suppression network, we conducted psychophysiological interaction (PPI) analyses in SPM12 (Wellcome Trust Centre for Neuroimaging, London), implemented in MATLAB vR2018a. These analyses were conducted to test two primary hypotheses concerning functional connectivity: the prefrontal-thalamic communication hypothesis (Hypothesis 3) and the thalamo-hippocampal modulation hypothesis (Hypothesis 4).

We followed a generalized seed-based PPI approach. For each participant, the de-convolved time series was extracted from a seed ROI which was predefined. The interaction term was computed as the product of this seed based physiological signal and the psychological contrast of interest (NT>T). This interaction term, along with the main effects of the seed time series and psychological regressor, was entered into a first-level general linear model (GLM) to predict the time series in the target ROI, consistent with an ROI-ROI PPI approach. Second-level models were then used to identify group-level effects of task-dependent connectivity.

Seed regions included the rDLPFC and HpC, as well as anatomically defined left, right, and combined NRe ROIs. As discussed earlier, the NRe is a midline structure and prior work has not provided conclusive evidence for hemispheric lateralization, primary analyses focused on the combined bilateral NRe mask to assess overall engagement. Exploratory analyses then examined the left and right NRe separately to probe for potential hemispheric differences.

Although some early studies suggested that inhibitory control may preferentially modulate the right HpC, more recent evidence does not robustly support this lateralization. Nevertheless, for consistency with prior PPI studies of memory suppression (Benoit & Anderson, 2012) we used the right HpC as the seed region in analyses involving NRe-HpC interactions. Supplementary analyses confirmed that results were qualitatively similar when the left HpC was used.

The Mega-TNT dataset comprises studies conducted across multiple laboratories using different scanners, stimulus materials, and acquisition parameters. To account for inter-study variability, dataset membership was included as a covariate in all second-level models. Statistical maps were corrected for multiple comparisons using cluster-level family-wise error (FWE) correction within an ICV mask.

In the ROI activation analyses, we modelled I and NI trials separately to examine how subjective differences in suppression success influenced regional activity. However, for the PPI analyses, all NT trials were modelled together, regardless of intrusion status. This decision reflected the primary aim of the PPI analyses, to characterize general task-dependent changes in connectivity associated with suppression demands rather than trial-by-trial variability. Subdividing trials by intrusion status would have substantially reduced trial numbers per condition, particularly given that intrusion rates varied across participants and could also differ across studies depending on the nature of the stimuli. Such reductions in trial numbers would compromise the reliability and interpretability of connectivity estimates.

3.3 Results

We first examined evidence for the Retrieval Stopping Hypothesis and the Reactive Control Hypothesis, both of which concerned univariate activation in the NRe. These hypotheses tested whether the NRe was involved in suppressing memory retrieval during NT trials and whether it was engaged reactively during memory intrusions. We then turned to functional connectivity analyses to evaluate the Prefrontal-Thalamic Communication Hypothesis and the Thalamo-Hippocampal Modulation Hypothesis, which assessed interactions among the rDLPFC, NRe, and HpC during suppression. Bonferroni correction was applied across all three ROI comparisons: Left, Right, and Whole NRe as each was analyzed and interpreted independently.

3.3.1 Retrieval Stopping Engages the Left and Right Nucleus Reuniens

The Retrieval Stopping Hypothesis posited that the NRe plays a role in suppressing memory retrieval, which would be reflected in increased univariate activation during NT relative to T trials. To examine this, we tested NT > T contrasts separately, in the left and right NRe.

As shown in Table 2 and Figure 3, both ROIs exhibited significantly greater activation during retrieval suppression, in accordance with the hypothesis. The right NRe showed a slightly larger mean beta difference, the left NRe yielded a higher t-value, reflecting lower variability across participants. The effect sizes were modest, d = 0.21 for the left NRe and 0.19 for the right, indicating small but consistent effects.

To complement the frequentist analyses, we conducted Bayesian paired-samples t-tests using directional priors consistent with the hypothesis. These analyses provided very strong evidence for the predicted effect in the left NRe (BF_{+0} = 182) and decisive

evidence in the right NRe (BF_{+0} = 681), indicating that the observed differences were far more likely under a model positing increased activation during retrieval suppression than under the null hypothesis. Together, these findings offer convergent support for the Retrieval Stopping Hypothesis and indicate robust bilateral engagement of the NRe during intentional suppression of memory retrieval.

Dataset	Left NRe			Right NRe		
	T-value	**p(unc.)**	**p(corr.)**	**T-value**	**p(unc.)**	**p(corr.)**
Mega-TNT	3.8942	**0.0001**	**0.0001**	3.4034	**0.0004**	**0.0007**

Table 2. T-values obtained from the NT > T contrast in the NRe from a Mega-TNT analysis across all datasets.

Note. p-values are Bonferroni-corrected to account for analysis of both left and right NRe ROIs.

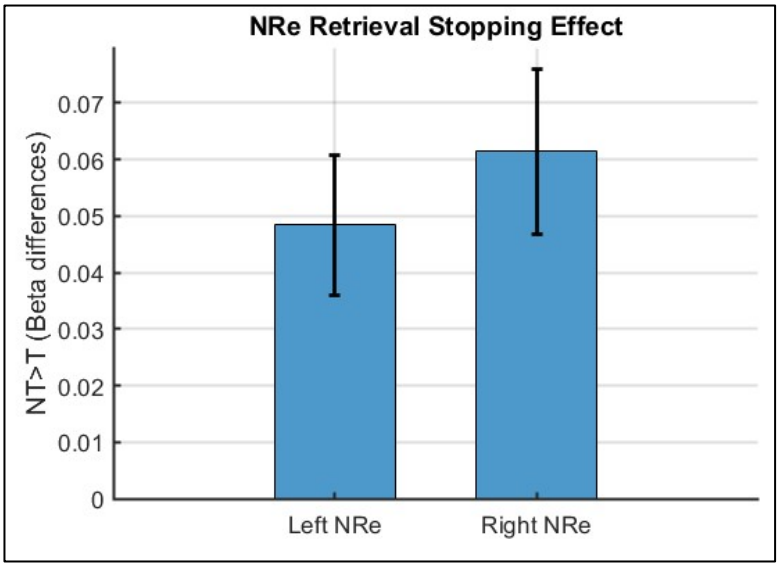

Figure 3. Mean beta differences (No-Think > Think) for the left and right NRe ROIs.

Both ROIs showed greater activation during retrieval suppression. Although the right NRe showed a slightly larger mean beta difference, the left NRe had a higher T-value, reflecting lower variability across participants.

3.3.2 Limited Evidence for Reactive Engagement of the NRe

The Reactive Control Hypothesis predicted that the NRe would show increased activation during trials in which unwanted memories intruded into awareness, reflecting its proposed role in reactive suppression once retrieval has begun. We tested this by comparing univariate activation between I and NI trials within the NT condition.

As shown in Table 3 and Figure 4, neither the left nor right NRe exhibited significant differences in activation between conditions. However, the left NRe showed a numerical increase in activation during intrusion trials, while the right NRe showed no

difference. Effect sizes were small, with d = 0.09 for the left NRe and approximately zero for the right.

Bayesian paired-samples t-tests further clarified the strength of evidence for these effects. The left NRe yielded anecdotal evidence for the null hypothesis (BF_{+0} = 0.48), indicating that the data were slightly more consistent with no intrusion-related modulation than with the predicted increase in activation. The right NRe showed moderate evidence for the null (BF_{+0} = 0.10), suggesting that activation was approximately ten times more likely to reflect no I > NI difference than to support the reactive control prediction. Robustness checks confirmed that these conclusions were stable across prior widths.

Taken together, these findings indicate no reliable evidence for increased NRe activation during memory intrusions in the current dataset and thus do not support the Reactive Control Hypothesis in this specific sample. Both frequentist and Bayesian analyses suggest that, for this dataset, the NRe was not strongly or consistently engaged in reactive control once retrieval had begun.

Dataset	Left NRe (I>NI)			Right NRe (I>NI)		
	T-value	p(unc.)	p(corr.)	T-value	p(unc.)	p(corr.)
Mega-TNT	1.5850	**0.0580**	0.1163	-0.0051	0.5020	1.0040

Table 3. T-values obtained from the comparison of Intrusion (I) vs. Non-Intrusion (NI) trials within the NRe collapsed across datasets in the Mega-TNT set that included intrusion ratings.

Note. Corrected p-values account for the inclusion of both left and right NRe ROIs.

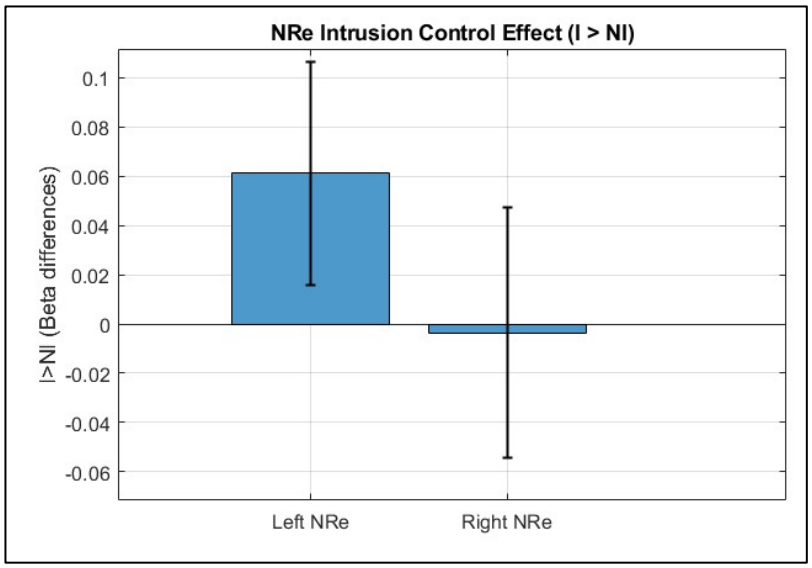

Figure 4. Mean beta differences (Intrusion > Non-Intrusion) for the left and right NRe ROIs. The left NRe showed numerically greater activation during intrusion trials compared to non-intrusion trials,

although this difference was not statistically significant. The right NRe showed no difference between conditions.

3.3.3 Right DLPFC exhibits greater connectivity with the NRe during retrieval suppression

The Prefrontal-Thalamic Communication Hypothesis predicted increased functional connectivity between the rDLPFC and the NRe during retrieval suppression. This is consistent with a top-down control mechanism mediated via the thalamus. We conducted PPI analyses to test this employing the rDLPFC as the seed region.

As shown in Table 4, connectivity with the bilateral NRe ROI showed a trend toward greater coupling during NT trials. This pattern was driven primarily by the right NRe, which exhibited significantly increased connectivity with the rDLPFC, ($t = 2.35$), an effect for which the analysis had moderate power ~65%. In contrast, the left NRe showed a similar but non-significant directional trend ($t = 0.95$), consistent with its low sensitivity ~16% power to detect effects of this magnitude. The whole NRe showed the same directional pattern ($t = 1.81$), with moderate ~44% power, although the effect did not survive correction.

These findings provide some support for the hypothesis, suggesting that retrieval suppression may involve increased prefrontal-thalamic communication, particularly in the right hemisphere. However, these effects reflect condition-dependent covariation and do not provide evidence of causal influence.

Region	T-Value	p (unc.)	p (corr.)
Left NRe	0.95	0.171	0.513
Right NRe	2.35	**0.009**	**0.028**
Whole NRe	1.81	**0.035**	0.105

Table 4. Results from the rDLPFC-NRe psychophysiological interaction analysis, showing condition-dependent connectivity increases during retrieval suppression for the Left, Right, and Whole NRe ROIs. Bonferroni-corrected p-values are reported.

3.3.4 Retrieval stopping reduces connectivity between the NRe and the hippocampus

The Thalamo-Hippocampal Modulation Hypothesis predicted reduced functional connectivity between the NRe and the HpC during retrieval suppression, consistent with a modulatory role in disrupting memory retrieval. To test this, we conducted PPI analyses using the right HpC as the seed and the NRe as the target.

As shown in Table 5, all analyses revealed a consistent trend toward decreased NRe–HpC connectivity during NT trials, regardless of whether the bilateral, left, or right NRe was used. Although none of the effects reached corrected significance, the directionality was uniform across ROIs, this pattern was driven primarily by the whole NRe ROI (t = –1.90), for which the analysis had moderate sensitivity ~48% power. The right NRe showed a similar directional effect (t = –1.88), ~47% power, and the left NRe likewise trended in the predicted direction (t = –1.78), ~43% power, although neither reached corrected significance. The modest size and non-significance of the effects, together with the limited sensitivity of the current analysis, highlight the need for further investigation using more sensitive connectivity measures.

Region	T-Value	p (unc.)	p (corr.)
Left NRe	-1.78	**0.037**	0.113
Right NRe	-1.88	**0.030**	0.091
Whole NRe	-1.90	**0.029**	0.088

Table 5. Results from the NRe-rHpC psychophysiological interaction analysis, showing condition-dependent connectivity decreases during retrieval suppression for the Left, Right, and Whole NRe ROIs. Bonferroni-corrected p-values are reported.

3.4 Discussion

This study tested four hypotheses concerning the role of the NRe in memory suppression during the TNT task using a mega-analysis of 330 participants. Broadly, the results support the hypothesis that the NRe contributes to retrieval suppression. We observed greater NRe activation during NT > T trials, and functional connectivity analyses indicated increased coupling between the rDLPFC and the NRe during suppression. However, evidence for reactive control and thalamo-hippocampal modulation was modest. This suggests that these mechanisms may be more context-sensitive or weaker in magnitude. These findings refine models of memory control by implicating subcortical structures in both initiation and potential maintenance of retrieval suppression.

3.4.1 Interpreting the Hypotheses

3.4.1.1 Retrieval Stopping Hypothesis

The finding of increased NRe activation during retrieval suppression aligns with its hypothesized role in retrieval suppression, suggesting that thalamic mechanisms may contribute to the suppression of unwanted memories. Previous work has emphasized cortical sources of inhibitory control, particularly in the rDLPFC and our results support the idea that subcortical structures like the NRe also play a role this process. This expands

existing models of memory control by positioning the NRe as a potential relay or modulatory node within the broader prefrontal-hippocampal circuit.

Activation was observed in both hemispheres, consistent with bilateral involvement. Although minor asymmetries emerged in statistical strength, these likely reflect sampling variability rather than meaningful lateralization. This reinforces the view that the NRe's role in retrieval suppression is more general rather than hemisphere specific.

3.4.1.2 Reactive Control Hypothesis

Although we anticipated stronger support for the Reactive Control Hypothesis, the observed effects were weaker and less consistent than expected. The left NRe exhibited a trend toward greater activation during memory intrusions, but this effect was not statistically significant, and Bayesian analyses provided only anecdotal support for the null hypothesis. Moreover, no reliable activation was detected in the right NRe, with Bayesian evidence moderately favouring the null.

One possibility is that the NRe's involvement in reactive control could emerge only under sufficiently high cognitive demand. Participants who experienced more frequent intrusions may have engaged greater suppression, possibly leading to stronger NRe recruitment, whereas participants with fewer intrusions could have contributed minimal signal to the contrast. This variability likely reduced power at the group level.

A second factor may be variability in stimulus content. Prior research suggests that emotionally salient or trauma-related stimuli are more resistant to suppression and could therefore elicit stronger intrusions and greater demand for reactive control. Because the included datasets varied in stimulus type and emotional content, such differences may have obscured reliable detection of NRe engagement.

These findings suggest that, if the NRe contributes to reactive suppression, its engagement is context-dependent and variable. Future studies should be designed to elicit more frequent or challenging intrusions and to standardize stimulus demands across participants.

3.4.1.3 Prefrontal-Thalamic Communication Hypothesis

PPI analyses provided support for the Prefrontal-Thalamic Communication Hypothesis, revealing increased functional connectivity between the rDLPFC and NRe during retrieval suppression. This effect was strongest in the right NRe, aligning with the well-established right-lateralized role of the rDLPFC in inhibitory control.

These findings suggest that suppression demands enhance communication between prefrontal executive systems and thalamic relay structures, supporting models in

which memory control depends on the dynamic recruitment of subcortical circuitry to modulate hippocampal retrieval.

3.4.1.4 Thalamo-Hippocampal Modulation Hypothesis

Partial support was found for the Thalamo-Hippocampal Modulation Hypothesis. Connectivity between the NRe and HpC decreased during suppression, however, these effects did not reach statistical significance.

The consistent directionality suggests that retrieval suppression may involve thalamic downregulation of hippocampal activity. The limited strength of these effects could reflect the complexity of hippocampal regulation, which could potentially involve multiple parallel control pathways. Alternatively, NRe-mediated modulation may be transient or state-dependent in ways that traditional PPI analyses are not well suited to capture. Future studies using effective connectivity methods may offer more precise insights into these dynamic interactions.

3.4.2 Methodological Considerations

While our results provide important insights, several methodological factors warrant consideration. First, intrusion-related analyses were conducted on a subset of the Mega-TNT dataset, limiting trial numbers and statistical power. Low intrusion rates for some participants could have reduced the ability to detect reactive control effects.

Second, the heterogeneous nature of the datasets could have introduced variability in imaging parameters, stimulus types, and study designs. Although we statistically accounted for inter-study differences by modelling dataset membership as a covariate, potential residual heterogeneity could have influenced effect sizes. More standardized designs and acquisition protocols could facilitate cleaner cross-study comparisons in future mega-analyses.

Third, standard-space NRe ROIs may have insufficiently captured individual anatomical variability, especially given the small size of the NRe and its proximity to surrounding thalamic nuclei. Hand-tracing participant-specific ROIs based on high-resolution anatomical scans could improve precision.

Finally, PPI analyses, while informative, are correlational. They do not establish causal directionality of influence between brain regions. More advanced methods such as DCM (Friston, Harrison, & Penny, 2003; Zeidman et al., 2019) are needed to infer effective connectivity and test mechanistic models of suppression.

3.5 Future Directions

Several promising avenues for future research emerge from these findings. First, task designs should be optimized to increase memory intrusions, using harder-to-suppress associations or emotionally salient stimuli, to better engage reactive suppression mechanisms.

Second, effective connectivity analyses, such as DCM, should be used to clarify the causal network architecture of the suppression network and the directionality of influences between rDLPFC, NRe, andHpC.

Third, given the NRe's small size, anatomical location, and potential variability across individuals, future studies would benefit from employing participant-specific, manually traced ROIs based on high-resolution structural MRI scans. Hand-tracing approaches can account for subtle inter-individual differences in thalamic anatomy that standard-space normalization may obscure. By precisely delineating the NRe relative to adjacent midline thalamic nuclei, such individualized ROI methods would enhance the anatomical specificity of analyses and improve the sensitivity to detect true NRe engagement during memory suppression.

While both the left and right NRe were engaged during retrieval suppression, asymmetries emerged across analyses: the right NRe showed stronger connectivity with the rDLPFC, whereas the left NRe showed numerically greater activation during memory intrusions. It is unclear whether these asymmetries could be functionally relevant. Asymmetry may reflect underlying differences in how each hemisphere contributes to retrieval suppression or simply could reflect variability across individuals. Future research should aim to clarify these possibilities, ideally using high-resolution imaging and individualized anatomical approaches.

Beyond establishing causal connectivity, future research should also shed light on the specific functional role of the NRe within the memory suppression network. One possibility, in accordance with traditional thalamic relay models, is that the NRe only transmits prefrontal control signals to the HpC. However, growing evidence from rodent studies suggests that the NRe may perform a more active modulatory role. For example, the NRe has been shown to regulate hippocampal oscillatory dynamics (Hallock et al., 2016, Jayachandran et al., 2023). By modulating hippocampal states, the NRe may act not merely as a passive conduit but as an active arbitrator, capable of disrupting or inhibiting retrieval-related processes.

Testing these hypotheses will require both advanced analysis techniques and improved imaging methods. DCM can infer the directionality of influences among rDLPFC, NRe, and HpC, distinguishing whether the NRe independently modulates hippocampal function or simply relays suppression signals. Furthermore, high field 7T MRI may enable more precise visualization of the NRe, allowing more accurate functional mapping. Recent studies have successfully imaged small subcortical structures such as the habenula and subthalamic nucleus at 7T using high-resolution T1-weighted sequences (Forstmann et al., 2017; Keuken et al., 2013). Applying similar imaging strategies to the midline thalamus could significantly enhance the ability to detect and characterize NRe activation during memory suppression.

Finally, causal manipulation studies are needed. Although TMS cannot directly reach deep thalamic structures like the NRe, applying TMS to rDLPFC could disrupt upstream control signals and indirectly test the necessity of the prefrontal-thalamic-hippocampal circuit. For example, if TMS-induced disruption of rDLPFC weakens NRe-HpC connectivity and impairs retrieval suppression, this would provide strong causal evidence for the functional importance of this pathway.

Understanding how thalamic circuits contribute to retrieval suppression will not only refine cognitive models but may also inform interventions for conditions involving intrusive memories, such as PTSD and depression.

Chapter 4

From Histology to Human MRI: Defining the Nucleus Reuniens

As discussed in previous chapters, the nucleus reuniens (NRe) of the thalamus has emerged as a critical node within the neural network supporting cognitive control and memory suppression. Positioned along the ventral midline of the thalamus, the NRe forms a key anatomical and functional bridge between the medial prefrontal cortex (mPFC) and the hippocampus (HpC). Converging evidence from animal studies and emerging human studies positions the NRe as a key node underlying memory and emotion regulation. Despite its growing theoretical and clinical relevance, the NRe remains poorly studied in the human brain.

Standard neuroimaging methods are currently unable to resolve the fine boundaries of small, midline thalamic nuclei like the NRe. It lies deep in the ventromedial thalamus, flanked by densely packed neighbouring nuclei and medially, it is bound by the third ventricle. Its small size makes it vulnerable to partial volume effects and signal contamination. These risks are further amplified by spatial normalization procedures, which can distort subcortical anatomy when warping brains into standard space, particularly in regions adjacent to cerebrospinal fluid. In the case of the NRe, minor misalignments or over smoothing may add voxels from the ventricles or even worse, from nearby nuclei, compromising anatomical specificity.

These challenges are not unique to the NRe alone but reflect a broader issue in human neuroimaging of the midline thalamus. While progress has been made in mapping large-scale cortical networks, midline thalamic nuclei are often grouped into undifferentiated “medial thalamus” regions. Conventional atlases do not capture subtle anatomical boundaries, and automated segmentation tools lack the resolution to identify them. For structures like the NRe, whose functional contributions are increasingly recognized as specific, this lack of anatomical resolution poses a barrier to studying this structure in the human brain.

This limitation became evident in the initial analyses presented in Chapter 3, which used a standardized, group-level ROI for the NRe developed in collaboration with Dr. Basilis Zikopoulos. While this method enabled the first large-scale functional

investigation of NRe involvement in human memory suppression, it also highlighted limitations: the ROI could not account for individual variability, nor could it guarantee exclusion of non-NRe voxels or ventricular voxels. These concerns strengthen the need for greater anatomical precision, especially when targeting subcortical structures that are only a few millimeters in diameter.

Accordingly, this chapter addresses the anatomical challenges of studying the NRe in humans by building a cross-species framework for segmentation. We first review the NRe's cytoarchitecture, molecular identity, and connectivity in rodents, where its structure is best understood. We then examine the primate NRe, which shares many conserved features with rodents but also incorporates new elements, possibly, more closely mirroring the human thalamus. Finally, we describe the methodology developed in this thesis to delineate the NRe in vivo in humans, including the anatomical criteria, manual segmentation pipeline, and quality control procedures used to ensure voxel-level precision.

This chapter aims to (1) to put together anatomical knowledge of the NRe from rodent and primate models, (2) to examine similarities and differences in the human brain and (3) to define the anatomical rationale and segmentation procedure developed for individualized human NRe ROIs.

4.1 The Nucleus Reuniens in Rodents

In rodents, the NRe is found to be situated along the ventral midline of the thalamus, and it borders the third ventricle. This structure has been commonly visualized using histological methods. It is composed almost entirely of glutamatergic neurons, it lacks intrinsic inhibitory interneurons and is found to exhibit a distinctive molecular profile which is different from neighbouring thalamic nuclei (Bokor et al., 2002).

Rodent studies have mapped the NRe's cytoarchitecture, neurochemistry and projection patterns with a resolution currently not possible in other species. These features have made rodents a key model organism for investigating thalamic contributions from regions like the NRe and map them onto cognitive attributes like memory, contextual updating, and executive control. Crucially, these studies have also provided anatomical priors which have guided efforts to localize its homolog in the primate and human brain. The next sections review these features in detail, emphasizing the ones most relevant for cross-species translation.

4.1.1 Cytoarchitecture and Neurochemical Profile

The rodent NRe is the largest nuclei of the ventral midline thalamus, directly above the third ventricle, extending longitudinally throughout the thalamus (Vertes et al., 2015). It is composed predominantly of glutamatergic projection neurons (Bokor et al., 2002) consistent with its role as an excitatory relay nucleus. Unlike many other thalamic regions, the NRe lacks intrinsic inhibitory interneurons and only consists of sparse and inconsistent axons, lacking cell bodies (Bokor et al., 2002). This arrangement distinguishes its local circuitry from primates, where GABAergic neurons are present (Joyce et al., 2022).

The NRe is marked by high expression of calbindin (CB) and calretinin (CR) and shows minimal parvalbumin (PV) labeling (Bokor et al., 2002). These markers are not evenly distributed, CB-rich zones tend to localize laterally and centrally, while CR-positive neurons cluster medially along the midline, forming a mosaic-like pattern that aids histological identification (Viena et al., 2020). This internal heterogeneity suggests that these distinct clusters could have functional translations.

Although these molecular features are not directly visible in human MRI, they have helped define its position in relation to other anatomical structures.

4.1.2 Connectivity of the Rodent NRe

The rodent NRe occupies a strategic position within prefronto-limbic–thalamic circuitry and is defined by a convergence of inputs along with topographically organized outputs. Its connectivity patterns have been extensively mapped using tract-tracing methods. These findings have delineated a clear anatomical profile that underpins its proposed role in coordinating medial prefrontal and hippocampal communication.

4.1.2.1 Afferent Inputs

The NRe receives extensive afferent inputs from the mPFC. Vertes (2002) demonstrated that the infralimbic cortex (IL), the prelimbic cortex (PL), the anterior cingulate cortex (ACC) and the medial agranular (AGm) cortex project to the NRe. The IL, PL and ACC are found to target the lateral NRe and AGm projects to the medial NRe. These projections were found to target the entire rostrocadual axis of the NRe. These findings were further confirmed by McKenna and Vertes in 2004, and in addition, they identified direct inputs from the hippocampal formation, specifically from the ventral subiculum and the CA1 region. These projections were found to be less dense as compared to projections from the mPFC.

It was also found that the NRe receives afferents from other cortical, subcortical and brainstem regions. Cortical inputs included the insular, retrosplenial cortex along with the perirhinal and ectorhinal cortices. The subcorticular structures included the lateral septum, the diagonal band of broca and the amygdala. The medial and basomedial nuclei of the amygdala were projected directly to the NRe. Further, the hypothalamus including the lateral, ventromedial, posterior and supramammillary nuclei also contributed to afferents. Lastly, the NRe received rich projections from various brainstem structures. These included that dorsal and median raphe, the periaqueductual gray nuclei and the reticular nuclei.

Taken together anatomically, in our context, the NRe can play a role in facilitating communication between the cortex, HpC and various subcortical structures including the amygdala.

4.1.2.2 Efferent Outputs

Vertes et al., (2006) demonstrated that the NRe projects densely to the CA1 region of the HpC as well as the stratum lacusnosum-moleculare (SLM) layer of the subiculum. These projections were found to be excitatory in nature and synapsed on pyramidal cells. Projections from the NRe were found across the dorsal and ventral HpC. Projections from the caudal NRe terminated at the lateral entorhinal cortex and CA1 whereas, projections from the rostral NRe were more abundant in the medial entorhinal cortex and subiculum. Thus, there was the presence of topographic specificity.

In addition to targets in the HpC, the NRe projects robustly to the mPFC, specifically to the IL, PL and the ACC. These IL and PL receive the densest projections. Thus, the NRe has outputs to both, the HpC and the NRe.

Further, the NRe was also found to project to the basolateral amygdala, the basomedial amygdala and the central nucleus of the amygdala. However, these projections were found to be less dense as compared to connections to the mPFC and the HpC. The NRe also projects to multiple nuclei within the hypothalamus such as the lateral hypothalamus, the paraventricular nucleus and the supramammillary nucleus.

4.1.2.3 Dual-Projecting Neurons

A key anatomical feature of the rodent NRe is the presence of dual-projecting neurons cells that send axon collaterals to both the mPFC and HpC. These neurons are typically immunonegative for CB and CR and are found to be molecularly distinct from these CB+ and CR+ clusters (Viena et al., 2020). However, these dual projecting neurons were encased in densely populated nests of CB+ and CR+ clusters which are referred to as

nests and are thought to gate or more broadly, modulate the activity of these dual projecting neurons. The existence of these dual projecting neurons provides a structural basis for bidirectional coordination between the mPFC and HpC through a shared relay. While this property is often highlighted in functional discussions, its anatomical specificity renders it translationally valuable, it enables a single thalamic population to engage distributed cortical and limbic targets via direct anatomical projections.

4.1.2.4 Topographic Organization and Laminar Targeting

The NRe exhibits a structured organization in both its topographic connectivity and laminar targeting. The observed patterns suggest that the NRe plays an active role in regulating and/or modulating information between the cortex, HpC and subcortical structures.

Inputs to the NRe are arranged along spatial gradients. As discussed earlier, within the mPFC, different subregions project to distinct parts of the NRe. The IL, PL, and ACC targets the lateral NRe, while the AGm projects to the medial NRe. These projections are found to span the full rostrocaudal axis of the NRe (McKenna & Vertes, 2004; Vertes, 2002).

Further subdivision has found around the rostrocaudal and mediolateral axes. The rostromedial NRe receives inputs primarily from the HpC and basal forebrain, the caudomedial NRe from cortical areas, and the rostrolateral NRe from the amygdala and basal forebrain (McKenna & Vertes, 2004).

Efferent projections of the NRe also exhibit an organized pattern. NRe neurons projecting to the HpC demonstrate rostrocaudal specificity. The rostral NRe projects predominantly to the medial entorhinal cortex and subiculum and the caudal NRe sends targets to the lateral entorhinal cortex and CA1 region (Vertes et al., 2006).

Along with spatial specificity, NRe efferents also exhibit laminar precision, particularly within the mPFC and HpC. In the mPFC, NRe projections terminate mainly in layers 1 and 5/6, with the IL and PL receiving the densest inputs (Vertes et al., 2006). Functionally, layer 1 is found to be important for modulating dendritic input onto pyramidal neurons, while layers 5/6 are associated with integrative and output processing.

In the HpC, NRe projections synapse onto pyramidal neurons in the SLM of the CA1 and the subiculum. These excitatory projections are well positioned to influence hippocampal function, as the SLM is a major input layer receiving convergent input from the entorhinal cortex and other associative regions (Vertes et al., 2006).

Although the resolution of these subdomains exceeds current capabilities in human neuroimaging, their specialised organization in rodents provides a template for hypothesizing functionally specific subregions in translational research. Further, laminar specificity reinforces the NRe's role not just as a relay, but as an integrative hub capable of modulating cortico-hippocampal communication.

