UNIVERSITY OF CALIFORNIA SAN DIEGO

Executive control over actions and long term memory retrieval

A dissertation submitted in partial satisfaction of
the requirements for the degree Doctor of Philosophy

in

Experimental Psychology

by

Kelsey Kathleen Sundby

Committee in charge:

Professor Adam R. Aron, Chair
Professor Timothy Brady
Professor Gedeon Deak
Professor Anastasia Kiyonaga
Professor John Serences

2021

The Dissertation of Kelsey Kathleen Sundby is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To my friends, family and mentors.

EPIGRAPH

Cognitive control processes live in the murky spaces between knowledge and action, influencing the translation from the first to the second while not being either one.

-David Badre

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I have a tremendous amount of people to acknowledge for helping me reach this milestone. I would first like to acknowledge my advisor Adam Aron who helped me grow more than I knew was possible over the past 5 years. Thank you for helping me reach my goals and for supporting me at every turn. Your insightfulness, work ethic, and passion for making the world a better place sets an example of the type of scientist I hope to become.

I would also like to sincerely thank the rest of my dissertation committee- Tim Brady, Gedeon Deak, Anastasia Kiyonaga, and John Serences. This was a challenging year for many reasons and I am very fortunate to have had this committee to help me continue to make progress. Your willingness to always find time to meet with me and your constructive and thoughtful feedback gave me the courage to branch outside of my comfort zone, explore new methods, and persevere.

My two collaborators, David Sutterer and Sirawaj Itthipuripat, were both incredible scientists to work with and always met me with compassion and patience despite my badgering e-mails. I would also like to thank Kirsten Adams, who I never formally worked with and who had no obligation to help me yet was incredibly generous with her time and was always willing to answer questions and brainstorm science.

Lastly, I certainly could not have gotten where I am today without the unwavering support of my friends and family. This includes my mother Katie Sundby and my father Scott Sundby, who always encouraged me to pursue what made me happiest and who spent countless hours discussing my research with me despite neither having a background in science. My two brothers, Taylor Sundby and Christopher Sundby, as well as my two sister-in-laws, Gillian Starkey and Taylor Peabody, were a constant source of support and comic relief. My fiancé, Matt

Rohn, whole heartedly supported me and kept me grounded, reminding me to appreciate this time even during the inevitable ups and downs of graduate school. I also have the whole Aron lab to thank including past and present members. In addition to being an incredible source of laughs, collaboration and scientific mentorship, this group became like a family to me that I knew I could always turn to. In particular, I would like to thank Dawn Finzi who is a good friend, a truly inspiring scientist, and somebody who was always there when I needed a hand. I also owe a huge thank you to an amazing group of friends outside the lab (a special thanks to Katie Van Alstyne, Tanushree Agrawal, Jonathan Keefe, and Taylor Butler) who made these past five years some of my best and who somehow always knew when I needed a cup of coffee or a glass of wine.

Chapter 1, in full, is a reprint of the material as it appears in the Journal of Neurophysiology 2020, Sundby, Kelsey; Jana, Sumitash; Aron, Adam. The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, is currently being prepared for submission for the publication of the material. Sundby Kelsey; Itthipuripat, Sirawaj; Aron, Adam. The dissertation author was the primary investigator and author of this paper.

Chapter 3 is coauthored with David W. Sutterer, Sirawaj Itthipuripat and Adam Aron. The dissertation author was the primary investigator and author of this paper.

VITA

2014 Bachelor of Arts, University of Virginia

2018 Master of Arts, Psychology, University of California San Diego

2021 Doctor of Philosophy, Experimental Psychology, University of California San Diego

PUBLICATIONS

Sundby K**,** Jana S, Aron AR (2021). Double blind disruption of right inferior frontal cortex with TMS reduces right frontal beta power for action-stopping. *Journal of Neurophysiology*

Hannah R, Muralhidaran V, Sundby K, Aron AR (2020). Temporally-precise disruption of prefrontal cortex informed by the timing of beta bursts impairs human action-stopping. *Neuroimage.*

Sundby, K**.**, Johanna Wagner, and Adam R. Aron (2019). The functional role of response suppression during an urge to relieve pain. *Journal of cognitive neuroscience* 31.9: 1404-1421.

Grodin, E. N., Sussman, L., Sundby, K**.,** Brennan, G. M., Diazgranados, N., Heilig, M., & Momenan, R. (2018). Neural correlates of compulsive alcohol seeking in heavy drinkers. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(12), 1022-1031.

Zhu, X.[1*], Sundby, K.[1*], Bjork, J.[2], Momenan, R.[1]. (2016). Alcohol Dependence and Altered Engagement of Brain Networks in Risky Decisions. *Frontiers in human neuroscience*, 10, 142.

ABSTRACT OF THE DISSERTATION

Executive control over actions and long term memory retrieval

by

Kelsey Kathleen Sundby

Doctor of Philosophy in Experimental Psychology

University of California San Diego, 2021

Professor Adam R. Aron, Chair

Environmental cues frequently elicit the retrieval of items stored in long term memory. In many cases, cued retrieval serves a valuable function, providing important reminders and connecting us to our past. At times, however, cues can elicit the retrieval of unwanted memories. A core question in cognitive psychology is how individuals are able to prevent these unwanted memory intrusions. Neuroimaging studies suggest that "canceling" retrieval requires a right prefrontal (PFC) network that is also recruited when we cancel actions. Our ability to study how this network exerts control over retrieval, however, is limited in several ways: 1) current studies use methods with poor temporal resolution that are unable to determine *when* this network

interrupts retrieval, and 2) measuring memory intrusions relies largely on self-report rather than objective measures. This dissertation aims to address these limitations. In Chapter 1, we address the first limitation by validating an EEG marker of the right PFC network. EEG has high temporal resolution, making it a valuable tool for testing the timing of network recruitment. In the two subsequent chapters, we address the second limitation by developing methods to objectively track the emergence of memories during efforts to suppress. We first show that an event related potential, known to track the amount of retrieved information, is reduced when trying to prevent retrieval (Chapter 2). We next used EEG-based decoding and tracked the specific memory rather than the quantity of retrieved information to more accurately determine whether participants were successful in avoiding retrieval (Chapter 3). Although we could still decode the prohibited memory, decoding accuracy was reduced later during retrieval, which may reflect efforts to "suppress" the intruding content. Lastly, by combining these tools, we found that early recruitment of the right PFC network corresponded to reduced decoding and thus, measureable changes in the prohibited memory representation (Chapter 4). Together these data may warrant a two-phase model of control that involves both early and late control processes. Overall, this dissertation contributes necessary tools for gaining a deeper understanding of the right PFC network and our ability to control when past experiences gain access to working memory.

GENERAL INTRODUCTION

Cues in our environment often trigger the retrieval of items stored in long term memory (LTM). In most cases, cued retrieval is beneficial, reminding us of impending dates, important tasks, and fond memories. At times, however, cues can elicit the retrieval of unwanted memories. A profound question in cognitive psychology is whether and how we are able to prevent the intrusion of unwanted memories upon encountering an associated cue. Several studies suggest that preventing cued retrieval requires an executive process analogous to how we cancel actions and that both processes are implemented via a right prefrontal (PFC) network (Brendan E Depue, Orr, Smolker, Naaz, & Banich, 2016; Guo, Schmitz, Mur, Ferreira, & Anderson, 2018). We currently, however, lack a detailed understanding of how the right PFC network disrupts a cognitive process like retrieval. There are at least two proposed mechanisms: one in which the right PFC acts to clear intruding information (Levy & Anderson, 2012) and another in which the right PFC "gates" working memory to prevent the influx of a cued memory (Castiglione, Wagner, Anderson, & Aron, 2019). Our ability to test these models is limited by the current set of tools for determining *when* the right PFC network intervenes and for objectively tracking the retrieval of memory information. This dissertation seeks to address each of these limitations in turn.

The current chapter reviews the extant literature on executive control over cued retrieval and expands on the limitations that this dissertation targets. We start with an overview of the "direct suppression" account, how it is studied in a lab setting and the behavioral data that support it. We then describe the timeline of long term memory retrieval which is important for considering when executive control might intervene. The subsequent section reviews several of the other existing methods for operationalizing executive control over memory and explains how

these differ from the Think/NoThink approach. We then move on to describe the candidate right PFC network in more detail, including how it supports action stopping and the current evidence linking it to control over cued retrieval. The chapter ends by summarizing the two most prominent theories for how the right PFC network exerts control over retrieval and the tools that are still needed to rigorously test these theories.

The Direct Suppression Account

Executive control refers to the cognitive processes used to select and enact actions adherent to our goals. Many psychological theories cite inhibition as an essential component of executive control (for review see Aron, 2007). According to these theories, we use inhibition to constrain the many rogue actions, emotions, and thoughts that would otherwise interfere with the appropriate response (e.g., Brendan E Depue et al., 2016; Hasher, Lustig, & Zacks, 2007; Levy & Anderson, 2002; Simpson & Kang, 1994). In the motor domain, the concept of inhibition is unequivocal. We can clearly, for example, rapidly inhibit motor output (i.e. stepping out into the road), if the situation demands it. In the case of motor control, there is clear evidence from TMS (Badry et al., 2009; Coxon, Stinear, & Byblow, 2006; van den Wildenberg et al., 2010), EEG (Swann et al., 2009; Wessel, 2020), and fMRI (Aron & Poldrack, 2006), of reduced activation in primary motor cortex upon inhibiting a response. Further, these inhibitory effects on motor cortex are thought to be driven by top-down signals from prefrontal cortex (for review, Aron, Robbins, & Poldrack, 2014; Jahanshahi, Obeso, Rothwell, & Obeso, 2015; Wiecki & Frank, 2013). Thus, evidence for inhibitory control in the motor system abides by the view that inhibition is a process, initiated by one region or network (e.g. the prefrontal cortex), that functions to reduce activity in a downstream target (e.g. motor cortex) (Aron, 2007).

A far more debated assertion is that we use inhibitory processes to control aspects of cognition including, for example, long term memory retrieval. Akin to canceling actions, we may often benefit from being able to cancel the retrieval of specific memories. This is true, for example, when we encounter a cue that prompts the retrieval of a painful or distracting memory. According to the direct suppression account, individuals are able to abort the retrieval process via a right prefrontal network that acts to inhibit the hippocampus, thereby disrupting the retrieval process (Benoit & Anderson, 2012). This phenomenon termed "retrieval suppression" is primarily studied using the Think/No-Think (TNT) paradigm (Anderson & Green, 2001).

The Think/NoThink Task (TNT)

TNT is typically a three phase behavioral paradigm. During Phase1 participants learn multiple word pairs[1] (e.g. OIL-PUMP) . The pairs presented in Phase1 are then assigned to one of three conditions: Think, NoThink, or Baseline. The Baseline word pairs are excluded from the subsequent Phase2 and only reappear in the final Phase3. In Phase2, participants are presented with the left-hand word from one of the previously learned pairs. The color of this cue word indicates if the trial is a "Think" trial, prompting the participant to silently retrieve the associated word from memory, or if it is a "NoThink" trial in which the participant must try to inhibit retrieval and prevent the memory from entering into consciousness. Lastly, in the original task version, participants are given a surprise recall test (Phase3) in which they are tested on their memory for all of the pairs learned in Phase1, including Baseline pairs.

Behavioral support for the direct suppression account

[1]The word pair version of TNT is the original and mostly widely used version of the task. Multiple researchers, however, have created versions of TNT that use richer stimuli that may better model our real world memory experience. For example, the NoThink effect has been observed for tasks using emotional stimuli (Brendan E Depue, Curran, & Banich, 2007; Küpper, Benoit, Dalgleish, & Anderson, 2014; Lambert, Good, & Kirk, 2010; van Schie, Geraerts, & Anderson, 2013), faces (Brendan E Depue, Banich, & Curran, 2006), and autobiographical content (Noreen & MacLeod, 2013; Stephens, Braid, & Hertel, 2013) as the memory pairs.

Several pieces of behavioral data have been interpreted as support for the direct suppression account. First, participants show suppression induced forgetting (SIF) or impaired recall for NoThink associates compared to both Think and Baseline associates that were learned in Phase1 but never underwent the Think/NoThink manipulation (e.g., Anderson & Green, 2001; Benoit & Anderson, 2012; Z. M. Bergström, J. W. de Fockert, & A. Richardson-Klavehn, 2009b; Castiglione et al., 2019; Brendan E Depue et al., 2007; Detre, Natarajan, Gershman, & Norman, 2013; Levy & Anderson, 2008; Racsmány, Conway, Keresztes, & Krajcsi, 2012; van Schie et al., 2013; Gerd Thomas Waldhauser, Lindgren, & Johansson, 2012).[2] This latter comparison demonstrates that the observed difference between Think and NoThink is not purely driven by improved memory for the Think items due to practice retrieving. The reduction in NoThink recall is instead thought to reflect an active inhibitory process that affects the memory trace itself.

There are, however, several non-inhibitory explanations for the decline in final recall. For example, people may avoid the NoThink associate by diverting attention to an alternative thought. These diversionary thoughts may form an association with the cue word that then produces interference during later efforts to retrieve the original associate (Hertel & Calcaterra, 2005; Lewis-Peacock & Norman, 2014; Tomlinson, Huber, Rieth, & Davelaar, 2009). Additionally, canceling retrieval may promote "unlearning" or a weakening of the association between the cue and the associate word (Melton & Irwin, 1940). Importantly, SIF effects are cue-independent, meaning participants fail to retrieve the NoThink associate even when prompted with a different cue. The Independent Probe (IP) test, an additional measure of final recall, reveals this effect. During an IP test, participants receive the first letter of one of the

[2] There are also, however, a number of studies that do not find the SIF effect (Bulevich, Roediger, Balota, & Butler, 2006; Hertel & Calcaterra, 2005; Hertel & Gerstle, 2003).

previously learned associates and a cue word that is semantically related to the associate yet new to the task. For example, if the original pair was “ordeal-roach”, participants might be tested with “insect r-.” Interestingly, participants continue to perform worse for NoThink items compared to Think and Baseline even for novel cue words (Anderson & Green, 2001). Neither an interference account (i.e. interference from an association between a diversionary thought and the original cue) nor an unlearning account (i.e. a weakened link between the associate and the original cue) predicts a forgetting effect with a novel cue.

Additional evidence of active inhibition comes from the pattern of reported intrusions during TNT. Intrusion ratings, a newer task addition, ask participants after each trial to report whether the associate came to mind “never”, “briefly”, or “often” (Levy & Anderson, 2012). Participants consistently report fewer intrusions for NoThink trials later in the task (Benoit, Hulbert, Huddleston, & Anderson, 2015; Castiglione et al., 2019; Gagnepain, Hulbert, & Anderson, 2017; Harrington, Ashton, Sankarasubramanian, Anderson, & Cairney, 2021; Legrand et al., 2020; Levy & Anderson, 2012; Mary et al., 2020). According to the direct suppression account, this decline in reported intrusions may reflect a combination of two things: 1) people becoming more practiced in exerting inhibitory control and 2) the cumulative consequence of repeatedly inhibiting the NoThink associate making it less retrievable and thus, less likely to intrude (Levy & Anderson, 2012).

Poor TNT performance is also linked to inhibitory deficits (for reviews see Levy & Anderson, 2008; Stramaccia, Meyer, Rischer, Fawcett, & Benoit, 2020). Populations including ADHD patients (Brendan E Depue, Burgess, Willcutt, Ruzic, & Banich, 2010) and individuals with alcohol abuse (Nemeth et al., 2014), both of which are commonly characterized by impaired inhibitory control, have difficulty suppressing retrieval. Impaired performance also occurs in

populations that are particularly vulnerable to intrusive thought patterns including PTSD (Catarino, Küpper, Werner-Seidler, Dalgleish, & Anderson, 2015; Mary et al., 2020; Sullivan et al., 2019; Gerd T Waldhauser et al., 2018), depression (Hertel & Gerstle, 2003; Joormann, Hertel, LeMoult, & Gotlib, 2009), and individuals with high trait anxiety (Marzi, Regina, & Righi, 2014) or an increased tendency to ruminate (Fawcett et al., 2015; Hertel, Maydon, Ogilvie, & Mor, 2018).