4.1.3 Summary and Translational Relevance

The rodent NRe is one of the most anatomically well-characterized amongst all midline thalamic nuclei. Its cytoarchitecture, molecular identity, and input-output connectivity has been mapped in detail, providing a foundational model for understanding how the thalamus integrates cortical, hippocampal, and subcortical signals.

From a translational lens, rodent data provide essential anatomical priors for identifying the NRe in the human brain. While human neuroimaging tools cannot currently resolve the NRe directly, rodent studies inform the spatial relationships that guide its inferred location, most notably its adjacency to the third ventricle, placement in the ventral thalamus, and separation from surrounding midline nuclei such as the rhomboid nuclei. The internal topography of rodent NRe outputs also supports the idea that different subregions may serve distinct functional pathways, even if these cannot yet be distinguished in vivo in humans.

Perhaps most significantly, the presence of dual-projecting neurons linking the mPFC and HpC provides a clear anatomical circuit that grounds human mechanistic hypotheses in tractable cellular architecture. While the existence of this cell type in humans remains unconfirmed, its robust documentation in rodents provides a framework to speculate cross-species comparisons.

In sum, the rodent NRe serves not only as a detailed anatomical model but also as a reference point for developing translational strategies. However, due to species specific differences in thalamic size, organization, and microcircuitry, further refinement requires bridging through non-human primate anatomy where thalamic architecture more closely approximates that of the human brain. This is the focus of the next section.

4.2 The Nucleus Reuniens in Non-Human Primates

Non-human primates, particularly macaques, provide an intermediate model for understanding the human NRe. In comparison to rodents, the primate thalamus exhibits greater structural complexity and closer correspondence to the human brain in scale and in shape. Although anatomical data in primates are limited, existing tract-tracing and

immunohistochemical studies provide insights into the NRe's composition, connectivity, and spatial layout. These features reveal conserved properties with the rodent brain along with species-specific adaptations which inform our assumptions when translating to human neuroimaging.

4.2.1 Cytoarchitecture and Molecular Profile

The NRe in primates is found to display a distinct cytoarchitecture and molecular profile that sets it apart from its rodent counterpart. Like the rodent NRe, the primate NRe is organized as a matrix-dominant nucleus, consisting of glutamatergic neurons which expresses CB and CR (Joyce et al., 2022). However, unlike the rodent NRe, the primate NRe consists of a sparse population of PV positive neurons. PV is typically associated with core thalamic nuclei that project to middle cortical layers and mediate focal, driver-type transmission (Jones, 1998). This suggests the presence of a partial core-like architecture in primates which is thought to be involved in more nuanced cortical interactions.

The most important deviation from the rodent NRe is the presence of GABAergic neurons in the primate NRe (Joyce et al., 2022). Using histochemistry, it was found that NRe neurons expressed GABA along with pleomorphic vesicles which are typically found in other thalamic nuclei capable of inhibitory neurotransmission. The existence of these inhibitory neurons within the NRe indicates greater local circuit complexity, which could translate to potential intrathalamic inhibition and modulation of activity within the nucleus itself.

These differences indicate evolutionary elaboration of the NRe in primates, potentially enabling more complex regulation of input and output signals to and from cortical and limbic targets. Currently, these microcircuit features are not observable in human imaging, their known presence in the primate NRe adds insight to translational inference to studying the human NRe.

4.2.2 Connectivity of the Primate NRe

The connectivity of the primate NRe reflects both its conserved role as a cortico-limbic relay and a degree of species-specific elaboration. While the overall projection pattern is broadly consistent with that of rodents, linking medial prefrontal and hippocampal structures, primate studies have revealed greater diversity in both afferent sources and internal organization. Tract-tracing experiments and immunohistochemical studies

provide information about these inputs and outputs, offering a more human-relevant anatomical templates.

4.2.2.1 Afferent Inputs

The primate NRe receives convergent inputs from several cortical and limbic regions regions reinforcing it as an integrative hub. Joyce et al., 2022 demonstrated that Area 25 or the subgenual mPFC is a primary source of input to the NRe. This region projects to the entire rostrocaudal axis of the NRe and these projections are topographically organized.

The NRe also receives afferents from the hippocampal formation, precisely, the subiculum and the CA1. These hippocampal projections include a multisynaptic arrangement involving both, excitatory and inhibitory interneurons projecting to the NRe. Like in the rodent brain, projections from Area 25 are denser as compared to afferents from the HpC.

The NRe also receives projections from the amygdala, particularly from the BLA and the basomedial nuclei. These projections are found to be stronger and more pervasive than in the rodent brain. Further, the amygdala axons which terminated in the NRe were found to be rich in mitochondria as compared to axons from Area 25 and the HpC. This indicates high synaptic activity at amygdala synapses with the NRe. However, axon bouton size, across inputs from Area 25, HpC and amygdala were found to be similar. Only HpC boutons formed multisynapse triads with the NRe connecting with CB+, CR+ and GABAergic dendrites. Thus, like in the rodent brain, the anatomic connections in the primate brain are conducive to prefrontal-hippocampal interactions via the thalamus.

4.2.2.2 Efferent Outputs

The primate NRe is found to project back to all regions from which it receives input supporting its role in bidirectional communication with these regions. The NRe sends projections to the hippocampal formation, specifically to the CA1 and subiculum and these projections are found to terminate on pyramidal neurons (Joyce et al., 2022). This suggests a direct means for inhibitory modulation of the HpC. The NRe sends its densest projections to Area 25 which is a part of the subgenual medial prefrontal cortex. Efferent projections also extend to the basomedial and basolateral nuclei of the amygdala. Thus, these bidirectional pathways reiterate the NRe's function as a central node in orchestrating activity between these various systems.

4.2.2.3 Dual-Projecting Neurons

To date, direct evidence for dual-projecting neurons in primates, cells that send axon collaterals to both the mPFC and HpC is lacking. This could be due to technical limitations in dual retrograde tracer experiments in large brains rather than true absence. However, the anatomical overlap between input and output zones, as well as the conserved topography of NRe efferents (discussed in the next section), suggests that such neuronal populations are plausible. The confirmation of their presence would substantially strengthen cross-species inferences and further support the proposed role of the NRe as a synchronizing and/or modulatory hub for prefrontal–hippocampal interactions.

4.2.2.4 Topographic Organization and Laminar Targeting

The primate NRe is found to exhibit evidence of topographic organization and laminar specificity. However, this organization is less characterized in primates as compared in rodents. Afferent inputs to the NRe were found to be spatially organised along the rostrocaudal and mediolateral axes. Projections from the HpC, Area 25 and amygdala terminate in overlapping but distinct zones within the NRe. A rostrocaudal gradient has been found to exist wherein, afferents from posterior medial prefrontal regions like Area 25 primarily target the rostral NRe whereas, projections from the anterior orbital regions target the caudal NRe. This spatial arrangement suggests that the NRe can integrate functionally diverse inputs in a region-specific manner. In terms of projections from the NRe, those to Area 25 are most dense and follow topographical routes, consistent with its efferents observed in this region.

In terms of laminar targeting, Joyce et al., (2022) demonstrated that NRe projections to the HpC were found to project to pyramidal neurons in the SLM layer of the CA1 and subiculum. However, evidence for laminar targeting with the PFC is absent. The NRe's neurochemical profile, consisting of CR+ and CB+ neurons suggests that these projections could terminate in superficial cortical layers, layers 1 to 3a which is typical of matrix-type thalamic connections (Jones, 1998). As discussed earlier, amygdalar projections to the NRe are primarily from the BLA and basomedial nuclei and form metabolically demanding synapses, their topography and laminar organization currently remain uncharacterised in primates.

4.2.4 Translational Implications

The primate NRe introduces several anatomical refinements that enhance its value for human inference. First, the presence of intrinsic inhibitory neurons and PV-positive cells suggests more elaborate gating mechanisms than in rodents, features that may also exist, but currently remain unconfirmed, in humans. Second, the strong BLA input positions the NRe to play a more direct role in affective regulation, consistent with its involvement in disorders like depression and PTSD.

Finally, the spatial relationships of the NRe to landmarks visible in structural imaging (e.g., third ventricle, interthalamic adhesion) are more homologous to humans than rodent data alone can provide. These conserved positional features, along with tract-tracing results, offer a structurally grounded basis for identifying the NRe in human neuroimaging.

Together, these insights position the primate NRe as an indispensable reference point bridging the high-resolution clarity of rodent studies with the spatial and connectional realism needed to define its human counterpart.

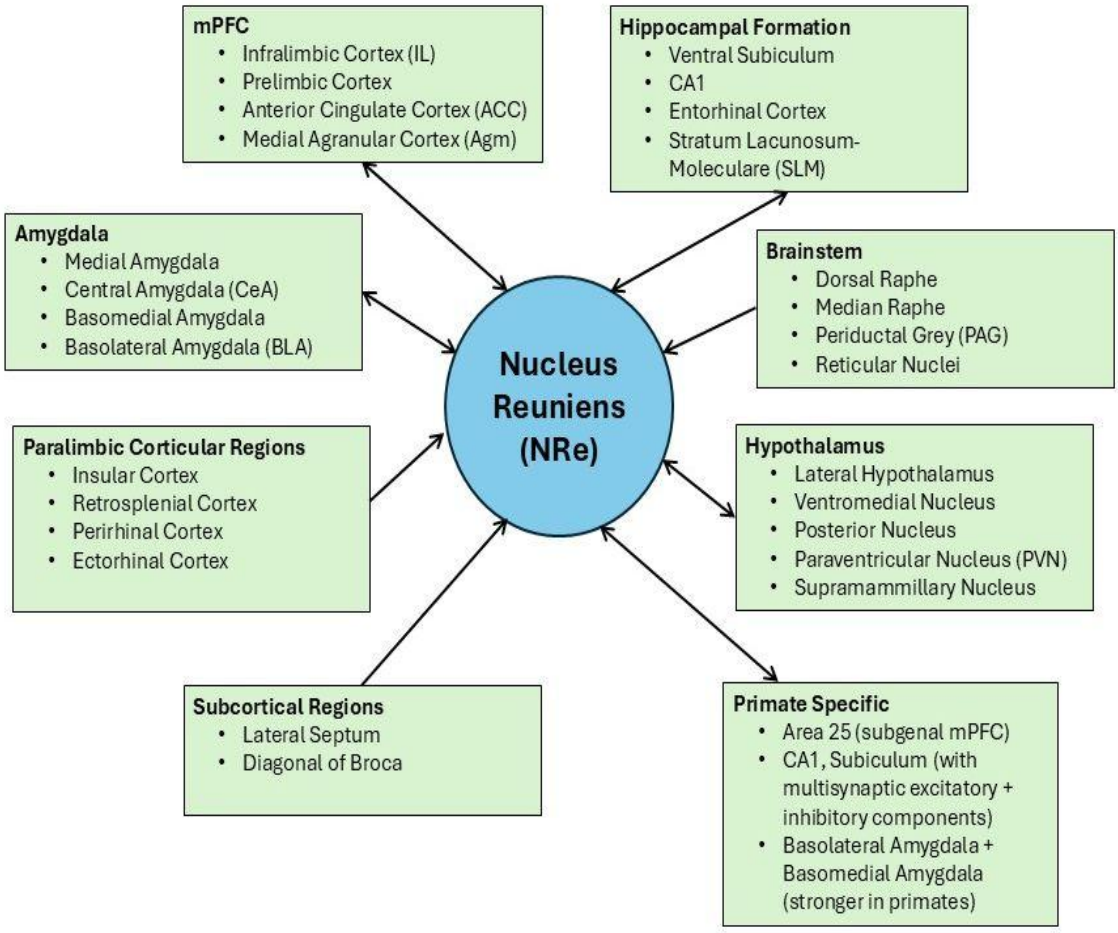

Figure 5. Cross-species overview of NRe connectivity, integrating rodent and primate evidence to illustrate major cortical, limbic, subcortical, and brainstem projections that inform human anatomical inference.

4.3 The Nucleus Reuniens in Humans: Anatomical Inference and Segmentation

Building on insights from rodent and non-human primate studies, this section addresses the challenge of defining the NRe in the human brain, where direct anatomical verification is limited. Unlike in preclinical models, human thalamic segmentation lacks histological evidence, and standard MRI lacks the resolution to visualize the NRe directly. Nonetheless, its consistent position relative to conserved midline structures enables a principled approach.

Recent work by Reeders et al. (2022) has demonstrated the feasibility of identifying nuclei present in the midline thalamus, including the likely location of the NRe using diffusion-weighted imaging and probabilistic tractography. Their approach revealed strong structural connectivity with limbic regions such as the mPFC, HpC and entorhinal cortex, broadly aligning with known NRe connectivity patterns discussed in

both, rodent and primate models. However, their results are derived from group-level clustering and do not allow direct anatomical localization at the individual level.

Rather than relying on coarse atlas labels or probabilistic templates, our present methodology adopts a semi-manual, anatomically constrained pipeline to localize the NRe at the individual level. This approach leverages spatial resolution obtained from MRI, cross-species anatomical correspondences, and validated spatial landmarks (e.g., third ventricle, interthalamic adhesion) to construct subject-specific ROIs that are both reproducible and anatomically plausible.

The following sections discusses a detailed description of the segmentation protocol developed in this thesis, including anatomical criteria, technical implementation, and validation procedures. The goal is not just to approximate the NRe's location, but to establish a framework that can be applied flexibly across datasets while preserving the structure's anatomical integrity.

4.3.1 Inferred Location and Anatomical Landmarks

The human NRe resides in the ventral midline thalamus, flanking the third ventricle and occupying the posterior half of the thalamic body in coronal sections. Its spatial location is inferred from comparative histology and connectional similarity to rodent and primate NRe, which show consistent placement adjacent to the third ventricle and below the interthalamic adhesion.

Four key anatomical landmarks constrain the NRe's location in the human brain:

- **Medial boundary:** The ventricular wall of the third ventricle. This boundary ensures exclusion of cerebrospinal fluid (CSF) and guarantees midline adherence.
- **Lateral boundary:** The internal medullary lamina, which separates medial thalamic nuclei from the mediodorsal and centromedian complexes.
- **Superior boundary:** The base of the interthalamic adhesion, when present. This dorsal limit and the NRe often terminates just below this structure.
- **Inferior boundary:** Ventral thalamic gray matter, bordering the dorsal hypothalamus and posterior hypothalamic nuclei.

4.3.2 Cross-Species Anchoring and Tractography-Based Inference

Localization of the human NRe is informed by cross-species anatomical comparison. Rodent data provides us with detailed cytoarchitecture and molecular identity, whereas primate data offers closer spatial homology with the human thalamus in terms of scale

and surrounding landmarks. The consistent position of the NRe relative to the third ventricle and interthalamic adhesion in rodents and macaques forms the basis for estimating its boundaries in humans.

Connectivity-based analyses further reinforce this localization. As discussed earlier, Reeders et al. (2022) employed diffusion MRI clustering to group thalamic voxels based on their probabilistic connections to the mPFC, medial temporal lobe, and nucleus accumbens. One ventral midline cluster, defined by strong connections to mPFC and HpC, aligned with the expected projection pattern of the NRe. This cluster was anatomically distinct from more dorsal paraventricular and paratenial nuclei, supporting its identification as the reuniens/rhomboid complex.

In parallel, our collaboration with Dr. Zikopoulos produced group-level NRe masks by registering macaque histological landmarks to human MNI space. These masks used observable MRI features including the third ventricle, interthalamic adhesion, and internal medullary lamina to constrain estimated NRe boundaries in the absence of direct contrast.

4.3 Segmentation Procedure: Delineating the NRe in the Human Brain

Translating anatomical insights from rodent and non-human primate studies into a replicable human protocol necessitated the development of a semi-manual, anatomically constrained segmentation pipeline. This method was designed to achieve voxel-level precision in localizing the NRe while accounting for its small volume, poor contrast on standard T1-weighted images, and proximity to the third ventricle and adjacent thalamic nuclei. The protocol integrates primate histological landmarks with in vivo structural neuroimaging, producing individualized ROIs suitable for anatomical and functional analysis.

4.3.1 Imaging Environment and Preparation

All segmentation procedures were performed on each participant's native-space T1-weighted anatomical scan, allowing direct visualization of individual thalamic anatomy without reliance on spatial normalization or group-level templates. Images were acquired on a 3T Siemens Prisma scanner using a multiband echo-planar imaging (EPI) sequence which an acceleration factor of 2. Using a multiband sequence was advantageous for anatomical localization due to improved slice coverage (thinner slices) and reduced acquisition time which benefit signal coverage from small structures like the NRe.

Initial inspection and boundary identification were conducted using the sagittal plane, which provided clear visibility of anterior–posterior landmarks. Segmentation and ROI drawing were performed in the coronal plane, offering optimal resolution of midline structures such as the third ventricle and internal thalamic borders. This dual-plane approach improved anatomical accuracy and reduced ambiguity in voxel selection near CSF-adjacent regions. Visualization and manual region marking were performed using MRIcron (https://www.nitrc.org/projects/mricron), which enabled detailed inspection of each participant's thalamus and the placement of consistent anatomical landmarks.

4.3.2 Delineation Protocol and Rater Procedure

Segmentation was conducted manually by two trained raters. Each rater independently examined the T1-weighted image in native space and identified the anterior and posterior boundaries of the thalamus, as well as the estimated extent of the NRe based on anatomical criteria. To promote consistency and minimize subjectivity, rater decisions were cross validated in a consensus meeting, and final masks were reviewed. This dual-rater protocol ensured anatomical rigor and reliability in ROI definition across participants.

4.3.3 Thalamic Boundary Identification

To localize the NRe within the correct rostrocaudal range of the thalamus, we first identified the full anterior–posterior extent of the thalamus in each participant using the sagittal plane in MRIcron. This view allowed for precise visualization of the third ventricle. The anterior boundary was defined as the first sagittal slice where thalamic gray matter became visible adjacent to the third ventricle, posterior to the disappearance of hypothalamic tissue. The posterior boundary was defined as the last slice where medial thalamic gray matter remained discernible before mixing with the surrounding white matter. These coordinates were used to define the coronal slice range for subsequent NRe segmentation.

These start and end slices were recorded via their Y-axis voxel coordinates and they served as a reference range for estimating the extent of the NRe. The NRe was defined as occupying approximately the posterior half of the thalamus. Operationally, we subtracted five slices from the anterior thalamic boundary to define the starting slice for the NRe, while the posterior boundary matched the final thalamic slice. These values were saved in a reference table for each participant and provided the slice indices used to generate the initial binary masks.

4.3.4 NRe Extent Localization

Following identification of the anterior and posterior thalamic boundaries in the sagittal plane, we calculated the corresponding coronal slice range for each participant and used these values to generate preliminary binary ROIs via a custom script. These initial ROIs defined the spatial bounds within which the NRe was likely to be located and were subsequently loaded into the MATLAB-based segmentation application for manual refinement. This ensured that all subsequent delineation occurred within an anatomically constrained and participant-specific ROI.

4.3.5 Anatomical Constraints and Landmark-Based Placement

A custom MATLAB-based segmentation application was developed in collaboration with Dr Crespo-García to enable semi-manual ROI refinement. The graphical user interface allowed synchronized viewing of T1-weighted coronal slices alongside digitized anatomical atlases, supporting anatomically constrained adjustments.

Within the app, the following anatomical boundaries were applied to restrict the NRe ROI:

- **Medial boundary:** The ventricular wall of the third ventricle, ensuring exclusion of CSF and midline adherence
- **Lateral boundary:** The internal medullary lamina and, when visible, the uncinate fasciculus (a bright white matter tract lateral to the NRe)
- **Superior boundary:** The base of the interthalamic adhesion, corresponding to the widest point of the third ventricle
- **Inferior boundary:** The dorsal margin of the hypothalamus and ventral thalamic gray matter

Midline symmetry was preserved by manually placing an anchor point at the center of the third ventricle on the coronal slice. This point enabled the app to compute mirrored left and right hemisphere ROIs. All ROI refinements were made within the slice range generated during the thalamic boundary analysis.

4.3.6 ROI Generation and Interpolation

After defining ROIs at the anterior and posterior boundaries of the NRe, the application performed interpolation which generated the ROI in all intermediate coronal slices. This ensured a smooth and anatomically grounded volumetric transition between these boundaries. This interpolation step preserved consistency across slices while respecting

voxelwise anatomical constraints applied. The output was a complete binary volume for each hemisphere, saved in native participant space.

4.3.7 Quality Control and Manual Correction

Each segmentation underwent detailed quality control in three stages:

1. **Visual inspection:** Raters reviewed all coronal slices containing NRe ROIs to confirm bilateral symmetry and anatomical plausibility.
2. **Landmark verification:** Adherence to anatomical boundaries (e.g., ventricular edge, interthalamic adhesion) was verified slice by slice.
3. **Voxel correction:** When needed, raters manually edited masks using built-in tools for voxel deletion, smoothing, or boundary refinement. These adjustments ensured exclusion of off-target voxels, such as intraventricular space or adjacent non-NRe gray matter.

All corrections were documented, and session data including slice ranges and edit logs—were saved.

4.3.8 ROI Export and Final Verification

Finalized ROIs were saved as binary NIfTI volumes in native participant space. These volumes were then reloaded in MRIcron for overlay with the original anatomical scan to confirm anatomical accuracy. Specifically, we verified that ROIs:

- Aligned correctly with midline gray matter
- Avoided the third ventricle and white matter tracts
- Fell within the expected anatomical region based on known landmarks

left and right ROIs were stored separately to allow hemispheric comparisons in downstream analyses.

4.3.9 Visual Walkthrough of the Segmentation Procedure

Supporting materials for the segmentation procedure are also included below. They aim to provide a step-by-step visual guide to the segmentation protocol, including screenshots from MRIcron and the MATLAB GUI, labelled anatomical landmarks, and example participant images. These resources replicate the original walkthrough but are consolidated here for reproducibility and transparency.

4.3.9.1 Anatomical Landmarks and Imaging Environment

We began by localizing the NRe using participant-specific T1-weighted MRI images. Anatomical landmarks derived from primate studies were used to guide localization. All imaging analyses were conducted in the coronal plane using MRIcron

(https://www.nitrc.org/projects/mricron), which provided convenient visibility of the third ventricle and surrounding structures.

4.3.9.2 Identifying the First and Last Coronal Slices of the Thalamus

To localize the NRe within the broader thalamus:

- The first coronal slice in which the thalamus appeared was identified independently for each hemisphere. The corresponding Y-coordinate was recorded (Figure 6).

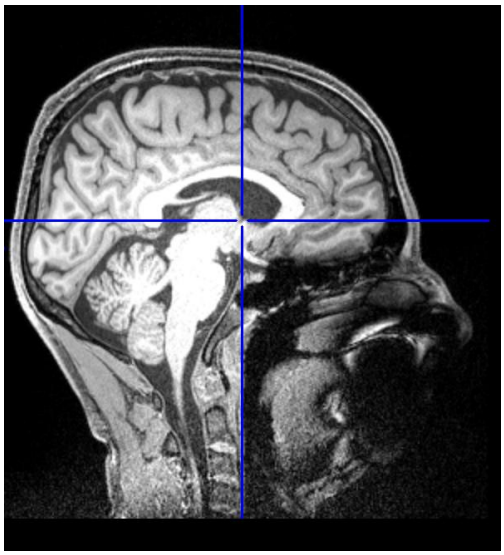

Figure 6. First Coronal Slice of the Thalamus

- Similarly, the last coronal slice of the thalamus was noted, and its Y-coordinate was also recorded (Figure 7). These coordinates defined the full anterior-posterior range of the thalamus.

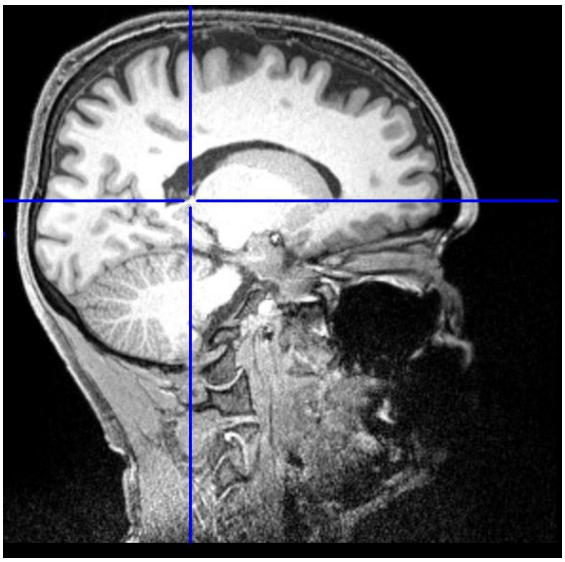

Figure 7. First Coronal Slice of the Thalamus

4.3.9.3 Defining the Start and End of the NRe

As the NRe is primarily located in the posterior half of the thalamus:

- We subtracted 5 Y-coordinate units from the first visible thalamic slice to define the starting slice of the NRe (Figure 8).

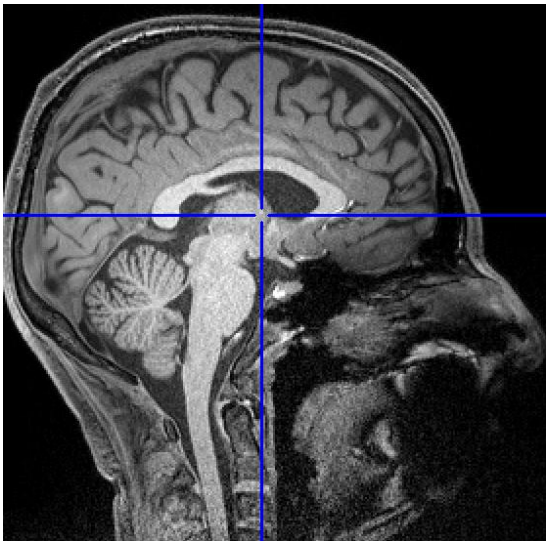

Figure 8. First Coronal Slice of the Nucleus Reuniens

- The ending slice was determined visually using anatomical cues from the T1 images and reference atlases (Figure 9).

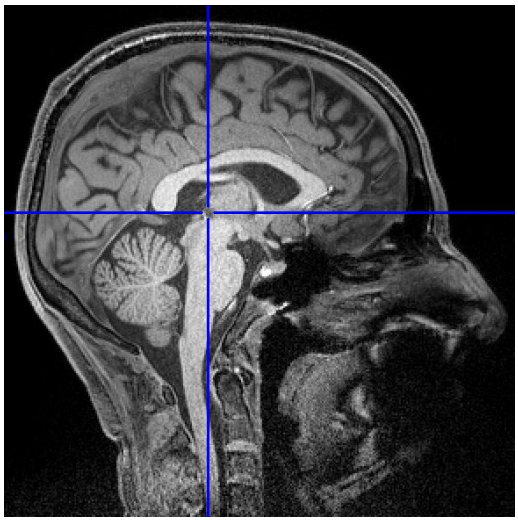

Figure 9. Last Coronal Slice of the Nucleus Reuniens

4.3.9.4 Custom MATLAB Application for NRe ROI Generation

In collaboration with Dr. Crespo-García, we developed a MATLAB-based application (MATLAB 2023b, MathWorks, MA) to create individualized NRe ROIs across participants. This tool integrated anatomical landmarks, visual aids, and user interaction to ensure high accuracy.

4.3.9.5 Loading Participant Data and Session Setup

Each participant's:

- Native T1-weighted MRI scan,
- First and last slice coordinates for left and right NRe,

were loaded into the application. Each participant was handled as a unique session, with settings saved for reproducibility (Figure 10).

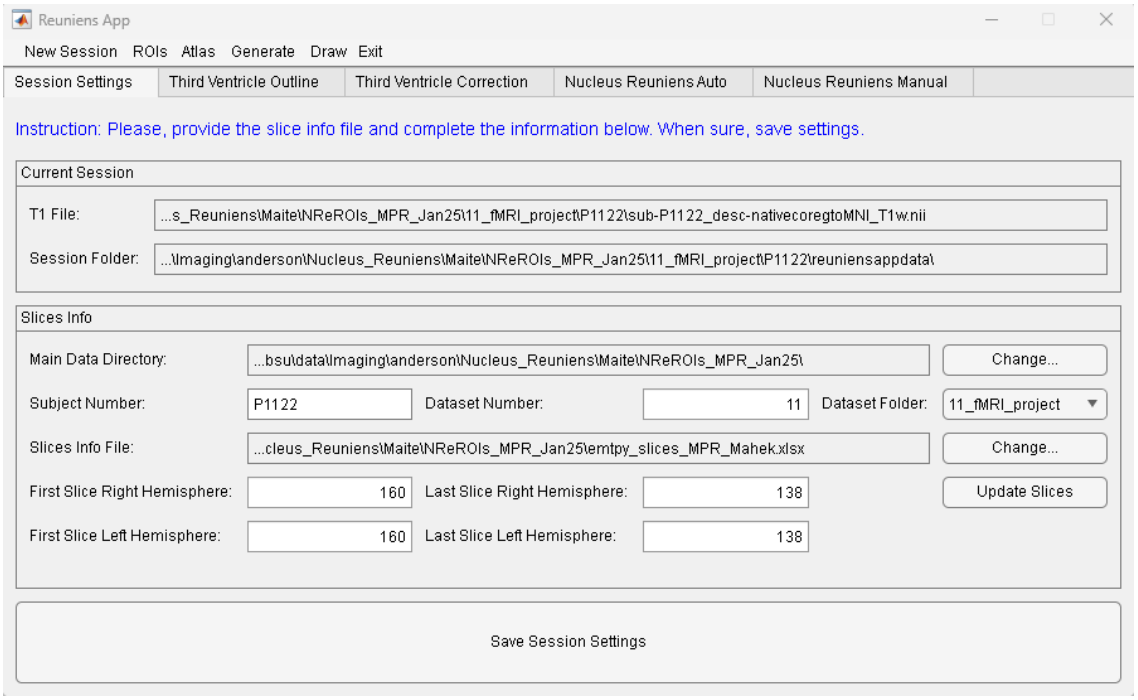

Figure 10. Loading participant data and session setup in MATLAB application

4.3.9.6 Midline Identification: Center of the Third Ventricle

Users were prompted to place a blue dot at the center of the third ventricle, a consistent midline structure (Figure 11). The NRe lies just lateral to this point, making it a crucial reference. The application automatically magnified this region to allow more precise ROI placement (Figure 12).

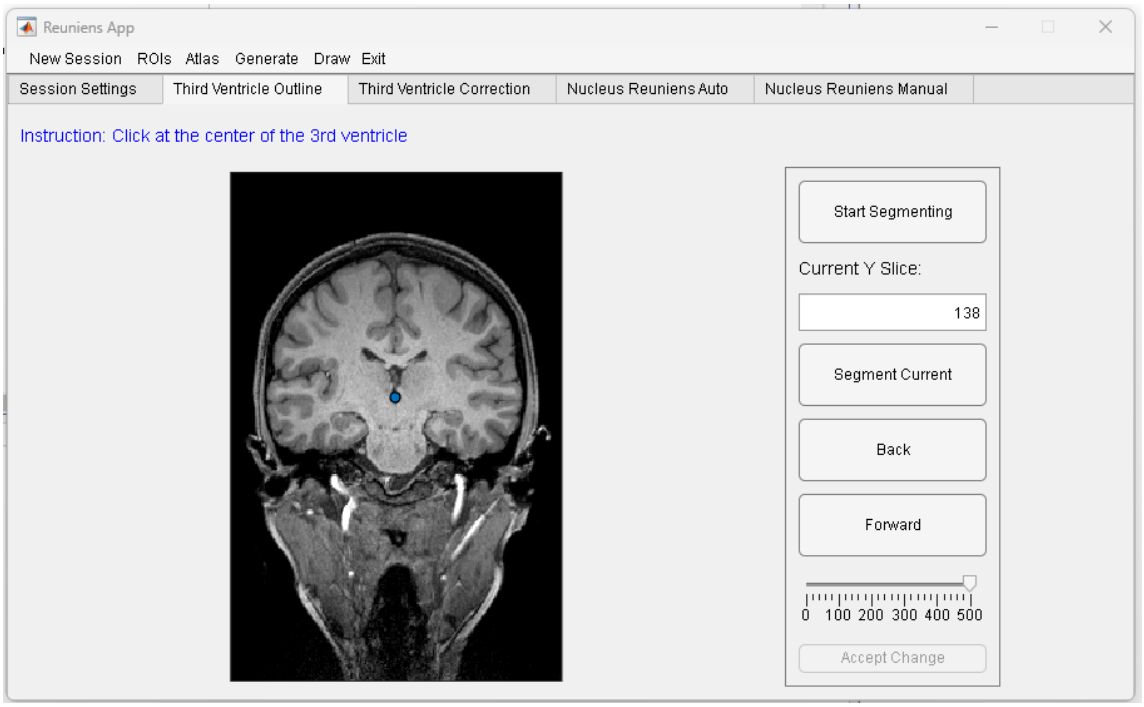

Figure 11. Selecting the centre of the third ventricle

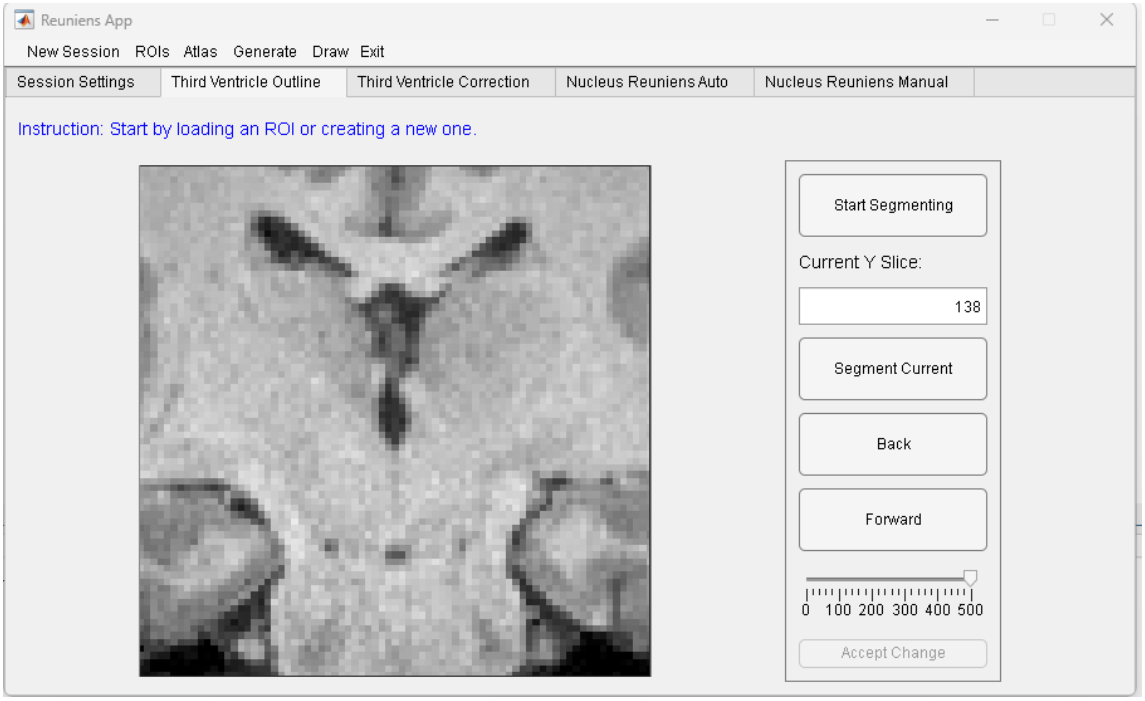

Figure 12. Magnified region around the third ventricle

4.3.9.7 Integration of Anatomical Atlases

Three anatomical reference atlases were embedded in the application for side-by-side viewing. These provided structural guidance, including white/gray matter differentiation and regional anatomical boundaries.

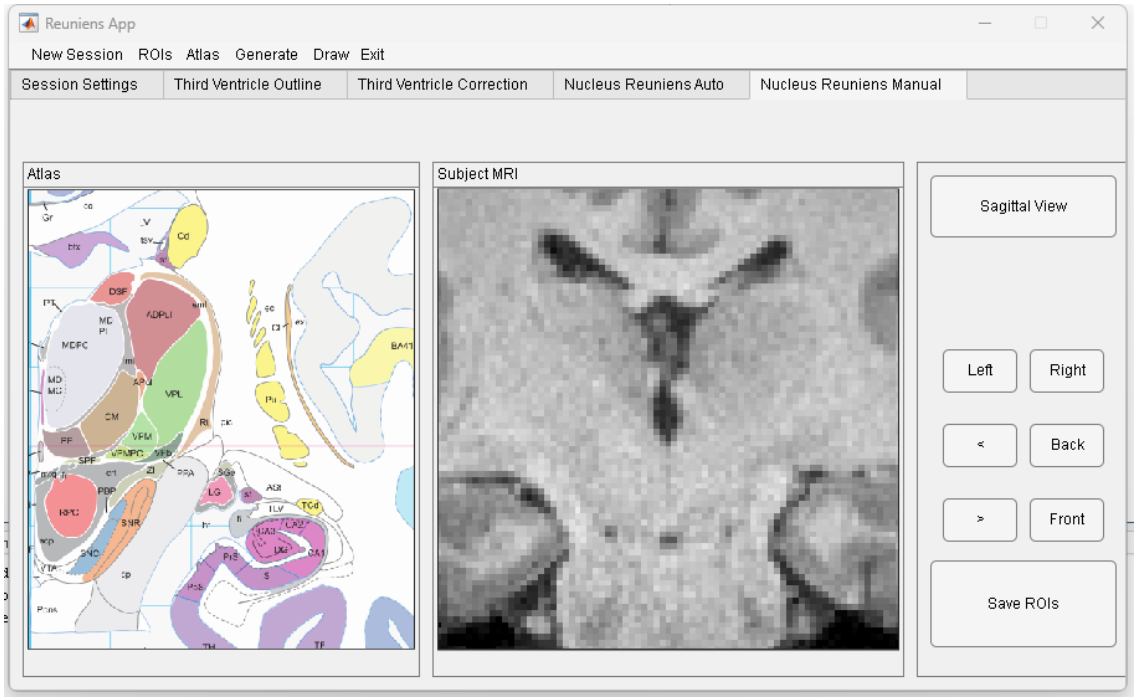

Figure 13. Paxinos & Franklin Mouse Brain Atlas – Coronal Plate

Figure 14. Paxinos & Watson Rat Brain Atlas – Histological Plate

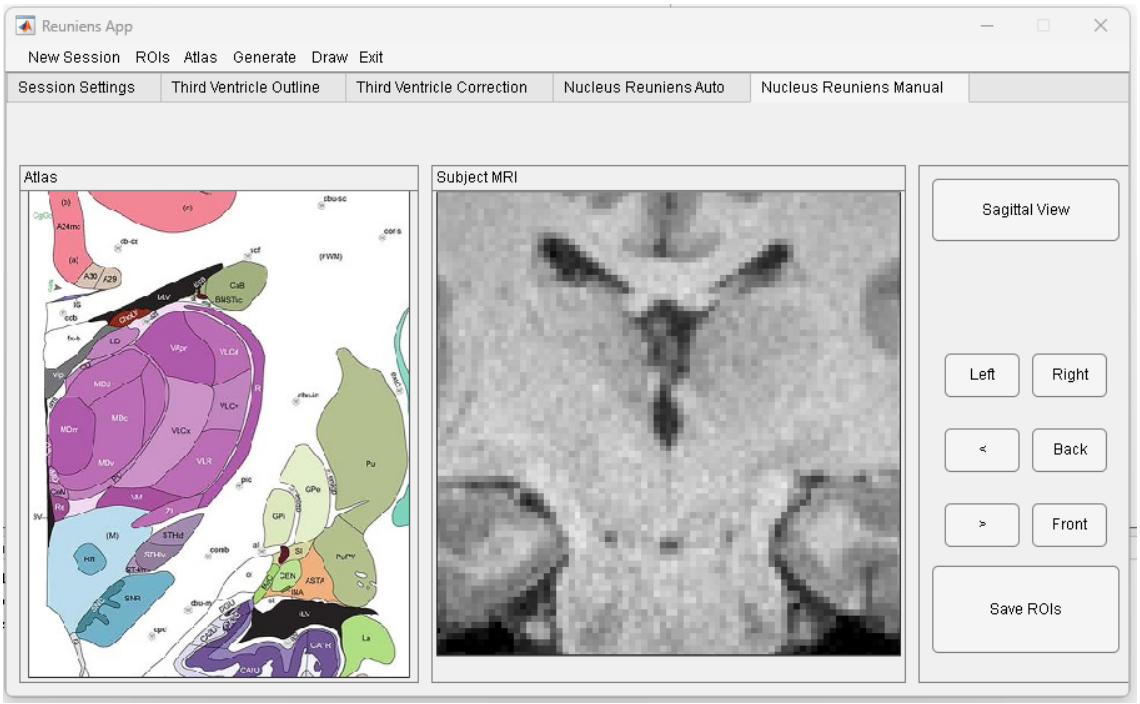

Figure 15. Paxinos & Watson Rat Brain Atlas – Thalamic Plate

4.3.9.8 Defining the Anatomical Constraints of the NRe

Placement of ROIs was guided by the following landmarks:

- **Inferior**: Posterior hypothalamus (lighter contrast),
- **Medial**: Adjacent to third ventricle,
- **Lateral**: Uncinate fasciculus (high-contrast white matter),
- **Superior**: Interthalamic adhesion (at maximum third ventricle width).

4.3.9.9 Placing the Anterior and Posterior Boundaries of the NRe

- Clicking **'Front'** navigated to the first NRe slice. ROIs were placed below the (Figure 16).
- Clicking **'Back'** moved to the last NRe slice. ROIs were placed below the widest part of the third ventricle (Figure 17).

Care was taken to avoid intrusion into neighbouring white matter or ventricular space.

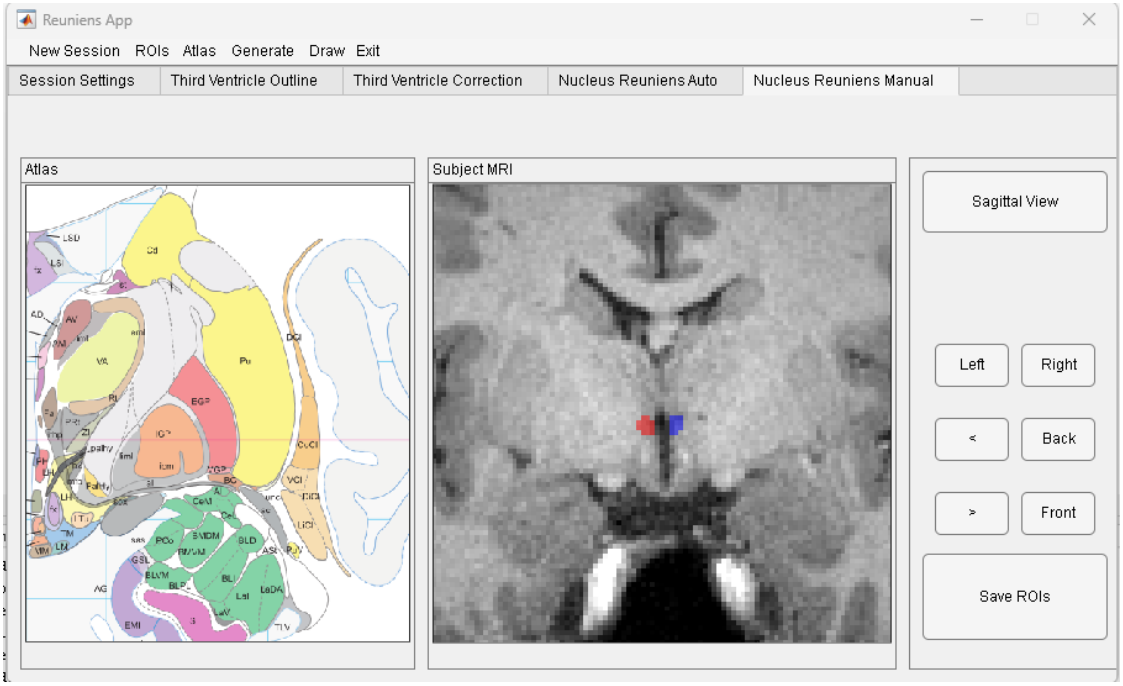

Figure 16. Placing voxels as part of the start of the NRe below the mouth of the third ventricle

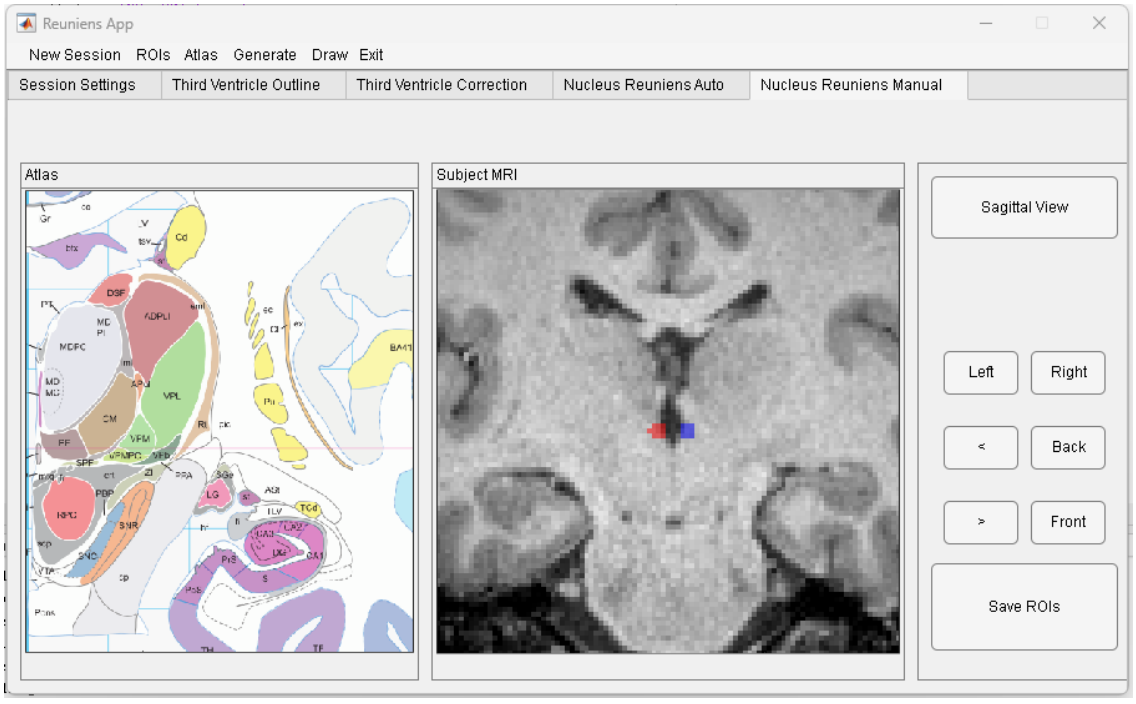

Figure 17. Placing voxels as part of the end of the NRe below the widest part of the third ventricle

4.3.9.10 ROI Generation

The **'Generate'** button created interpolated, continuous NRe ROIs between anterior and posterior boundaries for both hemispheres. This ensured smooth and anatomically valid segmentation.

4.3.9.11 Quality Control and Symmetry Check

Users scrolled through the slices using **'<' and '>'** to:

- Confirm bilateral symmetry,
- Verify proximity to the third ventricle,
- Check that voxels remained within gray matter regions and avoided ventricles.

4.3.9.12 Manual Corrections

If needed, the **'Draw'** panel provided tools for voxel deletion, edge smoothing, and realignment to refine the masks.

4.3.9.13 Saving and Exporting Finalized ROIs

Once finalized, ROIs for the left and right NRe were saved for each participant and stored along with session metadata for future analysis.

4.3.9.14 Final Confirmation in MRIcron

As an additional verification step, the finalized ROIs were overlaid onto the native T1 images in MRIcron**.** A thorough visual review ensured consistency and anatomical accuracy across the full anterior-posterior extent.

4.4 Summary and Future Directions

The NRe is a structurally elusive yet functionally relevant component of thalamocortical circuitry. In this chapter, we traced its anatomy from well-characterized rodent models through increasingly human-like primate brains to its inferred location in the human thalamus. This framework demonstrates the evolutionary continuity of the NRe and the translational challenges posed by differences within species, limited imaging resolution in humans and incomplete postmortem data in humans.

Drawing on conserved anatomical landmarks and connections we developed a semi-manual segmentation protocol to enable individualized identification of the NRe within structural MRI. While this method does not permit direct visualization of the nucleus, it provides a principled, voxel-level approximation grounded in comparative neuroanatomy and in vivo imaging constraints. In contrast to coarse, group-level atlases, this individualized approach offers enhanced anatomical specificity which is critical when targeting small midline structures like the NRe.

Crucially, the segmentation pipeline was developed in consultation with Dr. Zikopoulos, a primate neuroanatomist, whose expertise helped constrain the spatial boundaries of the human NRe based on macaque histology and conserved anatomical landmarks. His guidance ensured anatomical fidelity throughout the localization process which generated these individualised masks in humans.

Despite this progress, many questions remain. The internal organization of the human NRe, its cell types, microcircuit architecture, and subfield specialization is currently inaccessible to imaging technologies. Future work leveraging ultra-high-field MRI, postmortem histology, or high-resolution connectomics could help refine our understanding and yield answers to these questions.

In the meantime, these anatomically constrained, voxel-level localized ROIs is a necessary step toward integrating the NRe into functional human neuroscience. The segmentation protocol introduced in this chapter lays the anatomical groundwork for subsequent chapters, where we examine the NRe's engagement in task-related activity, its coupling with hippocampal and prefrontal circuits, and its relationship to memory suppression and fear extinction.

More broadly, this individualized approach may serve as a template for future automated tools. If applied consistently across participants and datasets, the resulting database of human-labelled NRe masks could be used to train supervised machine learning models for faster and more scalable NRe identification. Embedding the current protocol within a structured software interface and logging ROIs alongside anatomical data could result in a generalizable, data-driven segmentation platform. Such scalability would not only accelerate future NRe research but could also make precise thalamic segmentation more accessible across research domains.

Chapter 5

Prefrontal–Thalamic Contributions to the Suppression of Unwanted Memories

The Think/No-Think (TNT) paradigm provides a well-established framework for studying intentional memory suppression. While previous work has highlighted the role of a prefrontal–hippocampal circuit in this process, the pathway through which top-down control influences hippocampal activity remains incompletely understood. As discussed in Chapters 3 and 4, converging anatomical and translational evidence points to the nucleus reuniens (NRe) of the thalamus as a potential mediator of this interaction. In chapter 3, we discussed the first functional evidence implicating the NRe in retrieval suppression through a mega-analytic approach, identifying consistent engagement of this region, in the context of proactive control across multiple datasets. However, those findings were limited because they were conducted within a standard MNI space using a group-level region of interest (ROI). In the present study, we build on that foundation by testing NRe engagement at the individual level using subject-specific ROIs.

Here, we use fMRI and subject-specific anatomical masks (discussed in Chapter 4) to probe NRe activity during memory suppression. This chapter forms the first phase of a within-subject design in which the same participants later completed a fear conditioning and extinction protocol. This two-part design provides a rare opportunity to test whether common neural mechanisms, particularly involving the NRe, support suppression across domains. In this chapter, we focus on the TNT task as a controlled setting for isolating neural systems involved in the deliberate inhibition of retrieval. Chapter 6 extends this investigation to the regulation of conditioned fear.

5.1 Participants

Forty-one healthy young right-handed native English speakers were recruited for the study (M=25.0 years, SD= 4.88 years; 18 males). Data from 6 participants in total was excluded from analysis: three participants because their data were extremely noisy due to movement-related noise artifacts, data from one participant was lost, and two participants did not meet the learning criteria during the test-feedback phase. All participants had normal or corrected to normal vision, normal colour perception, and reported no learning, language or attention deficits. They were all MRI compatible and did not report

psychological or neurological impairments. Ethical approval for this study was granted by the Cambridge Psychology Research Ethics Committee (CPREC; Reference: PRE.2023.068) and all participants were reimbursed at a rate of £12 per hour.

5.2 Apparatus and Experimental Design

5.2.1 Overview of Experimental Sessions

5.2.2 Think/No-Think Task

All data were collected in 2024, at the MRC Cognition & Brain Sciences Unit. All participants came in for two sessions: one session per task. The experiment was a 2-day study. Participants performed the Think/No-Think and Fear Extinction tasks on separate days, with the order of these tasks counterbalanced across participants (Figure 18). Both sessions were conducted within a 5-day window. The TNT task was designed in PsychoPy (v2023.2.1) and adapted from a prior within-lab protocol (Sankarasubramanian, 2022) with modifications to suit aims of this current study.

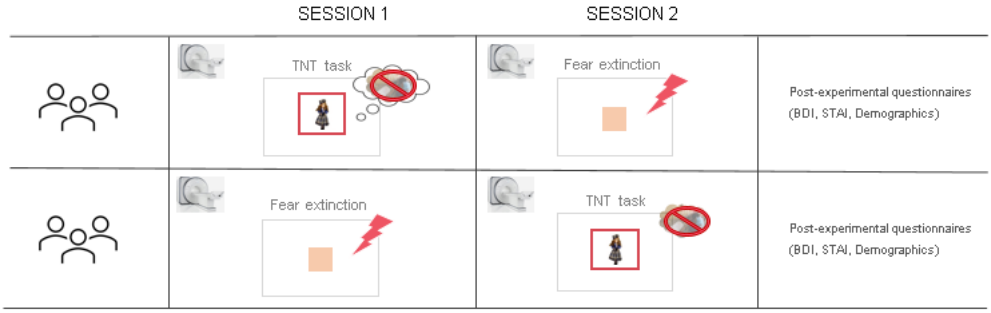

Figure 18. Schematic illustrating counterbalanced sessions on two separate days for the Think/No-Think and Fear Conditioning tasks respectively.

5.2.2.1 Stimuli

Eighty-four randomly paired cue-target pairs (e.g., a green crayon paired with a vicious looking barking dog) were employed as stimuli. The cues were familiar everyday objects, which were neutral in valence (e.g., a toothbrush, a pen, etc.) and chosen from the Bank of Standardized Stimuli repository (Brodeur et al., 2014). The targets were scenes which were either neutral or negative in valence and were chosen from the IAPS and NAPS database of images (Lang et al., 1997; Marchewka et al., 2014). Based on the Self-Assessment Manikin (SAM) valence and arousal rating scales (1= very negative, 5= neutral, and 9= very positive), half of the scenes were labelled negative with low valence scores (M = 3.87, SD = 1.54) and high arousal scores (M = 6.32, SD = 1.78). The other half of the scenes were labelled Neutral and had higher valence scores (M = 5.53, SD = 0.81), and low arousal scores (M = 1.27, SD = 0.54). The negative scenes were grouped

equally according to the emotions of disgust, fear, or sadness. Neutral scenes were inanimate objects, living beings, or landscapes. The complexity, luminance, and memorability scores for these scenes were obtained using a MemNet CNN (Khosla et al., 2015). For negative scenes, the mean CNN memory score was (M = 0.81, SD = 0.04), and for neutral scenes, the score was (M = 0.77, SD = 0.07). Scenes in each experimental condition were matched for their CNN memory score, and each condition had 50% negative and 50% neutral pairs with equal representation of feelings and categories.

Seventy-two cue-targets were critical pairs and twelve were filler pairs. The eighty-four pairs were divided into three groups counterbalanced across experimental conditions (Think, No-Think, and Baseline) with 24 pairs per group (four fillers in each group). There were 3 counterbalancing groups in total (A' to 'C'), for the 3 permutations of the 3 items sets and participants were pseudo-randomly assigned to the 3 counterbalancing groups.

Between trials, a white fixation cross was projected onto a black background. All text was presented in white colour and size 13 font for all trials.

5.2.2.2 Procedure

The TNT task consisted of an initial affect evaluation phase, a study phase, a test-feedback phase, a Think No-think phase, a surprise recall test phase and finally, a second affect evaluation phase (Figure 19).

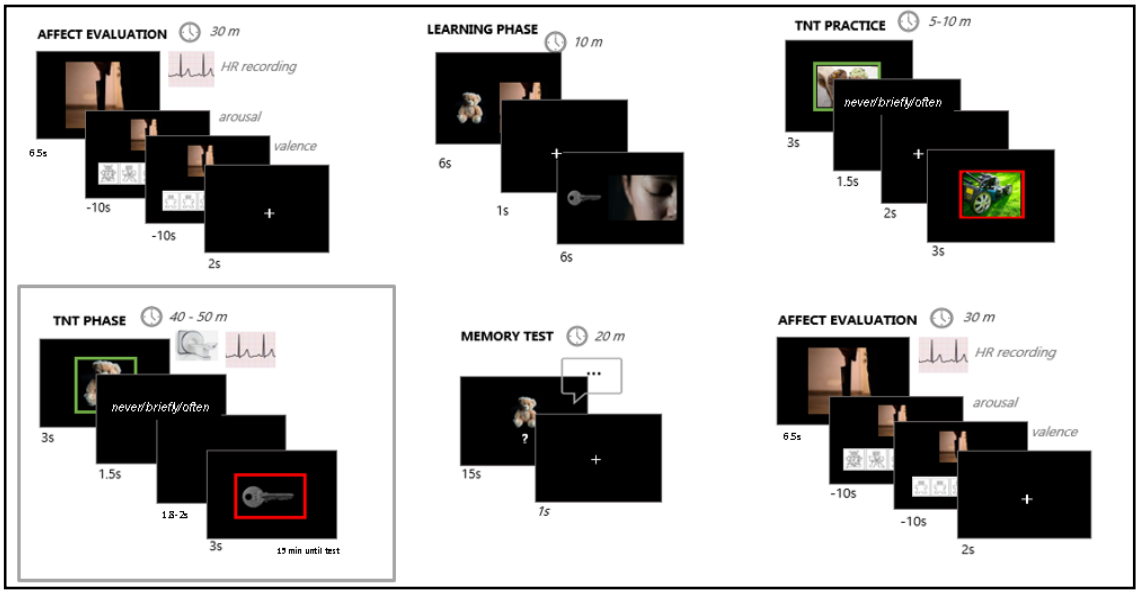

Figure 19. Schematic illustrating the various phases of the modified Think/No-Think task.

5.2.2.1 Pre-TNT Affect Evaluation

In the first phase, participants viewed all scenes sequentially, in an order optimised to ensure that each critical pair containing a negative scene did not immediately follow other

negative scene pairs. Negative filler scenes, however, could follow other critical negative scene pairs.

This sequencing was intended to reduce affective carryover effects, i.e the emotional impact of one scene influencing the perception of subsequent scenes. In particular, repeated exposure to negative content in close succession is known to lead to habituation or emotional blunting, which may dampen participants' responses and reduce the sensitivity of arousal and valence ratings. By spacing out critical negative pairs, we aimed to preserve their emotional salience and minimize confounding from prior affective states. There were no constraints on all neutral scenes; they could follow other neutral or negative scenes, regardless of whether they were critical pair scenes or fillers.

All scenes were also divided into blocks consisting of 14 scenes (critical pairs and fillers both), within each block, the order of critical negative scenes were optimised. It was also ensured that each such block had an equal number of negative and neutral scenes (critical pairs and fillers both). Each scene was presented for 6.5 seconds and participants were instructed to focus their attention and look at the screen throughout this period, even if some of these scenes were unpleasant or aversive to look at. During this period, heart rate data was collected via ECG recording electrodes, employing the BioPac ECG module, in Acknowledge (version 3.9), arranged in an Einthoven's triangle arrangement. After 6.5 seconds, the scene became smaller positioned in the middle of the screen with the SAM arousal scale (Figure 20) appearing right below it. Participants were instructed to rate their perceived arousal for each scene using a simplified 5-point scale— where 1 indicated lowest arousal, 3 indicated medium arousal, and 5 indicated highest arousal— by pressing the corresponding number key on the keyboard. Unlike the 9-point scale employed during the initial scene selection process, the 5-point scale was chosen here for its simplicity and efficiency during in-task judgments (Bradley & Lang, 1994; Grühn & Scheibe, 2008). Participants were told to make this rating without too much deliberation and within 10 seconds and the trial proceeded to the next screen as soon as a response was recorded.

After their arousal responses were recorded, the SAM valence scale (Figure 21) appeared on the screen. Like for the arousal rating, the scene was again reduced in size and centered on the screen, with the SAM valence scale displayed below it. Participants were instructed to give their valence rating for the given scene using the same 5-point scale. This rating was also made by pressing keys one to five on the keyboard. This

sequence of events continued until all 84 scenes were presented, each followed by arousal and then valence ratings using the SAM scales.

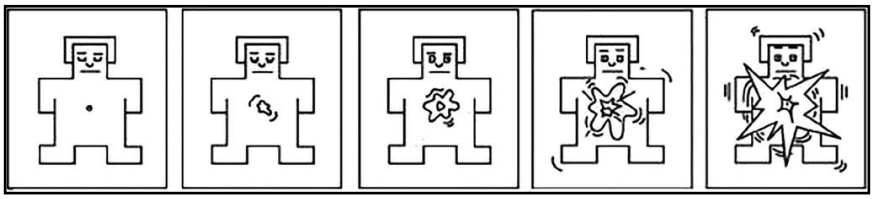

Figure 20. Self-Assessment Manikin (SAM) Arousal Scale.

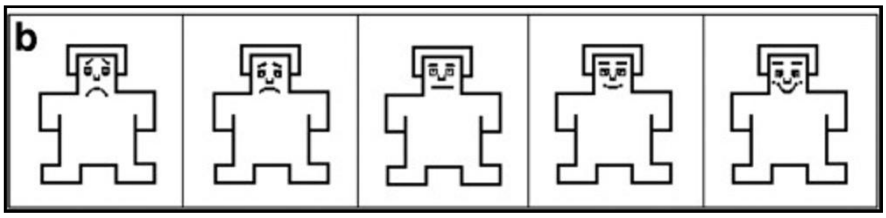

Figure 21. Self-Assessment Manikin (SAM) Valence Scale.

5.2.2.2 Learning Phase

In the study phase, participants were asked to associate neutral objects (serving as cues) with neutral or negative scenes (serving as targets). These cue-target pairs appeared together on the screen with the cue on the left side and the target on the right for 6 seconds with an Interstimulus interval (ISI) of 1 second.