It is noteworthy that the extant evidence for inhibitory control over retrieval relies on retroactive measures (i.e. later forgetting, post-hoc intrusion ratings, etc.). As a result, these measures fail to capture what happens in the moment that an individual must implement control. Direct evidence that the right PFC reduces activation of the specific prohibited memory during the "suppression" window remains to be shown.

The timeline of episodic memory retrieval

To better understand how memories become susceptible to manipulation it is useful to take a broader look at how episodic memories are formed and retrieved. Current research suggests that memory encoding occurs when an ensemble of hippocampal and cortical neurons co-activate during our initial experience. This unique and memory-specific pattern of hippocampal-cortical activation is sometimes referred to as the "engram" or the memory trace (Tonegawa, Liu, Ramirez, & Redondo, 2015). According to hippocampal indexing theory, the hippocampus stores an index of the cortical sites that participated in the original episode via the strengthening of synaptic weights (Teyler & Rudy, 2007). After encoding the initial experience, a related cue can then activate the hippocampal index (i.e. intrahippocampal pattern completion) which subsequently reactivates the entire cortical representation (i.e. cortical reinstatement) (Staresina & Wimber, 2019).

Electrophysiological data provide a timeline of cued retrieval, beginning with cue processing and ending with our phenomenological experience of the past episode (for review see Staresina & Wimber, 2019). These data suggest that cue-specific processing transpires across the first 200ms of encountering a cue. Within ~500ms, cue information arrives at the medial temporal lobe (MTL), where an 'old/new' signal is generated. If an old signal is detected, the hippocampus performs pattern completion which reactivates the hippocampal index. Single neuron recordings from humans show that pattern completion occurs approximately 500ms after encountering the cue. Finally, cortical reinstatement or the reactivation of the full cortical engram emerges 500-1500ms post cue and leads to our conscious experience of remembering (Staresina et al., 2019).

As discussed in more detail below (see Two accounts for executive control over LTM), a control process could intervene at multiple points in the retrieval timeline. The target of executive control (e.g. cue processing, cortical reinstatement, etc.) and the precise function (e.g. blocking or clearing information) likely depends on when control interjects.

Operationalizing executive control over LTM

Executive control over LTM is studied in a number of ways. Although this dissertation focuses on "stopping" cued retrieval, here we briefly review other methods to operationalize and study executive control over LTM apart from the TNT task (Figure 0.1). This is not an exhaustive list but instead focuses on a few prominent examples and how they differ from the TNT approach. This body of research largely supports the general principle that memory can be actively tuned by executive processes.

1. ***Memory Inhibition During Selective Retrieval.*** Memory inhibition is when the inhibitory process targets the active neural representation of a memory (Anderson &

Hulbert, 2021). This may be particularly useful during selective retrieval when a cue activates multiple competing memories and we must only retrieve one. For example, when standing at the wedding altar it is critical that we retrieve our current partner's name as opposed to a former partner. A growing body of evidence suggests that we actively inhibit competing memories during selective retrieval and that this produces retrieval induced forgetting (RIF) or impaired recall for the competing items (e.g., Anderson & Bell, 2001; Anderson, Bjork, & Bjork, 1994; Cinel, Cortis Mack, & Ward, 2018; Ciranni & Shimamura, 1999; Glynn, Salmon, & Low, 2019; Johnson & Anderson, 2004). A standard RIF task has participants study multiple categories and specific exemplars within those categories (e.g. Fruit-Orange, Fruit-Banana or Drinks-Vodka, Drinks-Wine etc.). Participants are then prompted to practice the retrieval for one of the studied categories and importantly, for only a subset of the exemplars shown for that category. For example, participants may practice the retrieval for "Orange" (but not "Banana") in the category Fruit (e.g. prompted with Fruit-O). Theories of inhibition assume that the cue "Fruit" elicits activation of multiple category members and that correctly selecting the target memory requires inhibition of the active competing memories. It follows then, that selectively retrieving Orange (and not Banana) requires the active inhibition of the memory for Banana. Support for this comes from a final recall test that instructs participants to report all of the remembered items in the practiced category. Participants retrieve fewer of the unpracticed competitors compared to the retrieved items in the practiced category. Further, memory for the unpracticed competitors is impaired even compared to memory for items in a baseline category that was studied in Phase1 but did not undergo any retrieval practice in Phase2 (e.g. items in

the category Drinks). A large literature on selective retrieval examines other non-inhibitory explanations for the drop in recall as well as several counterarguments (e.g. cue-independence) (for review Anderson & Hulbert, 2021) that continue to point to inhibition as a key contributor (but also see Raaijmakers & Jakab, 2013).

A core difference between selective retrieval and retrieval suppression is that inhibition during selective retrieval is an unintentional byproduct of efforts to retrieve the target memory. Alternatively, TNT creates an instance in which an individual intends to suppress retrieval. In this sense, studies on selective retrieval have parallels with the concept of lateral inhibition in the motor system, a process by which boosting activation of a target response results in the automatic inhibition of competing response representations during action selection (Duque, Lew, Mazzocchio, Olivier, & Ivry, 2010).

2. ***Directed Forgetting.*** Another line of work provides an example of process inhibition. In this case the target of inhibition is not a specific memory representation but rather a memory process. Directed forgetting paradigms reveal voluntary control over memory encoding. In these tasks, researchers present stimuli (either lists or one item at a time) to participants and instruct participants after each trial to either remember or forget the last seen item/list. In a subsequent final recall test, participants tend to recall more Remember items compared to Forget items. Again, the difference in recall could occur for a number of reasons. For instance, participants may perform selective rehearsal or more elaborate encoding of an item when instructed to remember it. Instead, if instructed to forget, people may simply abandon rehearsal (Basden, Basden, & Gargano, 1993). Alternatively, encoding may be actively suppressed upon receiving a forget cue (Zacks, Radvansky, & Hasher, 1996). Although researchers still debate the mechanism underlying directed

forgetting, the overall results demonstrate an ability to voluntarily exert control over memory and the processes that support it. Retrieval suppression and TNT, however, target a very different question than directed forgetting. In the case of directed forgetting, control processes (inhibition or otherwise) intervene during memory encoding when the memory may be in a more malleable state. Retrieval suppression instead proposes a theory for how memories can be altered even after the memory has been successfully encoded into LTM.

3. ***The White Bear Effect.*** At first glance, the TNT literature appears to severely contradict one of the earliest studies on thought suppression which concluded that it was a largely futile and perhaps even counterproductive strategy. In the iconic White Bear paradigm, Wegner, Schneider, Carter, and White (1987) instructed participants to not think about a white bear for five minutes. Participants were told to speak their stream of consciousness and to ring a bell to report any moments in which the prohibited thought of a white bear came to mind. Wegner et al. (1987) discovered that participants frequently reported thinking of the white bear despite the task instruction. Further, participants who were instructed to suppress thoughts of a white bear for 5 minutes and then had a 5 minute free expression period, during which thoughts of the white bear were permissible, reported far *more* white bear thoughts during free expression compared to participants who performed the task in the opposite order (i.e. free express and then suppress). The term paradoxical rebound effect refers to this uptick in the occurrence of a thought post suppression and is thought to reflect a later preoccupation with the suppressed thought. Wegner et al. (1987) concluded that suppression strategies risk a maladaptive cycle that may actually magnify the consequence of unwanted thoughts.

Critiques of the White Bear paradigm note, however, that the design conflates the task goal with the task instruction (Visser et al., 2020). In other words, to abide by the instruction to "not think of a white bear" and to report violations, requires that the participant repeatedly reactivate the thought and continuously check, "Am I thinking of a white bear?" Therefore, the task instruction both places the to-be-suppressed item in mind and requires that the participant defy the task goal to assess whether or not they are succeeding.

Conversely, retrieval suppression includes an intermediate step (i.e. retrieval), that must occur before the prohibited item is in mind. The memory cue itself is not the to-be-suppressed item. Thus, this intermediate step provides an opportunity for control mechanisms to intervene either before the completion of retrieval or concurrently, shutting down activation of the emerging information.

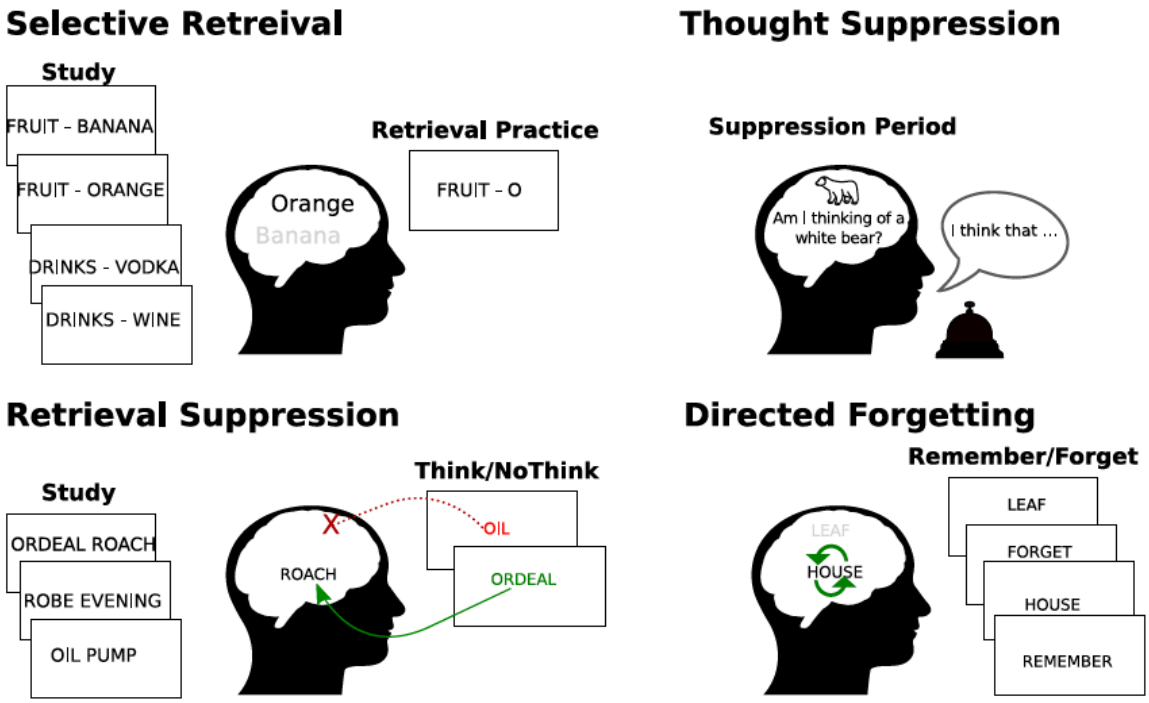

Figure 0.1. An illustration of several common methods for operationalizing executive control over long term memory.

A right prefrontal network for controlling actions and LTM retrieval

The current section provides more detail on the candidate right PFC network. Many aspects of motor control ranging from action selection to action stopping rely on fronto-basal-ganglia networks (Aron, 2011; Mink, 1996; Wiecki & Frank, 2013). In the case of action stopping, the process is thought to operate via the hyperdirect pathway, a "shortcut" connecting the right inferior frontal gyrus (rIFG) with the subthalamic nucleus (STN) (Aron, Behrens, Smith, Frank, & Poldrack, 2007; W. Chen et al., 2020; Nambu, Tokuno, & Takada, 2002). With direct excitatory projections from the right PFC to the STN, the hyperdirect pathway enables rapid inhibition of the thalamus that then cuts off drive to motor cortex.

The majority of work on action-stopping and the right PFC network uses the stop signal task (Logan & Cowan, 1984). In a basic version of the task, participants press a button in response to the direction of a "Go cue", an arrow pointing either leftward or rightward. On a minority of trials, typically 25%, the Go cue turns red after a specified delay period, called the stop signal delay. The red arrow is the stop signal and prompts the participant to try to inhibit the initiated button response. This task has provided many insights into executive control over actions despite being a very simple model. In addition to mapping out the network architecture, research using the task helped identify beta oscillations as a potential electrophysiological signature of the network and a channel of communication (Enz, Ruddy, Rueda-Delgado, & Whelan, 2021; Hannah, Muralidharan, Sundby, & Aron, 2020; Schaum et al., 2021; Swann et al., 2009; Wagner, Wessel, Ghahremani, & Aron, 2018; Wessel et al., 2016).

The direct suppression account, contends that suppressing cued retrieval also engages key regions of the right PFC network that has been well established in the action-stopping literature. Support for this comes from fMRI studies that show shared cortical and subcortical nodes of

activation during both action-stopping tasks and the TNT task (Benoit & Anderson, 2012; Benoit et al., 2015; Brendan E Depue et al., 2016; Guo et al., 2018; Levy & Anderson, 2012). The major point of divergence is thought to occur downstream, i.e. the target of inhibitory control. During TNT, instead of targeting motor cortex, activation of the right PFC corresponds to reduced activity in the hippocampus (Anderson et al., 2004; Benoit & Anderson, 2012; Benoit et al., 2015; Brendan E Depue et al., 2007; Gagnepain, Henson, & Anderson, 2014; Schmitz, Correia, Ferreira, Prescot, & Anderson, 2017)[3].The degree of hippocampal downregulation relates to the magnitude of forgetting during final recall and may be driven by GABAergic mechanisms in the hippocampus (Levy & Anderson, 2012; Schmitz et al., 2017).

Retrieval suppression recruits two regions of the right lateral PFC: the right ventral lateral prefrontal cortex (rVLPFC) and the right dorsolateral prefrontal cortex (rDLPFC)[4]. Although both regions are also involved in action-stopping, the rVLPFC, and especially the right inferior frontal gyrus (rIFG), is a critical node (Aron et al., 2014; Bari & Robbins, 2013; Cai, Ryali, Chen, Li, & Menon, 2014; Garavan, Ross, & Stein, 1999; Jahanshahi et al., 2015). Human lesion studies (Aron, Fletcher, Bullmore, Sahakian, & Robbins, 2003) and disruptive TMS (Chambers et al., 2006) both support the rIFG as a core region for inhibiting actions. There is no clear consensus on whether or how the rDLPFC and rVLPFC uniquely contribute to retrieval suppression. Whereas several studies report a strong link between rDLPFC and hippocampal modulation (Benoit et al., 2015; Levy & Anderson, 2012; Schmitz et al., 2017), other work

[3] Reduced activity in the hippocampus may also reflect increased task difficulty (Reas & Brewer, 2013). However, the relationship between behavior (i.e. final forgetting and reported intrusions) and the observed hippocampal downregulation helps to discount difficulty as the most likely explanation.

[4] Both the rVLPFC and rDLPFC are also implicated in selective retrieval and directed forgetting. Although this dissertation focuses on the overlap between action-stopping and retrieval suppression, a broader hypothesis termed the "prefrontal control hypothesis" extends the idea of a supramodal inhibitory control network to these other memory domains (Anderson & Hulbert, 2021).

highlights both as key contributing nodes that flexibly target the hippocampus or the motor cortex depending on task demands (Apšvalka, Ferreira, Schmitz, Rowe, & Anderson, 2020).

Without further support, however, the claim that overlapping nodes of activation denotes a common inhibitory process across domains (e.g. control of actions and control of memory) is problematic because it relies on reverse inference (Poldrack, 2006). That is, the mere observation that the same brain region is recruited during two types of behaviors (i.e. stopping actions and stopping retrieval) does not imply that the region is performing the same cognitive function (i.e. inhibition) in each case. This logic would only be accurate if the region of interest only contributes to one cognitive function. Despite compelling evidence that the right PFC is critical for action-stopping, few would claim that action-stopping is its sole function[5].