5.2.2.3 Test-Feedback Phase

After participants had been exposed to all object–scene pairs, they completed a test-feedback phase. Cue objects were presented one at a time, sequentially, in the center of the screen. Participants had 4 seconds to indicate whether they knew the associated scene. They responded by pressing '1' on the keyboard to indicate that they knew the associated scene and '2' to indicate that they did not. If they pressed '1', they were then shown three scene options presented below the cue object and asked to select the correct one by pressing 'z' for the left option, 'x' for the middle, or 'c' for the right. Each trial included the target scene and two lures. The lures were matched to the target by valence category, such that negative scenes were paired with negative lures and neutral scenes with neutral lures, in order to minimize affective salience biases during selection (Talmi et al., 2008). These lures were not set according to condition (i.e., Think, No-Think or Baseline) and could come from any condition regardless of the condition of the target. The lure sets were not fixed across participants and were randomly generated for every participant. The order of these test trials were randomized and were not dependent on condition or valence. Target items from all conditions and valence categories were randomized during the test phase so as to minimize any potential order effects. Regardless of the participant's initial

response, whether they indicated knowledge of the scene or not, the correct scene was subsequently presented in a blue frame for 3 seconds before the next trial began. After all items had been tested, this procedure was repeated two additional times: once with feedback (as described above) and once without any feedback. To proceed to the next phase, participants were required to reach a learning criterion of at least 80% of correct recognition followed by accurate identification of the target from the recognition set, separately for scenes with negative valence and for those with neutral valence. Participants who did not meet this criterion were stopped from progressing to the next phase.

5.2.2.4 Practice Think/No-Think Phase

Participants completed two practice runs of the TNT task using twelve filler items. These items were distinct from the critical experimental stimuli and were used solely for practice before the main task. The first practice run did not include intrusion ratings. In this phase, an object cue filler was presented from a Think or No-Think pair for 3 seconds. This cue was surrounded with a frame which was green for Think trials and red for No-Think trials. Participants were given direct suppression instructions. Specifically, they were told to actively recall the associated scene when presented with an object cue inside a green frame (Think trials), and to suppress retrieval when the object appeared inside a red frame (No-Think trials). For No-Think trials, participants were explicitly instructed not to allow the associated scene to enter awareness. They were told to not engage in distracting thoughts or think of something else, but instead keep their focus on the cue and prevent the memory from coming to mind. If the scene did intrude into awareness, they were instructed to push it out of mind immediately and return their attention to the object cue. These instructions reflect the standard direct suppression procedure used in Think/No-Think paradigms (Anderson & Green, 2001). Following this first run, participants were administered a diagnostic questionnaire presented verbally. The questions assessed whether participants kept their eyes on the cue, avoided retrieving the scene, and refrained from using thought substitution or imaginary related strategies. Participants' responses were recorded and used to identify any misunderstandings or non-compliant strategies (e.g., thought substitution). Based on their answers, participants received immediate, personalised feedback to clarify expectations, especially the instruction to suppress the associated scene without replacing it with other material.

The second practice run also used twelve filler items and was identical in structure to the first, with the addition of intrusion ratings. After each trial, participants were

presented with an intrusion rating question: “Did the associated scene come to mind?” Participants were instructed to make this judgement rapidly and intuitively without reflecting on the original event. Participants indicated this by pressing ‘1’ for ‘Never’, ‘2’ for ‘Briefly’, and ‘3’ for ‘Often’ using the keyboard, similar to intrusion rating trials employed in TNT studies (Gagnepain et al., 2017; Benoit et al., 2015; Levy & Anderson, 2012; Mary et al. 2020). Participants had 1.5 seconds to make their response. If a response was made before the deadline, a white fixation cross appeared immediately and functioned as the inter-trial interval (ITI). The fixation remained onscreen for a brief, variable duration before the next trial began. Since the ITI began immediately after a response was registered, the total trial duration varied depending on response time.

After the second practice run, the diagnostic questionnaire was administered again. In addition to the earlier questions, participants were also asked about their experience with the intrusion rating task, specifically whether they understood the rating scale and responded intuitively as instructed. Feedback was once again provided based on their responses to confirm comprehension and to correct any deviations from the required suppression strategy. This process of providing individual feedback process ensured that all participants fully understood and correctly implemented the direct suppression instructions before proceeding to the main task

5.2.2.5 Main TNT Phase

Participants were then taken to the MRI facility where the main TNT phase was conducted while participants were lying in the scanner. Before the TNT phase commenced, participants underwent a refresher phase during which they passively viewed each object-scene pair for 6 seconds with an ISI of 1 second just as they did during the learning phase. This was done to familiarize participants with the MRI environment. After this, participants went on to TNT phase which was exactly the same as the practice phase but this time, with critical cues. Intrusion ratings were made like during the second phase of practice but this time, participants reported the frequency of their intrusions using the button box inside the scanner.

The TNT phase consisted of four blocks, containing 96 trials each. A total of 48 unique object cues (24 Think and 24 No-Think) were used and each cue was repeated 8 times across the entire phase. Object cues were repeated only after all cues in a block had been presented once. After each trial, participants provided an intrusion rating indicating the extent to which the associated scene came to mind. Participants had 1.5 seconds to respond, and if a response was made before the deadline, the rating scale was removed

immediately and the trial advanced to ITI. The ITI duration was jittered between 1800 and 2400 ms, sampled in 200 ms increments, and was not adjusted based on response time. The average duration of the TNT phase, including all trial events and breaks, was approximately 39.76 minutes (SD = 2.07 minutes, range = 36.03–44.46 minutes). Between each of the four blocks (96 trials per block), participants received a self-paced break lasting approximately 1.00–1.50 minutes, contributing to expected variability in total duration across participants.

During each block, heart rate data were continuously recorded using the Siemens pulse rate recorder, which was clipped onto the participant's left index finger. During these breaks, participants were reminded of the task instructions, including the need to maintain focus on the cue, avoid retrieval during No-Think trials, and refrain from using substitute thoughts. After the second block, the diagnostic questionnaire was re-administered to monitor strategy use and reinforce compliance with the suppression instructions.

5.2.2.6 Surprise Memory Test

After this phase, participants were taken out from the scanner and were given a surprise memory test. During this phase, every object cue was presented sequentially on the screen, this time without any green or red frame. Participants were asked to indicate if they were able to recall the associated scene which had been paired with the object. They had to indicate 'Yes' or 'No' by pressing keys 1 and 2, respectively, on the keyboard. Participants had 5 seconds to make this response. If a participant pressed '1' to indicate that they recalled the scene associated with the cue object on the screen, they had 15 seconds during which they were instructed to verbally generate a response and describe the scene in as much detail as possible. If participants pressed '2' to indicate that they could not recall the scene associated with the object on the screen, they had to wait for 15 seconds until the next trial appeared on the screen. This was done to make sure that participants did not have the inherent motivation to indicate that they did not recall the associated scene so as to shorten the task. However, in some instances, participants spontaneously recalled the associated scene during this delay period, despite initially indicating failure of memory. If the recalled scene met the scoring criteria, the response was accepted and scored accordingly. Between these trials, an ITI of 1 second was employed. These oral responses were recorded using a recorder and were later scored. The scoring was based on two criteria: a) Does the description meaningfully fit any of the scenes? b) Does the description uniquely match the correct scene, compared to all other

scenes which are part of the task? If the answer to both questions was 'Yes', then the scene was coded as having been recalled correctly, otherwise it did not count as a successful recall

5.2.2.7 Post-TNT Affect Evaluation

Finally, all participants performed the second affect evaluation phase, which was the same as the first affect evaluation Phase, including the order of presentation of the scenes. Participants were reminded to give their ratings based on how the scene made them feel at that point in time, without too much deliberation.

After this final session, participants were asked questions in order to obtain a subjective account of their experience during the TNT task, specifically during the No-Think trials. These questions included asking participants to report how often they were able to completely prevent the associated scene from coming into mind during the No-Think trials and whether this ability became easier or harder across repetitions. Participants were asked whether they suspected that they would later be tested on the No-Think scenes and if so, whether they used any strategies which helped them to retain the associations despite direct suppression instructions. Following this, participants were administered with the compliance questionnaire to assess their compliance with task instructions. Participants rated, on a scale from 0 ("Never") to 4 ("Always"), the frequency with which they engaged in three specific behaviours that violated the suppression instructions:

1. **Memory checking during the trial:** Intentionally checking whether they remembered the associated scene while viewing the No-Think cue.
2. **Memory checking after the trial:** Intentionally checking their memory for the associated scene with the No-Think cue once it disappeared from the screen.
3. **Deliberate rehearsal:** Intentionally thinking about the associated scene to strengthen their memory for it.

A non-compliance score was calculated by adding up the ratings to these three questions, yielding a score ranging from 0 to 12. This score reflects the extent to which a participant may have intentionally violated the direct suppression instructions. In accordance with recommendations from Liu et al. (2021), a cut-off score of 5 or higher was used as an indicator of significant non-compliance. This threshold was selected, based on evidence showing that SIF is observed only among participants with scores of 3 or below, whereas participants with scores of 4 or more show diminished or absent SIF. Therefore, participants scoring 5 or above were considered non-compliant, and their data

were eliminated from analysis. However, no participants in this present study met the exclusion criterion, and thus no data were removed for non-compliance.

Participants also completed a demographic and background questionnaire. Standard demographic items included age, gender, handedness, native language, and level of education.

5.3 fMRI Acquisition and Analysis

5.3.1 MRI Parameters

All scanning took place in the MRC CBU 3T Siemens Magnetom Prisma MRI system using a 32-channel whole head coil. Functional data were acquired using a multi-band gradient-echo, echo-planar pulse sequence (EPI; 192×192×96; 2 mm^3 isotropic voxels; repetition time = 1762 ms; echo time = 31.6 ms; flip angle = 67 degrees; slice number= 48 horizontal slices; slice order = interleaved slice acquisition; multi-band acceleration factor = 2).

5.3.2 Pre-processing and Modelling

Statistical Parametric Mapping (SPM12, University College London, London, UK; http://www.fil.ion.ucl.ac.uk/spm/software/spm12/) in Matlab 2023b (Mathworks, MA, USA) was used to perform the fMRI analysis. All acquired images were realigned, across and within the 4 runs, and slice-timed corrected to the first slice acquired in time. The images were co-registered in two steps: 1) functional images were rigid-body co-registered to structural images, and 2) both were then co-registered to SPM's T1 MNI template. Next, the structural image was segmented and normalised to MNI space, and the same normalisation warps applied to the functional images. For the whole-brain analysis, these normalized images were smoothed by 8 mm FWHM. For the subsequent ROI analysis, smoothing was omitted. This decision was intentional: smoothing could lead to the blurring of anatomical boundaries and the misattribution of signal from neighbouring regions which could translate to reduced spatial specificity.

5.3.3 ROI Definitions

5.3.3.1 Right Dorsolateral Prefrontal Cortex (rDLPFC)

The rDLPFC ROI was functionally defined based on a meta-analytic conjunction analysis combining the TNT and Stop-signal tasks as reported in Apšvalka et al. (2022). The rDLPFC mask corresponded to the anterior portion of Brodmann areas 9, 10, and 46, derived from the ALE conjunction map in MNI space. Its role in domain-general

inhibitory control has been validated through multivoxel pattern analysis (MVPA) and dynamic causal modeling (DCM).

5.3.3.2 Right Ventrolateral Prefrontal Cortex (rVLPFC)

The rVLPFC ROI was also defined from the Apšvalka et al. (2022) meta-analysis. This region encompassed Brodmann areas 44 and 45, extending into the anterior insula. Like the rDLPFC, this ROI was identified based on consistent recruitment across both action and memory stopping tasks and validated through MVPA and DCM.

5.3.3.3 Anterior Cingulate Cortex (ACC)

The ACC ROI, specifically the dorsal anterior cingulate cortex (dACC) was functionally defined based on conjunction analyses was functionally defined based on a conjunction map of the No-Think > Think and Stop > Go contrasts (Apšvalka et al.,2022). This map was derived from ALE meta-analyses of independent fMRI datasets and identified ACC subregions consistently co-activated during both memory inhibition and action cancellation

5.3.3.4 Inhibitory Control Network

To assess domain-general inhibitory control at a network level, we additionally employed a composite Inhibitory Control Network ROI provided by Apšvalka et al. (2022). This ROI integrates regions consistently recruited across both action cancellation (Stop > Go) and memory inhibition (No-Think > Think) tasks. These regions are:

- Right dorsolateral prefrontal cortex (BA 9/10/46)
- Right ventrolateral prefrontal cortex and anterior insula (BA 44/45)
- Supplementary motor area (BA 6/8)
- Right precentral gyrus (BA 6)
- Anterior cingulate cortex (ACC)
- Right basal ganglia

These regions were identified as part of a domain-general inhibitory control network, primarily within the cingulo-opercular network, with some overlap into the frontoparietal network.

5.3.3.5 Hippocampus

The hippocampus ROI was derived from the 10-study Mega-TNT dataset, which included 330 manually traced hippocampi, each traced in the participant's native space. These masks were normalized to 1×1×1 mm MNI space, binarized, and summed across participants. The resulting overlap map was divided by 330 and multiplied by 100 to yield

a voxel-wise percentage overlap. The final bilateral hippocampus mask included only voxels identified as hippocampus in at least 20% of subjects (i.e., ≥66 participants). We focused on the anterior hippocampus given its preferential role in emotional memory encoding, contextual fear learning, and its involvement in top-down suppression mechanisms, particularly when aversive or motivationally salient stimuli are involved (Anderson & Floresco, 2022).

5.3.3.6 Amygdala

The amygdala ROI was anatomically defined bilaterally using the Harvard-Oxford Subcortical Structural Atlas, a probabilistic atlas derived from manually segmented amygdalae in 21 participants, registered to MNI152 space. The resulting map was thresholded at 50% to include voxels labelled as amygdala in at least half the sample.

5.3.3.7 Bilateral Nucleus Reuniens (NRe)

The NRe ROI was defined bilaterally for each participant using the manual segmentation procedure developed in Chapter 4. Guided by conserved anatomical landmarks, including the third ventricle, interthalamic adhesion, and internal medullary lamina, this method enables voxel-level localization of the NRe within each participant's native-space T1-weighted MRI, while minimizing signal contamination from adjacent thalamic structures. Segmentation was performed independently by two trained raters using a custom MATLAB interface, followed by interpolation and quality-controlled mask refinement.

5.3.4 fMRI Univariate Analysis

5.3.4.1 Common Analysis parameters

A general linear model (GLM) was constructed for each participant to model BOLD responses during the tasks. Neural predictors (described for each task below) were convolved with a canonical hemodynamic response function (HRF). Six further nuisance regressors representing head motion (three translations and three rotations) were included to account for residual motion-related artifacts. The GLM was fit using an autoregressive AR(1)-plus-white-noise model of autocorrelated error, with a high-pass filter (cutoff = 1/128 Hz) implemented via a set of discrete cosine basis functions.

For whole-brain analyses, the GLM was estimated voxel wise. For region-of-interest (ROI) analyses, a single time course was extracted using the first temporal component of a singular value decomposition (SVD) across all voxels in the ROI. Parameter estimates for each event type were averaged across runs, weighted by the

number of events contributing to each condition in each run. These parameter estimates were then contrasted between conditions, as detailed for each task below.

5.3.4.2 Details specific to the TNT analysis

Events were modeled as 3-second boxcar functions, corresponding to the duration the object cue remained on screen. Separate regressors were included for Think, Intrusion, and Non-Intrusion trials, each for both negative and neutral scenes. An additional regressor captured motor responses from button presses.

Although both negative and neutral conditions were modelled in the GLM, the primary analytic goal was not to examine the effects of emotional valence. Therefore, no statistical contrasts between negative and neutral conditions were conducted in neuroimaging analyses. However, to enable direct comparison with the Fear Conditioning (FC) task in which we were particularly interested in the stimuli associated with the aversive shock—separate contrasts were also constructed using only the negative TNT trials. These contrasts allowed us to isolate suppression-related neural activity under affectively matched conditions across both tasks.

The following contrasts were tested:

1. No-Think > Think: to replicate previous results within the inhibitory network
2. No-Think > Think (Negative only): as above, but to more closely match with the FC task
3. Intrusions > Non-Intrusions: to replicate previous results of reactivation of memories within the inhibitory network
4. Intrusions > Non-Intrusions (Negative only): as above, but to more closely match with the FC task

5.3.5 Second-Level Analysis

For each of the task-specific contrasts above, the contrast estimate for each participant was entered into a one-sample T-test against zero. A conjunction analysis was used to identify common voxels across tasks. For voxel-wise analysis, FDR correction for multiple comparisons was used, unless otherwise stated. For ROIs, significance levels were Bonferroni corrected for the number of ROIs.

5.4 TNT Results

5.4.1 Behavioural Results

5.4.1.1 Pre-TNT v/s Post-TNT differences for arousal ratings

Participants viewed neutral and negative scenes from the object-scene pairs and made arousal and valence ratings, as described earlier. These ratings were obtained before and again after the TNT task in the scanner. Depending on the counterbalancing group (A, B or C), scenes from these object-scene pairs were either assigned to the Think, No-Think or Baseline condition, and segregated into Negative and Neutral categories.

The pre-TNT arousal and valences scores were subtracted from the post-TNT scores to test the prediction that arousal and valence will decrease for No-Think scenes as compared to scenes in the Baseline condition. We predicted that this decrease in the No-Think condition would be more pronounced for Negative scenes as compared to Neutral scenes.

5.4.1.1.1 Arousal Ratings Decreased Across Time, and Were Higher for Negative Stimuli

Figure 22 presents mean pre-to-post arousal change (Post – Pre) for each condition, with individual participants plotted as points. Negative values indicate a reduction in arousal from pre-TNT to post-TNT.

Across all three conditions (Baseline, No-Think, Think), arousal ratings tended to decrease slightly (mean change values < 0), with this reduction numerically larger for negative scenes than neutral scenes. A 2x2x2x3 linear mixed-effects ANOVA on raw arousal scores was conducted with within-participants factors of Time (Pre vs Post TNT), Valence (Negative vs Neutral scene) and Condition (Baseline vs No-think), and between-participant factor of Counterbalancing. Fixed effects were tested using Satterthwaite-approximated dfs, resulting in non-integer df values. For this analysis, the sample size was n=33 because post-TNT rating data was lost for 3 participants.

There was a significant main effect of Time, $F(1, 2918) = 31.04$, $p < .0001$, $\eta^2 = .033$ which indicated a reduction of arousal ratings from pre-TNT ($M = 2.59$, $SD = 0.73$) to post-TNT ($M = 2.35$, $SD = 0.75$). A significant main effect of valence was also found $F(1, 2906) = 122.78$, $p < .001$, $\eta^2 = .056$. Negative stimuli ($M = 2.67$, $SD = 0.78$) were rated as more arousing compared to neutral stimuli ($M = 2.21$, $SD = 0.66$).

5.4.1.1.2 No Evidence for Differential Arousal Reduction by Valence

The interaction between Time and Valence was not significant $F(1, 2906) = 1.55$, $p = .214$. This suggests that the reduction in arousal from pre to post TNT did not differ between

negative and neutral scenes. Though the main effects of time and valence were significant, there was no evidence for a pre-post arousal reduction based on valence.

5.4.1.1.3 Reduced Arousal Trends in the No-Think Condition Demonstrated the Effect of Suppression-Induced Forgetting

The main effect of Condition was not significant $F(1, 2906) = 1.89$, $p = .169$ but exhibited a non-significant trend. Arousal ratings were slightly higher for baseline stimuli as compared to no-think stimuli. The means showed a trend consistent with a suppression effect, but this did not reach significance. The interactions between Valence x Condition, $F(1, 2906) = 1.07$, $p = .301$ and between Time x Condition, $F(1, 2906) = 0.00$, $p = .986$ were both not significant.

5.4.1.1.4 Effects Varied Across Counterbalancing Groups

Several higher-order interactions involving Counterbalancing group were significant, indicating variability in the direction and magnitude of suppression effects depending on stimulus assignment. For instance, Time × Counterbalancing, $F(2, 2918) = 3.69$, $p = .025$, $\eta^2 = .010$, Valence × Counterbalancing, $F(2, 2906) = 13.12$, $p < .001$, $\eta^2 = .020$ and the Condition × Counterbalancing, interaction $F(2, 2906) = 10.86$, $p < .001$, $\eta^2 = .018$. These effects reflect item-specific differences across counterbalanced stimulus sets rather than meaningful psychological effects. Counterbalancing was included to distribute such item variance across participants, and significant interactions indicate that some stimulus sets were inherently more arousing or negative than others. Since counterbalancing groups do not correspond to any theoretical manipulation, these interactions are reported for completeness but are not interpreted further.

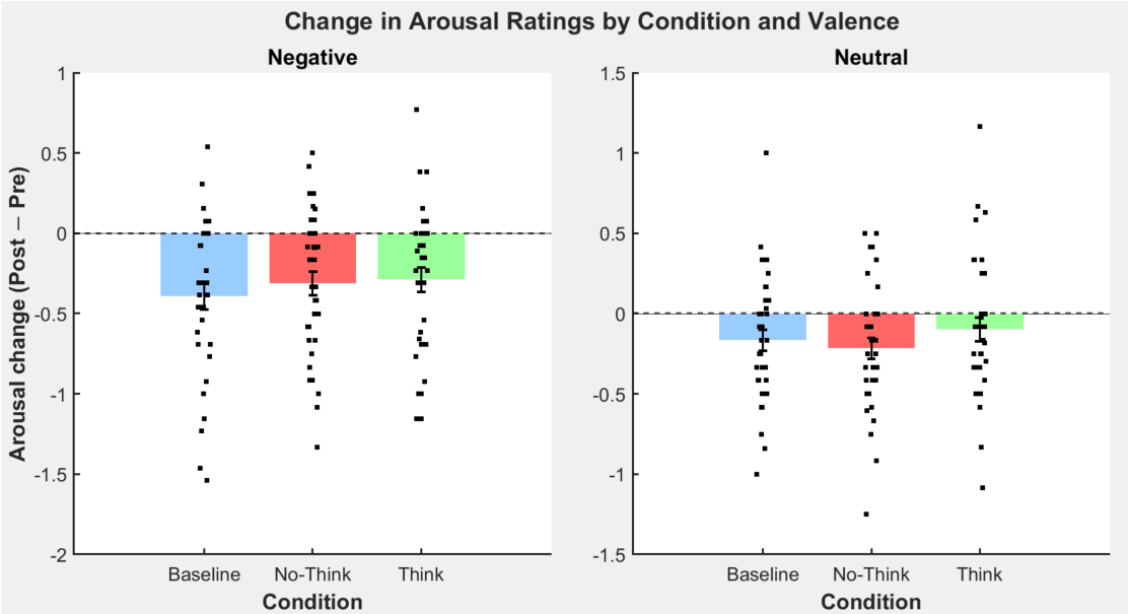

Figure 22. Change in arousal ratings (Post – Pre) for Negative and Neutral scenes across the Baseline, No-Think, and Think conditions. Bars show mean arousal change and error bars show ±1 SEM. Black points represent individual participants' change scores. Negative values indicate reduced arousal from pre to post-TNT.

5.4.1.2 Pre-TNT v/s Post-TNT differences for Valence ratings

5.4.1.2.1 Valence Ratings Remained Stable across Time and Conditions

Figure 23 presents mean pre-to-post valence change (Post – Pre) for each condition, with individual participants plotted as points. Positive values indicate an increase in valence ratings (i.e., scenes judged as slightly more positive), whereas negative values indicate a reduction. Across all three conditions (Baseline, No-Think, Think), valence ratings remained relatively stable, with only small changes observed from pre-TNT to post-TNT. These changes did not show a systematic pattern across experimental conditions or valence categories.

Another 2 Time (Pre vs Post TNT) × 2 Valence (Negative vs Neutral scene) × 2 Condition (Baseline vs No-think) x 3 (Counterbalancing group) linear mixed-effects ANOVA was conducted on raw valence scores this time. Again, fixed effects were tested using Satterthwaite-approximated dfs, resulting in non-integer df values.

There was no significant main effect of Time, $F(1, 2948) = 0.58$, $p = .447$, indicating that valence ratings did not significantly differ from pre to post TNT. However, there was a significant main effect of Valence, $F(1, 2919.35) = 54.33$, $p < .001$, $\eta^2 = .019$, with negative stimuli ($M = 2.77$) rated less positively as compared to neutral stimuli ($M = 3.06$), regardless of time or condition.

5.4.1.1.2 No Evidence for Time × Valence Interaction

The Time × Valence interaction was not significant, $F(1, 2919.35) < 0.01$, $p = .996$, suggesting that valence ratings for negative and neutral stimuli did not change from pre- to post-TNT. This result indicates that the TNT manipulation did not significantly alter the emotional valence of the stimuli.

5.4.1.1.3 Items in Baseline and No-Think groups Did Not Differ in their Valence Ratings

The main effect of Condition was not significant, $F(1, 2919.37) = 0.24$, $p = .621$ indicating that valence for No-Think items did not differ and emotional tone did not dampen as compared to items in the Baseline group. Similarly, there were no significant Time × Condition or Valence × Condition interaction.

5.4.1.1.4 Counterbalancing Influenced Valence Ratings

Although several interactions involving Counterbalancing group reached significance including a Valence X Counterbalancing, $F(2, 2919.35) = 10.29$, $p < .001$, $\eta^2 = .007$ and a Condition × Valence × Counterbalancing interaction, $F(2, 29) = 10.97$, $p < .001$, $\eta^2 = .031$ and lastly, a Valence × Condition × Counterbalancing: $F(2, 2919.35) = 6.25$, $p = .002$. They reflect item-specific variability across the counterbalanced stimulus sets rather than theoretically meaningful psychological effects. Counterbalancing was included specifically to distribute such item variance across participants, and therefore these interactions are not interpreted further.

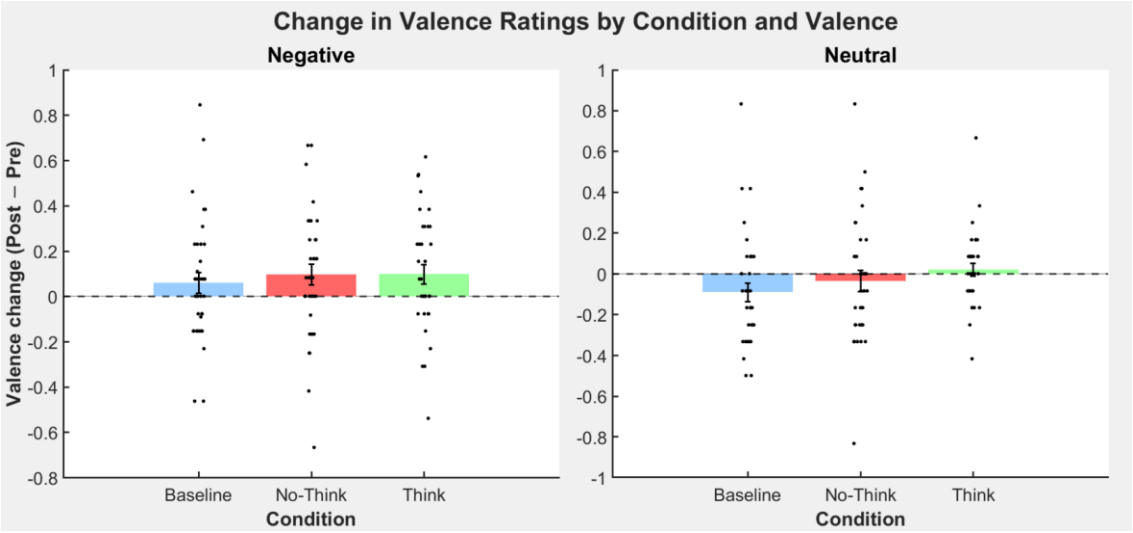

Figure 23. Change in valence ratings (Post – Pre) for Negative and Neutral scenes across the Baseline, No-Think, and Think conditions. Bars show mean valence change and error bars show ±1 SEM. Black points represent individual participants' change scores. Values near zero indicate that valence ratings remained largely stable from pre- to post-TNT, with no consistent differences across conditions.

5.4.2 Suppression Reduces Intrusion Frequency: Overall and Across Time

Participants reported whether the associated scene came to mind following each trial during the TNT task, using a 3-point scale ("never," "briefly," or "often"). Following established TNT protocols (Benoit et al., 2015; Gagnepain et al., 2017; Levy & Anderson, 2012; Mary et al. 2020) any response indicating awareness ("briefly" or "often") was classified as an intrusion.

As shown in Figure 24, participants reported significantly less awareness of the associate on No-Think trials (M = 28.91%, SE = 2.25) than on Think trials (M = 83.07%, SE = 1.74), $t(34) = 14.65$, $p < .001$. These values reflect aggregate intrusion frequencies averaged across all TNT repetitions. This large effect demonstrates that participants were able to modulate memory retrieval based on task demands, replicating a hallmark behavioural signature of the TNT paradigm.

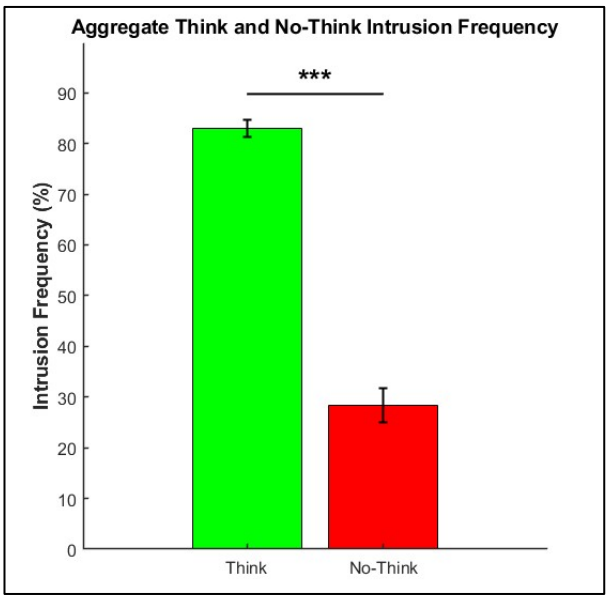

Figure 24. Aggregate intrusion frequency for Think and No-Think trials across repetitions during the TNT task.

To assess suppression-related change across repetitions, we analyzed No-Think intrusion frequency over time. As demonstrated by Figure 25, participants exhibited a significant reduction in No-Think intrusions from the first run (M = 42.79%, SE = 3.88) to the final run (M = 20.62%, SE = 3.50), $t(34) = 6.46$, $p < .001$. This decrease suggests that repeated suppression attempts increased participants' ability to prevent unwanted memories from entering awareness, consistent with prior findings (Hellerstedt et al., 2016; Gagnepain et al. 2017; Levy & Anderson, 2012; Van Schie & Anderson, 2017). Consistent with this, a repeated-measures ANOVA revealed a main effect of repetition,

F(7, 245) = 21.66, p < .001, confirming that intrusion frequency declined significantly across suppression repetitions.