Importantly, however, canceling retrieval produces other established hallmarks of the action-stopping network. Like suppressing actions, suppressing retrieval yields an increase in right frontal beta power, a proposed electrophysiological marker of the underlying right PFC network (Castiglione et al., 2019; Wagner et al., 2018). Further, using transcranial magnetic stimulation (TMS), Castiglione and Aron (2021) recently found that suppressing retrieval in the TNT task resulted in reduced motor excitability, measured by motor evoked potentials (MEP), in a task irrelevant muscle. Smaller MEPs reflect reduced corticospinal excitability (i.e. inhibition of motor cortex). Reduced MEPs in a task irrelevant muscle may reflect global motor suppression, a phenomenon consistently observed during reactive action-stopping. For instance, stopping saccades results in reduced MEPs in the hand (Wessel, Reynoso, & Aron, 2013), stopping speech produces reduced MEPs of the hand (Cai, Oldenkamp, & Aron, 2012; Wessel & Aron, 2017; Wessel et al., 2016), and stopping the hand results in reduced MEPs in the leg

[5] Again, the relationships observed between the neural activity and behavioral data help to support these claims as they no longer rely purely on reverse inference.

(Badry et al., 2009; Majid, Cai, George, Verbruggen, & Aron, 2012). Castiglione and Aron (2021) propose that having a global suppressive process that shutdowns responses across all modalities may provide the quickest and most efficient route to canceling an inappropriate response regardless of modality. Alternatively, global MEP suppression may simply be a byproduct of the right PFC network that is observed regardless of the type of process being stopped.

A recent study (not yet peer reviewed) provides further support for a shared right PFC network. Apšvalka et al. (2020) used multivoxel patterns of BOLD activity to train a classifier to discriminate between inhibit versus respond conditions in one modality (i.e. Stop vs. Go or Think vs. NoThink) and then tested whether the classifier could distinguish between inhibit and respond conditions in the other modality. The classification was performed separately for voxels in the rDLPFC and rVLPFC. The cross-modality classifier performed above chance, succeeding in discriminating respond vs. inhibit for both the rVLPFC and the rDLPFC seeds. These results provide further support for a common cognitive process shared between action-stopping and retrieval suppression that is supported by a right PFC network.[6]

Two accounts for executive control over LTM

We currently have a much deeper understanding of how the right PFC network operates to cancel actions. In contrast, it remains unclear precisely how or when the right PFC network exerts control to prevent memory intrusions. There are at least two potential accounts (Figure 0.2): 1) the right PFC is recruited in response to a memory intrusion and functions to push the memory information out of mind, or 2) the right PFC intervenes immediately after the NoThink

[6] Conversely, the cross-modality classifiers failed when they were applied to the target regions (i.e. motor cortex and hippocampus) (e.g. M1 for action-stopping, and hippocampus for NoThink). This result is consistent with the idea that the right PFC network harbors inhibitory control for both domains while the downstream target differs according to task demands. However, it seems at odds with the recent TMS result showing suppression in M1 even during the TNT task (Castiglione & Aron, 2021).

instruction to *prevent* any intrusion from entering into conscious working memory. These accounts are not necessarily mutually exclusive but gaining a deeper understanding of whether one is more common or effective will help guide future investigations.

1. ***Clearing Working Memory.*** Levy and Anderson (2012) posit that inhibitory control is recruited *when* an intrusion occurs and must be suppressed. Consistent with this conjecture, downregulation of hippocampus BOLD activity is greater on NoThink trials with reported intrusions. Similarly, MEP reductions are greater on NoThink trials with reported intrusions (Castiglione & Aron, 2021)[7]. The clearing account, sometimes described as a "purge" function, points to reconsolidation theory to explain how inhibiting an intruding memory might produce later forgetting effects. Reconsolidation theory suggests that reactivating a memory trace places the memory in a labile state (Dudai, 2004; Lee, 2009; Nader, Schafe, & Le Doux, 2000). Therefore, a memory intrusion may in fact be necessary for an encoded memory to become susceptible to manipulation (Detre et al., 2013).

   Support for this theory is limited in several ways. Active inhibition and disruption of the memory trace is a core feature of this account. Yet there is currently no direct evidence that the right PFC network actively inhibits memory information on NoThink trials. Indirect support for this idea comes from the observation that decreased hippocampal BOLD activity, the presumed target of inhibitory control, is linked to both increased right PFC activity and behavioral metrics of retrieval suppression. Modulation of hippocampal BOLD activity could, however, reflect a number of processes (i.e. task

[7] The timing of this MEP reduction is earlier than what we would expect (300-500ms post cue) if the MEP reduction reflects a clearing of retrieved information. Based on the retrieval timeline discussed earlier, the reduction in MEP occurs prior to pattern completion (~500ms) and cortical reinstatement (500-1500ms). Although the recordings from hippocampal cells show increases in hippocampal firing beginning prior to 500ms post cue, it is unlikely that a participant would be cognizant of an intrusion at this point (Staresina et al., 2019).

difficulty, a failure to initiate retrieval as opposed to active inhibition, etc.). Further, the majority of data linking the right PFC network to downstream effects on the hippocampus uses fMRI, a method with good spatial resolution but poor temporal resolution. Assigning the right PFC network a clearing function requires tools with better temporal resolution to determine precisely when the network interrupts relative to the retrieved memory content. Lastly, unlike the hyperdirect pathway, there is little structural evidence for an anatomical route by which the right PFC could rapidly inhibit hippocampal activity. [8]

2. ***Gating Working Memory.*** A more recent study suggests that inhibitory control can also be recruited much earlier during retrieval. Castiglione et al. (2019) showed that preventing an intrusion (indicated by self report) produced a rapid (within 500ms) increase in right frontal beta power, the same EEG signature observed when participants successfully cancel actions in the stop signal task (also within 500ms of the stop cue). Assuming modulations of right frontal beta reflect a common inhibitory process, these data suggest that control is recruited almost immediately upon seeing a memory cue and the NoThink instruction. In this case, executive control may function like a gate that pre-emptively stops an intrusion from entering into conscious working memory.

Accordingly, our lab proposes a novel theory for how the right PFC network may prevent intrusions. Specifically, engaging the right PFC network may excite the STN[9] which then impedes thalamic drive back to cortex and disrupts reinstatement of the cued

[8] Data from rodents and non-human primates reveal pathways through the anterior cingulate cortex (ACC) which may function to link the DLPFC to the medial temporal lobe (Anderson, Bunce, & Barbas, 2016). However, these data are not yet confirmed in human neuroimaging. Further, with only polysynaptic connections from the right PFC network to the hippocampus, this pathway may not allow for rapid inhibitory control equivalent to what is observed in the motor system.

[9] While the majority of data focuses on cortical overlap between action-stopping and retrieval suppression, a recent meta-analysis revealed subcortical co-localization within the basal ganglia (Guo et al., 2018).

memory. Thus, rapidly suppressing basal ganglia output to cortex may disrupt the transition from pattern completion to cortical reinstatement or the "loading" of information from LTM into working memory.

This theory makes several assumptions. First, we assume that LTM retrieval and maintenance of working memory representations are at least partly supported by thalamocortical drive. Support for this idea comes from studies that show involvement of the basal ganglia in the maintenance for verbal working memory (Chang, Crottaz-Herbette, & Menon, 2007) and from theories that propose a fronto-basal ganglia gating system for working memory that evolved from gating functions in the motor system (Chatham & Badre, 2015; Scimeca & Badre, 2012). Further, the theory hinges on the ability for the right PFC-basal ganglia network to disrupt not only motor representations but also cognitive representations. Evidence for this comes from studies that examine instances in which unexpected events trigger the right PFC network and results in the disruption of working memory content (Wessel & Aron, 2017). Although we believe this theory provides a viable mechanism for how we prevent intrusions when provoked by a cue (the core focus of this dissertation), it does not provide a clear explanation for the final forgetting effect observed across many studies.

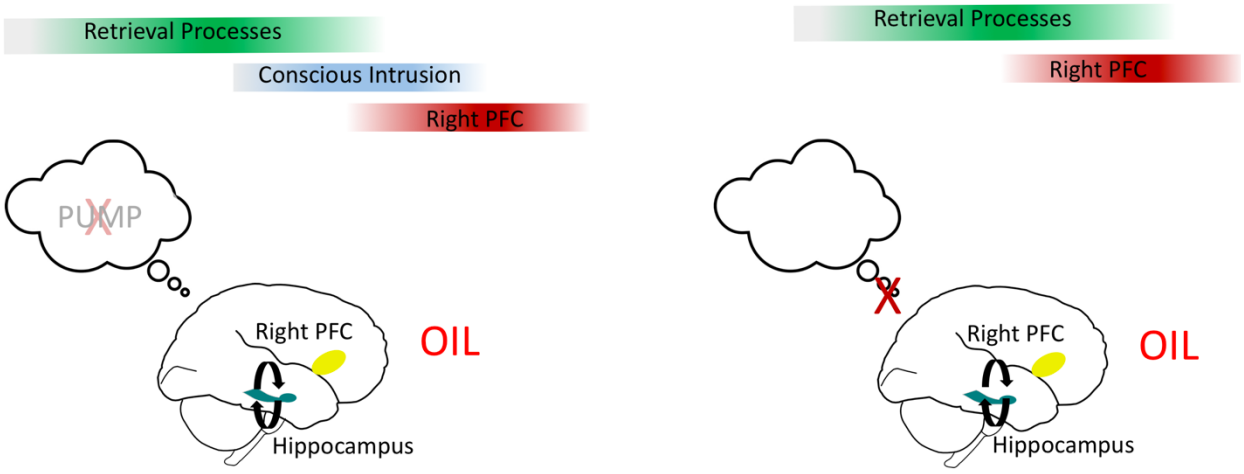

Figure 0.2. There are at least two proposed mechanisms for how the right PFC network exerts control over retrieval: one in which the right PFC acts to clear intruding information (left) and another in which the right PFC "gates" working memory to prevent the influx of a cued memory (right).

Summary and Dissertation Aims

A growing body of evidence supports the theory of a shared executive network that functions to inhibit either actions or memories. From an evolutionary perspective, this model of the brain seems logical, providing a single efficient and multipurpose "emergency brake." The current data, however, leave many unresolved questions. In particular, we lack a full understanding of when and how the right PFC network exerts control over a cognitive process like retrieval. Our ability to study how this network enacts control is limited in several critical ways: 1) current neuroimaging studies use methods with poor temporal resolution that are unable to determine *when* the right PFC network interrupts retrieval, and 2) the majority of evidence for retrieval suppression relies largely on self-report rather than an objective measure of the prohibited memory.

This dissertation seeks to address each of these limitations. We first aim to validate a high temporal resolution marker of the right PFC network using scalp EEG (Chapter 1). Second, we aim to develop an objective tool for tracking the emergence of retrieved memories during efforts to prevent retrieval (Chapters 2 and 3). And finally, we aim to combine these tools to test

whether recruitment of the right PFC network, measured with EEG, corresponds to a reduction in an objective metric of the prohibited memory (Chapter 4). By addressing these limitations, we hope to contribute tools that will help to advance our understanding of executive control over retrieval and our ability to cope with unwanted reminders of the past.

Chapter 1:

Double blind disruption of right inferior frontal cortex with TMS reduces right frontal beta power for action-stopping

As it appears in

Journal of Neurophysiology

2020

J Neurophysiol 125: 140–153, 2021.
First published October 28, 2020; doi:10.1152/jn.00459.2020

american physiological society

JNP JOURNAL OF NEUROPHYSIOLOGY.

RESEARCH ARTICLE

Double-blind disruption of right inferior frontal cortex with TMS reduces right frontal beta power for action stopping

Kelsey K. Sundby, Sumitash Jana, and Adam R. Aron

Department of Psychology, University of California San Diego, La Jolla, California

Abstract

Stopping action depends on the integrity of the right inferior frontal gyrus (rIFG). Electrocorticography from the rIFG shows an increase in beta power during action stopping. Scalp EEG shows a similar right frontal beta increase, but it is unknown whether this beta modulation relates to the underlying rIFG network. Demonstrating a causal relationship between the rIFG and right frontal beta in EEG during action stopping is important for putting this electrophysiological marker on a firmer footing. In a double-blind study with a true sham coil, we used fMRI-guided 1-Hz repetitive transcranial magnetic stimulation (rTMS) to disrupt the rIFG and to test whether this reduced right frontal beta and impaired action stopping. We found that rTMS selectively slowed stop signal reaction time (SSRT) (no effect on Go) and reduced right frontal beta (no effect on sensorimotor mu/beta related to Go); it also reduced the variance of a single-trial muscle marker of stopping. Surprisingly, sham stimulation also slowed SSRTs and reduced beta. Part of this effect, however, resulted from carryover of real stimulation in participants who received real stimulation first. A post hoc between-group comparison of those participants who received real first compared with those who received sham first showed that real stimulation reduced beta significantly more. Thus, real rTMS uniquely affected metrics of stopping in the muscle and resulted in a stronger erosion of beta. We argue that this causal test validates right frontal beta as a functional marker of action stopping.

NEW & NOTEWORTHY Action stopping recruits the right inferior frontal gyrus (rIFG) and elicits increases in right frontal beta. The present study now provides causal evidence linking these stopping-related beta oscillations to the integrity of the underlying rIFG network. One-hertz transcranial magnetic stimulation (TMS) over the rIFG impaired stopping and reduced right frontal beta during a stop-signal task. Furthermore, the effect on neural oscillations was specific to stopping-related beta, with no change in sensorimotor mu/beta corresponding to the Go response.

inhibitory control; oscillations; stop-signal task; transcranial magnetic stimulation

INTRODUCTION

Action stopping is a critical component of everyday cognitive control that allows us to rapidly cancel inappropriate responses and adapt to changes in our goals and environment. The ability to quickly stop an action is thought to require a right lateralized fronto-basal ganglia network in which the right inferior frontal gyrus (rIFG) is a critical node (reviewed by Aron et al. 2014; Bari and Robbins 2013; Cai et al. 2014; Coxon et al. 2009, 2012; Garavan et al. 1999; Jahanshahi et al. 2015). Identifying a scalp electrophysiological marker of this network would provide a high-temporal resolution, portable, and inexpensive tool. For example, it could be used for testing recent theories that a right frontal executive system is also involved in stopping memory retrieval, emotional responsiveness, and gait (Berkman et al. 2009; Castiglione et al. 2019; Depue et al. 2007; Guo et al. 2018; Wagner et al. 2016).

An apparent electrophysiological marker of the rIFG-mediated action-stopping process is increased beta band power (~16 Hz). Evidence for this comes from both electrocorticography (ECoG) recorded from the rIFG (Swann et al. 2009; Wessel et al. 2013) and scalp EEG studies that recorded neural activity during a stop-signal task (Hannah et al. 2020; Jana et al. 2020; Wagner et al. 2018). In a stop-signal task, participants press a button to respond to Go cues (i.e., a rightward or leftward pointing arrow). On a minority of trials (i.e., 25%), a stop signal occurs, requiring participants to try to rapidly cancel their prepotent response. Critically, for the abovementioned EEG and ECoG studies, the beta increase occurred in a specific time window important for stopping, i.e., after the stop signal

Correspondence: K. K. Sundby (ksundby@ucsd.edu).
Submitted 3 August 2020 / Revised 6 October 2020 / Accepted 27 October 2020

appears and before the participant's stop signal reaction time (SSRT), a latent measure of the stopping process. Although ECoG provides a more anatomically specific link between stopping-related beta and the rIFG compared with EEG, neither method shows that these beta modulations depend on the integrity of the rIFG during stopping. Another benefit of better understanding the role of this oscillatory signature and how it relates to the underlying network is the potential to entrain neural oscillations and hypothetically affect inhibitory control.

Here we used fMRI-guided 1-Hz repetitive transcranial magnetic stimulation (rTMS) to disrupt the rIFG and to test whether this selectively reduces right frontal beta power and stopping performance in a simple stop-signal task. Although prior studies have shown impaired stopping after 1-Hz rTMS over rIFG (Chambers et al. 2006, 2007; for meta-analysis see Yang et al. 2018), here we present the first study to use a "true sham" double-blind approach (that also elicits facial discomfort) and to test how disruption of the rIFG affects right frontal beta in EEG during action stopping.

In the present study, each participant first underwent fMRI while performing the stop-signal task. We used participant-specific activation maps and structural scans to localize our rIFG target stimulation site. On a subsequent visit, the participants underwent a TMS-EEG protocol. This started with a baseline session of the stop-signal task with EEG recorded. There then followed, in each participant, two TMS sessions, either in the order real TMS then sham TMS or the order sham TMS then real TMS, on the same day for each participant, with a 30-min "washout period." For each of these TMS sessions, we first delivered 1-Hz rTMS over the rIFG and then recorded scalp EEG during the stop-signal task. Although ideally sham and real stimulation would have been conducted on two separate days, we did them here on the same day in order to have a single right frontal EEG spatial filter per subject that we could compare across task sessions.[1] Furthermore, although the duration of 1-Hz rTMS effects varies across studies, stimulation sites, and TMS protocols, multiple studies report effects that last <30 min (Boroojerdi et al. 2000; Chen et al. 1997; Lang et al. 2008; Nyffeler et al. 2006; but also see Siebner et al. 2003). Even so, we did anticipate, in our preregistration document, that we might have to deal with an order effect, i.e., that real TMS would "leak" into the sham session.

Our neural measures were beta activity in a right frontal spatial filter to track changes in stopping-related beta and mu/beta activity in a left sensorimotor spatial filter (contralateral to the responding hand) to test for effects on Go-related neural activity.

Behaviorally, our measures were stop signal reaction time (SSRT) and Go latency. We also included an electromyography (EMG) metric of stopping based on recent work (Jana et al. 2020) showing that this indexes a pseudo-single-trial metric of stopping latency called "CancelTime" (also see Raud et al. 2020; Raud and Huster 2017; Thunberg et al. 2019). EMG CancelTime reflects the time at which EMG activity is shut down on stop trials that show a small burst of muscle activity that is subsequently silenced as the action is successfully withheld.

We preregistered the following key predictions for real versus sham stimulation: *1*) a reduction in stopping-related right frontal beta; *2*) a prolongation of both behavioral measures of stopping (SSRT and CancelTime); *3*) no change in Go behavior or Go-related left sensorimotor mu/beta activity.

MATERIALS AND METHODS

Methods

The study design and predictions were preregistered as an Open Science Framework (OSF) document (https://osf.io/9vynp/?view_only=7ecf843dfcd247eebe4885f4aff39bc5).

Power.

Two prior studies showed that 1-Hz rTMS over rIFG compared with control sites increased SSRT with a combined effect size of 0.76 (Chambers et al. 2006, 2007). Based on this effect size, we estimated a required sample size of 17 (to achieve 90% power) to detect a real vs. sham effect on SSRT with a one-tailed test. However, because we did not have sufficient information for a power analysis of our EEG hypothesis, we planned to collect 30 participants.

Participants.

Per our preregistration document, we aimed to analyze 30 participants who had a right frontal component to test our EEG hypothesis. However, we anticipated running more participants, as a right frontal spatial filter is typically found in ~85% of people for this task (Wagner et al. 2018). Participants completed three visits: an orientation visit with a rTMS demonstration (*visit 1*), a fMRI visit to localize the site of stimulation (*visit 2*), and the EEG-rTMS visit (*visit 3*). Five participants withdrew from the study after the rTMS demonstration on *visit 1* because of a level of discomfort that would be intolerable on *visit 3* (see VISIT 1: SAFETY SCREENING AND RTMS DEMONSTRATION below for more details). One participant withdrew after *visit 2* because of a head injury that occurred outside of the study. In total, we collected fMRI, EEG, and behavioral data in 33 participants to achieve the 30-participant goal for EEG analysis. All 33 were included in the behavioral analysis (18 women, 15 men; right-handed; mean age = 20 yr, SD = 1.91 yr). All participants provided written informed consent according to a protocol approved by the University of California San Diego (UCSD). They were compensated $20/h.

[1]Testing our hypothesis about right frontal beta critically requires first deriving a right frontal spatial filter from each participant's EEG data, yet our various publications have shown that it is not possible to derive a right frontal spatial filter in every participant (here we used generalized eigenvalue decomposition, but the same applies for other methods such as independent component analysis). We worried that if we had conducted the TMS-EEG protocol on two separate visits per participant this would have resulted in too much subject attrition per expended effort (i.e., we would have to exclude participants for whom we could not identify a reliable filter in either of the sessions). We reasoned that having a single right frontal spatial filter per participant to better compare the effects of real versus sham stimulation trumped the worry about a possible order effect, with real TMS "leaking" into the sham session.

Visit details.

VISIT 1: SAFETY SCREENING AND RTMS DEMONSTRATION. Participants came to the laboratory for an fMRI and TMS safety screening and for a demonstration of the 1-Hz rTMS procedure. Safety screening confirmed that all participants met our fMRI and TMS safety requirements. To demonstrate the rTMS procedure, we first delivered single-pulse TMS over left primary motor cortex (M1) to estimate each participant's resting motor threshold (RMT). We then delivered a brief train of rTMS over rIFG at 110% of this RMT value to ascertain whether it would be tolerable for them in *visit 3*. If it was too uncomfortable, the participant was withdrawn from the study.

VISIT 2: FMRI SCANNING. fMRI was used to identify an rTMS target for each participant for *visit 3* (Fig. 1*C*). We scanned participants with a 3T GE scanner at the Center for Functional Magnetic Resonance Imaging at UCSD. Each scanning session included an anatomical scan and two 6-min blocks of a visual stop-signal task (Fig. 1*A*) that reliably activates the pars opercularis of the rIFG (Aron and Poldrack 2006). We collected 182 functional T2-weighted echo planar images (EPIs) for each of the task blocks [repetition time (TR) 2 s].

We used FSL software (https://www.fmrib.ox.ac.uk/fsl) for all preprocessing and analysis steps. To identify the optimal target for rTMS stimulation, we created activation maps by contrasting stop trials versus go trials for each participant. We derived these activation maps with the exact steps described in Aron and Poldrack (2006) except that we included all stop trials for the contrast Stop vs. Go to obtain more observations, as we had only two runs per participant (the typical contrast is Successful Stop vs. Go). We were able to pool across all stop trials for our target localization because the pars opercularis shows activation on both successful and failed stop trials (Aron and Poldrack 2006). Each participant's activation map was aligned into the participant's native anatomical space and then masked with the pars opercularis region of interest from the Harvard-Oxford Atlas (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases). We selected our specific target by identifying the voxels within the pars opercularis mask that showed peak activation related to stopping. For one participant who had no discernible right pars opercularis activation, apparently because of poor task performance, we used the Montreal Neurological Institute (MNI) coordinates (*x*: 61, *y*: 21, *z*: 13) reported by Chambers et al. (2006), a prior study that used similar rTMS procedures to target the rIFG.

VISIT 3: RTMS AND EEG. The *visit 3* procedure used two TMS coils (PowerMag Lab100; MAG & More GmbH, Munich, Germany): a real coil (Double coil PMD70-pCool) and a sham coil (Double coil PMD70-pCool-Sham). The position of the coil was guided with neuronavigation (Brainsight, Rogue Research Inc., Montreal, QC, Canada; Brain Science Tools, The Netherlands) to the pars opercularis activation spot. The sham coil is designed to produce physical sensations (i.e., facial twitches) that are comparable to the real coil, allowing for a true double-blind procedure. At the start of *visit 3*, we first placed an EEG cap on each participant. Next, we determined the RMT by delivering single-pulse TMS over left M1, over the EEG cap. RMT was defined as the lowest pulse intensity to produce a motor evoked potential (MEP) of at least 0.05 mV in 5 of 10 consecutive pulses. The subsequent rTMS intensity was set to 110% of each participant's RMT. After EEG capping and the TMS threshold procedure, participants performed several minutes of practice (80 total practice trials) of the stop-signal task. After practice, the main experimental procedure was as follows: *1*) A designated laboratory member (not the experimenter delivering TMS) labeled the

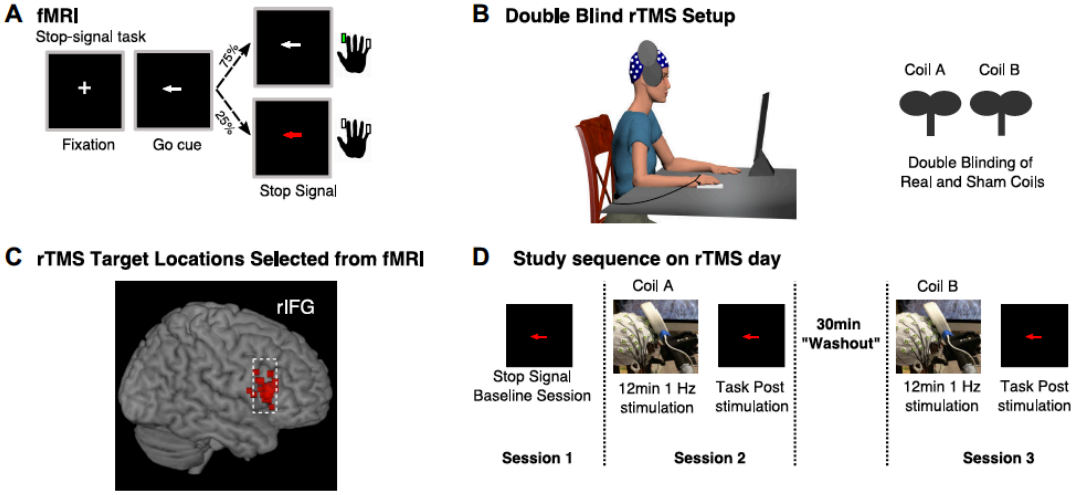

Figure 1. Task, fMRI targeting, and EEG-transcranial magnetic stimulation (TMS) procedure. *A*: diagram of the stop-signal task. *B*: example of EEG-repetitive TMS (rTMS) setup and double-blinding procedure. *C*: rTMS targets selected from voxels in right inferior frontal gyrus (rIFG) with peak activation during the stop-signal task in fMRI scanner. *D*: sequence of events during the EEG-rTMS study visit.

sham and real TMS coils (*coil A* and *coil B* according to the blinding document) while the experimenter waited outside of the room (Fig. 1*B*). *2*) The participant completed a 12-min baseline session of the stop-signal task. This first task session provided baseline measurements for stopping performance as measured by SSRT and EMG CancelTimes and the stopping-related right frontal beta signature. *3*) Twelve minutes of 1-Hz rTMS was delivered with *coil A*. *4*) Immediately afterwards, the participant performed another 12-min session of the stop-signal task as we recorded EEG. *5*) The participant then rested for a 30-min "washout" period (ideally this would be much longer, to ensure that the effects of the real coil dissipate, but the overall session was already several hours long). *6*) Twelve minutes of 1-Hz rTMS stimulation was then done with *coil B*. *7*) Immediately afterwards, the participant performed another 12-min session of the stop-signal task as we recorded EEG (Fig. 1*D*). Overall, *visit 3* typically lasted between 3.5 and 4.5 h depending on a number of factors (especially the time it took to cap with EEG, to identify the M1 hotspot, to find the resting motor threshold, and to implement the Brainsight TMS coil localization method).

In all three task sessions, participants responded to the Go cues on a button box with their right first dorsal interosseus (FDI; index finger) for rightward arrows and their right abductor digiti minimi (ADM; pinky finger) for leftward arrows. We placed surface EMG electrodes on both of these task-relevant muscles to record EMG as participants performed the task (see Fig. 6*A*). Requiring participants to respond with their pinky and index fingers was important because it allowed us to record EMG data from two task-relevant muscles in the same hand.

Stop-signal task (visits 2 and 3).

We used a visual stop-signal task (Logan and Cowan 1984) programmed in MATLAB 2015b (The MathWorks, United States) and Psychtoolbox (Brainard 1997). On each trial participants saw a left or right white arrow on the screen and pressed a button according to the direction of the arrow (Go trials). However, on 25% of the trials (stop trials), the arrow suddenly turned red after a stop signal delay (SSD), indicating that the participant must try to cancel their imminent response. We used an algorithm to adjust the SSD to achieve ~50% successful stops; this reduced or increased the SSD by 50 ms depending on the outcome of the previous stop trial. Participants were instructed to respond as quickly and as accurately as possible.

On *visit 2* (fMRI) there were two blocks of the task, of 128 trials each, of which 32 trials (25%) were stop signal trials and 96 were Go (for more details see Aron and Poldrack 2006). Each of the task sessions during *visit 3* (baseline, post *coil A*, and post *coil B*) was 12 min long. Each of these sessions had 320 trials, of which 25% were stop signal trials (i.e., 80 stop trials and 240 Go trials per session). Participants received a 10-s break after every 80 trials.

EEG recording (visit 3).

We used a 64-electrode EasyCap slim electrode system (EasyCap and BrainVision actiCHamp amplifier; Brain Products GmbH, Gilching, Germany) with electrode placement in the International 10-20 System and PyCorder (Brain Products) to record the EEG data. Electrode Fpz served as the ground. We recorded the data reference free with EEG electrodes placed bilaterally over mastoids for later off-line re-referencing. We reduced electrode impedances to <10 kΩ before recording. We placed additional electrodes at each canthus and one electrode below the right eye to monitor for eye movements. The EEG and EOG data were sampled at 1,000 Hz.

EMG recording (visit 3).

We used a Grass QP511 AC amplifier (Glass Technologies, West Warwick, RI) with a frequency cutoff between 30 and 1,000 Hz to collect the EMG data. Data were sampled at 2 kHz with a CED Micro 1401 mk II acquisition system and recorded in CED Signal v4 software (Cambridge Electronic Design Limited, Cambridge, UK).

Preregistered Predictions

Behavior.

RTMS EFFECTS ON STOP. We predicted that real rTMS stimulation would impair stopping, reflected by a lengthening of both SSRT and EMG CancelTimes.

RTMS EFFECTS ON GO. We expected no effect on the Go process as indicated by both no change in Go response times and EMG onset times. Alternatively, we speculated that real rTMS could quicken Go responding, on the view that rIFG also acts as a precautionary brake in contexts where stopping might be required (Swann et al. 2012; Wessel et al. 2013). Thus, eroding this brake could result in quicker overall responding.

EEG.

BASELINE SESSION. Consistent with our recent report (Wagner et al. 2018), we expected right frontal beta during the baseline session of the stop-signal task (before any stimulation) to increase in power after the stop signal and before SSRT. Furthermore, we expected the stopping-related increase in right frontal beta power to be greater for successful compared with failed stop trials.

RTMS EFFECTS ON STOP. Our core hypothesis was that disrupting rIFG would reduce a putative electrophysiological marker of the executive system, namely, right frontal beta power. Specifically, we expected reduced right frontal beta power in a time window critical for stopping (i.e., after the stop signal and before SSRT) in the task session following real stimulation.

RTMS EFFECTS ON GO. We predicted no change in left sensorimotor mu/beta (7–25 Hz) band desynchronization associated with the Go process for the three task sessions.

Analysis

Behavior.

BASELINE SESSION EXCLUSION CRITERIA. We preregistered our plan to exclude any participants who violated the following sanity checks during the baseline session of the stop-signal task: *1*) Stop % > 30% and < 70% (percentage of successful stops should be close to 50%), *2*) mean response time (RT) < 700 ms (>700-ms RTs indicate a strategy of waiting for the stop signal to appear), *3*) Go error rate < 10%, *4*)

Go RT > Failed Stop RT (this latter is a required assumption of the race model). On the basis of these criteria, all participants were included.

CALCULATING SSRT. Stop signal reaction time (SSRT) was calculated with the integration method (Verbruggen et al. 2019). This estimates when the stop process finishes by integrating the response time (RT) distribution and identifying the point at which the integral equals *P*(respond|stop signal). When using a tracking method to adjust stop signal delays during the task, this point in the distribution corresponds to the *n*th RT, where *n* = the number of RTs in the distribution multiplied by the overall probability of stopping. SSRT is calculated by subtracting the participants' mean SSD from the *n*th RT. For the most reliable and unbiased SSRTs, we included all Go trials with a response in the RT distribution (including incorrect responses). Furthermore, per the recommended methods (Verbruggen et al. 2019), we replaced any Go omission trials with the participant's maximum RT.