We also tested whether valence modulated intrusion frequency using a 2 (Valence: Negative, Neutral) × 8 (Repetition) repeated-measures ANOVA. There was no significant main effect of valence, F(1, 35) = 0.38, p = .541, indicating that negative (M = 28.26%, SE = 3.46) and neutral (M = 27.24%, SE = 3.39) scenes elicited comparable intrusion rates. The Valence × Repetition interaction was also not significant, F(7, 245) = 0.59, p = .762, showing that negative and neutral items exhibited similar intrusion-decline trajectories over time. Although our neuroimaging analyses focus on collapsed or negative-only trials, we display intrusion frequency separately for neutral and negative items (Figure 25) to demonstrate that the decline in intrusions over time is not driven by either valence condition alone.

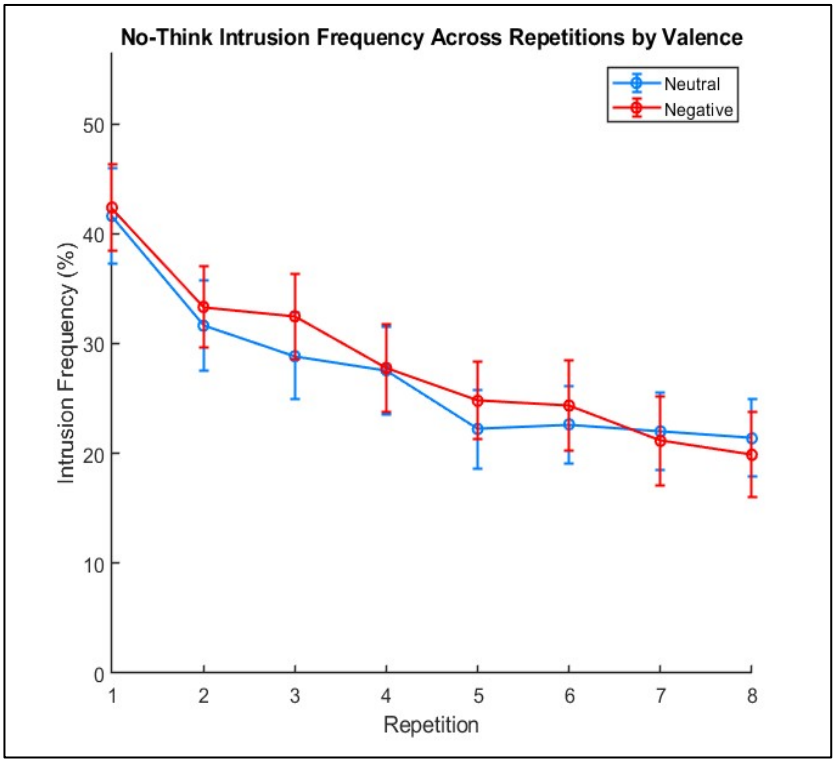

Figure 25. No-Think intrusion frequency across repetitions as a function of scene valence.

5.4.3 Suppression Reduces Final Recall Performance

Recall performance from the final memory test is shown in Figure 26. To evaluate suppression-induced forgetting (SIF) across all items, a 2 (Condition: No-Think, Baseline) × 2 (Valence: Negative, Neutral) × 3 (Counterbalancing Group: A, B, C) mixed-effects ANOVA was conducted. This analysis revealed a significant main effect of Condition, F(1, 1718) = 55.60, p < .001, η^2 = .032, indicating reduced memory for No-

Think items compared to Baseline items. No significant main effects of Valence, F(1, 1718) = 0.13, p = .716, or Counterbalancing Group, F(2, 1718) = 0.22, p = .805, were found. Furthermore, no significant interactions emerged between Condition and Valence, F(1, 1718) = 0.12, p = .727, Condition and Counterbalancing, F(2, 1718) = 2.15, p = .117, or Valence and Counterbalancing, F(2, 1718) = 1.52, p = .219.

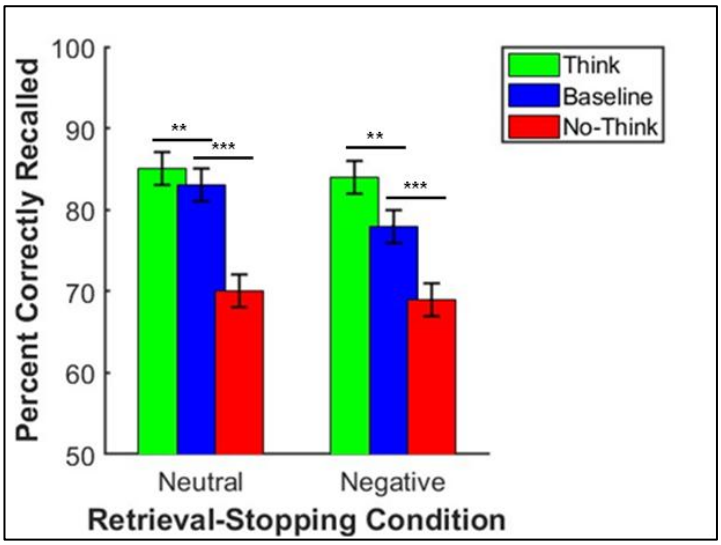

Figure 26. Final recall accuracy as a function of retrieval condition and stimulus valence.

To enable direct comparison with the CS+ condition of the Fear Conditioning (FC) task, we also constructed separate contrasts using only the negative TNT trials. Nevertheless, in Figure 24, we display recall performance for both neutral and negative scenes to demonstrate that suppression and facilitation effects are observed across both valence conditions. This ensures that the observed memory effects are not driven solely by either neutral or negative stimuli but instead reflect a generalizable pattern across affective content. While the main analyses focus on collapsed data, presenting both valences provides confidence that neither valence particularly drives the effects.

Follow-up comparisons confirmed both facilitation and suppression effects. These contrasts were tested collapsed across valence, revealing that recall was significantly better for Think items. As expected, recall was significantly better for Think items as compared to Baseline items, ($t(34) = 2.52$, $p = .012$), indicating retrieval-induced facilitation. In contrast, recall was significantly worse for No-Think items compared to Baseline items, ($t(34) = –4.80$, $p < .001$), confirming suppression-induced forgetting. These results replicate the core pattern observed in prior TNT studies (Anderson & Green, 2001) demonstrating the ability to suppress memory through motivated retrieval control. Although cardiac data were collected, they are not analysed in the present thesis and will not be discussed further.

5.4.4 Neuroimaging Results

For planned contrasts, we first conducted ROI analyses to test regions previously implicated in retrieval suppression and reactive control during memory intrusions, applying Bonferroni correction for the n ROIs tested contrasts within each ROI ($\alpha = .05/7 = 0.0071$). To check for any additional effects across the brain, we then report whole-brain analyses corrected for multiple comparisons across voxels using FDR $p<.05$.

5.4.4.1 Suppressing Retrieval Engaged Canonical Inhibitory Control Networks

We hypothesized that direct suppression would recruit a set of prefrontal control regions known to support retrieval stopping, including right anterior DLPFC, VLPFC, ACC, and regions within the broader cognitive control network. Consistent with this, ROI analyses revealed significantly greater activation for No-Think > Think trials in DLPFC ($t = 5.69$, $p < .001$), VLPFC ($t = 9.57$, $p < .001$), ACC ($t = 9.92$, $p < .001$), and the control network ($t = 8.53$, $p < .001$), all of which survived Bonferroni correction. These results replicate prior work showing robust prefrontal engagement during memory suppression (Apšvalka et al., 2022).

The HpC showed a suppression-related decrease ($t = –2.52$, $p = .017$), though this did not survive correction for the number of ROIs tested. Amygdala activation was not significant ($t = –1.56$, $p = .127$), consistent with previous findings suggesting that suppression does not broadly downregulate affective regions when collapsing across valence (Gagnepain et al., 2017).

Crucially, the NRe also showed greater activation during No-Think compared to Think trials ($t = 2.18$, $p = .036$). Although this effect was significant at the uncorrected threshold, it did not survive Bonferroni correction for the number of ROIs tested. To complement the frequentist analyses, we conducted a Bayesian one-sample t-test using a directional prior consistent with the hypothesis. This analysis yielded evidence that was directionally consistent with the predicted effect, with the Bayes factor indicating that the observed increase in NRe activation was approximately 1.5 times more likely under a model assuming greater activation during suppression than under the null hypothesis ($BF_{10} = 1.47$). The posterior median effect size was positive and modest ($\delta = 0.34$, 95% CI [0.01, 0.68]), and the robustness check showed that across a range of reasonable prior widths, the evidence consistently favoured the presence of an effect, even though its strength remained in the anecdotal range. Thus, the Bayesian analysis aligns with the frequentist findings in supporting the predicted direction of the effect, suggesting that the

NRe is likely engaged during retrieval suppression in this dataset, albeit with limited evidential strength.

Nonetheless, this finding converges with the results from Chapter 3, offering further support for the role of the NRe in prefrontal–hippocampal inhibitory control during retrieval suppression.

Whole-brain analyses revealed significantly greater activation for No-Think> Think in a few other regions including the right insula, right Frontal Eye Fields and left insula, as well as occipital and parietal areas (Table 6). These findings suggest the engagement of a broader network, consistent with prior work implicating the insula and FEF in salience detection, task-set switching, and inhibitory control processes (Menon & Uddin, 2010).

Region	X	Y	Z	Cluster Size (voxels)	Peak t-value	p_FDR
Right Insula	38	26	-2	2920	7.84	p < .001
Right Frontal Eye Fields	6	24	36	2787	7.55	p < .001
Left Insula	-40	14	0	1126	6.54	p < .001
Left Visual Association Area	-42	-72	-6	1032	5.65	p < .001
Right Visual-Motor Area	20	-68	50	1014	5.37	p < .001
Right Fusiform Gyrus	44	-62	-10	262	5.22	0.035

Table 6. Regions showing greater activation for No-Think compared to Think trials (NT >T). Significant clusters from the whole-brain univariate contrast of No-Think > Think, thresholded at p < .05, FDR-corrected. Coordinates refer to peak voxel locations in MNI space.

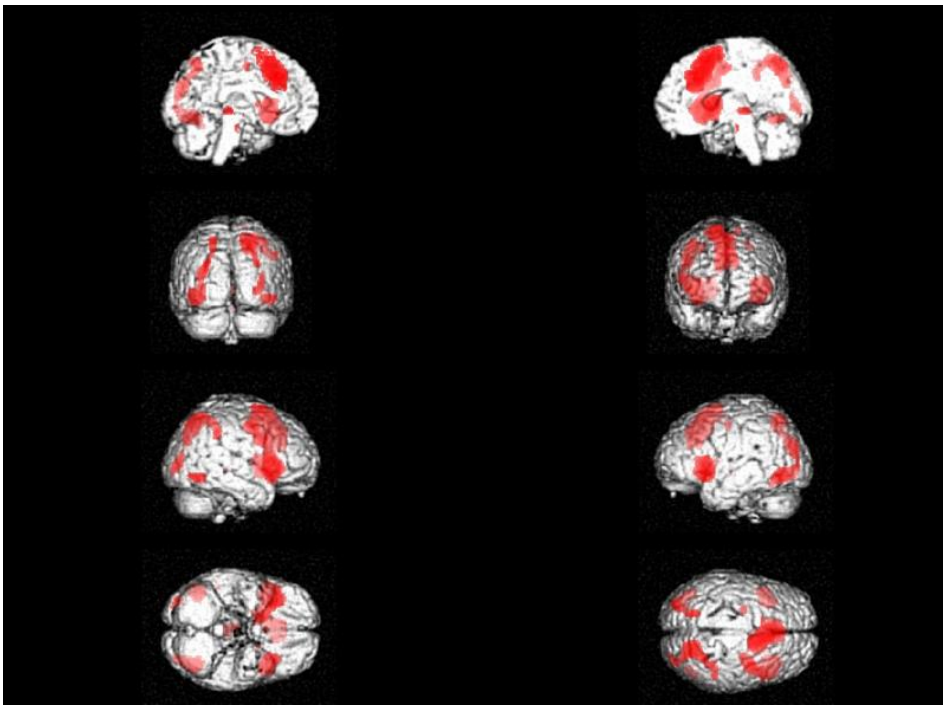

Figure 27. Brain regions showing greater activation for No-Think compared to Think trials (NT > T).

Surface renderings display significant clusters from the whole-brain univariate contrast thresholded at $p < .05$, FDR-corrected. Red clusters indicate regions with significantly increased activation during memory suppression. Coordinates correspond to peak voxel locations in MNI space (see Table 7 for details).

5.4.4.2 Control Networks Were Reactively Engaged During Intrusions

If the control system is re-engaged when suppression fails, then we would expect increased activity in control-related regions during intrusions relative to non-intrusions. ROI analyses confirmed this prediction: in DLPFC ($t = 3.69$, $p < .001$), VLPFC ($t = 4.62$, $p < .001$), ACC ($t = 4.88$, $p < .001$) and the control network ($t = 4.02$, $p < .001$), all surviving Bonferroni correction.

The HpC also showed increased activation during intrusions ($t = 2.89$, $p = .007$), consistent with the interpretation that intrusions reflect reactivation of unwanted memory traces. However, the simultaneous engagement of prefrontal regions, including the rDLPFC, rVLPFC, and ACC along with the broader inhibitory control network suggests that these intrusions were not passively experienced but instead triggered reactive attempts to suppress the retrieved content.

In contrast, the amygdala did not show significant activation during intrusions ($t = 1.53$, $p = .137$), suggesting that the intrusions, on average, did not elicit strong affective responses, or that emotional salience was not reliably reactivated. This was not surprising, given that the stimuli were not particularly arousing or strongly negative.

Crucially, the NRe also showed significantly greater activation for Intrusions > Non-Intrusions ($t = 3.48$, $p = 001$), surviving Bonferroni correction. This finding indicates that the NRe is not only engaged during local suppression attempts (as seen in No-Think trials) but is also recruited reactively when control must be reasserted in response to emergent mnemonic conflict. To complement these frequentist findings, we conducted a Bayesian one-sample t-test using a directional prior consistent with the hypothesis that intrusions elicit greater NRe activation than non-intrusions. The Bayesian analysis yielded strong evidence for the predicted effect ($BF_{10} = 23.36$), indicating that the observed increase in NRe activation during intrusions was far more likely under the alternative hypothesis than under the null. The posterior distribution supported a positive and reliable effect, with a median effect size of $\delta = 0.55$ (95% CI [0.20, 0.91]). Importantly, the Bayes Factor robustness check showed that evidence in favour of H_1 remained strong across a broad range of prior widths, bolstering the stability of the inference. Together, these Bayesian results converge with the frequentist analyses, providing strong support for the conclusion that the NRe is robustly engaged during intrusive recollection and participates in reactive control processes when suppression fails.

Whole-brain univariate analyses also revealed greater activation for Intrusions compared to Non-Intrusions in regions including the left insula, right inferior frontal gyrus, right frontal eye fields and parietal and visual-motor areas (Table 7). This is consistent with prior work which has implicated the right inferior frontal cortex in response stopping (Aron et al., 2014).

Region	X	Y	Z	Cluster Size (voxels)	Peak t-value	p_FDR
Left Insula	-34	18	4	1164	6.29	p < .001
Left Visuomotor Cortex	-4	-72	32	3414	5.86	p < .001
Right Inferior Frontal Gyrus (IFG)	34	14	-2	777	5.85	0.001
Right Frontal Eye Fields (FEF)	0	26	50	2600	5.74	p < .001
Left Angular Gyrus	-50	-52	36	1112	5.32	0.056

Table 7. Regions showing greater activation for Intrusions compared to Non-Intrusion trials (I > NI).
Significant clusters from the whole-brain univariate contrast of Intrusions > Non-intrusions, thresholded at p < .05, FDR-corrected. Coordinates refer to peak voxel locations in MNI space.

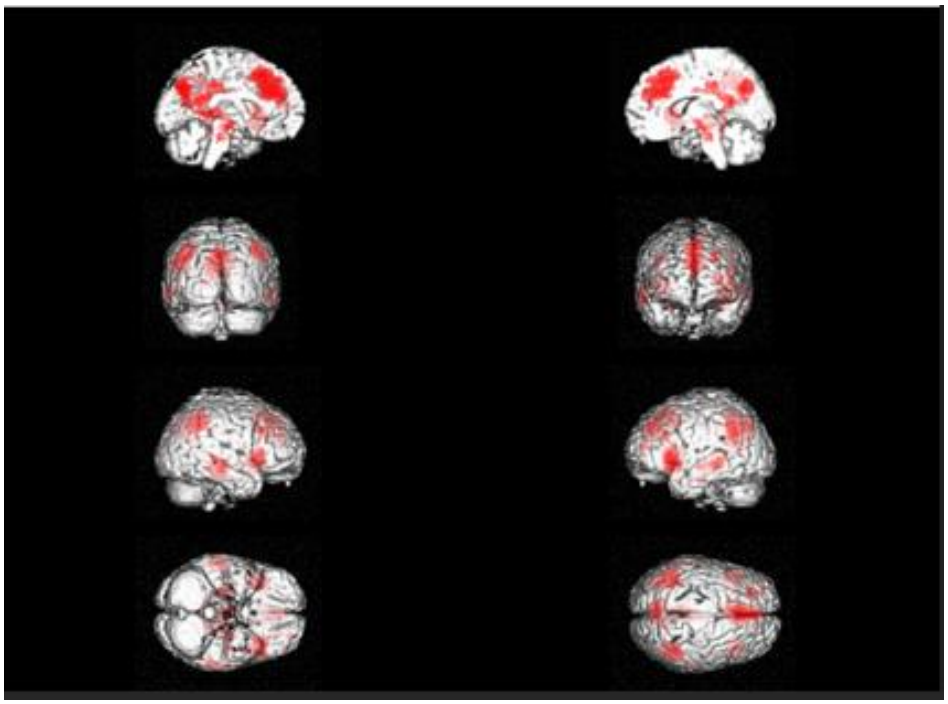

Figure 28. Brain regions showing greater activation for Intrusions compared to Non-Intrusions (I > NI).
Surface renderings display significant clusters from the whole-brain univariate contrast, thresholded at p < .05, FDR-corrected. Red clusters indicate regions with significantly increased activation during memory suppression. Coordinates correspond to peak voxel locations in MNI space (see Table 7 for details).

5.4.4.3 Suppression of Negative Memories Recruited Similar Control Regions

We next examined whether suppression-related recruitment of control regions remained reliable for negative scenes alone, to allow affectively matched comparisons with the Fear Conditioning task considered later. In the No-Think > Think contrast restricted to negative trials, ROI analyses again revealed significant increases in rDLPFC ($t = 4.66$, $p < .001$), VLPFC ($t = 7.96$, $p < .001$), ACC ($t = 8.10$, $p < .001$), and the control network ($t = 6.32$, $p < .001$), all passing correction.

The HpC showed a non-significant trend toward reduced activation ($t = –1.98$, $p = .056$), while amygdala activity remained non-significant ($t = –0.80$, $p = .427$).

In contrast to the full-condition analysis, the NRe did not show significant activation in the No-Think > Think contrast for negative scenes ($t = 1.38$, $p = .178$). This result is likely due to reduced statistical power, as the number of trials was halved in this analysis. Indeed, based on the observed effect size, $d \approx 0.23$, the post hoc power for this contrast was low, ≈28% indicating that the analysis had limited sensitivity to detect effects of this magnitude. To complement these frequentist analyses, we conducted a Bayesian one-sample t-test using a directional prior consistent with the hypothesis. The Bayesian analysis provided anecdotal evidence for the null as compared to the hypothesised increase during suppression ($BF_{10} = 0.43$). Although the posterior median effect size was positive ($\delta = 0.21$), the 95% credible interval included zero [–0.11, 0.54], indicating substantial uncertainty in the magnitude of the effect. The Bayes factor robustness check further showed that the evidence remained consistently on the side of the null across a range of plausible prior widths. Thus, unlike the full-condition analysis, the Bayesian results for negative scenes alone do not support reliable suppression-related engagement of the NRe. This pattern is consistent with the reduced statistical power inherent in the valence-restricted analysis and suggests that any NRe involvement during suppression of negative memories is weaker or less consistently expressed in this dataset. To complement these findings, whole-brain analyses were also conducted, and these results are presented in Table 8.

Region	X	Y	Z	Cluster Size (voxels)	Peak t-value	p_FDR
Right Frontal Eye Fields (FEF)	8	24	38	2849	8.17	p < .001
Right Insula	34	28	0	1872	7.13	p < .001
Right Inferior Frontal Gyrus (IFG)	-32	20	8	883	5.97	p < .001
Left Visual Association Cortex	-42	-70	-8	663	5.93	0.001
Right Visuo-Motor Area	22	-66	52	580	5.22	0.002

Table 8. Whole-brain results for the No-Think compared to Think (Negative only) trials. Significant clusters from the whole-brain univariate contrast of No-Think > Think, thresholded at p < .05, FDR-corrected. Coordinates refer to peak voxel locations in MNI space.

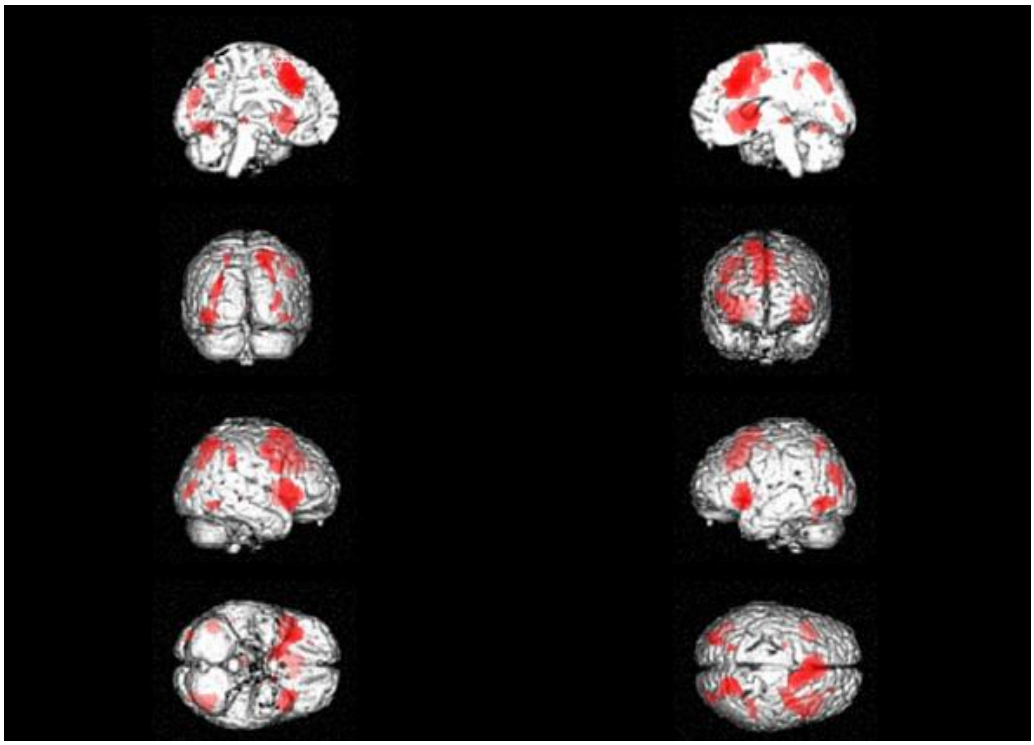

Figure 29. Brain regions showing greater activation for No-Think compared to Think trials for Negative only trials (NT > T_Negative).

Surface renderings display significant clusters from the whole-brain univariate contrast, thresholded at p < .05, FDR-corrected. Red clusters indicate regions with significantly increased activation during memory suppression. Coordinates correspond to peak voxel locations in MNI space (see Table 8 for details).

5.4.4.4 Negative Intrusions Recruited Both Control and Memory Systems

Finally, we examined the neural response to intrusions of negative memories. ROI analyses showed significantly greater activity for Intrusions than Non-Intrusions in rDLPFC ($t = 3.01$, $p = .005$), VLPFC ($t = 4.18$, $p < .001$), ACC ($t = 4.75$, $p < .001$), and the control network ($t = 3.83$, $p = .001$), all of which survived Bonferroni correction.

Both the HpC ($t = 2.49$, $p = .018$) and amygdala ($t = 2.08$, $p = .045$) showed increased activity, though these effects did not survive correction. This pattern suggests that intrusions of negative content may partially re-engage mnemonic and affective systems, even when suppression is attempted. Crucially, the NRe showed significantly greater activation for Intrusions > Non-Intrusions ($t = 3.03$, $p = .005$), surviving Bonferroni correction. This finding reinforces the role of the NRe in reactive control, demonstrating its recruitment in response to the emergence of emotionally salient intrusive memories. To complement the frequentist analyses, we conducted a Bayesian one-sample t-test using a directional prior consistent with the hypothesis. The Bayesian analysis yielded positive but modest evidence for increased NRe activation during retrieval suppression ($BF_{10} = 1.47$). The posterior median effect size was consistent with the hypothesised effect ($\delta = 0.34$, 95% CI [0.01, 0.68]), indicating a small but credible positive shift. Importantly, the Bayes Factor robustness check showed that the evidence in favour of the alternative hypothesis remained stable across a reasonable range of prior widths. Thus, the Bayesian analysis converges with the frequentist findings in supporting the predicted direction of the effect, suggesting that the NRe may be engaged during retrieval suppression in this dataset. Whole-brain results are presented in Table 9.

Region	X	Y	Z	Cluster Size (voxels)	Peak t-value	p_FDR
Left Insula	-36	20	4	1067	5.77	p < .001
Anterior Temporal Pole	34	14	-4	1872	5.26	0.012
Right Supramarginal Gyrus	46	-50	50	560	5.01	0.005
Right Frontal Eye Fields	4	28	42	1422	5.01	p < .001
Left Angular Gyrus	-42	-52	46	419	4.46	0.012

Table 9. Whole-brain results for the Intrusions compared to Non- Intrusions (Negative only). Significant clusters from the whole-brain univariate contrast of Intrusions > Non-Intrusions (Negative only), thresholded at p < .05, FDR-corrected. Coordinates refer to peak voxel locations in MNI space.

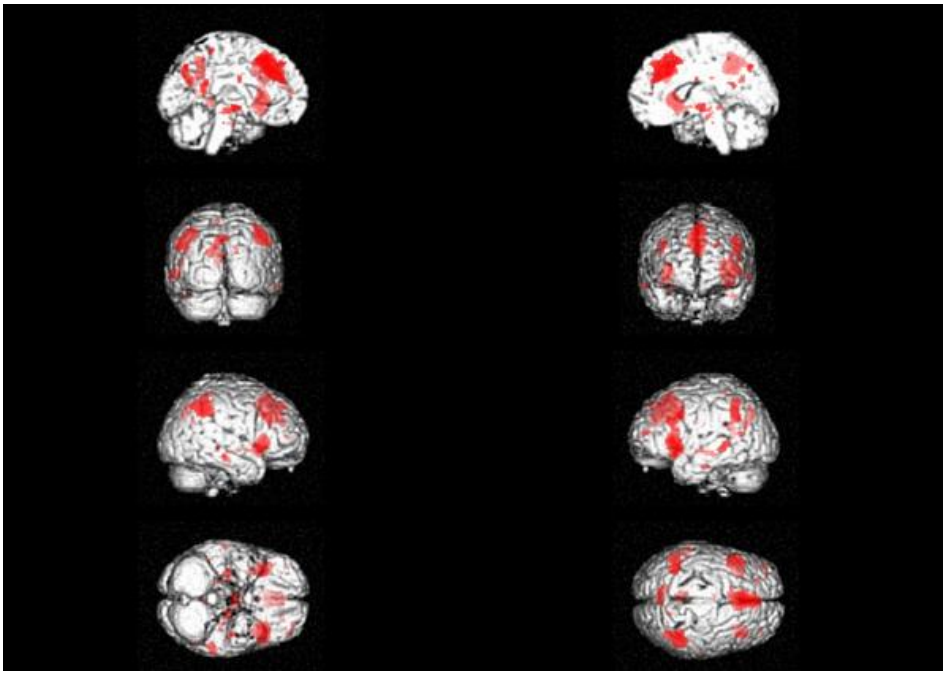

Figure 30. Brain regions showing greater activation for Intrusions compared to Non-Intrusions for Negative only trials (I >NI_Negative).

Surface renderings display significant clusters from the whole-brain univariate contrast, thresholded at p < .05, FDR-corrected. Red clusters indicate regions with significantly increased activation during memory suppression. Coordinates correspond to peak voxel locations in MNI space (see Table 10 for details).

5.5 Discussion

This study aimed to clarify the neural mechanisms underlying retrieval suppression, with a particular focus on the role of the NRe in the TNT paradigm. By combining a well-established memory control task, we examined whether the NRe is recruited during voluntary memory suppression in humans, and whether its engagement differs in reactive control contexts. These analyses extended our prior mega-analytic work (Chapter 3) and offered a more fine-grained account of the thalamic contributions to memory control.

5.5.1 Memory suppression engages a flexible control network

Participants demonstrated reliable suppression-induced forgetting, recalling fewer No-Think items than Baseline items. This behavioural pattern confirms that repeated efforts to prevent retrieval can reduce later memory accessibility. At the neural level, suppression trials elicited greater activation in prefrontal control regions, specifically the anterior rDLPFC, rVLPFC and the ACC, alongside reduced hippocampal activity.

These effects were observed regardless of negative or neutral valence. Results from both, recall nor intrusion frequency did not demonstrate a significant valence interaction, and the direction of the results were consistent across conditions. This reinforces the fact that the memory control system operates across affective contexts.

5.5.2 The NRe contributes to suppression, with stronger involvement during reactive control

A primary contribution of this study lies in advancing our understanding of NRe function in memory suppression. In chapter 3, we identified group-level NRe recruitment during suppression across datasets. Here, we used participant-specific anatomical ROIs to replicate this finding at the individual level, improving anatomical precision for this small midline thalamic structure.

The NRe showed significant activity during No-Think trials, however, this effect was not significant when analyses were limited to negative items. This could be due to reduced statistical power from halving the number of trials in this condition, which may have obscured an otherwise reliable effect. Importantly in accordance with the thalamo-hippocampal modulation hypothesis (Anderson et al., 2016) NRe activity was consistently elevated during memory intrusions, even when analyzing only negative items. This reliable engagement during suppression failure points to reactive control, where the NRe is recruited once unwanted content has intruded into awareness.

Importantly, hippocampal activity was also elevated during trials marked by intrusions, consistent with the reactivation of suppressed memories. At first instance, this might appear to contradict the suppression goal, suggesting a failure of control. However, this increase in hippocampal activity was accompanied by heightened engagement of the broader control network, including the anterior rDLPFC and rVLPFC along with the ACC. This pattern is more consistent with a reactive control response wherein the memory intrudes into awareness despite initial suppression attempts, prompting a secondary recruitment of regulatory processes to counteract retrieval.

These findings suggest that the NRe may not directly inhibit hippocampal retrieval per se but instead participate in re-engaging or coordinating control mechanisms when retrieval suppression must be reasserted. One possibility is that the NRe acts as a relay, supporting communication between prefrontal and hippocampal regions under conditions of conflict or intrusion. While our current data cannot establish causality or directional connectivity, the observed co-activation is consistent with rodent models in which the NRe facilitates top-down modulation of hippocampal activity, particularly under cognitively demanding or high-interference circumstances.

5.5.3 Advances over previous work and theoretical implications

Compared to Chapter 3, which used group-level ROIs and focused on consistent patterns across multiple datasets, the current study offers several advances. First, the use of subject-specific NRe masks enhances anatomical precision and sensitivity for detecting activity in this small midline structure. Second, by modelling memory intrusions directly within a single, standardized dataset, this chapter provides evidence for reactive NRe engagement- a pattern that was not observed in Chapter 3. While Chapter 3 did test for NRe activation during intrusions, the marginally significant result was interpreted cautiously, given variability in intrusion measures across studies and imprecise group-level NRe ROI masks. The present findings thus build on that work by identifying a more consistent pattern of NRe recruitment during retrieval failure, expanding the scope of its involvement from anticipatory control to possible re-engagement following intrusions.

The findings in this study support a dual-process model of memory suppression: one in which prefrontal regions and the NRe are engaged prevent retrieval altogether, and another in which similar systems are recruited reactively when suppression fails. The NRe may contribute to both phases, but in our study, its role appears more consistent and reliable during reactive control, under conditions of mnemonic conflict.

5.5.4 Limitations and future directions

This study has some limitations. While the use of participant-specific ROIs improved anatomical accuracy, our univariate analyses cannot address the temporal dynamics or directionality of prefrontal-NRe-hippocampal interactions. Future work using effective connectivity or dynamic modelling techniques will be necessary to determine whether the NRe initiates, modulates, or simply co-activates with control processes.

Finally, this chapter represents the cognitive half of a within-subject design that also includes affective regulation. In Chapter 6, we turn to the second phase of this study: a fear conditioning and extinction protocol designed to test whether similar neural systems, including the NRe are involved in suppressing learned emotional responses. Together, these two chapters allow us to ask whether a common set of control mechanisms supports regulation across retrieval stopping and fear extinction domains. The evidence presented here lays the foundation for that investigation by establishing the NRe as a candidate structure for facilitating memory suppression, one that may also play a broader role in the flexible control of mental content.

Chapter 6

Do Retrieval Stopping Mechanisms Suppress Fear? An fMRI Investigation of Extinction

This chapter is the second part of a within-subject investigation studying the neural mechanisms underlying inhibitory control over unwanted thoughts and fear. While Chapter 5 (Part 1) focused on the suppression of unwanted memories through the Think/No-Think (TNT) paradigm, the current chapter (Part 2) tests whether similar neural systems support the extinction of conditioned fear via a differential fear conditioning and extinction protocol.

As discussed in Chapter 2, the retrieval stopping model of extinction challenges the traditional views of extinction which states that extinction is merely associative learning and instead proposes that fear reduction and extinction rely on active, top-down suppression of aversive memory traces. According to this framework, extinction is not only about forming new safety associations, but also about suppressing the retrieval of the original fear memory, potentially engaging the same prefrontal–thalamic–hippocampal pathways involved in memory control.

To directly test this hypothesis, we have examined whether the NRe, previously implicated in memory suppression (Chapters 3 and 5), also contributes to fear extinction in humans. In the TNT task, intrusion frequencies index the success of memory suppression; in the extinction task, expectancy ratings track fear regulation. As in Chapter 5, we again used participant-specific anatomical NRe ROIs to ensure anatomical precision employing the manual segmentation pipeline as discussed in Chapter 4.

In addition to the NRe, this chapter examines activation in the right dorsolateral prefrontal cortex (rDLPFC), right ventrolateral prefrontal (rVLPFC), anterior cingulate cortex (ACC), and hippocampus (HpC), which constitute key components of the broader inhibitory control network.

6.1 Apparatus and Experimental Design

6.1.1 Fear Conditioning Task

The task was designed in PsychoPy (v2022.1.4). Shocks were delivered to the non-dominant wrist using a Digitimer DS7A electrical stimulator.

6.1.1.1 Participants

In this phase, data from sixteen participants were excluded because due to a programming error, they received an unintended shock during one CS− trial. Since this compromised the integrity of the CS− as a safety signal, these participants were not analyzed in the current study. This left a final sample of nineteen participants (n = 19) whose data met all experimental criteria. All participants had normal or corrected-to-normal vision, normal colour perception, and reported no learning, language, or attention deficits. All were MRI-compatible and reported no history of psychological or neurological disorders. Ethical approval for this study was granted by the Cambridge Psychology Research Ethics Committee (CPREC; Reference: PRE.2023.068) and all participants were reimbursed at a rate of £12 per hour.

6.1.1.2 Stimuli

The Fear Conditioning (FC) task involved the presentation of two conditioned stimuli (CS): a CS+ and a CS-. The CS+ was associated with an aversive unconditioned stimulus, a mild electric shock, delivered on 50% of CS+ trials. In contrast, the CS- was never paired with a shock. A blue circle and a yellow square served as the two CSs. For half the participants, the yellow square was designated as the CS+ and the blue circle as the CS-, whereas this assignment was reversed for the other half.

Each stimulus was displayed on a black background for 3 seconds per trial. Between stimuli, a fixation cross appeared for a variable ITI, jittered between 1800-2400ms in 200ms increments to prevent temporal predictability. The intensity of the electric shock was individually calibrated to ensure it was uncomfortable but not painful.

6.1.1.3 Procedure

Shock-naïve participants first underwent a shock thresholding procedure to determine their individual tolerance level. This was defined as the voltage at which shocks were perceived as extremely annoying but not painful, in line with methodological recommendations for human fear conditioning (Lonsdorf et al., 2017). While lying in the scanner, two adhesive scanner-compatible electrodes were attached to their left wrist, connected to the shock box.

The procedure began with a 20V shock, after which participants rated its intensity. Voltage was then gradually increased in 5V increments until participants reported the shock as extremely annoying yet importantly, not painful. To finalize this calibration, the voltage was raised one step beyond this level and participants were given a choice between the two intensities. The selected voltage remained fixed throughout the

experiment, neither increasing nor decreasing and participants were informed about the same. The shock began 100 ms after CS+ onset and lasted for 200 ms, embedded within a 1-second digital pulse sequence (50 pulses/second). Participants were not explicitly instructed about the transition from conditioning to extinction and thus the shock was delivered early enough during CS+ presentation so that participants had sufficient time to engage in voluntary memory suppression during the reminder presentation of the cue. To minimize habituation effects, the chosen shock level was received only once during the US calibration procedure.

Following calibration, participants were instructed that the experiment involved passively viewing shapes on the screen. The experiment consisted of two shapes but only a single shape appeared per trial. One shape (CS+) was occasionally paired with a shock, while the other (CS-) was never associated with a shock. The assignment of CS+ and CS- remained consistent throughout the experiment (e.g., if the blue circle was designated as CS+, it could be paired with a shock, while the yellow square never could). The assignment of the CS+ shape was counterbalanced across participants. However, due to the encountered programming error as discussed earlier, the final sample of participants is not counterbalanced.

At specific intervals, participants were asked to rate their perceived likelihood of receiving a shock when viewing a given shape (Lonsdorf et al., 2017). This rating scale ranged from 1 to 5, where: 1 indicated a 100% certainty that no shock would occur, 3 indicated that the participant was unsure and 5 indicated a 100% certainty that a shock would occur. Participants were instructed to make this rating using the button box in the scanner.

To examine how participants perceived the CS+ and CS- across different phases of the experiment, we analysed their shock expectancy ratings. These ratings are widely used in the fear conditioning and extinction literature as an index of participants' explicit awareness of learned threat contingencies. They provide insight into the participant's model of the CS+ – US relationship and are often used to assess the acquisition and updating of threat associations across learning and extinction phases (Boddez et al., 2013; Lonsdorf et al., 2017; Phelps et al., 2004).

The experiment consisted of three main phases, the conditioning phase, the extinction phase and the reinstatement phase (Figure 31). All three phases were conducted within the scanner wherein; the conditioning phase and the extinction phase were

conducted in the first block, and the reinstatement phase was conducted in the second block. Heart rate data were recorded during both blocks, similar to the TNT task.

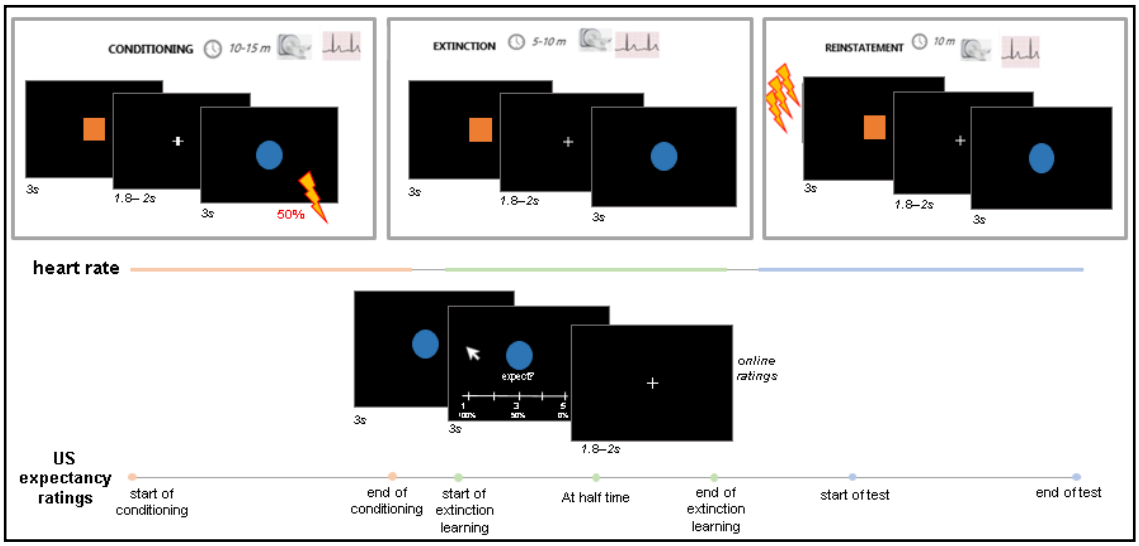

Figure 31. Schematic illustrating the various phases in the Fear Conditioning Task.

The conditioning phase consisted of 40 trials in total, and 20 were CS- trials and the other 20 were CS+ trials. One half (10) of the CS+ trials were paired with the electric shock (the US). This was followed directly by the extinction phase within the same block and participants were not informed about the distinction between the two phases to ensure that if fear extinction actually employs memory suppression mechanisms, then, this happens in an uninstructed manner. The extinction trials were also 40 in number; 20 trials were CS+ stimuli and the other 20 were CS- stimuli. The key difference this time was that none of the CS+ stimuli were paired with an electric shock. Participants made shock expectancy ratings four times for each shape (CS+ and CS–), with two ratings collected during the conditioning phase (one early and one late) and two during the extinction phase (early and late). After this block ended, there was a break of 15 minutes during which participants were instructed to lay still in the scanner and Diffusion Tensor Imaging (DTI) data were acquired during this period.

Finally, participants performed the reinstatement phase. At the beginning of this phase, participants received 4 un-signaled electric shocks in quick succession. Following these shocks, this phase had 16 trials in total, 8 were CS+ trials and the other 8 were CS- trials. This time, none of the CS+ trials were paired with electric shocks. Participants were also not required to make any shock expectancy ratings.

6.2 fMRI Acquisition and Analysis

6.2.1–6.2.4 MRI Acquisition and Analysis

Since this is a within-subject design, the imaging acquisition parameters, preprocessing steps, the ROIs employed and the univariate analysis procedures followed in this chapter are identical to those described in sections 5.3.1 through 5.3.4 which are part of Chapter 5.

6.2.5 Details specific to the FC analysis

For the Fear Conditioning (FC) task, separate GLMs were specified for the acquisition/extinction and reinstatement phases. Events were modelled as boxcar functions (3-second duration) time-locked to stimulus onset. Regressors included were CS+ and CS− trials during conditioning and extinction as well as unconditioned stimulus (US) presentations. For the reinstatement phase, separate regressors modeled CS+ and CS− trials following re-exposure to the US.

The primary contrasts of interest were CS+ > CS− comparisons across each learning phase: conditioning, extinction and reinstatement. These contrasts were designed to parallel suppression-related contrasts in the TNT task, thereby enabling cross-task comparisons of affectively matched control mechanisms.

6.2.6 Second-Level Analysis

For each of the task-specific contrasts above, the contrast estimate for each participant was entered into a one-sample T-test against zero. A conjunction analysis was used to identify common voxels across tasks. For voxel-wise analysis, FDR correction for multiple comparisons was used, unless otherwise stated. For ROIs, significance levels were Bonferroni corrected for the number of ROIs.

6.3 Fear Conditioning Results

6.3.1 Behavioural Results

6.3.1.1 Evidence for Associative Learning and Extinction from Shock Expectancy Ratings

Our main focus was on difference scores (CS+ minus CS−), to quantify differential expectancy between conditioned and unconditioned stimuli. However, to better interpret the source of these differences, we also plotted raw expectancy scores for CS+ and CS− across all phases. Expectancy ratings were collected on a 5-point scale (1 = “not at all likely”, 5 = “very likely”), such that higher values correspond to greater expected likelihood of receiving a shock.

As shown in Figure 32, CS− expectancy remained low and stable across all phases, consistent with its role as a safety signal. In contrast, CS+ expectancy changed robustly over time. From Pre-Conditioning to Post-Conditioning, CS+ expectancy increased substantially, indicating successful acquisition of the CS+–US association. Expectancy then decreased during Early and Late Extinction, reflecting participants' updating of contingency knowledge once reinforcement ceased.

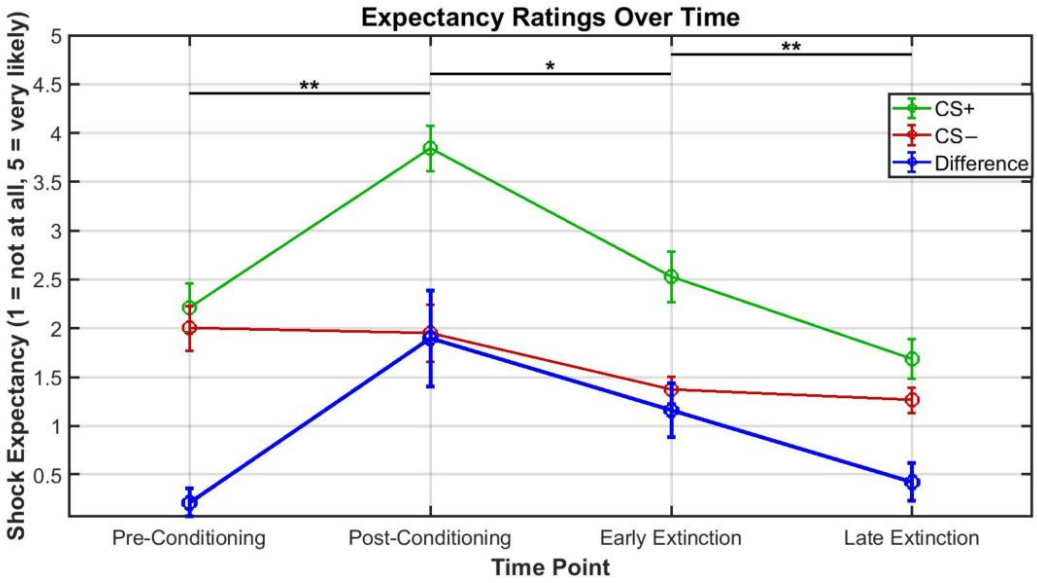

Figure 32. Expectancy ratings across four experimental time points: Pre-Conditioning, Post-Conditioning, Early Extinction, and Late Extinction.

Mean ratings are shown separately for CS+ (green), CS− (red), and the CS difference scores (blue). Error bars indicate the standard error of the mean.

A repeated-measures ANOVA on CS+–CS− difference scores showed a significant main effect of Time, $F(3, 54) = 9.39$, $p < .001$, $\eta^2 = .23$, demonstrating clear within-subject changes across learning phases. To clarify these changes, three planned paired-sample t-tests were conducted with Holm–Bonferroni correction.

Shock expectancy increased from Pre-Conditioning to Post-Conditioning, $t(18) = -3.40$, $p = .0032$, confirming successful associative learning. Expectancy then decreased from Post-Conditioning to Early Extinction, $t(18) = 2.22$, $p = .039$, and decreased further from Early to Late Extinction, $t(18) = 3.68$, $p = .002$. All comparisons survived the corrected significance thresholds, indicating a reliable pattern of acquisition followed by extinction of shock expectancy.

Together, these results demonstrate that participants successfully acquired, updated, and extinguished threat-related expectations in accordance with the

experimental contingencies. The sharp rise in CS+ expectancy during conditioning, followed by its systematic decline during extinction, aligns with theoretical models of contingency learning and replicates prior work showing that expectancy ratings provide a sensitive index of explicit fear learning (Boddez et al., 2012; Lonsdorf et al., 2017).

6.3.2 Neuroimaging Results

We conducted ROI and whole-brain analyses to evaluate activity across key structures known to be implicated in fear conditioning, fear extinction and those involved in memory control. Based on our a priori hypotheses, we tested the contrast CS+ > CS− within each of three main task phases: Conditioning and Extinction and Reinstatement. However, no clusters survived correction for multiple comparisons at the whole-brain level (cluster-level FDR corrected $p < .05$).

To control for multiple comparisons across the seven ROIs tested in the experiment, a Bonferroni-corrected significance threshold of $\alpha = .0071$ was applied uniformly across both models and contrasts.

6.3.2.1 Amygdala and Parahippocampal Activation Without Prefrontal Involvement in Conditioning

As predicted, CS+>CS- activation was found in amygdala ($t = 2.216$, $p = 0.040$) which indicated affective associative learning. However, this activity exhibited a non-significant trend. Further, significant activity was also found in the parahippocampus ($t = 3.353$, $p = 0.004$) suggesting that contextual encoding of the CS+ successfully took place.

However, there was no reliable activation found in prefrontal control-related regions. The rDLPFC ($t = -0.236$, $p = 0.816$), rVLPFC ($t = 0.715$, $p = 0.484$) and the ACC ($t = 0.688$, $p = 0.500$) all failed to show significant activity. The broader inhibitory control network also did not show significant activity ($t = 1.433$, $p = 0.169$) but exhibited a trend, considering the uncorrected threshold.

The absence of prefrontal recruitment is broadly consistent with findings from the meta-analysis conducted by Fullana et al. (2016) where they report that prefrontal activations are less reliably observed. Such regions tend to show stronger and more consistent activation during extinction and/or emotion regulation.

The NRe ($t= -0.211$, $p = 0.836$) also did not exhibit significant activity. In rodent studies, (as discussed in chapter 2) the NRe is implicated in supporting the precision of contextual fear memory acquisition (Xu and Südhof, 2013; Ramanathan et al., 2018). The

absence of NRe activity here could indicate species related differences or due to less statistical power, given the small sample size.

6.3.2.2 Extinction Engaged Neither Fear Expression Nor Suppression Mechanisms

During extinction, the CS+ stimuli was presented without reinforcement, immediately after the conditioning trials with no explicit instructions which communicated the transition. The amygdala ($t = 1.211$, $p = 0.242$) and the parahippocampus ($t = 0.017$, $p = 0.987$) both did not show significant activity.

Similarly, there was no reliable significant activity in prefrontal control regions. The rDLPFC ($t = 1.626$, $p = 0.121$), the rVLPFC ($t = 1.814$, $p = 0.086$) exhibited a trend to significance but did not survive correction. The ACC ($t = 1.268$, $p = 0.221$) also failed to exhibit significant activity. However, the broader inhibitory control network did approach significance ($t = 2.049$, $p = 0.055$) but did not survive correction. This trend indicates that suppression mechanisms could be at play during extinction.

This pattern of activity does not seem to stem from failed fear acquisition since the amygdala and parahippocampus exhibited significant activity during conditioning. The absence of significant activity from the extinction phase could be due to limited reactivation of threat associations, potentially since the shock occurred with CS+ onset. The temporal association between the cue and outcome could have been less salient, weakening the CS+ as a retrieval cue. In addition, the small sample size might have led to low statistical power leading to these results.

Like in conditioning, even during extinction, the NRe ($t = 1.026$, $p = 0.318$) did not exhibit significant activity. If retrieval stopping is involved during extinction, given that the NRe was implicated in retrieval stopping (as discussed in Chapter 5), activation was expected here. Moreover, rodent studies have consistently demonstrated that the NRe supports extinction encoding and retrieval (Ramanathan et al., 2018; Ratigan et al., 2023; Totty et al., 2023). However, lack of its involvement could be due to species related differences, sensitive effects which are not detectable with the current sample size or that the task design did not recruit thalamocortical pathways.

6.3.2.3 Reinstatement Failed to Recover Memory-Related or Affective Activity

The reinstatement phase included un-signaled shocks 15 minutes after extinction took place. During reinstatement, the CS+ was presented without any further reinforcement. There were no reliable activation in the amygdala ($t = -1.300$, $p = 0.212$) and parahippocampus ($t = -1.266$, $p = 0.223$). Further, there was no significant activation even in the prefrontal control regions, rDLPFC ($t = -1.546$, $p = 0.142$), rVLPFC ($t = 0.287$, p

= 0.778) and the ACC (t = 0.642, p = 0.530) or thr broader inhibitory control network (t = 0.351, p = 0.730).

The negative trend in the rDLPFC could reflect disengagement for the CS+ given that it ceases to exist as an aversive stimulus under conditions of reinstatement. Overall, reinstatement did not reactivate regions underlying fear memory or fear extinction indicating that reinstatement failed to occur and even if it did, the associations were weak and inaccessible.

The NRe did not exhibit significant activity (t = -0.221, p = 0.828). However, this finding is not surprising, given that reinstatement failed to occur in this experimental design. Thus, we cannot draw any conclusions about NRe activity during fear reinstatement.

6.5 Discussion

This chapter aimed to investigate the retrieval stopping model of fear extinction by testing whether fear extinction and memory suppression share common neural pathways. In doing so, we focused on brain regions known to be implicated in memory suppression including prefrontal regions, the amygdala and the HpC. In addition, we also tested the NRe since, rodent studies have implicated this structure in both, conditioning and extinction. Further, recent evidence (from Chapter 3 and Chapter 5) suggest that it is involved in memory control, further reinforcing its inclusion in this study.

6.5.1 Dissociation Between Behavioural Learning and Neural Engagement

Participants exhibited significant associative learning at the behavioural level. Expectancy ratings demonstrated significant increases for CS+ stimuli after conditioning along with reliable decreases during extinction which indicated successful acquisition and updating of threat contingencies. Thus, participants were able to reliably adjust their expectations based on the change in reinforcement contingencies.

However, in comparison, neural evidence underlying fear learning, extinction and reinstatement were limited. During conditioning, there was a trend to significant CS+ > CS- activation in the amygdala and significant activation in the parahippocampus which provides some evidence that there was encoding of threat related representations at the neural level. Yet, there was no significant activity in the prefrontal regions including the ACC where activity is typically found. The NRe did not exhibit any meaningful activity during conditioning.

During extinction or reinstatement, no significant activity was found in any of the ROIs, including the amygdala which is typically involved in fear reduction during extinction and fear expression during reinstatement. The HpC also failed to exhibit activity as expected in both, extinction and reinstatement phases. There was no activity in the prefrontal regions thought to be involved during fear extinction. The broader inhibitory control network showed a trend toward significance during extinction but did not survive correction. Lastly, there was no activity in observed in the NRe either during extinction or during reinstatement.

These findings suggest a dissociation between participants' explicit behavioural learning and the underlying neural systems that typically support fear memory regulation.

6.5.2 Methodological Considerations Likely Contributing to Neural Disengagement

Two aspects concerning task design could have contributed to the weak or even absent neural activations observed. Firstly, the shock was delivered immediately with the onset of the shock (CS+) which could have eliminated the anticipatory delay and thus prediction error. Without this temporal separation, participants may have experienced the CS+ and the US as a co-occurring event as against a predictive association. If this were the case, it could have reduced the salience of the CS+ as a cue in order to trigger retrieval. Thus, prefrontal top-down mechanisms modulating the amygdala and the HpC may not have been recruited, potentially because of the lack of this predictive window. This could have resulted in attenuated neural activity despite contingency learning.

Secondly, the shock administered to individual participants was calibrated to be "highly annoying but not painful." This threshold was determined whilst the participants were inside the scanner, by gradually increasing the voltage until the participant's selected threshold was achieved and it was administered only once and in a non-threatening context, prior to the task. Some participants may have become habituated to the shock or reappraised it as tolerable. This would diminish the emotional salience of the US itself, weakening the degree to which typical aversive-learning circuits, especially concerning the amygdala, the HpC and its regulatory prefrontal inputs are recruited during conditioning.

Together, the absence of a predictive delay and the reduced aversiveness of the shock provide an explanation for the unexpected neural findings. Participants may have successfully encoded the CS–US contingency at a cognitive or expectancy level, as

reflected by the behavioural data while failing to engage the neural systems that typically support anticipatory threat processing and emotional regulation during fear conditioning.

6.5.3 Evaluating the Retrieval Stopping Model During Fear Extinction

As discussed earlier, the retrieval stopping model of fear extinction (Anderson & Floresco, 2022) proposes that during extinction, exposure to the CS+ may involuntarily trigger the retrieval of the aversive US. This could act as an intrusion and in order to counter this, the brain recruits top-down inhibitory control processes to inhibit the reactivated fear memory trace, analogous to suppression engaged as observed during the TNT paradigm.

The current findings do not provide evidence for such recruitment. During extinction, only significant trends but no significant activation was observed in either the prefrontal control regions or in the broader inhibitory control network. This absence could be due to three reasons. First, as mentioned earlier, the small sample size could have led to low statistical power. Then, this absence could reflect from a genuine failure to recruit the network underlying retrieval suppression or that suppression was not required since these threat memories were not strongly reactivated during extinction.

As discussed, if memory traces were weak as a result of poor encoding due to the co-occurrence of the CS+ and the US or because of the low salience of the shock then, extinction will have not required the recruitment of inhibitory systems potentially involved in the modulation of aversive memories. In this light, the absence of prefrontal or hippocampal engagement is more plausibly interpreted as insufficient mnemonic conflict rather than as evidence which supports the retrieval stopping model. As discussed in TNT literature, suppression is likely to come into play when interference is high, and unwanted memories interfere requires more active control (Levy & Anderson, 2012). There does not seem to be meaningful reactivation of the to-be-suppressed memory during extinction in this paradigm and thus, suppression mechanisms may not have been recruited at all.

6.5.4 Absence of NRe Engagement and its Implications for Thalamocortical Control in Fear Regulation

In chapter 5, we found that the NRe was activated during memory suppression during the TNT task. Specifically, the NRe was significantly recruited during memory intrusions which suggested its role during reactive control. Alongside this, evidence that the NRe is involved during conditioning and necessary for extinction learning and extinction retrieval in rodent models (Ramanathan et al., 2018; Ratigan et al., 2023) led us to

hypothesize that the NRe would be similarly engaged whilst conditioning, and during extinction and reinstatement, when fear memories might re-emerge.

However, the NRe did not exhibit significant activity during any of these phases. The absence of any meaningful activity could reflect species-specific differences in functional attributes of the NRe. The human NRe may function differently as compared to its rodent counterpart. Secondly and most plausibly, the failure in the reactivation of fear memories during both, extinction and reinstatement could mean that the engagement of thalamocortical control mechanisms was not required.

The pattern of activity observed here contrasts with Chapter 5 where the NRe was significantly engaged during reactive control. This could also suggest that the NRe might only be involved when there is strong mnemonic conflict rather than the mere presence of aversive or emotionally salient content.

6.5.5 Limitations and Future Directions

This study has several limitations which limit the interpretation of the current findings. The exclusion of participants due to incorrect shock delivery during CS- trials greatly reduced the sample size thus, limiting statistical power. Further, the timing of the shock, with the onset of the CS+ potentially stunted associative learning by reducing the prediction error signal. Finally, the calibration of the shock to the participants' individual thresholds in a non-threatening environment may have led to habituation effects thus diminishing the aversiveness of the shock, early on.

Going ahead, future studies could benefit from various methodological adjustments. Firstly, studies should employ a longer anticipatory delay between the CS+ and the US during conditioning which could facilitate predictive learning. Thus, threat expectancy could be enhanced and during extinction, the absence of the US could lead to a prediction error which could then enable the recruitment of top-down control mechanisms.

Another important avenue will be to understand whether explicitly instructing participants to suppress the memory of the US during extinction could enhance extinction learning outcomes by leading to greater engagement of the prefrontal inhibitory control regions. This should be directly compared to a similar well-powered study in which no explicit suppression instructions are administered to understand whether suppression mechanisms are spontaneously engaged. It will be interesting to compare the extinction

learning outcomes between these two procedures by quantifying the extent to which prefrontal mechanisms are employed.

Thirdly, the extinction test could be delayed. Instead of conducting extinction immediately after conditioning as we did (we employed a 15-minute break), a delayed extinction test i.e., conducting conditioning on day 1 and extinction on day 2 would enable better consolidation of the fear memory. This increased consolidation could potentially lead to greater reactivation during extinction. This in turn could lead to greater mnemonic interference which could recruit prefrontal inhibitory control driven mechanisms. It would be interesting to corelate the strength of conditioned responses with the degree of engagement (if engaged) of suppression related mechanisms to study whether individuals with greater fear memories recruit top-down inhibitory processes.

Further, care should be taken that the administration of the US is not prone to habituation effects. A potential way to reduce this risk is to employ aversive stimuli within a more engaging and ecologically valid context, for example by employing virtual reality (VR). Though the MRI environment restricts movement and using standard consumer VR headsets is not possible, several studies have successfully implemented VR inside the scanner using MR-compatible stereoscopic goggles, screen–mirror projection systems or adapted VR head-mounted displays that enable visual immersion while keeping the participant's head still. Lenormand and Piolino (2022) provide an extensive review of such VR-fMRI systems, discussing that MRI-compatible goggles, joysticks, button boxes, and motion-tracking gloves can be used to maintain interactivity while adhering to strict movement constraints required for high-quality MRI. Other technological developments have increased immersion: for example, Gauthier et al. (2021) describe an MRI-compatible immersive VR platform employing high-resolution MR-safe stereoscopic goggles and optical motion capture. Their system creates a virtual replica of the MRI bore and uses hand-tracked movements to control a virtual avatar, producing high embodied presence with minimal movement-related artefacts. Although vestibular cues cannot be reproduced when participants lie supine in the MRI scanner, Reggente et al. (2018) discuss that VR can support immersive, navigable, and context-rich task designs that enhance ecological validity compared to traditional 2D paradigms. In addition, VR-fMRI studies have been found to reliably recruit canonical activation of regions involved in episodic memory (Lenormand and Piolino, 2022).

Finally, future studies must employ effective connectivity analysis like DCM to understand the directionality of communication between prefrontal inhibitory control

regions and the NRe along with the amygdala and the HpC. Applying DCM within a naturalistic VR context may clarify whether top-down prefrontal suppression mechanisms are recruited during extinction, in environments that more closely resemble real-world threat contexts.

Chapter 7

General Discussion

This thesis aimed to test whether in humans, the nucleus reuniens (NRe) of the thalamus plays a role in the suppression of unwanted memories and fear and whether fear extinction and memory suppression share common neural pathways. Across a set of converging studies which began with meta-analytic evidence of the NRe's engagement during retrieval suppression, extending this through the development of a segmentation protocol to delineate subject-specific NRe ROIs and finally culminated in a within-subject fMRI study to investigate the neural mechanisms along with the role of the NRe underlying retrieval suppression and fear extinction. Thus, this thesis provides empirical and anatomical evidence for thalamic contribution in inhibitory control mechanisms in the human brain. This chapter, a general discussion integrates findings from previous chapters and evaluates its contributions and limitations along with suggesting avenues for future research.