CALCULATING EMG CANCELTIME AND EMG ONSET. EMG data were collected to determine EMG onsets, a secondary measure of the Go process, and EMG CancelTimes, a pseudo-single-trial metric of the Stop process. We analyzed the EMG data according to methods described in Jana et al. (2020). EMG data were filtered with a fourth-order Butterworth filter (roll-off 24 dB/octave) to eliminate 60-Hz line noise. The data were then full-wave rectified, and the root mean square (RMS) was computed with a 50 ms centered window. For each trial we selected the EMG peak by identifying periods of EMG activity that exceeded a trial specific threshold. This threshold was set to 8 standard deviations above the trial's mean baseline EMG activity between the fixation cross and the Go cue. If more than one section of EMG activity exceeded the threshold, we selected the maximum as the peak.

We next examined the EMG activity before the peak and marked the EMG onset as the moment at which EMG activity dipped below 20% of peak amplitude and remained below this threshold for at least 5 ms. Using these trial-specific thresholds provides superior detection of onsets compared with a fixed threshold (e.g., *X* number of standard deviations above an average) that is applied to all trials and does not account for trial-by-trial noise in the EMG data. Our approach is particularly useful for detecting EMG onsets on the successful stop trials with smaller partial EMG bursts. Next, EMG offset was marked as the moment after the peak EMG when EMG activity decreased for five consecutive milliseconds. We removed trials as outliers if the EMG onset or the EMG offset were 1.5 times the interquartile range (IQR) of the first and third quartiles of their respective distribution. Finally, we marked EMG CancelTimes on all successful stop trials with a partial EMG response. CancelTime reflects the period of time between the stop signal and the point at which the EMG burst starts declining consistently. More specifically, the moment of EMG decline is defined as the first point after the EMG peak where EMG declines consistently for at least five consecutive milliseconds. For CancelTimes, outliers were rejected if they fell below a 50-ms cutoff or above Q3 + 1.5 × IQR.

Because EMG amplitude for the FDI and ADM muscles varied considerably, we normalized the muscle activity by the peak activity for each muscle before averaging the EMG data across the muscles.

Our laboratory developed these EMG analysis methods after we had begun data collection for the present study. Consequently, EMG data were only collected for 29 of the 33 participants who completed the study.

EEG.

PREPROCESSING. We analyzed the EEG data with EEGLAB 14.1.1b (Delorme and Makeig 2004) in MATLAB 2015b (The MathWorks). Data were downsampled to 512 Hz and low-pass filtered with a MATLAB polyphaser antialiasing filter. We then re-referenced the data to an average of the two mastoid electrodes. We applied a high-pass filter at 2 Hz [finite impulse response (FIR) order 3,300] and a notch filter at 60 and 180 Hz (FIR order 846) to remove electrical noise. We then visually inspected the continuous data and manually rejected any stretches contaminated with excessive artifact. After preprocessing, we concatenated the three sessions of EEG data and submitted it to independent component analysis (ICA) decomposition (Makeig et al. 1996). We identified and removed artifact components (e.g., blinks, muscle tension) before further analysis. After preprocessing, we identified a left sensorimotor and a right frontal spatial filter to test our EEG hypotheses related to the Go and stop processes, respectively. We used linear spatial filters rather than select electrodes to boost the signal-to-noise ratio in these analyses (Cohen 2017).

LEFT SENSORIMOTOR FILTER AND GO FREQUENCY OF INTEREST. Based on the EEG data for Go trials in the stop-signal task (Wagner et al. 2018) and also a substantial literature on EEG correlates of action execution (Kilavik et al. 2013), we expected the Go response to elicit a mu/beta desynchronization in a left sensorimotor component contralateral to the responding hand. Furthermore, we predicted that the effects of rTMS would be specific to the stopping system and not impact this signature of the Go process. To test this prediction, we first identified a left sensorimotor component from the ICA decomposition for each participant. This component was selected according to spatial topography (left central). We then epoched the data relative to the Go cue on all Go trials and generated event-related perturbations (ERSPs) (see Fig. 2*A* for group ERSP). Relative changes in spectral power were computed by calculating the mean difference between each single-trial log spectrogram and the mean baseline spectrum (the log spectrum 1,000 ms before the Go cue averaged across all trials). We used the ERSP from each participant's baseline EEG session to identify the frequency of interest, the participant's mu/beta (7–25 Hz) frequency band that showed the greatest reduction in the time window surrounding the Go response (Go cue to the participant's average Go RT) (across participants, mean frequency of interest = 13.83, SD = 5.19). We narrowed our time window of interest by identifying all time points between the Go cue and the participant's average Go response time (using the baseline EEG session) in which the participant's selected frequency (±2) dropped below zero, indicative of a Go-related desynchronization.

Extracting average power. We then extracted the power for the frequency of interest in the time window of interest for each EEG session (baseline, post real stimulation, post sham stimulation). We conducted a one-way repeated-

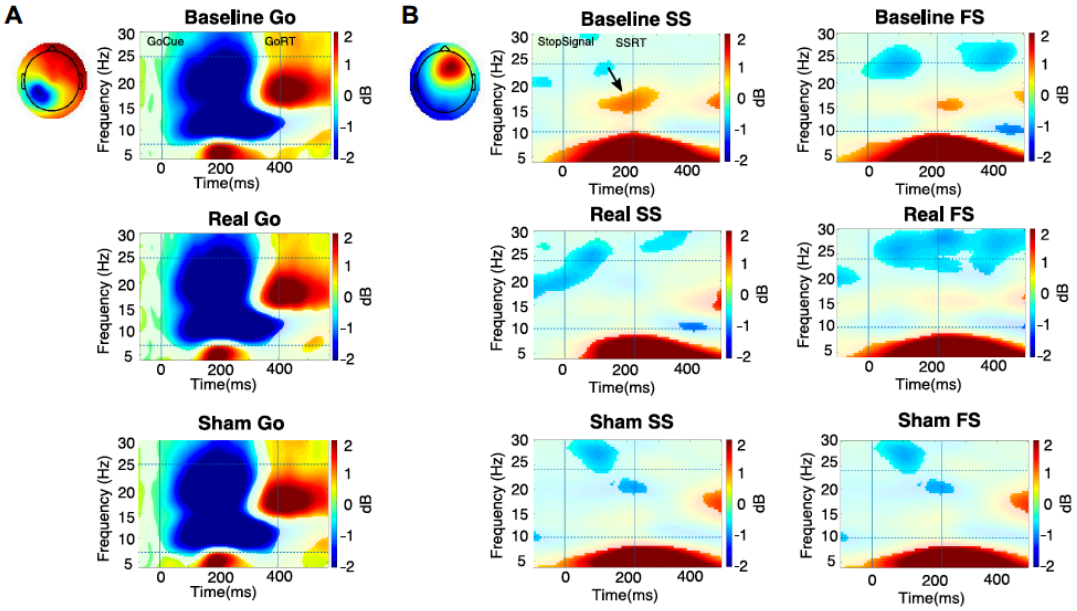

Figure 2. Event-related spectral perturbation (ERSP) results for Going and Stopping. *A*: group ERSPs for the left sensorimotor filter for Go trials in all 3 task sessions. *Time 0* on the *x*-axis is the time of the Go cue. The ERSP is generated relative to the baseline, which is −1,000 ms before the Go cue. For each of baseline and real and sham transcranial magnetic stimulation (TMS), there is a typical event-related desynchronization in mu/beta followed by a postmovement beta rebound. RT, response time. *B*: group ERSPs for the right frontal filter for stop trials (SS, successful stop; FS, failed stop) in all 3 task sessions. *Time 0* on the *x*-axis is the time of the stop signal. SSRT, stop signal reaction time. The ERSP is generated relative to the baseline, which is from −1,000 to −500 ms before the stop signal. For the baseline session, successful stop trials show an increase in beta band power (indicated by arrow) between the stop signal and SSRT, consistent with several prior studies, whereas this is absent after real and sham TMS.

measures ANOVA to test for changes in left sensorimotor mu/beta power across task sessions.

RIGHT FRONTAL FILTER AND STOP FREQUENCY OF INTEREST. We expected stopping-related modulations of beta to occur in right frontal electrodes, consistent with prior work (Castiglione et al. 2019; Wagner et al. 2018) and the putative right-lateralized anatomy of the underlying network. Based on these studies, we planned to identify a right frontal spatial filter (a weighting applied to all electrodes) from each participant's EEG data (being mindful, as explained above, that not every participant would provide such a filter). Before unblinding the data, we used two different source separation techniques to try to identify a right frontal spatial filter in the first 30 participants who completed the study. As specified in the OSF document, we planned to adopt the source separation technique that provided the largest number of participants with a right frontal filter to test our hypothesis.

1. ICA clustering approach. We first submitted the data to the ICA clustering approach used in Wagner et al. (2018). This approach yielded 18 participants with a right frontal IC of the initial 30 collected.

2. GED approach. Next we applied a guided multivariate source separation method using generalized eigenvalue decomposition (GED) (Cohen 2017). This approach is "guided" because it uses priors about topography (i.e., right frontal), frequency of interest (i.e., beta band 12–24 Hz), and time window of interest (i.e., time between stop signal and SSRT). This approach allowed us to compute the weights or spatial filters over the right prefrontal cortex that optimally separated a time window critical for stopping from a baseline time window. We used successful stop trials from the baseline EEG session to identify this filter in a manner that was independent of our core hypothesis (i.e., right frontal beta would be modulated after real stimulation).

We ultimately used the GED approach, as it was able to identify filters for 27 of the initial 30 participants, only requiring data collection for 3 more. Next, we computed ERSPs for failed stop trials and successful stop trials for each session (baseline, real, sham) (see Fig. 2*B* for group ERSPs). Relative changes in spectral power were computed by calculating the mean difference between each single-trial log spectrogram and the mean baseline spectrum (the log spectrum 1,000 ms to 500 ms before the stop signal averaged across all trials). Using the baseline session alone, we then selected a participant-specific frequency of interest and time window of interest between the stop signal and SSRT on successful stop trials. Specifically, we selected the frequency of interest as the participant's beta frequency band (12–24 Hz)

that showed the greatest increase in the time between the stop signal and SSRT (across participants, mean frequency of interest = 16.53, SD = 3.70). We narrowed our time window of interest by identifying all time points between the stop signal and the participant's SSRT (using the successful stop trials in the baseline EEG session) in which the participant's selected frequency (±2) increased above zero.

1. Extracting average power. We then extracted the mean power of the frequency of interest from baseline, post-sham and post-real EEG sessions on both successful and failed stop trials. We conducted a 3 × 2 ANOVA with session (baseline, post real, post sham) and stop trial outcome (successful, failed) as factors and mean power of the frequency of interest as the dependent measure.

2. Beta bursts. Recent studies on beta oscillations show that changes in average power are the result of brief bursting events in the beta band rather than a prolonged modulation. Analyses of beta bursts provide a rich set of features including burst timing relative to an event of interest (see Hannah et al. 2020; Jana et al. 2020; Little et al. 2019; Wessel 2019). A study that examined beta bursts during stopping showed that earlier bursts were associated with quicker SSRTs and CancelTimes (Jana et al. 2020). Although our experiment was designed to focus on beta power, we extracted beta bursts in the baseline task session to test whether we could replicate this relationship between beta burst timing and stopping behavior in our data. We only conducted this analysis on the baseline data because we only had 12 min of data per session and we expected beta bursts to occur less frequently after real stimulation. [Measures like burst timing require many observations to achieve a reliable estimate.] We extracted beta bursts from the baseline task session in the epoched data after filtering at the participant's frequency of interest (±2.5) (frequency-domain Gaussian, fwhm = 5 Hz). We then defined the burst threshold according to the beta amplitude in a baseline period (−1,000 to −500 ms before the Stop Signal; mean SSD −1,000 ms to mean SSD −500 ms before Go cue). A Hilbert transform was used to derive the complex analytic time series. We took the absolute of the analytic signal to compute beta amplitude during the baseline period. We calculated the median and standard deviation of the baseline period beta after pooling across all trial types (successful stop, failed stop, and Go trials). Bursts were then defined as any period in a trial when beta amplitude exceeded the baseline median value + 1.5 SD (for diagram see supplemental material of Jana et al. 2020). After identifying the bursts, we marked burst time as the time (relative to the stop signal) of each burst's peak amplitude.

Statistical analyses.

For pairwise comparisons, the data were first checked for normality with a Shapiro–Wilk test, and if the data were normally distributed a *t* test was performed. In the few cases where the data were not normally distributed, when comparing two independent samples we used a Mann–Whitney *U* test. For analyses with multiple levels, we performed repeated-measures ANOVAs and Bonferroni-corrected *t* tests for the subsequent pairwise comparisons. All analyses were conducted as two-tailed tests.

RESULTS

Preregistered Analyses

Behavior.

BASELINE SESSION. All participants met the baseline behavior sanity checks (see BASELINE SESSION EXCLUSION CRITERIA above) (Table 1).

EFFICACY OF DOUBLE BLIND. To test for consistent coil placement for both sham and real coils on *visit 3*, we measured the distance between the coil and the target (locus on pars opercularis) throughout each stimulation session. Unfortunately, this software update only became available midway through data collection, allowing us to collect measurements for 14 of the participants. In this sample, the coil placement was highly similar (mean ± SE: real = 2.32 ± 0.13 mm, sham = 2.35 ± 0.11), with no significant difference [$t(13) = 0.213$, $P = 0.834$].

Furthermore, after *visit 3*, participants were also asked to describe their experience during each TMS session "including any sensations in facial muscles, how intense the stimulation felt, or anything else that comes to mind." Participants were not explicitly informed that there were two different coils but simply that they would be receiving two sessions of rTMS. The free response answers were then recoded into three categories: discomfort, intensity, or twitching. Multiple participants did distinguish between discomfort and intensity (the level of pounding or strength of the pulse). Only seven participants reported a difference in discomfort between the two coils, with four participants reporting greater discomfort during real stimulation and three reporting greater discomfort during sham stimulation. Thirteen participants experienced a difference in "intensity," in which 10 described the real stimulation as more intense. Finally, 12 participants mentioned facial twitching, with 9 reporting more facial twitching during the real stimulation. Importantly, in the free response answers, no participant mentioned any suspicion of a sham coil or even two different coils. Together these data demonstrate a rather successful double-blind procedure.

METRICS OF THE STOP PROCESS.

1. SSRT. As predicted, a repeated-measures ANOVA showed a main effect of task session (baseline, post sham, post real) on SSRT (see Table 1 for session values) [$F(2,64) = 6.818$, $P = 0.002$]. We next conducted two paired-sample *t* tests with a Bonferroni adjusted alpha level of 0.025 (0.05/2). SSRT was prolonged after real stimulation compared with the baseline session [$t(32) = 2.885$, $P = 0.007$, 2-tailed, $d = 0.502$]. However, surprisingly, SSRT was also prolonged after sham stimulation compared with baseline SSRT [$t(32) = 3.111$, $P = 0.004$, 2-tailed, $d = 0.542$], and there was no

Table 1. *Behavior*

Session	Go RT	Failed Stop RT	Stop %	Correct Go %	SSRT
Baseline	404 (7)	375 (6)	48 (1)	99 (0)	222 (5)
Real	402 (8)	378 (7)	49 (1)	99 (0)	232 (4)
Sham	401 (7)	373 (6)	47 (1)	99 (0)	231 (4)

Values are means (SE). Stop %, percent of successful stops; RT, response time; SSRT, stop signal reaction time.

difference between SSRT after real stimulation and after sham stimulation [$t(32) = 0.416$, $P = 0.680$] (Fig. 3*A*).

2. EMG. Contrary to our prediction, we found no main effect of task session on mean CancelTimes [$F(2,56) = 0.783$, $P = 0.462$; Fig. 6*B*, *left*; mean ± SE: baseline = 165.36 ± 5.74 ms; sham = 167.17 ± 6.28 ms, real = 169.73 ± 5.65 ms].