7.1 Chapter-by-Chapter summary of findings

Chapters 1 and 2 consisted of the literature review which discussed the theoretical foundations for investigating common neural mechanisms underlying memory suppression and fear extinction and the potential role of the NRe in these processes.

Chapter 1 discussed memory suppression as an active, voluntary inhibitory control mechanism which potentially relies on prefrontal-thalamic-hippocampal pathways. Converging evidence from studying memory suppression in humans via the Think/No-Think (TNT) paradigm has implicated the right dorsolateral prefrontal cortex (rDLPFC) in suppressing the retrieval of unwanted and intrusive memories by modulating (inhibiting) the activity of the hippocampus (HpC). Further, evidence from across species, specifically rodents and primates has implicated the NRe as an evolutionary conserved anatomical relay which could facilitate such prefrontal regulation of the hippocampus.

Chapter 2 discussed fear extinction which has traditionally been conceptualised as a form of inhibitory associative learning from the lens of the retrieval stopping model of fear extinction. According to this model, fear extinction could involve the same inhibitory control processes which are implicated in retrieval suppression. In this case,

prefrontal mechanisms would actively suppress the retrieval of fear memories. This chapter focused on overlapping neural regions implicated in both these processes, including prefrontal regions, the HpC and potentially the NRe. In rodents, the NRe is widely found to be implicated during fear conditioning and extinction learning as well as the retrieval of extinction memories. Establishing whether such an overlap actually exists could lead to significant theoretical as well as clinical implications especially in conditions which involve intrusive thoughts and memories.

Chapters 3 through 6 test the hypothesized contributions of the NRe to retrieval suppression and fear extinction, employing a meta-analysis concerning data from ten fMRI studies employing the TNT paradigm, increasingly precise anatomical masks and lastly, a within-subject fMRI study testing memory suppression and fear extinction in two separate sessions.

Chapter 3 consisted of a meta-analysis of ten existing fMRI datasets which employed the TNT paradigm but consisted of different types of stimuli, number of repetitions and varied scanning protocols providing a rare opportunity to test if the NRe is involved in memory suppression. The NRe MNI masks used in this study were developed in collaboration with primate anatomist Dr Basilis Zikopolous and is the first attempt to study this structure in relation to memory suppression. This analysis revealed reliable activation of the bilateral NRe during suppression (No-Think) trials as compared to Think trials which provided initial evidence for consistent thalamic activation during voluntary memory suppression. However, there was limited support for NRe engagement during reactive memory suppression which was characterized by reported intrusions. Further, it demonstrated that functional connectivity, as measured via Psychophysiological Interaction Analysis increased between the rDLPFC and NRe whereas, connectivity between the NRe and HpC decreased. The ROIs used for these analyses were generic MNI ROIs and thus, limitations in spatial resolution hindered voxel-level confident localization, specifically to the NRe.

In order to overcome this limitation, **Chapter 4** developed a novel procedure, informed by rodent and primate histology, thus enabling individualized segmentation of the NRe using standard resolution (3T) human MRI. The advancement of this methodological approach enabled precise identification of subject-specific individual NRe ROIs along with an essential anatomical foundation for further functional analyses.

Chapter 5 used this protocol to individually define subject-specific NRe ROIs in an fMRI study of the TNT task in a within subject design. Results from this study

demonstrated that the NRe was activated during retrieval suppression however, there was strongest activation on trials which involved memory intrusions, even for negatively valenced stimuli. These findings support the thalamo-hippocampal modulation hypothesis wherein, the NRe seems to be capable of re-engaging inhibitory control mechanisms to counteract unwanted intrusions in situations of high mnemonic conflict.

Finally, **Chapter 6** applied the same subject-specific NRe ROIs in the second part of the within-subject design to test whether the same NRe mediated retrieval suppression mechanisms are implicated during fear extinction. Participants exhibited behavioural evidence of extinction (as index via reduced shock expectancy ratings), neuroimaging analysis revealed no significant activation of the NRe, of associated regions as part of the inhibitory control network nor typical regions involved during conditioning, extinction or during reinstatement.

7.2 Theoretical Implications

7.2.1 Is the NRe a relay structure, or modulator, or both?

An ongoing question which remains is whether the NRe merely acts as a relay, communicating signals between the mPFC and HpC or whether it also is modulatory in nature, dynamically influencing how and when information is transmitted. As discussed, the NRe occupies a strategic position to participate as a relay. Tract tracing studies in rodents (Vertes et al., 2006) and primates (Joyce et al. 2022) and a more recent diffusion-weighted study in humans (Reeders et al., 2022) have confirmed robust connectivity between the mPFC and the HpC. The meta-analysis in Chapter 3 demonstrated increased NRe activity during memory suppression along with increased connectivity to the rDLPFC and decreased connectivity to the HpC. This observed pattern of activity is consistent with the NRe functioning as a relay structure.

However, evidence suggests that the NRe also plays a modulatory role. Stimulation of the rodent NRe can lead to both excitatory and inhibitory effects in the CA1 via the recruitment of hippocampal interneurons (Dolleman-Van der Weel et al., 1997). The facilitation of these excitatory and inhibitory effects via the NRe could indicate that it is capable of gating hippocampal output based on task demands. Additionally, laminar specificity in the rodent NRe reinforces the idea that the NRe could perform integrative modulatory functions (Vertes et al., 2006). Furthermore, Totty et al. (2023) demonstrated that theta synchrony between the mPFC and the HpC, modulated by the NRe, decreased during fear recall and increased during extinction recall. Inactivation

of the NRe was found to prevent the suppression of theta coupling during extinction recall which then led to fear relapse. Similarly, Xu and Südof (2013) demonstrated that disrupting mPFC→NRe communication led to contextual overgeneralization of fear. This can be described as a modulatory failure of the NRe to constrain retrieval to the relevant context alone. Then, the presence of GABAergic interneurons in the primate NRe (Joyce et al., 2022) suggests that it could perform modulatory functions, filtering inputs when required. Secondly, the primate NRe also receives inputs from the basolateral amygdala which could indicate that the NRe could integrate affective salience to factor in modulatory decisions.

Our findings also lend some support to this narrative. In the TNT study discussed in Chapter 5, univariate analysis revealed that strongest NRe activation was found when participants reported memory intrusions. This suggests that the NRe was reactively recruited to re-engage inhibitory control mechanisms when mnemonic conflict is high. In the fear conditioning (FC) study in Chapter 6, the NRe demonstrated no engagement of the NRe during extinction, despite behavioural evidence of extinction learning. Whilst this could be attributed to methodological limitations (as discussed in chapter 6), it could also imply that NRe might be context-sensitive and may come into play only during mnemonic interference.

Recent work conducted by Rivera Núñez et al. (2025) found that the NRe was engaged during both, the encoding and retrieval of overgeneralized emotional memories in peri-adolescents suffering from anxiety. Functional connectivity of the NRe with the mPFC and CA1 varied based on anxiety severity and emotional valence of the stimuli, indicating that it does not operate as a relay alone. The NRe was found to be active during false alarms, when participants incorrectly identified emotionally similar lures as ones they had previously encountered. This reinforces the potential role of the NRe during mnemonic conflict. Further, it was found that functional connectivity between the mPFC and the NRe weakened with anxiety severity, and this was particularly observed for neutral stimuli. This pattern could indicate impaired ability of the NRe to filter out irrelevant information about non-threatening stimuli and facilitate a bias towards emotionally overgeneralized stimuli in such clinical populations. Similarly, increased NRe-CA1 connectivity during false alarms to negative stimuli supporting increased retrieval of overgeneralized memory traces. Thus, these findings support the idea the NRe functions as a dynamic gatekeeper and selectively facilitates or suppresses memory related signals based on context and emotional state.

Vantomme et al. (2025) found further evidence to support the idea that the NRe indeed functions as both, a relay and a modulator within the mPFC-HpC circuit. Their study did not involve a behavioural task but it consisted of optogenetic and electrophysiological recordings in mice. Stimulation of the NRe led to fast monosynaptic EPSCs in the mPFC which is consistent with a relay function. However, sustained polysynaptic excitation was also observed. This suggests that the NRe can modulate excitability the mPFC over extended time periods and it supports a more dynamic modulatory role.

Taken together, current evidence lends support to a dual-role functional ability of the NRe, functioning both, as a relay and modulator, depending on the context. For disorders like PTSD and anxiety characterized by intrusive cognition and disrupted cognitive control mechanisms, the dysfunction may lie in the NRe's modulatory abilities. Framing the NRe as both, a relay structure and a modulator opens doors to studying the NRe as a regulatory region for understanding its role in mental health disorders wherein the gating and/or suppression of intrusive memory and emotions is awry.

7.2.2 Proactive or Reactive? The Proposed Role of the NRe in Inhibitory Control

Another key question is whether the NRe participates in inhibitory control proactively by preventing mnemonic content from entering awareness to begin with or does it partake reactively by supressing mnemonic content after it has already entered awareness. While our current study could not directly test for proactive v/s reactive control, our findings provide indirect evidence, supporting a reactive role. This is in accordance with Anderson, Bunce and Barbas (2016) who identified the NRe-HpC pathway suited to interrupting retrieval once it is already underway. As discussed above, results from Chapter 5 demonstrated most robust activations in the NRe were in trials involving intrusions indicative of its role in reactive control. In Chapter 6, no significant NRe engagement was found during fear extinction despite behavioural evidence of extinction learning which led us to conclude that the NRe is potentially context sensitive and is only recruited during conditions of sufficient mnemonic conflict.

Evidence for a reactive role also comes from Tuna et al. (2025), who across humans and rodents demonstrated that NRe activity is consistently evoked after the presentation of threat-predictive cues. Human BOLD responses showed short-latency increases to the CS^+ and rat photometry also revealed rapid increases in NRe calcium

immediately following CS onset, indicating that the NRe engages only once a threat representation has already been activated. In rodents, NRe calcium activity decreased during freezing and then rose approximately 500 ms before freezing terminated, showing that the NRe helps to suppress defensive states after they have been engaged. Single-unit recordings replicated this pattern: the majority of NRe neurons showed reduced firing during freezing and a rebound in firing just prior to its cessation. Together, these convergent results demonstrate that NRe activity tracks and contributes to resolving ongoing retrieval and defensive states, consistent with a reactive inhibitory mechanism.

Electrophysiological evidence in rodents throws light on the NRe's role in reactive control. Likhtik et al. (2014) demonstrated that infralimbic input could rapidly entrain the basolateral amygdala, as quick as within hundreds of milliseconds of a safety cue, once a threat representation has already been activated. By analogy, NRe engagement during intrusion trials may reflect a similar rapid gating mechanism. Additional support comes from Vasudevan et al. (2022) who found that pharmacological inactivation of the NRe immediately after extinction training or reactivation did not impair later retrieval. However, inactivation immediately before retrieval impaired extinction expression. These results reinforce the view that the NRe is engaged during online retrieval processes in line with its role in reactive control.

However, the NRe's potential involvement in proactive control cannot be dismissed. Xu and Südhof (2013) also demonstrated that disrupting mPFC → NRe communication impaired contextual control of fear, leading to overgeneralisation, which implies a role for the NRe in filtering retrieval according to context before inappropriate traces are expressed. Crucially, the anatomical debate also bears on this issue since, historically, it was assumed that the mPFC lacks direct connections to the HpC and the NRe was assumed as the only relay. However, more recent evidence from Malik et al. (2022) identified sparse direct mPFC→CA1 projections. This now puts into question the role of the NRe within the inhibitory network all together. It could suggest that proactive control could bypass the NRe and the NRe could be recruited as a flexible modulator when control needs to be applied reactively, under mnemonic conflict. However, this is speculative, and it remains to be tested. Recent work has also revealed direct projections from the NRe to the medial septum, with silencing of this pathway selectively impairing extinction of remote but not recent fear memories (Tomaszewski et al., 2024). Given the medial septum's established role in modulating hippocampal theta rhythms (Tsanov,

2018), this pathway may provide an interface through which septal dynamics bias hippocampal-prefrontal states, while the NRe contributes more directly to retrieval gating.

Understanding whether the NRe contributes primarily to proactive or reactive control is important with different implications for how circuits prevent versus suppress intrusive memories. Future work employing temporally precise methods (e.g., simultaneous EEG–fMRI, intracranial recordings, or pathway-specific stimulation via TMS) will be essential to disentangle how these pathways are differentially recruited across proactive and reactive control demands.

7.2.3 Shared vs. Dissociable Mechanisms of Memory Suppression and Fear Extinction

Another crucial question which is at the heart of this thesis is whether memory suppression and fear extinction actually rely on shared inhibitory mechanisms or whether they are supported by dissociable neural networks. Our findings indicate that suppression and extinction both seem to be dependent on prefrontal regulation but however they diverge in their downstream circuitry. In our data, in both, Chapters 3 and 5, the NRe was activated during retrieval suppression. By contrast, extinction as discussed in Chapter 6 was successful at the behavioural level but there was no evidence of NRe recruitment. This dissociation could suggest failure of our design in evoking extinction and thus NRe activity or that indeed, extinction does not engage suppression circuitry. Recent cross species evidence by Tuna et al. (2025) agrees with this view. They demonstrated that NRe activity in both, humans and rodents encode the associative value of conditioned stimuli during fear acquisition showing stronger responses to threat-related cues. During extinction and extinction retrieval, NRe cue-evoked activity decreases, consistent with reduced threat value. Moreover, in rodents, distinct NRe ensembles were found to encode fear memories and extinction memories potentially indicating that even within the NRe, inhibitory control over fear and safety learning can be catered to by distinct neuronal populations.

Another factor to be taken under consideration is the distinction between voluntary suppression and the spontaneous recruitment of inhibitory network. Voluntary suppression as in the TNT paradigm requires deliberate engagement of the rDLPFC-NRe-HpC pathways to suppress intrusions. By contrast, extinction could engage suppression like processes spontaneously, when intrusive fear representations might overwhelm the retrieval of safety associations. Thus, suppression and extinction may differ in the

conditions under which inhibition is employed, suppression under deliberate effort and extinction when mnemonic conflict is high. Relatedly, the retrieval stopping model argues that fear is regulated via suppression when it is considered inappropriate. One possible explanation could be that suppression is only employed when aversive fear memories are consciously experienced as intrusions. Taken in this light, without perceived intrusions, extinction could depend more on competitive retrieval dynamics than on direct inhibition. This could also explain how it could proceed without the engagement of the NRe.

Suppression could primarily engage rDLPFC-HpC pathways whereas extinction could be more reliant on rDLPFC-vmPFC-amygdala interactions. Hennings et al (2022) found that vmPFC reinstated extinction memories whereas the dACC and insula reinstated fear memories with dissociation within the long axis of the HpC, shaping which trace dominated behaviour. These findings suggest that extinction depends on competitive retrieval dynamics and is an example of extinction proceeding without suppression related activity. Further, in a behavioural study, Hennings et al. (2021) found that pairing extinction with suppression led to thought generalisation, and they claimed that this was likely because suppression downregulated hippocampal involvement which is crucial for transferring safety associations across contexts. However, a crucial detail, discussed earlier is that in their design, participants were made to supress the CS+ instead of the memory of the US which may have altered the retrieval dynamics which underlie these findings. Quaedflieg et al. (2025) also found that voluntary suppression during extinction reduced negative affect but did not carry any benefit over standard extinction.

A translational lens throws light upon these potential distinct pathways. Neuromodulation work indicated that stimulation of the left PFC connected to the vmPFC enhanced extinction recall (Raij et al., 2018), while in PTSD patients rDLPFC stimulation predicted better treatment outcomes through greater top-down inhibition of the amygdala (Fonzo et al., 2017). Thus, vmPFC-amygdala interactions are important for strengthening extinction memories and the rDLPFC can exert control over the amygdala.

Taken together, evidence supports a partially overlapping but functionally dissociable network. Suppression and extinction both rely on prefrontal control but differ in how and when inhibition might be employed. Suppression depends on the voluntary engagement of the rDLPFC to inhibit the HpC whereas, extinction takes place through prefrontal-amygdala pathways and hippocampal context signals with suppression like recruitment occurring potentially when fear intrusions overwhelm the retrieval of safety signals.

7.3 Limitations and Future Directions

The findings in this thesis offer a solid foundation for the understanding of the NRe in memory suppression in humans and raise new questions about its involvement in fear conditioning and fear extinction. At the same time, they highlight several gaps, both methodological and conceptual which shape agenda for future work. Some of these directions are discussed below.

7.3.1 Establishing Anatomical Precision: Moving Beyond Inference

A key limitation of this thesis lies in the difficulty of being able to conclusively verify whether the structure labelled as the NRe in our analyses precisely corresponds to its histologically defined counterpart in the human brain. Given the NRe's small size, location in the deep midline and poor visibility on standard MRI, there remains uncertainty. Although this thesis employed a novel anatomically informed significant pipeline, grounded in cross species comparisons, direct anatomical validation in humans is not currently possible. As discussed in Chapter 4, this is a challenge faced by all neuroimaging studies targeting small midline thalamic nuclei.

This being said, the approach employed in this thesis could offer greater anatomical specificity in comparison to current methods used in other recent human NRe-based studies. These studies localised the NRe using group level connectivity clusters or probabilistic thalamic masks which do not account for anatomical variability between individuals. In contrast, this thesis defined subject-specific NRe ROIs in native space, thus avoiding spatial normalization artifacts and offering finner grain anatomical precision. This approach thus can be considered as a methodological advancement and offers a foundation for investigating the NRe in functional neuroimaging in humans.

Going forward, ultra-high field 7T MRI could allow for more direct visualization or more accurate segmentation of small structures like the NRe. Its combination with diffusion tractography or postmortem MRI histology alignment could advance the anatomical and functional study of this structure. Secondly, the segmentation application developed in this thesis could be further refined through automated machine-learning extensions. After initial expert validation across a number of datasets, such an approach could streamline ROI identification across large datasets, preserving the anatomical rigor of the landmark-based method. Thus, the application could provide both, an immediate methodological contribution and a platform for future technical development, integrating information obtained from 7T and potential postmortem histology.

7.3.2 Causal Modelling of NRe Connectivity

While this thesis found functional engagement of the NRe during memory suppression, its precise causal role within the broader prefrontal-hippocampal network remains unknown. Thus, in effect, the current analyses cannot determine whether NRe activity is necessary for suppression and whether it actually mediates top-down signals from the rDLPFC to the HpC or whether it is engaged in a non-essential capacity. This distinction will also throw light upon whether the NRe acts as a relay or a dynamic modulator.

Future research should incorporate Dynamic Causal Modelling (DCM) or other effective connectivity analyses. DCM could enable for formal comparison of relevant models, and we could test if suppression is better explained by a rDLPFC→NRe→HpC pathway as comparted to a rDLPFC→HpC pathway. These models could also test whether the NRe is recruited preferentially during mnemonic conflict i.e., when participants report that they experience intrusions. This in turn could also clarify the NRe's role in proactive control v/s reactive control. Further, a model with the ACC as the upstream driver recruiting the NRe could be tested. Sierra et al. (2017) demonstrated that the formation of both, recent and remote memories became inaccessible when the ACC was inhibited and that retrieval could only be restored when both the ACC and the NRe were reactivated. This could mean that the ACC and the NRe could operate together in concert with the ACC detecting the need to control and the NRe relaying these signals downstream.

These questions are also relevant in the context of fear extinction. Applying a model-based connectivity approach could aid our understanding in whether the absence of observable NRe activity reflects true dissociation in the fear extinction circuitry or whether the NRe contributes in a more subtle, context-dependent manner such as modulating the HpC and/or the amygdala when intrusive fear representations dominate retrieval. DCM could shed light on whether fear extinction and memory suppression actually share the same neural pathways. Then, aberrant prefrontal-thalamic-hippocampal communication could be a hallmark of psychiatric disorders such as PTSD, anxiety and OCD. DCM could reveal whether disruptions stem from weakened top-down recruitment of the NRe, impaired NRe-hippocampal gating or hyperconnectivity which can amplify overgeneralized threat memories. For example, as discussed earlier, Rivera Núñez et al. (2025) found that in anxious peri-adolescents mPFC-NRe connectivity weakened for neutral stimuli biasing retrieval towards emotionally overgeneralised memories. Future

work combining within subject suppression and extinction paradigms with effective connectivity analysis could thus also identify network-related biomarkers of NRe related vulnerability.

7.3.3 Causal Manipulation and Oscillatory signatures NRe-driven inhibitory control

EEG and MEG could be employed during TNT task and extinction tasks to test whether oscillatory markers such as frontal-midline theta ramping (proactive gating) or intrusion-locked theta/beta bursts (reactive control) emerge across both tasks. As discussed in Rowlands (2024), event-related potential (ERP) markers of suppression, such as the N2 component linked to conflict monitoring (Streb et al., 2016), could also be assessed to determine whether inhibitory control processes are spontaneously recruited during extinction. These readouts could clarify whether suppression and extinction share a unified oscillatory blueprint.

Then, the NRe itself cannot be directly targeted with non-invasive stimulation, TMS/tES to hypothesized upstream cortical inputs such as the rDLPFC or ACC could indirectly disrupt prefrontal–thalamic–hippocampal interactions. If perturbing these regions reduces NRe–hippocampal connectivity and impairs suppression or extinction performance, this would provide causal evidence for the role of the NRe in these processes. Combining such perturbation with concurrent EEG or fMRI would allow assessment of whether the oscillations identified in observational work are abolished or reorganized when prefrontal input is disrupted.

Finally, in clinical contexts, intracranial EEG could provide direct validation of NRe involvement in human inhibitory control, bridging non-invasive neuroimaging with invasive recordings.

7.3.4 Recommendations for Future Studies to Understand if Extinction employs Suppression Strategies

The FC study in this thesis has several limitations. The exclusion of participants due to incorrect shock delivery during CS- trials led to a significant reduction in the sample size, thus, reducing statistical power. Further, the timing of the shock delivered with the onset of the CS+ could have led to stunted associative learning by reducing the prediction error signal. Importantly, the calibration of the shock to participants' individual thresholds in a non-threatening environment could have led to habituation early on and diminished the aversiveness of the US.

Learning from this example, future studies could benefit from several adjustments. Firstly, employing a longer delay period between the CS+ and the US during conditioning could strengthen associative learning and enhance threat expectancy during extinction. The absence of the US during this phase could lead to a stronger prediction error and cause the need for the recruitment of top-down inhibitory control mechanisms. Another key avenue is to test whether explicitly instructing participants to suppress the memory of the US during extinction could facilitate extinction learning by leading to the recruitment of prefrontal inhibitory control regions. Directly comparing this approach to a well-powered study without these explicit suppression instructions could help throw light upon whether suppression mechanisms can be spontaneously engaged or require explicit instruction. Additionally, extinction could be delayed, instead of conducting extinction shortly after conditioning, as we did, implementing a 24-hour interval would potentially enable for better consolidation of the fear memory thus strengthening its reactivation during extinction. This could increase mnemonic interference and thus intrusions potentially amplifying the need for the recruitment of suppression related networks.

Future work should also take into consideration individual differences by correlating the strength of the conditioned response with the degree of engagement of suppression-related mechanisms to understand if individuals with stronger fear memories preferentially recruit top-down inhibitory control mechanisms. Also, care should be taken to ensure that the administration of the US is not prone to habituation effects as we suspect did take place in our study. The aversive stimuli could be administered via a more engaging context such as virtual reality which is more immersive.

A further limitation lies in the fact that the current study did not employ more fine-grained behavioural measures of extinction. Future FC studies employing instructed suppression could quantify trial-by-trial dynamics such as expectancy ratings or intrusion reports, thereby, also throwing light upon proactive and reactive control strategies and the recruitment of the NRe. Further, it would be interesting for longitudinal studies to test whether repeated practice of suppression could generalise to improved extinction retention or reduced relapse and whether these effects differ in high-risk populations suffering from disorders characterised by intrusive thought. Finally, computational modelling approaches such as reinforcement learning or Bayesian models of control allocation may be able to capture how participants allocate inhibitory resources across trials and contexts providing a bridge between cognitive measures and connectivity

metrics. For instance, Leone et al. (2022) demonstrated that modelling hidden belief and prediction-error dynamics revealed a predictive-reactive control imbalance in PTSD which could not be detected by standard connectivity analyses.

7.4 Clinical Implications and Concluding Thoughts

Intrusive, distressing memories and thoughts represent a hallmark of disorders such as PTSD, OCD, depression and anxiety. By demonstrating reliable NRe engagement during memory suppression in humans, this work has identified a midline thalamic hub which could contribute to the pathophysiology of these conditions. In PTSD, persistent intrusions and impaired extinction could reflect failures of NRe mediated prefrontal-hippocampal synchrony. Compulsive thoughts in OCD, and ruminative tendencies in depression and excessive worry in anxiety could stem from intrusive thoughts due to impaired frontal-thalamic-hippocampal inhibition. Taken together, the NRe could underlie a transdiagnostic mechanisms by which disrupted inhibitory control over memory contributes to intrusive cognition across various disorders.

If this is the case then, cognitive training tasks, such as the TNT paradigm could be adapted to strengthen inhibitory control and integrated into exposure-based therapies with suppression framed as an active regulatory process as compared to avoidance. As discussed, neuromodulatory techniques such as TMS to the rDLPFC or ACC could engage NRe-mediated pathways indirectly, while future technologies like deep brain ultrasound stimulation could enable more direct modulation of midline thalamic nuclei. Then, the role of GABAergic mechanisms in both, suppression and extinction indicate a potential pharmacological route. In humans, higher GABA concentration predicted stronger prefrontal-hippocampal coupling and more effective suppression (Schmitz et al., 2017), while is rodents, GABAergic somatostatin-expressing interneurons were found to be recruited during extinction to gate relapse (Lacagnina et al., 2023).

Finally, NRe activity and connectivity could provide biomarkers of resilience and vulnerability to psychopathology dominated by intrusions. Trauma survivors who effectively recruited suppression were found to remain resilient whereas individuals who failed to engaged suppression were more likely to develop PTSD (Mary et al., 2020). Individual differences in NRe recruitment during suppression tasks could help identify at-risk populations before symptoms flare up, enabling early intervention.

This thesis provided the first evidence of NRe involvement in human memory suppression employing a subject-specific anatomically grounded segmentation technique

but its causal role in the inhibitory control network and generalisation to clinical population remains to be studied. However, by implicating the NRe at the intersection of prefrontal-hippocampal control, this work lays a mechanistic foundation and opens doors for studying subcortical involvement in inhibitory control and thus in intrusive cognition.

References

Agustín-Pavón, C., Braesicke, K., Shiba, Y., Santangelo, A. M., Mikheenko, Y., Cockroft, G., Asma, F., Clarke, H., Man, M.-S., & Roberts, A. C. (2012). Lesions of Ventrolateral Prefrontal or Anterior Orbitofrontal Cortex in Primates Heighten Negative Emotion. *Biological Psychiatry*, *72*(4), 266–272. https://doi.org/10.1016/j.biopsych.2012.03.007

Alkemade, A., & Forstmann, B. U. (2021). Imaging of the human subthalamic nucleus. In *Handbook of Clinical Neurology* (Vol. 180, pp. 403–416). Elsevier. https://doi.org/10.1016/B978-0-12-820107-7.00025-2

Anderson, M. C., Bunce, J. G., & Barbas, H. (2016). Prefrontal–hippocampal pathways underlying inhibitory control over memory. *Neurobiology of Learning and Memory*, *134*, 145–161. https://doi.org/10.1016/j.nlm.2015.11.008

Anderson, M. C., & Floresco, S. B. (2022). Prefrontal-hippocampal interactions supporting the extinction of emotional memories: The retrieval stopping model. *Neuropsychopharmacology*, *47*(1), 180–195. https://doi.org/10.1038/s41386-021-01131-1

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, *410*(6826), 366–369. https://doi.org/10.1038/35066572

Anderson, M. C., & Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. *Trends in Cognitive Sciences*, *18*(6), 279–292. https://doi.org/10.1016/j.tics.2014.03.002

Anderson, M. C., & Hulbert, J. C. (2021). Active Forgetting: Adaptation of Memory by Prefrontal Control. *Annual Review of Psychology*, *72*(1), 1–36. https://doi.org/10.1146/annurev-psych-072720-094140

Anderson, M. C., Ochsner, K. N., Kuhl, B., Cooper, J., Robertson, E., Gabrieli, S. W.,

Glover, G. H., & Gabrieli, J. D. E. (2004). Neural Systems Underlying the Suppression of Unwanted Memories. *Science*, *303*(5655), 232–235. https://doi.org/10.1126/science.1089504

Apšvalka, D., Ferreira, C. S., Schmitz, T. W., Rowe, J. B., & Anderson, M. C. (2022). Dynamic targeting enables domain-general inhibitory control over action and thought by the prefrontal cortex. *Nature Communications*, *13*(1), 274. https://doi.org/10.1038/s41467-021-27926-w

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. https://doi.org/10.1016/j.tics.2013.12.003

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, *38*(1), 95–113. https://doi.org/10.1016/j.neuroimage.2007.07.007

Banks, P. J., Warburton, E. C., & Bashir, Z. I. (2021). Plasticity in Prefrontal Cortex Induced by Coordinated Synaptic Transmission Arising from Reuniens/Rhomboid Nuclei and Hippocampus. *Cerebral Cortex Communications*, *2*(2), tgab029. https://doi.org/10.1093/texcom/tgab029

Barbas, H., & Blatt, G. J. (1995). Topographically specific hippocampal projections target functionally distinct prefrontal areas in the rhesus monkey. *Hippocampus*, *5*(6), 511–533. https://doi.org/10.1002/hipo.450050604

Barbas, H., & Pandya, D. N. (1989). Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. *Journal of Comparative Neurology*, *286*(3), 353–375. https://doi.org/10.1002/cne.902860306

Barker, G. R. I., & Warburton, E. C. (2018). A Critical Role for the Nucleus Reuniens in Long-Term, But Not Short-Term Associative Recognition Memory Formation. *The Journal of Neuroscience*, *38*(13), 3208–3217. https://doi.org/10.1523/JNEUROSCI.1802-17.2017

Basile, B. M., & Hampton, R. R. (2017). Dissociation of item and source memory in rhesus monkeys. *Cognition*, *166*, 398–406. https://doi.org/10.1016/j.cognition.2017.06.009

Bekinschtein, P., Weisstaub, N. V., Gallo, F., Renner, M., & Anderson, M. C. (2018). A retrieval-specific mechanism of adaptive forgetting in the mammalian brain. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-07128-7

Benoit, R. G., & Anderson, M. C. (2012). Opposing Mechanisms Support the Voluntary Forgetting of Unwanted Memories. *Neuron*, *76*(2), 450–460. https://doi.org/10.1016/j.neuron.2012.07.025

Boddez, Y., Baeyens, F., Luyten, L., Vansteenwegen, D., Hermans, D., & Beckers, T. (2013). Rating data are underrated: Validity of US expectancy in human fear conditioning. *Journal of Behavior Therapy and Experimental Psychiatry*, *44*(2), 201–206. https://doi.org/10.1016/j.jbtep.2012.08.003

Bokor, H., Csáki, Á., Kocsis, K., & Kiss, J. (2002). Cellular architecture of the nucleus reuniens thalami and its putative aspartatergic/glutamatergic projection to the hippocampus and medial septum in the rat. *European Journal of Neuroscience*, *16*(7), 1227–1239. https://doi.org/10.1046/j.1460-9568.2002.02189.x