METRICS OF THE GO PROCESS. We performed a repeated-measures ANOVA on Go response times and found no change in response times across the three task sessions (see Table 1 for session values) [$F(2,64) = 0.206$, $P = 0.814$; Fig. 3*B*]. We also did not observe any change in EMG onset, a secondary behavioral measure of the Go process [$F(2,56) = 1.618$, $P = 0.207$].

EEG.

BASELINE PREDICTIONS. Consistent with prior work, we expected to observe a greater increase in right frontal beta for successful stop trials compared with failed stop trials in the baseline task session data (i.e., task data before any stimulation) (Castiglione et al. 2019; Wagner et al. 2018). This result did replicate in our data [$t(29) = 4.062$, $P < 0.001$, 2-tailed, $d = 0.742$; Fig. 4*A*] (bearing in mind that this result is somewhat biased by the fact that we had selected the right frontal (RF) filter, time window of interest, and frequency of interest based on the successful stop trials alone).

METRIC OF THE STOP PROCESS. Average power. As described above, we selected a participant-specific beta band by identifying the frequency with peak power between the stop signal and SSRT on successful stop trials during the baseline EEG session. To test whether disrupting the rIFG with 1-Hz stimulation degrades right frontal beta power, we then conducted a 2 × 3 repeated-measures ANOVA with stop trial outcome (successful, failed) and session (baseline, post sham, post real) as the factors. There was a main effect of session [$F(2,58) = 9.907$, $P < 0.001$ and a outcome × session interaction [$F(2,58) = 8.085$, $P = 0.001$; Fig. 4*A*]. We focused on successful stop trials and conducted two paired-samples t tests to examine specific effects of real stimulation compared with no stimulation (baseline task session) and compared with sham stimulation. Using a Bonferroni-corrected alpha of 0.025, we found a significant reduction in beta power on successful stop trials after real rTMS stimulation compared with the baseline task session [$t(29) = -6.129$, $P < 0.001$, 2-tailed, $d = 1.119$]. There was not, however, a significant difference between beta power on successful stop trials after real rTMS stimulation and after sham stimulation [$t(29) = -1.452$, $P = 0.157$, 2-tailed]. As predicted, these data show that disrupting rIFG with real rTMS results in a reduction of right frontal beta power. Contrary to our hypothesis, however, sham stimulation also had an effect [$t(29) = -4.841$, $P < 0.001$, $d = 0.884$].

METRICS OF THE GO PROCESS. Average power. We expected 1-Hz stimulation over rIFG to have effects specific to the stopping process, and not the Go process. To examine Going, we conducted a repeated-measures ANOVA with session (baseline, post sham, post real) as the factor and left sensorimotor mu/beta power on Go trials as the dependent measure. There was no significant change across sessions [$F(2,56) = 2.797$, $P = 0.070$; Fig. 4*B*]; if anything, the event-

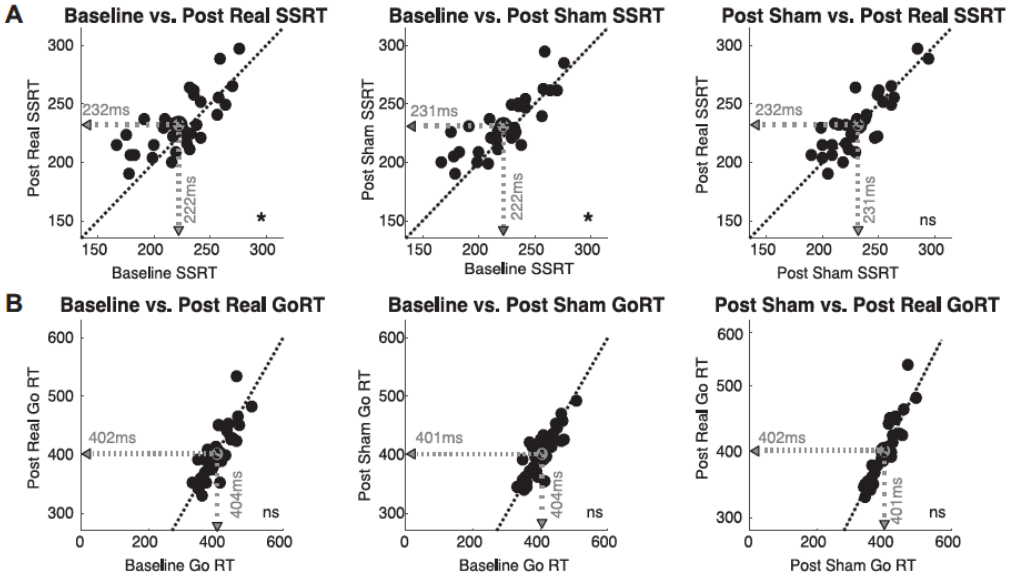

Figure 3. Effect of real vs. sham transcranial magnetic stimulation (TMS) on Stopping and Going. *A*: the change in stop signal reaction time (SSRT) across task sessions. After real TMS, SSRT is lengthened relative to baseline, but SSRT is also lengthened after sham TMS relative to baseline. There is no difference in SSRT following real vs. sham TMS. *B*: Go response times (RTs) plotted for each task session. Neither real nor sham TMS affected Go RT relative to baseline. *$P < 0.05$; ns, Not significant.

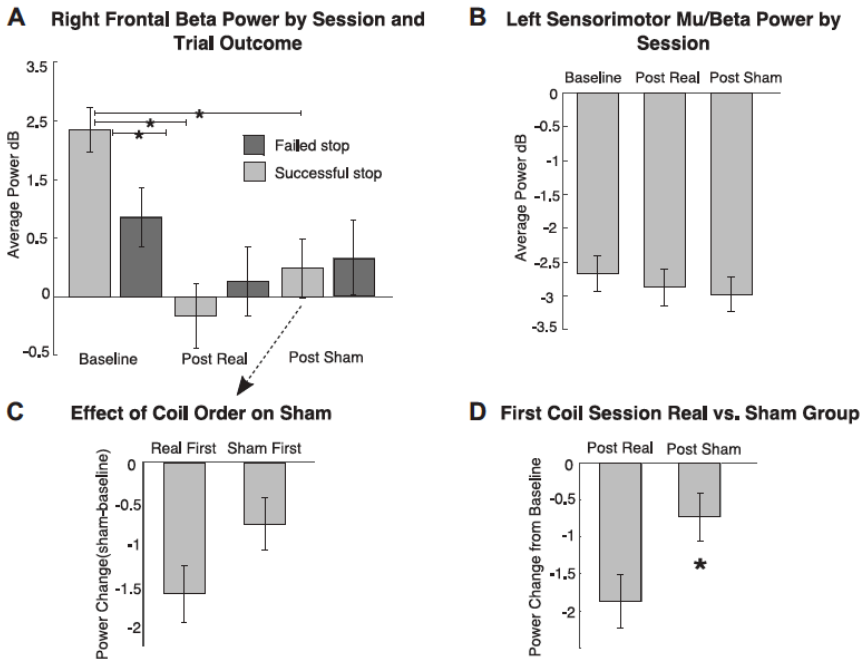

Figure 4. Effect of real vs. sham repetitive transcranial magnetic stimulation (rTMS) on right frontal beta power for Stopping and on left sensorimotor mu/beta power for Going. *A*: change in right frontal beta power across task sessions and for the comparison of successful and failed stop trials. For each participant, the time and frequency band to extract beta power were determined from the baseline successful stop trials alone and then applied to all other conditions. *B*: average left sensorimotor mu/beta power related to the Go response for each task session. For each participant, the time and frequency band to extract mu/beta power were determined from the baseline Go trials alone and then applied to the other conditions. *C*: a comparison of the change in right frontal beta power after sham stimulation for participants who had the real coil first vs those with the sham coil first. As a between-participant analysis, these are expressed relative to each participant's baseline session: thus in participants with real stimulation first, beta power was much more reduced (from baseline) in the post-sham session compared with participants who had sham first. *D*: a comparison of the change in right frontal beta after the first stimulation session only, by which we could compare the effects of real vs. sham stimulation without any risk of real stimulation leaking into sham. As a between-participant analysis, these data again represent the change from the baseline session and show that beta power was reduced (from baseline) more for participants in the real coil group (i.e., real first) than participants in the sham group (i.e., sham first). *$P < 0.05$.

related desynchronization relative to baseline on each Go trial was even greater in the real and sham task sessions compared with the baseline session. This is the opposite of what one would predict if rTMS reduced an event-related response. Taken together with the stopping analysis above, these results show that the observed changes in neural oscillations were specific to stopping-related right frontal beta (right frontal beta was reduced for stop trials, and left sensorimotor mu/beta was not reduced on Go trials).

Post Hoc Results

Testing coil order effects.

One reason why the difference between real and sham did not come out in the above might have been related to an order effect, wherein real TMS lingered over into the sham session (indeed, our preregistration report anticipated that the 30-min washout period might not be sufficient). This makes the prediction that beta power after sham TMS would be more reduced if the sham session occurred second (i.e., after real) compared with first (i.e., before real): indeed, this prediction held out, albeit at trend level with a two-tailed test: $t(28) = 1.78$, $P = 0.0836$, $d = 0.651$ (Fig. 4*C*). One concern when interpreting this result is that later task sessions (e.g., the sham sessions that occur after real) may be more susceptible to boredom, fatigue, or lapses in attention that might increase as the visit progresses. However, the same analysis using the real session data (i.e., comparing the effect of real stimulation in those participants who had real first vs. those with real last) did not show a greater reduction in beta for those participants who received real as the final session ($U = 87$, $P = 0.313$). We also did not observe any temporal effect on behavior when we compared Go RTs [$t(32) = -1.675$, $P = 0.104$] and SSRTs [$t(32) = 1.02$, $P = 0.313$] for post-TMS *session 1* vs. post-TMS *session 2* regardless of coil type (i.e., the final session does not have the longest SSRTs or Go RTs simply because it was last).

This result shows that running both real and sham TMS on the same day introduced a confound—a lingering effect of real TMS on sham TMS. [As noted above, conducting both in 1 day allowed us to use the same EEG spatial filter to compare across sessions and was in line with several estimates

in the literature that show effects dissipating within 30 min.] However, one way to test our core idea, that there is a real vs. sham difference, is to do a between-subject test that examines only the first poststimulation session, i.e., to compare the effects of real stimulation in those participants who had real first (n = 16) versus the effects of sham in those who had sham first (n = 14) (for each participant we "normalized" the data by taking the power difference for real/sham minus baseline). We used a nonparametric test because we observed a significant departure from normality after splitting the participants. Indeed, there was a significant between-group difference and beta was reduced on successful stop trials significantly more for the real stimulation group compared with the sham stimulation group [U = 64.0, P = 0.047, 2-tailed test; Fig. 4*D*]. This between-group result is an important piece of evidence for the idea that the rIFG and underlying circuitry is critical for stopping-related beta oscillations. For completeness, we also conducted this analysis on SSRT (P = 0.704) and CancelTimes (P = 0.425) for the full behavioral sample but found no significant difference between the real and sham groups. Finally, we tested whether there was a correlation across participants between changes in beta power with changes in SSRT and CancelTime. None of the correlations was significant (all $P > 0.05$).

Beta bursts and stopping behavior in baseline task session.

Based on a recent report linking beta burst timing to stopping behavior, we conducted a post hoc analysis of beta bursts in our baseline task data. We expected to replicate the prior finding showing a correlation between stopping behavior (i.e., SSRTs and EMG CancelTimes) and the timing of beta bursts on stop trials (Jana et al. 2020), although this was not a prediction at the time of writing our preregistered document. Consistent with Jana et al. (2020), SSRT was correlated with burst timing (mean ± SE beta burst time = 121 ± 5.56 ms) on successful stop trials (R = 0.549, P = 0.002; Fig. 5*A*), showing a robust relationship between the standard metric of the stop process (i.e., SSRT) and the timing of stopping-related beta modulations. We also observed a similar relationship between burst timing and EMG CancelTime in the baseline task session (R = 0.370, P = 0.058; Fig. 5*B*). We did not perform these correlations for the other two task sessions because beta oscillations were reduced and resulted in too few trials with bursts to reliably estimate burst timing (see *METRIC OF THE STOP PROCESS. Average power.* under *EEG* above). Finally, we correlated EMG CancelTimes against SSRT, separately for each session. These were significantly correlated in all cases (R = 0.774, $P < 0.001$ in the baseline session; R = 0.793, $P < 0.001$ after real stimulation; R = 0.744, $P < 0.001$ after sham stimulation).

Shift in EMG CancelTime distribution.

Although mean EMG CancelTime was not different between real and sham stimulation, a different feature of CancelTime was affected, i.e., CancelTime variability. Specifically, we calculated each participant's coefficient of variation (COV) for each session (COV = mean of session CancelTime/standard deviation of session CancelTime). An ANOVA showed that CancelTime COV was significantly different across sessions [$F(2,56)$ = 5.058, P = 0.010; Fig. 6*B*, *right*). Follow-up tests showed that CancelTime COV was reduced after real stimulation compared with both baseline and sham sessions [$t(28) = -2.646$, P = 0.013, d = 0.491 and $t(28) = -3.153$, P = 0.004, d = 0.584, respectively]. This suggests that specific portions of the CancelTime distribution were disproportionately affected by real rIFG stimulation. We next aimed to identify the portion of the CancelTime distribution that was affected by real stimulation. Because of the low number of partial EMG trials (mean count = 26.07, SD = 6.11 per participant per session), we did not have a sufficient number of observations to produce reliable session distributions for single participants. Instead, we collapsed across all participants, resulting in a single CancelTime distribution for each of the three sessions: baseline session, post real stimulation, and post sham stimulation. We then produced a cumulative distribution function (CDF) to statistically compare the shape of each distribution. We conducted Kolmogorov–Smirnov tests to compare CDFs for the real session versus baseline, the real session versus sham, and the sham session versus baseline. There was no difference between baseline and sham (P = 0.262). However, there was a significant difference for the real session versus baseline ($P < 0.001$) and for the real session versus sham (P = 0.004). Specifically, the leftward tail of the distribution (i.e., the quickest CancelTimes) was cut after real stimulation (see Fig. 6, *C* and *D*). This suggests that assuming a constant speed for the stopping process may be erroneous and that disrupting rIFG does not simply shift the entire distribution. Instead, our data suggest that rTMS over rIFG may specifically interfere with the quickest/most efficient instances of stopping.

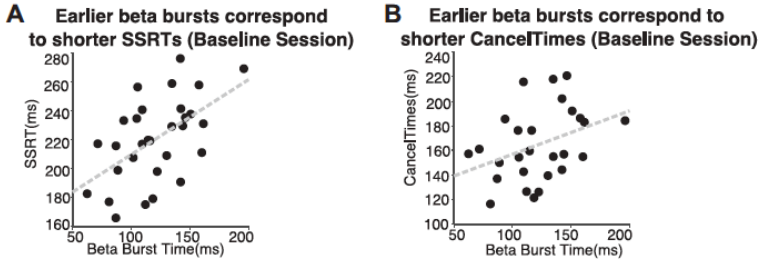

Figure 5. Relationship between the timing of beta modulations and stopping behavior in the baseline task session. *A*: consistent with prior work, we found a positive correlation between the timing of beta bursts (relative to the stop signal) and stop signal reaction time (SSRT). *B*: we observed a similar relationship between burst time and EMG CancelTime.