Bouton, M. E., Maren, S., & McNally, G. P. (2021). Behavioral and neurobiological mechanisms of pavlovian and instrumental extinction learning. *Physiological Reviews*, *101*(2), 611–681. https://doi.org/10.1152/physrev.00016.2020

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Bukalo, O., Pinard, C. R., Silverstein, S., Brehm, C., Hartley, N. D., Whittle, N., Colacicco, G., Busch, E., Patel, S., Singewald, N., & Holmes, A. (2015). Prefrontal inputs to the amygdala instruct fear extinction memory formation. *Science Advances*, *1*(6), e1500251. https://doi.org/10.1126/sciadv.1500251

Bunce, J. G., Zikopoulos, B., Feinberg, M., & Barbas, H. (2013). Parallel prefrontal pathways reach distinct excitatory and inhibitory systems in memory-related rhinal cortices. *Journal of Comparative Neurology*, *521*(18), 4260–4283. https://doi.org/10.1002/cne.23413

Carmichael, S. T., & Price, J. L. (1995). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *Journal of Comparative Neurology*, *363*(4), 615–641. https://doi.org/10.1002/cne.903630408

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. https://doi.org/10.1177/0956797615569889

Chalkia, A., Vanhasbroeck, N., Van Oudenhove, L., Kindt, M., & Beckers, T. (2023). Emotional associative memory is disrupted by directed forgetting. *Communications Psychology*, *1*(1), 24. https://doi.org/10.1038/s44271-023-00024-x

Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy*, *58*, 10–23. https://doi.org/10.1016/j.brat.2014.04.006

Crespo-García, M., Wang, Y., Jiang, M., Anderson, M. C., & Lei, X. (2022). Anterior Cingulate Cortex Signals the Need to Control Intrusive Thoughts during Motivated Forgetting. *The Journal of Neuroscience*, *42*(21), 4342–4359. https://doi.org/10.1523/JNEUROSCI.1711-21.2022

Delgado, M. R., Nearing, K. I., LeDoux, J. E., & Phelps, E. A. (2008). Neural Circuitry Underlying the Regulation of Conditioned Fear and Its Relation to Extinction. *Neuron*, *59*(5), 829–838. https://doi.org/10.1016/j.neuron.2008.06.029

Depue, B. E., Banich, M. T., & Curran, T. (2006). Suppression of Emotional and

Nonemotional Content in Memory: Effects of Repetition on Cognitive Control. *Psychological Science*, *17*(5), 441–447. https://doi.org/10.1111/j.1467-9280.2006.01725.x

Depue, B. E., Curran, T., & Banich, M. T. (2007). Prefrontal Regions Orchestrate Suppression of Emotional Memories via a Two-Phase Process. *Science*, *317*(5835), 215–219. https://doi.org/10.1126/science.1139560

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. https://doi.org/10.1093/cercor/bhu324

Dolleman-van Der Weel, M. J., Griffin, A. L., Ito, H. T., Shapiro, M. L., Witter, M. P., Vertes, R. P., & Allen, T. A. (2019). The nucleus reuniens of the thalamus sits at the nexus of a hippocampus and medial prefrontal cortex circuit enabling memory and behavior. *Learning & Memory*, *26*(7), 191–205. https://doi.org/10.1101/lm.048389.118

Fanselow, M. S. (1980). Conditional and unconditional components of post-shock freezing. *The Pavlovian Journal of Biological Science*, *15*(4), 177–182. https://doi.org/10.1007/BF03001163

Fonzo, G. A., Goodkind, M. S., Oathes, D. J., Zaiko, Y. V., Harvey, M., Peng, K. K., Weiss, M. E., Thompson, A. L., Zack, S. E., Lindley, S. E., Arnow, B. A., Jo, B., Gross, J. J., Rothbaum, B. O., & Etkin, A. (2017). PTSD Psychotherapy Outcome Predicted by Brain Activation During Emotional Reactivity and Regulation. *American Journal of Psychiatry*, *174*(12), 1163–1174. https://doi.org/10.1176/appi.ajp.2017.16091072

Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience*, *6*(2), 119–130. https://doi.org/10.1038/nrn1607

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*(4), 1273–1302. https://doi.org/10.1016/S1053-8119(03)00202-7

Fullana, M. A., Albajes-Eizagirre, A., Soriano-Mas, C., Vervliet, B., Cardoner, N., Benet, O., Radua, J., & Harrison, B. J. (2018). Fear extinction in the human brain: A meta-analysis of fMRI studies in healthy participants. *Neuroscience & Biobehavioral Reviews*, *88*, 16–25. https://doi.org/10.1016/j.neubiorev.2018.03.002

Fullana, M. A., Harrison, B. J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A., & Radua, J. (2016). Neural signatures of human fear conditioning: An updated and extended meta-analysis of fMRI studies. *Molecular Psychiatry*, *21*(4), 500–508. https://doi.org/10.1038/mp.2015.88

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. https://doi.org/10.1523/JNEUROSCI.2732-16.2017

Gauthier, B., Albert, L., Martuzzi, R., Herbelin, B., & Blanke, O. (2021). Virtual Reality platform for functional magnetic resonance imaging in ecologically valid conditions. *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, 1–12. https://doi.org/10.1145/3489849.3489894

Gothard, K. M., Erickson, C. A., & Amaral, D. G. (2004). How do rhesus monkeys (Macaca mulatta) scan faces in a visual paired comparison task? *Animal Cognition*, *7*(1), 25–36. https://doi.org/10.1007/s10071-003-0179-6

Grühn, D., & Scheibe, S. (2008). Age-related differences in valence and arousal ratings of pictures from the International Affective Picture System (IAPS): Do ratings become more extreme with age? *Behavior Research Methods*, *40*(2), 512–521. https://doi.org/10.3758/brm.40.2.512

Guise, K. G., & Shapiro, M. L. (2017). Medial Prefrontal Cortex Reduces Memory

Interference by Modifying Hippocampal Encoding. *Neuron*, *94*(1), 183-192.e8. https://doi.org/10.1016/j.neuron.2017.03.011

Hallock, H. L., Wang, A., & Griffin, A. L. (2016). Ventral Midline Thalamus Is Critical for Hippocampal-Prefrontal Synchrony and Spatial Working Memory. *Journal of Neuroscience*, *36*(32), 8372–8389. https://doi.org/10.1523/JNEUROSCI.0991-16.2016

Hampshire, A. (2015). Putting the brakes on inhibitory models of frontal lobe function. *NeuroImage*, *113*, 340–355. https://doi.org/10.1016/j.neuroimage.2015.03.053

Hellerstedt, R., Johansson, M., & Anderson, M. C. (2016). Tracking the intrusion of unwanted memories into awareness with event-related potentials. *Neuropsychologia*, *89*, 510–523. https://doi.org/10.1016/j.neuropsychologia.2016.07.008

Hennings, A. C., Bibb, S. A., Lewis-Peacock, J. A., & Dunsmoor, J. E. (2021). Thought suppression inhibits the generalization of fear extinction. *Behavioural Brain Research*, *398*, 112931. https://doi.org/10.1016/j.bbr.2020.112931

Hennings, A. C., McClay, M., Drew, M. R., Lewis-Peacock, J. A., & Dunsmoor, J. E. (2022). Neural reinstatement reveals divided organization of fear and extinction memories in the human brain. *Current Biology*, *32*(2), 304-314.e5. https://doi.org/10.1016/j.cub.2021.11.004

Hurley, K. M., Herbert, H., Moga, M. M., & Saper, C. B. (1991). Efferent projections of the infralimbic cortex of the rat. *Journal of Comparative Neurology*, *308*(2), 249–276. https://doi.org/10.1002/cne.903080210

Ito, H. T., Zhang, S.-J., Witter, M. P., Moser, E. I., & Moser, M.-B. (2015). A prefrontal–thalamo–hippocampal circuit for goal-directed spatial navigation. *Nature*, *522*(7554), 50–55. https://doi.org/10.1038/nature14396

Jayachandran, M., Linley, S. B., Schlecht, M., Mahler, S. V., Vertes, R. P., & Allen, T.

A. (2019). Prefrontal Pathways Provide Top-Down Control of Memory for Sequences of Events. *Cell Reports*, *28*(3), 640-654.e6. https://doi.org/10.1016/j.celrep.2019.06.053

Jayachandran, M., Viena, T. D., Garcia, A., Veliz, A. V., Leyva, S., Roldan, V., Vertes, R. P., & Allen, T. A. (2023). Nucleus reuniens transiently synchronizes memory networks at beta frequencies. *Nature Communications*, *14*(1), 4326. https://doi.org/10.1038/s41467-023-40044-z

Ji, J., & Maren, S. (2007). Hippocampal involvement in contextual modulation of fear extinction. *Hippocampus*, *17*(9), 749–758. https://doi.org/10.1002/hipo.20331

Jones, E. G. (1998). Viewpoint: The core and matrix of thalamic organization. *Neuroscience*, *85*(2), 331–345. https://doi.org/10.1016/S0306-4522(97)00581-2

Jovanovic, T., Ely, T., Fani, N., Glover, E. M., Gutman, D., Tone, E. B., Norrholm, S. D., Bradley, B., & Ressler, K. J. (2013). Reduced neural activation during an inhibition task is associated with impaired fear inhibition in a traumatized civilian sample. *Cortex*, *49*(7), 1884–1891. https://doi.org/10.1016/j.cortex.2012.08.011

Joyce, M. K. P., García-Cabezas, M. Á., John, Y. J., & Barbas, H. (2020). Serial Prefrontal Pathways Are Positioned to Balance Cognition and Emotion in Primates. *The Journal of Neuroscience*, *40*(43), 8306–8328. https://doi.org/10.1523/jneurosci.0860-20.2020

Joyce, M. K. P., Marshall, L. G., Banik, S. L., Wang, J., Xiao, D., Bunce, J. G., & Barbas, H. (2022). Pathways for Memory, Cognition and Emotional Context: Hippocampal, Subgenual Area 25, and Amygdalar Axons Show Unique Interactions in the Primate Thalamic Reuniens Nucleus. *The Journal of Neuroscience*, *42*(6), 1068–1089. https://doi.org/10.1523/JNEUROSCI.1724-21.2021

Keuken, M. C., Bazin, P.-L., Schäfer, A., Neumann, J., Turner, R., & Forstmann, B. U. (2013). Ultra-High 7T MRI of Structural Age-Related Changes of the

Subthalamic Nucleus. *The Journal of Neuroscience*, *33*(11), 4896–4900. https://doi.org/10.1523/JNEUROSCI.3241-12.2013

Lacagnina, A. F., Dong, T. N., Iyer, R. R., Khan, S., Mohamed, M. K., & Clem, R. L. (2023). *Ventral hippocampal interneurons govern extinction and relapse of contextual associations*. https://doi.org/10.1101/2023.11.28.568835

Lenormand, D., & Piolino, P. (2022). In search of a naturalistic neuroimaging approach: Exploration of general feasibility through the case of VR-fMRI and application in the domain of episodic memory. *Neuroscience & Biobehavioral Reviews*, *133*, 104499. https://doi.org/10.1016/j.neubiorev.2021.12.022

Leone, G., Postel, C., Mary, A., Fraisse, F., Vallée, T., Viader, F., De La Sayette, V., Peschanski, D., Dayan, J., Eustache, F., & Gagnepain, P. (2022). Altered predictive control during memory suppression in PTSD. *Nature Communications*, *13*(1), 3300. https://doi.org/10.1038/s41467-022-30855-x

Lesting, J., Narayanan, R. T., Kluge, C., Sangha, S., Seidenbecher, T., & Pape, H.-C. (2011). Patterns of Coupled Theta Activity in Amygdala-Hippocampal-Prefrontal Cortical Circuits during Fear Extinction. *PLoS ONE*, *6*(6), e21714. https://doi.org/10.1371/journal.pone.0021714

Levy, B. (2002). Inhibitory processes and the control of memory retrieval. *Trends in Cognitive Sciences*, *6*(7), 299–305. https://doi.org/10.1016/S1364-6613(02)01923-X

Levy, B. J., & Anderson, M. C. (2012). Purging of Memories from Conscious Awareness Tracked in the Human Brain. *The Journal of Neuroscience*, *32*(47), 16785–16794. https://doi.org/10.1523/JNEUROSCI.2640-12.2012

Likhtik, E., Popa, D., Apergis-Schoute, J., Fidacaro, G. A., & Paré, D. (2008). Amygdala intercalated neurons are required for expression of fear extinction. *Nature*, *454*(7204), 642–645. https://doi.org/10.1038/nature07167

Likhtik, E., Stujenske, J. M., A Topiwala, M., Harris, A. Z., & Gordon, J. A. (2014). Prefrontal entrainment of amygdala activity signals safety in learned fear and innate anxiety. *Nature Neuroscience*, *17*(1), 106–113. https://doi.org/10.1038/nn.3582

Liu, P., Hulbert, J. C., Yang, W., Guo, Y., Qiu, J., & Anderson, M. C. (2021). Task compliance predicts suppression-induced forgetting in a large sample. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-99806-8

Lonsdorf, T. B., Klingelhöfer-Jens, M., Andreatta, M., Beckers, T., Chalkia, A., Gerlicher, A., Jentsch, V. L., Meir Drexler, S., Mertens, G., Richter, J., Sjouwerman, R., Wendt, J., & Merz, C. J. (2019). Navigating the garden of forking paths for data exclusions in fear conditioning research. *eLife*, *8*, e52465. https://doi.org/10.7554/eLife.52465

Malik, R., Li, Y., Schamiloglu, S., & Sohal, V. S. (2022). Top-down control of hippocampal signal-to-noise by prefrontal long-range inhibition. *Cell*, *185*(9), 1602-1617.e17. https://doi.org/10.1016/j.cell.2022.04.001

Mamad, O., McNamara, H. M., Reilly, R. B., & Tsanov, M. (2015). Medial septum regulates the hippocampal spatial representation. *Frontiers in Behavioral Neuroscience*, *9*. https://doi.org/10.3389/fnbeh.2015.00166

Mamat, Z., & Anderson, M. C. (2023). Improving mental health by training the suppression of unwanted thoughts. *Science Advances*, *9*(38), eadh5292. https://doi.org/10.1126/sciadv.adh5292

Marchewka, A., Żurawski, Ł., Jednoróg, K., & Grabowska, A. (2014). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods*, *46*(2), 596–610. https://doi.org/10.3758/s13428-013-0379-1

Marek, R., Sun, Y., & Sah, P. (2019). Neural circuits for a top-down control of fear and extinction. *Psychopharmacology*, *236*(1), 313–320.

https://doi.org/10.1007/s00213-018-5033-2

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. https://doi.org/10.1126/science.aay8477

McKenna, J. T., & Vertes, R. P. (2004). Afferent projections to nucleus reuniens of the thalamus. *Journal of Comparative Neurology*, *480*(2), 115–142. https://doi.org/10.1002/cne.20342

Meehan, S. M., & Schechter, M. D. (1994). Conditioned place preference/aversion to fenfluramine in fawn hooded and Sprague-Dawlet rats. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *18*(3), 575–584. https://doi.org/10.1016/0278-5846(94)90014-0

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: A network model of insula function. *Brain Structure and Function*, *214*(5–6), 655–667. https://doi.org/10.1007/s00429-010-0262-0

Milad, M. R., Pitman, R. K., Ellis, C. B., Gold, A. L., Shin, L. M., Lasko, N. B., Zeidan, M. A., Handwerger, K., Orr, S. P., & Rauch, S. L. (2009). Neurobiological Basis of Failure to Recall Extinction Memory in Posttraumatic Stress Disorder. *Biological Psychiatry*, *66*(12), 1075–1082. https://doi.org/10.1016/j.biopsych.2009.06.026

Milad, M. R., & Quirk, G. J. (2002). Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature*, *420*(6911), 70–74. https://doi.org/10.1038/nature01138

Milad, M. R., & Quirk, G. J. (2012). Fear Extinction as a Model for Translational Neuroscience: Ten Years of Progress. *Annual Review of Psychology*, *63*(1), 129–151. https://doi.org/10.1146/annurev.psych.121208.131631

Milad, M. R., Vidal-Gonzalez, I., & Quirk, G. J. (2004). Electrical Stimulation of Medial Prefrontal Cortex Reduces Conditioned Fear in a Temporally Specific Manner. *Behavioral Neuroscience*, *118*(2), 389–394. https://doi.org/10.1037/0735-7044.118.2.389

Milad, M. R., Wright, C. I., Orr, S. P., Pitman, R. K., Quirk, G. J., & Rauch, S. L. (2007). Recall of Fear Extinction in Humans Activates the Ventromedial Prefrontal Cortex and Hippocampus in Concert. *Biological Psychiatry*, *62*(5), 446–454. https://doi.org/10.1016/j.biopsych.2006.10.011

Moscarello, J. M. (2020). Prefrontal cortex projections to the nucleus reuniens suppress freezing following two-way signaled avoidance training. *Learning & Memory*, *27*(3), 119–123. https://doi.org/10.1101/lm.050377.119

Navawongse, R., & Eichenbaum, H. (2013). Distinct Pathways for Rule-Based Retrieval and Spatial Mapping of Memory Representations in Hippocampal Neurons. *The Journal of Neuroscience*, *33*(3), 1002–1013. https://doi.org/10.1523/JNEUROSCI.3891-12.2013

Norberg, M. M., Krystal, J. H., & Tolin, D. F. (2008). A Meta-Analysis of D-Cycloserine and the Facilitation of Fear Extinction and Exposure Therapy. *Biological Psychiatry*, *63*(12), 1118–1126. https://doi.org/10.1016/j.biopsych.2008.01.012

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. https://doi.org/10.1037/a0028888

Oyarzún, J. P., Lopez-Barroso, D., Fuentemilla, L., Cucurell, D., Pedraza, C., Rodriguez-Fornells, A., & De Diego-Balaguer, R. (2012). Updating Fearful Memories with Extinction Training during Reconsolidation: A Human Study Using Auditory Aversive Stimuli. *PLoS ONE*, *7*(6), e38849. https://doi.org/10.1371/journal.pone.0038849

Pavlov, I. P. (2010). Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex. *Annals of Neurosciences*, *17*(3). https://doi.org/10.5214/ans.0972-7531.1017309

Paxinos, G., & Watson, C. (1982). *The rat brain in stereotaxic coordinates*. Academic Pr.

Petrides, M., & Pandya, D. N. (1999). Dorsolateral prefrontal cortex: Comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. *European Journal of Neuroscience*, *11*(3), 1011–1036. https://doi.org/10.1046/j.1460-9568.1999.00518.x

Phelps, E. A., Delgado, M. R., Nearing, K. I., & LeDoux, J. E. (2004). Extinction Learning in Humans. *Neuron*, *43*(6), 897–905. https://doi.org/10.1016/j.neuron.2004.08.042

Prather, M. D., Lavenex, P., Mauldin-Jourdain, M. L., Mason, W. A., Capitanio, J. P., Mendoza, S. P., & Amaral, D. G. (2001). Increased social fear and decreased fear of objects in monkeys with neonatal amygdala lesions. *Neuroscience*, *106*(4), 653–658. https://doi.org/10.1016/s0306-4522(01)00445-6

Preuss, T. M. (1995). Do Rats Have Prefrontal Cortex? The Rose-Woolsey-Akert Program Reconsidered. *Journal of Cognitive Neuroscience*, *7*(1), 1–24. https://doi.org/10.1162/jocn.1995.7.1.1

Quaedflieg, C. W. E. M., Ashton, S. M., Beckers, T., & Timmers, I. (2025). Special Issue Registered Report: Intentional suppression as a method to boost fear extinction. *Journal of Behavior Therapy and Experimental Psychiatry*, *87*, 102018. https://doi.org/10.1016/j.jbtep.2025.102018

Quirk, G. J., & Mueller, D. (2008). Neural Mechanisms of Extinction Learning and Retrieval. *Neuropsychopharmacology*, *33*(1), 56–72. https://doi.org/10.1038/sj.npp.1301555

Raij, T., Nummenmaa, A., Marin, M.-F., Porter, D., Furtak, S., Setsompop, K., & Milad,

M. R. (2018). Prefrontal Cortex Stimulation Enhances Fear Extinction Memory in Humans. *Biological Psychiatry*, *84*(2), 129–137. https://doi.org/10.1016/j.biopsych.2017.10.022

Ramanathan, K. R., Jin, J., Giustino, T. F., Payne, M. R., & Maren, S. (2018). Prefrontal projections to the thalamic nucleus reuniens mediate fear extinction. *Nature Communications*, *9*(1), 4527. https://doi.org/10.1038/s41467-018-06970-z

Ramanathan, K. R., & Maren, S. (2019). Nucleus reuniens mediates the extinction of contextual fear conditioning. *Behavioural Brain Research*, *374*, 112114. https://doi.org/10.1016/j.bbr.2019.112114

Ramanathan, K. R., Ressler, R. L., Jin, J., & Maren, S. (2018). Nucleus Reuniens Is Required for Encoding and Retrieving Precise, Hippocampal-Dependent Contextual Fear Memories in Rats. *The Journal of Neuroscience*, *38*(46), 9925–9933. https://doi.org/10.1523/JNEUROSCI.1429-18.2018

Ratigan, H. C., Krishnan, S., Smith, S., & Sheffield, M. E. J. (2023). A thalamic-hippocampal CA1 signal for contextual fear memory suppression, extinction, and discrimination. *Nature Communications*, *14*(1), 6758. https://doi.org/10.1038/s41467-023-42429-6

Reeders, P. C., Rivera N., M. V., Vertes, R. P., Mattfeld, A. T., & Allen, T. A. (2022). *Identifying the midline thalamus in humans* in vivo. https://doi.org/10.1101/2022.02.20.481099

Reggente, N., Essoe, J. K.-Y., Aghajan, Z. M., Tavakoli, A. V., McGuire, J. F., Suthana, N. A., & Rissman, J. (2018). Enhancing the Ecological Validity of fMRI Memory Research Using Virtual Reality. *Frontiers in Neuroscience*, *12*, 408. https://doi.org/10.3389/fnins.2018.00408

Rescorla, R. A. (1972). Informational Variables in Pavlovian Conditioning. In *Psychology of Learning and Motivation* (Vol. 6, pp. 1–46). Elsevier. https://doi.org/10.1016/S0079-7421(08)60383-7

Rescorla, R., & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II: Current Research and Theory: Vol. Vol. 2*.

Rivera Núñez, M. V., McMakin, D. L., & Mattfeld, A. T. (2025). Nucleus reuniens: Modulating emotional overgeneralization in peri-adolescents with anxiety. *Cognitive, Affective, & Behavioral Neuroscience*, *25*(1), 173–187. https://doi.org/10.3758/s13415-024-01226-4

Rowlands, M. (2025). *Fear Not: Exploring the Role of Memory Suppression During Pavlovian Fear Extinction* [Apollo - University of Cambridge Repository]. https://doi.org/10.17863/CAM.114679

Roy, A., Svensson, F. P., Mazeh, A., & Kocsis, B. (2017). Prefrontal-hippocampal coupling by theta rhythm and by 2–5 Hz oscillation in the delta band: The role of the nucleus reuniens of the thalamus. *Brain Structure and Function*, *222*(6), 2819–2830. https://doi.org/10.1007/s00429-017-1374-6

Sankarasubramanian, S. (2022). *Domain-general control mechanisms underlying stopping of thought and action* [Apollo - University of Cambridge Repository]. https://doi.org/10.17863/CAM.91796

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. https://doi.org/10.1038/nature08637

Schmitz, T. W., Correia, M. M., Ferreira, C. S., Prescot, A. P., & Anderson, M. C. (2017). Hippocampal GABA enables inhibitory control over unwanted thoughts. *Nature Communications*, *8*(1), 1311. https://doi.org/10.1038/s41467-017-00956-z

Sierra, R. O., Pedraza, L. K., Zanona, Q. K., Santana, F., Boos, F. Z., Crestani, A. P., Haubrich, J., De Oliveira Alvares, L., Calcagnotto, M. E., & Quillfeldt, J. A. (2017). Reconsolidation-induced rescue of a remote fear memory blocked by an

early cortical inhibition: Involvement of the anterior cingulate cortex and the mediation by the thalamic nucleus reuniens. *Hippocampus*, *27*(5), 596–607. https://doi.org/10.1002/hipo.22715

Stramaccia, D. F., Meyer, A.-K., Rischer, K. M., Fawcett, J. M., & Benoit, R. G. (2021). Memory suppression and its deficiency in psychological disorders: A focused meta-analysis. *Journal of Experimental Psychology: General*, *150*(5), 828–850. https://doi.org/10.1037/xge0000971

Streb, M., Mecklinger, A., Anderson, M. C., Lass-Hennemann, J., & Michael, T. (2016). Memory control ability modulates intrusive memories after analogue trauma. *Journal of Affective Disorders*, *192*, 134–142. https://doi.org/10.1016/j.jad.2015.12.032

Suzuki, W., & Amaral, D. (1994). Topographic organization of the reciprocal connections between the monkey entorhinal cortex and the perirhinal and parahippocampal cortices. *The Journal of Neuroscience*, *14*(3), 1856–1877. https://doi.org/10.1523/JNEUROSCI.14-03-01856.1994

Talmi, D., Anderson, A. K., Riggs, L., Caplan, J. B., & Moscovitch, M. (2008). Immediate memory consequences of the effect of emotion on attention to pictures. *Learning & Memory*, *15*(3), 172–182. https://doi.org/10.1101/lm.722908

Tomaszewski, K. F., Ziółkowska, M., Łukasiewicz, K., Cały, A., Sotoudeh, N., Puchalska, M., Salamian, A., & Radwanska, K. (2024). *Projections from thalamic nucleus reuniens to medial septum enable extinction of remote fear memory*. https://doi.org/10.1101/2024.05.20.594930

Totty, M. S., Tuna, T., Ramanathan, K. R., Jin, J., Peters, S. E., & Maren, S. (2023). Thalamic nucleus reuniens coordinates prefrontal-hippocampal synchrony to suppress extinguished fear. *Nature Communications*, *14*(1), 6565. https://doi.org/10.1038/s41467-023-42315-1

Troyner, F., Bicca, M. A., & Bertoglio, L. J. (2018a). Nucleus reuniens of the thalamus

controls fear memory intensity, specificity and long-term maintenance during consolidation. *Hippocampus*, *28*(8), 602–616. https://doi.org/10.1002/hipo.22964

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology / Psychologie Canadienne*, *26*(1), 1–12. https://doi.org/10.1037/h0080017

Tuna, T., Totty, M. S., Badarnee, M., Afonso Gonçalves Mourão, F., Peters, S., Milad, M. R., & Maren, S. (2025). *Associative coding of conditioned fear in the thalamic nucleus reuniens in rodents and humans*. https://doi.org/10.1101/2025.03.18.643915

Uylings, H. B. M., Groenewegen, H. J., & Kolb, B. (2003). Do rats have a prefrontal cortex? *Behavioural Brain Research*, *146*(1–2), 3–17. https://doi.org/10.1016/j.bbr.2003.09.028

Van Schie, K., & Anderson, M. C. (2017). Successfully controlling intrusive memories is harder when control must be sustained. *Memory*, *25*(9), 1201–1216. https://doi.org/10.1080/09658211.2017.1282518

Vantomme, G., Devienne, G., Hull, J. M., & Huguenard, J. R. (2024). *Reuniens thalamus recruits recurrent excitation in medial prefrontal cortex*. https://doi.org/10.1101/2024.05.31.596906

Varela, C., Kumar, S., Yang, J. Y., & Wilson, M. A. (2014). Anatomical substrates for direct interactions between hippocampus, medial prefrontal cortex, and the thalamic nucleus reuniens. *Brain Structure and Function*, *219*(3), 911–929. https://doi.org/10.1007/s00429-013-0543-5

Vasudevan, K., Ramanathan, K. R., Vierkant, V., & Maren, S. (2022). Nucleus reuniens inactivation does not impair consolidation or reconsolidation of fear extinction. *Learning & Memory*, *29*(8), 216–222. https://doi.org/10.1101/lm.053611.122

Vertes, R. P. (2002). Analysis of projections from the medial prefrontal cortex to the thalamus in the rat, with emphasis on nucleus reuniens. *Journal of Comparative*

Neurology, *442*(2), 163–187. https://doi.org/10.1002/cne.10083

Vertes, R. P. (2004). Differential projections of the infralimbic and prelimbic cortex in the rat. *Synapse*, *51*(1), 32–58. https://doi.org/10.1002/syn.10279

Vertes, R. P. (2006). Interactions among the medial prefrontal cortex, hippocampus and midline thalamus in emotional and cognitive processing in the rat. *Neuroscience*, *142*(1), 1–20. https://doi.org/10.1016/j.neuroscience.2006.06.027

Vertes, R. P. (2015). Major diencephalic inputs to the hippocampus. In *Progress in Brain Research* (Vol. 219, pp. 121–144). Elsevier. https://doi.org/10.1016/bs.pbr.2015.03.008

Vertes, R. P., Hoover, W. B., Szigeti-Buck, K., & Leranth, C. (2007). Nucleus reuniens of the midline thalamus: Link between the medial prefrontal cortex and the hippocampus. *Brain Research Bulletin*, *71*(6), 601–609. https://doi.org/10.1016/j.brainresbull.2006.12.002

Viena, T. D., Linley, S. B., & Vertes, R. P. (2018). Inactivation of nucleus reuniens impairs spatial working memory and behavioral flexibility in the rat. *Hippocampus*, *28*(4), 297–311. https://doi.org/10.1002/hipo.22831

Viena, T. D., Rasch, G. E., Silva, D., & Allen, T. A. (2021). Calretinin and calbindin architecture of the midline thalamus associated with prefrontal–hippocampal circuitry. *Hippocampus*, *31*(7), 770–789. https://doi.org/10.1002/hipo.23271

Visser, R. M., Bathelt, J., Scholte, H. S., & Kindt, M. (2021). Robust BOLD Responses to Faces But Not to Conditioned Threat: Challenging the Amygdala's Reputation in Human Fear and Extinction Learning. *The Journal of Neuroscience*, *41*(50), 10278–10292. https://doi.org/10.1523/JNEUROSCI.0857-21.2021

Wang, J., John, Y., & Barbas, H. (2021). Pathways for Contextual Memory: The Primate Hippocampal Pathway to Anterior Cingulate Cortex. *Cerebral Cortex*, *31*(3), 1807–1826. https://doi.org/10.1093/cercor/bhaa333

Wang, Y., Zhu, Z., Hu, J., Schiller, D., & Li, J. (2021). Active suppression prevents the return of threat memory in humans. *Communications Biology*, *4*(1), 609. https://doi.org/10.1038/s42003-021-02120-2

Xu, W., & Südhof, T. C. (2013). A Neural Circuit for Memory Specificity and Generalization. *Science*, *339*(6125), 1290–1295. https://doi.org/10.1126/science.1229534

Zeidman, P., Jafarian, A., Corbin, N., Seghier, M. L., Razi, A., Price, C. J., & Friston, K. J. (2019). A guide to group effective connectivity analysis, part 1: First level analysis with DCM for fMRI. *NeuroImage*, *200*, 174–190. https://doi.org/10.1016/j.neuroimage.2019.06.031