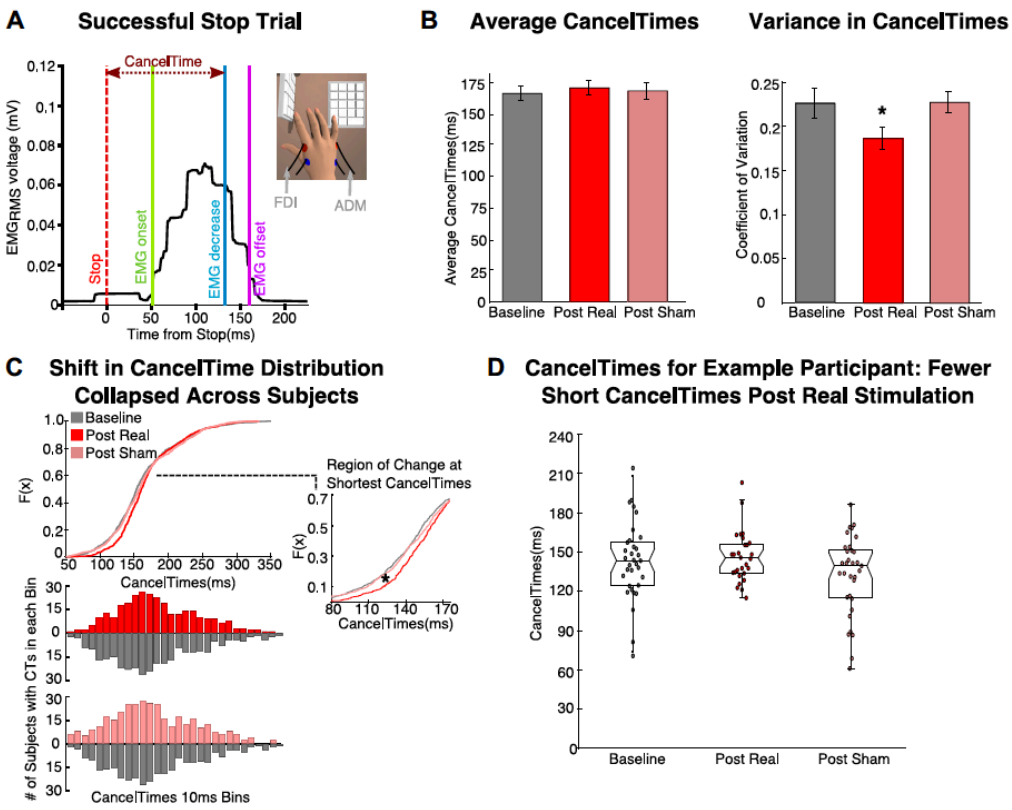

Figure 6. Analysis of the CancelTime metric of stopping from electromyography. *A*: example EMG recording from the responding hand on a successful stop trial with partial muscle activity. EMG onset marks the earliest detection of muscle activity after the stop signal. The CancelTime is the point at which EMG activity begins to decrease from the peak activity on trials when the action was successfully canceled. CancelTimes served as an alternative metric of stopping speed. RMS, root mean square. *B*: there was no difference between sessions in average CancelTimes (*left*), but there was in CancelTime variability measured by the coefficient of variation (*right*). *C*, *top*: the cumulative distribution function (CDF) illustrates the distribution of CancelTimes for each session (collapsed across all participants). The *inset* shows the portion of the CDF with a significant shift after real stimulation (i.e., the shortest CancelTimes are less frequent after real stimulation). *Bottom*: the number of participants contributing to each CancelTime (CT) bin included in the CDF. These data show that fewer participants overall have data points in the earliest CancelTime bins after real stimulation and that the effect on the CDF is not driven by only a few participants. *D*: example participant illustrating how the shortest CancelTimes are largely abolished after real stimulation.

DISCUSSION

In a preregistered double-blind study, we used 1-Hz rTMS to disrupt rIFG and/or the underlying right fronto-basal ganglia stopping network to test the effects on stopping-related beta and stopping performance. In the baseline session before rTMS, we observed the typical increase in right frontal beta power on stop trials that was greater for successful versus failed stop trials. Consistent with our predictions, after real rTMS SSRT was lengthened, and there was a reduction of the stopping-related beta power in a time window critical for stopping. Furthermore, this was a selective effect, as real rTMS did not reliably change Go RT or EEG mu/beta desynchronization in a left sensorimotor region. Surprisingly, however, sham stimulation also led to an increase in SSRT and a reduction in right frontal beta. However, a post hoc between-group analysis that considered the coil order showed that part of this reduction for sham might be due to a lingering effect of real stimulation, when real came first. Another between-group analysis, focused on the first rTMS session only, showed that the real stimulation group had a greater reduction of beta compared with the sham stimulation group. This latter result is one of the more important results of this report, as it shows that real stimulation reduces right frontal beta power significantly more than sham stimulation. We interpret this as evidence that real rTMS altered the neural architecture of the rIFG and/or underlying connected circuitry in a way that selectively effected the generation or propagation of beta oscillations and the event-related recruitment of the stopping system.

To date, there are more than half a dozen electrophysiological studies in humans pointing to an increase of right

frontal beta power related to action stopping or other forms of putative inhibitory control (Castiglione et al. 2019; Hannah et al. 2020; Jha et al. 2015; Picazio et al. 2014; Schaum et al. 2020; Swann et al. 2009; Wagner et al. 2016, 2018; Wessel et al. 2013). Changes in beta are also observed in other key regions of the stopping network during inhibitory control, further supporting beta as a functionally relevant signature of the stopping process (e.g., Brittain et al. 2012; but also see Errington et al. 2020). The present study, however, is the first to use a causal test to demonstrate that stopping-related modulations of beta are related to the integrity of the rIFG and/or the underlying circuitry. This has several important implications. First, it motivates the use of right frontal beta recorded in scalp EEG as a tool for testing theories of a putative right IFG network in other domains of inhibitory control such as emotional responsiveness and memory retrieval (Berkman et al. 2009; Castiglione et al. 2019; Depue et al. 2007; Guo et al. 2018). Scalp EEG has the advantage, over fMRI, of temporal resolution and portability. Furthermore, right frontal beta can be decomposed into beta bursts that have richer features such as trial-specific timing that other studies have used in research on action stopping (Hannah et al. 2020; Jana et al. 2020; Little et al. 2019; Wessel 2019). Although the present experiment was designed to focus on beta power, we did replicate prior work showing a relationship between the timing of beta bursts and stopping behavior in our baseline task session. Second, this work joins a growing body of literature that uses causal methods to test how changing neural circuits affects oscillations. By showing that disrupting a node in the stopping network reduced beta oscillations, our results support the idea that these oscillations are functionally relevant rather than a mere by-product of cognitive or behavioral processing (e.g., Fetz 2013; Hanslmayr et al. 2014, 2019; Joundi et al. 2012).

Although real rTMS did have selective effects on electrophysiology, the effects on behavior were only subtle. However, in both task sessions (sham and real) in which we observed a reduction in beta power relative to the baseline task session, we also found impaired stopping performance. As predicted, we observed an increase in SSRT after real stimulation. However, we also observed an effect after sham stimulation, and, unlike the EEG results, there was no clear distinction in the behavioral effect when we conducted a between-group analysis that removed coil order as a potential confound or when we analyzed the first session only (those who got real TMS first vs. those who got sham TMS first). We suspect that sham rTMS affected subsequent stopping behavior, at least in part, as a side effect of physical discomfort. Unlike earlier rIFG TMS studies that used a sham where the coil is turned away from the head (Chambers et al. 2006, 2007), our sham coil also activated the facial muscles, which is quite uncomfortable for 12 min. This discomfort may have put participants into a subsequently altered cognitive set that affected stopping performance. Indeed, other studies have shown that discomfort from stimulation can affect motor and cognitive task performance (Abler et al. 2005; Meteyard and Holmes 2018). Another reason that SSRT may not have been sensitive to coil type is that it is a single number calculated per participant. Although our alternative behavioral measure of stopping speed, mean CancelTime measured with EMG, was not different between coil types, a post hoc analysis showed differences in the CancelTime distribution. Specifically, there was a lengthening of the shortest CancelTimes after real versus sham stimulation, whereas the remainder of the distribution was unchanged. We consider two accounts of why only the shortest CancelTimes were affected. First, although 1-Hz rTMS is thought to have inhibitory effects on the underlying cortex, the affected region may still function, albeit less efficiently. As a result of a sluggish rIFG, the stopping process may show a slight lag, pushing the quickest instances of stopping into the next "bin" but leaving the remainder of the distribution largely intact. Specifically, because the longer CancelTimes are already relatively slow, the effect may get washed out for longer CancelTimes and affect only the quickest instances of stopping (i.e., the shortest CancelTimes). Second, the rIFG is only one node in a putative neural architecture for action stopping. Although other brain regions may compensate for an impaired rIFG, this may not happen quickly enough to affect the shortest CancelTimes.

One interpretational limitation is that we only stimulated one site, the rIFG, with the sham and real coils. Therefore, we cannot rule out that stimulating other prefrontal areas or nodes of the stopping network might also have yielded similar effects. Indeed, several studies have shown effects of TMS on stopping behavior when stimulating other nodes of the stopping network (e.g., Obeso et al. 2017). Our study, however, was hypothesis driven based on prior scalp EEG and ECoG studies (Swann et al. 2009; Wagner et al. 2018; Wessel et al. 2013), and going beyond one prefrontal site would have made our study intractable. It already required three visits per participant (and the TMS visit, involving EEG capping, was already very long). Future work should probe other stimulation sites and test for changes in right frontal beta to examine region specificity.

Finally, it is pertinent to note that several studies also implicate the rIFG in attention orienting and stimulus detection (Corbetta et al. 2002; Sharp et al. 2010) and it is possible that impaired stopping occurred because rTMS interfered with the role of the rIFG in detecting the stop signal as opposed to implementing the stopping process. Although we cannot definitively rule out this possibility, evidence from selective stopping tasks that control for attentional capture clearly show that rIFG recruitment and beta modulations are linked specifically to the stopping process, in contrast to other right prefrontal subregions that contribute more to attentional capture (Schaum et al. 2020; Sebastian et al. 2016). There is also now further anatomical evidence for a hyperdirect pathway in humans that provides a route for the rIFG to rapidly inhibit thalamocortical drive to the motor cortex via the subthalamic nucleus and cancel actions (Chen et al. 2020). Based on the current literature, our effects of rTMS over rIFG are likely the result of disrupting a stopping-specific process. Future work, however, should continue to tease apart the relationship between right frontal beta and action stopping versus attention and stimulus detection.

In conclusion, our study showed a real vs. sham effect on stopping-related beta, and it also showed across participants that the timing of beta bursts correlated with SSRT (just like Hannah et al. 2020; Jana et al. 2020). But we did not see a specific influence of real versus sham TMS on SSRT, nor did

we find that those participants with more TMS-reduced beta had longer SSRTs, so our study does still leave open the important question of whether beta is causal to stopping ability. To our knowledge, however, this is the first causal evidence linking stopping-related changes in right frontal beta to the integrity of the underlying rIFG-related network. This advances our understanding of the role of beta in inhibitory control over actions and supports the use of this EEG signature as a tool for testing the role of a rIFG prefrontal network in other domains of executive control.

ACKNOWLEDGMENTS

We thank John Rothwell for technical advice on procedures and Seraphina Solders for helping to develop this project as part of a rotation in the laboratory. We also thank Vignesh Muralidharan and Ricci Hannah for advice on study analyses and interpretation and Cynthia Yu for assistance with data collection.

GRANTS

This work was supported by NIH Grant DA-026452, NIMH Grant MH-020002, and the James S. McDonnell Foundation (220020375).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

K.K.S. and A.R.A. conceived and designed research; K.K.S. and S.J. performed experiments; K.K.S. analyzed data; K.K.S. and S.J. interpreted results of experiments; K.K.S. and S.J. prepared figures; K.K.S. drafted manuscript; K.K.S., S.J., and A.R.A. edited and revised manuscript; K.K.S. and A.R.A. approved final version of manuscript.

REFERENCES

Abler B, Walter H, Wunderlich A, Grothe J, Schönfeldt-Lecuona C, Spitzer M, Herwig U. Side effects of transcranial magnetic stimulation biased task performance in a cognitive neuroscience study. *Brain Topogr* 17: 193–196, 2005. doi:10.1007/s10548-005-6028-y.

Aron AR, Poldrack RA. Cortical and subcortical contributions to Stop signal response inhibition: role of the subthalamic nucleus. *J Neurosci* 26: 2424–2433, 2006. doi:10.1523/JNEUROSCI.4682-05.2006.

Aron AR, Robbins TW, Poldrack RA. Inhibition and the right inferior frontal cortex: one decade on. *Trends Cogn Sci* 18: 177–185, 2014. doi:10.1016/j.tics.2013.12.003.

Bari A, Robbins TW. Inhibition and impulsivity: behavioral and neural basis of response control. *Prog Neurobiol* 108: 44–79, 2013. doi:10.1016/j.pneurobio.2013.06.005.

Berkman ET, Burklund L, Lieberman MD. Inhibitory spillover: intentional motor inhibition produces incidental limbic inhibition via right inferior frontal cortex. *Neuroimage* 47: 705–712, 2009. doi:10.1016/j.neuroimage.2009.04.084.

Boroojerdi B, Prager A, Muellbacher W, Cohen LG. Reduction of human visual cortex excitability using 1-Hz transcranial magnetic stimulation. *Neurology* 54: 1529–1531, 2000. doi:10.1212/WNL.54.7.1529.

Brainard DH. The Psychophysics Toolbox. *Spat Vis* 10: 433–436, 1997. doi:10.1163/156856897X00357.

Brittain JS, Watkins KE, Joundi RA, Ray NJ, Holland P, Green AL, Aziz TZ, Jenkinson N. A role for the subthalamic nucleus in response inhibition during conflict. *J Neurosci* 32: 13396–13401, 2012. doi:10.1523/JNEUROSCI.2259-12.2012.

Cai W, Ryali S, Chen T, Li CS, Menon V. Dissociable roles of right inferior frontal cortex and anterior insula in inhibitory control: evidence from intrinsic and task-related functional parcellation, connectivity, and response profile analyses across multiple datasets. *J Neurosci* 34: 14652–14667, 2014. doi:10.1523/JNEUROSCI.3048-14.2014.

Castiglione A, Wagner J, Anderson M, Aron AR. Preventing a thought from coming to mind elicits increased right frontal beta just as stopping action does. *Cereb Cortex* 29: 2160–2172, 2019. doi:10.1093/cercor/bhz017.

Chambers CD, Bellgrove MA, Gould IC, English T, Garavan H, McNaught E, Kamke M, Mattingley JB. Dissociable mechanisms of cognitive control in prefrontal and premotor cortex. *J Neurophysiol* 98: 3638–3647, 2007. doi:10.1152/jn.00685.2007.

Chambers CD, Bellgrove MA, Stokes MG, Henderson TR, Garavan H, Robertson IH, Morris AP, Mattingley JB. Executive "brake failure" following deactivation of human frontal lobe. *J Cogn Neurosci* 18: 444–455, 2006. doi:10.1162/089892906775990606.

Chen R, Classen J, Gerloff C, Celnik P, Wassermann EM, Hallett M, Cohen LG. Depression of motor cortex excitability by low-frequency transcranial magnetic stimulation. *Neurology* 48: 1398–1403, 1997. doi:10.1212/WNL.48.5.1398.

Chen W, de Hemptinne C, Miller AM, Leibbrand M, Little SJ, Lim DA, Larson PS, Starr PA. Prefrontal-subthalamic hyperdirect pathway modulates movement inhibition in humans. *Neuron* 106: 579–588.e3, 2020. doi:10.1016/j.neuron.2020.02.012.

Cohen MX. Comparison of linear spatial filters for identifying oscillatory activity in multichannel data. *J Neurosci Methods* 278: 1–12, 2017. doi:10.1016/j.jneumeth.2016.12.016.

Corbetta M, Kincade JM, Shulman GL. Neural systems for visual orienting and their relationships to spatial working memory. *J Cogn Neurosci* 14: 508–523, 2002. doi:10.1162/089892902317362029.

Coxon JP, Stinear CM, Byblow WD. Stop and go: the neural basis of selective movement prevention. *J Cogn Neurosci* 21: 1193–1203, 2009. doi:10.1162/jocn.2009.21081.

Coxon JP, Van Impe A, Wenderoth N, Swinnen SP. Aging and inhibitory control of action: cortico-subthalamic connection strength predicts stopping performance. *J Neurosci* 32: 8401–8412, 2012. doi:10.1523/JNEUROSCI.6360-11.2012.

Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 134: 9–21, 2004. doi:10.1016/j.jneumeth.2003.10.009.

Depue BE, Curran T, Banich MT. Prefrontal regions orchestrate suppression of emotional memories via a two-phase process. *Science* 317: 215–219, 2007. doi:10.1126/science.1139560.

Errington SP, Woodman GF, Schall JD. Dissociation of medial frontal β-bursts and executive control. *J Neurosci* JN-RM-2072-20, 2020. doi:10.1523/JNEUROSCI.2072-20.2020.

Fetz EE. Volitional control of cortical oscillations and synchrony. *Neuron* 77: 216–218, 2013. doi:10.1016/j.neuron.2013.01.003.

Garavan H, Ross TJ, Stein EA. Right hemispheric dominance of inhibitory control: an event-related functional MRI study. *Proc Natl Acad Sci USA* 96: 8301–8306, 1999. doi:10.1073/pnas.96.14.8301.

Guo Y, Schmitz TW, Mur M, Ferreira CS, Anderson MC. A supramodal role of the basal ganglia in memory and motor inhibition: meta-analytic evidence. *Neuropsychologia* 108: 117–134, 2018. doi:10.1016/j.neuropsychologia.2017.11.033.

Hannah R, Muralidharan V, Sundby KK, Aron AR. Temporally-precise disruption of prefrontal cortex informed by the timing of beta bursts impairs human action-stopping. *Neuroimage* 222: 117222, 2020. doi:10.1016/j.neuroimage.2020.117222.

Hanslmayr S, Axmacher N, Inman CS. Modulating human memory via entrainment of brain oscillations. *Trends Neurosci* 42: 485–499, 2019. doi:10.1016/j.tins.2019.04.004.

Hanslmayr S, Matuschek J, Fellner M-C. Entrainment of prefrontal beta oscillations induces an endogenous echo and impairs memory formation. *Curr Biol* 24: 904–909, 2014. doi:10.1016/j.cub.2014.03.007.

Jahanshahi M, Obeso I, Rothwell JC, Obeso JA. A fronto-striato-subthalamic-pallidal network for goal-directed and habitual inhibition. *Nat Rev Neurosci* 16: 719–732, 2015. doi:10.1038/nrn4038.

Jana S, Hannah R, Muralidharan V, Aron AR. Temporal cascade of frontal, motor and muscle processes underlying human action-stopping. *eLife* 9: e50371, 2020. doi:10.7554/eLife.50371.

Jha A, Nachev P, Barnes G, Husain M, Brown P, Litvak V. The frontal control of stopping. *Cereb Cortex* 25: 4392–4406, 2015. doi:10.1093/cercor/bhv027.

Joundi RA, Jenkinson N, Brittain J-S, Aziz TZ, Brown P. Driving oscillatory activity in the human cortex enhances motor performance. *Curr Biol* 22: 403–407, 2012. doi:10.1016/j.cub.2012.01.024.

Kilavik BE, Zaepffel M, Brovelli A, MacKay WA, Riehle A. The ups and downs of β oscillations in sensorimotor cortex. *Exp Neurol* 245: 15–26, 2013. doi:10.1016/j.expneurol.2012.09.014.

Lang N, Speck S, Harms J, Rothkegel H, Paulus W, Sommer M. Dopaminergic potentiation of rTMS-induced motor cortex inhibition. *Biol Psychiatry* 63: 231–233, 2008. doi:10.1016/j.biopsych.2007.04.033.

Little S, Bonaiuto J, Barnes G, Bestmann S. Human motor cortical beta bursts relate to movement planning and response errors. *PLoS Biol* 17: e3000479, 2019. doi:10.1371/journal.pbio.3000479.

Logan GD, Cowan WB. On the ability to inhibit thought and action: a theory of an act of control. *Psychol Rev* 91: 295–327, 1984. doi:10.1037/0033-295X.91.3.295.

Makeig S, Bell AJ, Jung TP, Sejnowski TJ. Independent component analysis of electroencephalographic data. In: *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 1996, p. 145–151.

Meteyard L, Holmes NP. TMS SMART—scalp mapping of annoyance ratings and twitches caused by transcranial magnetic stimulation. *J Neurosci Methods* 299: 34–44, 2018. doi:10.1016/j.jneumeth.2018.02.008.

Nyffeler T, Wurtz P, Lüscher H-R, Hess CW, Senn W, Pflugshaupt T, von Wartburg R, Lüthi M, Müri RM. Repetitive TMS over the human oculomotor cortex: comparison of 1-Hz and theta burst stimulation. *Neurosci Lett* 409: 57–60, 2006. doi:10.1016/j.neulet.2006.09.011.

Obeso I, Wilkinson L, Teo JT, Talelli P, Rothwell JC, Jahanshahi M. Theta burst magnetic stimulation over the pre-supplementary motor area improves motor inhibition. *Brain Stimul* 10: 944–951, 2017. doi:10.1016/j.brs.2017.05.008.

Picazio S, Veniero D, Ponzo V, Caltagirone C, Gross J, Thut G, Koch G. Prefrontal control over motor cortex cycles at beta frequency during movement inhibition. *Curr Biol* 24: 2940–2945, 2014. doi:10.1016/j.cub.2014.10.043.

Raud L, Huster RJ. The temporal dynamics of response inhibition and their modulation by cognitive control. *Brain Topogr* 30: 486–501, 2017. doi:10.1007/s10548-017-0566-y.

Raud L, Westerhausen R, Dooley N, Huster RJ. Differences in unity: the go/no-go and stop signal tasks rely on different mechanisms. *Neuroimage* 210: 116582, 2020. doi:10.1016/j.neuroimage.2020.116582.

Schaum M, Pinzuti E, Sebastian A, Lieb K, Fries P, Mobascher A, Jung P, Wibral M, Tuescher O. Cortical network mechanisms of response inhibition (Preprint). *bioRxiv* 940841, 2020. doi:10.1101/2020.02.09.940841.

Sebastian A, Jung P, Neuhoff J, Wibral M, Fox PT, Lieb K, Fries P, Eickhoff SB, Tüscher O, Mobascher A. Dissociable attentional and inhibitory networks of dorsal and ventral areas of the right inferior frontal cortex: a combined task-specific and coordinate-based meta-analytic fMRI study. *Brain Struct Funct* 221: 1635–1651, 2016. doi:10.1007/s00429-015-0994-y.

Sharp DJ, Bonnelle V, De Boissezon X, Beckmann CF, James SG, Patel MC, Mehta MA. Distinct frontal systems for response inhibition, attentional capture, and error processing. *Proc Natl Acad Sci USA* 107: 6106–6111, 2010. doi:10.1073/pnas.1000175107.

Siebner HR, Filipovic SR, Rowe JB, Cordivari C, Gerschlager W, Rothwell JC, Frackowiak RS, Bhatia KP. Patients with focal arm dystonia have increased sensitivity to slow-frequency repetitive TMS of the dorsal premotor cortex. *Brain* 126: 2710–2725, 2003. doi:10.1093/brain/awg282.

Swann N, Tandon N, Canolty R, Ellmore TM, McEvoy LK, Dreyer S, DiSano M, Aron AR. Intracranial EEG reveals a time- and frequency-specific role for the right inferior frontal gyrus and primary motor cortex in stopping initiated responses. *J Neurosci* 29: 12675–12685, 2009. doi:10.1523/JNEUROSCI.3359-09.2009.

Swann NC, Cai W, Conner CR, Pieters TA, Claffey MP, George JS, Aron AR, Tandon N. Roles for the pre-supplementary motor area and the right inferior frontal gyrus in stopping action: electrophysiological responses and functional and structural connectivity. *Neuroimage* 59: 2860–2870, 2012. doi:10.1016/j.neuroimage.2011.09.049.

Thunberg C, Messel MS, Raud L, Huster R. Stimulating stopping? Investigating the effects of tDCS over the inferior frontal gyri and visual cortices (Preprint). *bioRxiv* 723296, 2019. doi:10.1101/723296.

Verbruggen F, Aron AR, Band GP, Beste C, Bissett PG, Brockett AT, et al. A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. *eLife* 8: e46323, 2019. doi:10.7554/eLife.46323.

Wagner J, Makeig S, Gola M, Neuper C, Müller-Putz G. Distinct β band oscillatory networks subserving motor and cognitive control during gait adaptation. *J Neurosci* 36: 2212–2226, 2016. doi:10.1523/JNEUROSCI.3543-15.2016.

Wagner J, Wessel JR, Ghahremani A, Aron AR. Establishing a right frontal beta signature for stopping action in scalp EEG: implications for testing inhibitory control in other task contexts. *J Cogn Neurosci* 30: 107–118, 2018. doi:10.1162/jocn_a_01183.

Wessel JR. β-Bursts reveal the trial-to-trial dynamics of movement initiation and cancellation (Preprint). *bioRxiv* 644682, 2019. doi:10.1101/644682.

Wessel JR, Conner CR, Aron AR, Tandon N. Chronometric electrical stimulation of right inferior frontal cortex increases motor braking. *J Neurosci* 33: 19611–19619, 2013. doi:10.1523/JNEUROSCI.3468-13.2013.

Yang CC, Völlm B, Khalifa N. The effects of rTMS on impulsivity in normal adults: a systematic review and meta-analysis. *Neuropsychol Rev* 28: 377–392, 2018. doi:10.1007/s11065-018-9376-6.

Chapter 1, in full, is a reprint of the material as it appears in the Journal of Neurophysiology 2020, Sundby, Kelsey; Jana, Sumitash; Aron, Adam. The dissertation author was the primary investigator and author of this paper.

Chapter 2:

Using the negative slow wave potential to measure memory intrusions

Abstract

Cued retrieval is the process by which an external cue evokes the retrieval of an associated item stored in long term memory. This memory process becomes problematic if the retrieved memory is distracting or harmful. Prior work suggests that individuals can avoid unwanted memory "intrusions" via a top-down inhibitory process that halts retrieval (Anderson & Green, 2001). The majority of evidence in support of this theory, however, relies largely on self-report and fails to objectively determine whether individuals are successful in preventing retrieval. Here we used the negative slow wave (NSW), an event related potential shown to track the number of items in working memory, as a putative objective metric of memory intrusions. In a modified version of the Think/NoThink task, participants learned associations between objects and contexts, i.e. large colored backgrounds, as we recorded EEG. Participants were then given a context as a memory cue and instructed to either retrieve (Think trials) or prevent retrieval (Not Think trials) of the associated object. Consistent with our hypothesis, NSW amplitude was reduced for NoThink compared to Think trials. We interpret this as evidence that participants were largely successful in preventing the associate from coming to mind. We did not, however, observe a reliable modulation of NSW amplitude on "catch trials" when participants were prompted to retrieve in a context that was typically paired with the NoThink instruction. This suggests that despite prior suppression episodes, participants maintain access to the NoThink associate. Although prior research demonstrates that repeated suppression erodes the NoThink memory trace, this is one of the first studies to intermix retrieval and no retrieval cues for NoThink items into the same task phase. Our result, therefore, suggests that consistent efforts to suppress may be necessary to have long term effects on the memory itself. Overall, these results

support the use of the NSW as a more objective metric for tracking retrieval success and for testing theories of executive control over long term memory retrieval.

Introduction

Unwanted thoughts can be unpleasant and disruptive to our daily behavior. Avoiding intrusive thoughts is particularly challenging when we encounter cues that prompt the retrieval of an unwanted memory. According to several prior studies, individuals may be able to voluntarily prevent cued retrieval to avoid specific memories (Anderson, 2005; Anderson & Green, 2001; Anderson & Hanslmayr, 2014). This is thought to require a right prefrontal network that acts to suppress the retrieval process (Anderson et al., 2004). There remain, however, several unresolved questions regarding the proposed mechanism and whether or not information is in fact excluded from working memory. An objective method for determining whether cued long term memories intrude is necessary for addressing these questions.

The Think/NoThink task (TNT) was designed to study executive control over retrieval. The standard version of TNT includes three phases. In Phase1 participants learn to associate multiple word pairs (e.g. oil-pump). In the subsequent Phase 2, participants undergo the Think/NoThink manipulation. Each trial begins with a memory cue, i.e. the left-hand word from one of the previously learned pairs (e.g. “oil”). The color of the cue word indicates if the participant should “Think” of the associate (i.e. retrieve “pump”), or try to prevent retrieval of the associate (i.e. “NoThink” trials). Recent versions of TNT also instruct participants to report whether or not the memory intruded after each trial: never, briefly, or often (Levy & Anderson, 2012). Participants are then given a surprise recall test in the final Phase 3. Dozens of TNT studies report reduced final recall for NoThink associates (e.g., Anderson & Green, 2001; Castiglione et al., 2019; Levy & Anderson, 2012; Racsmány et al., 2012; Gerd Thomas

Waldhauser et al., 2012) and a decline in the frequency of self-reported intrusions for NoThink trials across the duration of the task (Benoit et al., 2015; Castiglione & Aron, 2021; Castiglione et al., 2019; Levy & Anderson, 2012). Together, these behavioral data are interpreted as evidence of "suppression induced forgetting," or an erosion of the memory trace via an inhibitory process.

Intrusion ratings provide trial-by-trial insight into a participant's phenomenological experience but are limited in several critical ways. First, the rating is performed after each trial, providing only retroactive information and is, therefore, unable to measure the retrieval period itself. Second, self-reported ratings assume that participants are accurate in their introspection and not influenced by demand characteristics from the task instruction. Event related potentials (ERPs) measured with scalp EEG provide a more objective method for tracking retrieval success. Multiple TNT studies report a reduction in the amplitude of several retrieval related ERPs during NoThink trials, including the left parietal positivity (LPP) and the FN400 (Z. M. Bergström, J. de Fockert, & A. Richardson-Klavehn, 2009a; Bergström, Velmans, de Fockert, & Richardson-Klavehn, 2007; C. Chen et al., 2012; Brendan Eliot Depue et al., 2013; Hanslmayr, Leipold, Pastötter, & Bäuml, 2009; Mecklinger, Parra, & Waldhauser, 2009; Gerd Thomas Waldhauser et al., 2012). Although both the LPP and FN400 increase in amplitude during successful retrieval attempts, it is unclear how these ERPs relate to the precise content or quantity of information being retrieved. For instance, the FN400 has been associated other cognitive processes including cue familiarity (for review seeRugg & Curran, 2007). Thus, the attenuation of ERPs broadly linked to retrieval success lack specificity and may reflect a number of retrieval related processes: e.g., intent of retrieval, degree of conscious awareness, a sense of familiarity etc. (for review see Friedman & Johnson Jr, 2000). Consequently, even unsuccessful efforts to prevent

retrieval could modulate these EEG signatures regardless of whether or not the prohibited information intrudes. Hellerstedt, Johansson, and Anderson (2016) present one of the first studies to use the negative slow wave potential (NSW) to more precisely track memory information during the standard TNT. The NSW is a negative ERP that indexes the quantity of information in working memory (for review see, Drew, McCollough, & Vogel, 2006), increasing in amplitude with the number of items being held in working memory (e.g., Fukuda, Mance, & Vogel, 2015; Itthipuripat & Woodman, 2018; Mecklinger & Pfeifer, 1996). Hellerstedt et al. (2016) reported a reduction in NSW amplitude for NoThink trials. These data suggest that NSW amplitude may function as a more objective "readout" of NoThink trial outcomes. In the present study we sought to replicate this result in a modified version of TNT that more closely models our real-world experience. Instead of word-pairs, participants learned associations between real-world objects and colored backgrounds called "contexts." The task design aimed to provide a simplified model of the tendency for a context (e.g. a coffee shop) to evoke associated memories (e.g. that first cup of coffee shared with our partner). Additionally, we expanded the prior work by using the NSW to test whether repeated attempts to inhibit retrieval hampers later efforts to retrieve the previously suppressed information.

We predicted a reduction in NSW amplitude for NoThink compared to Think trials, indicating an exclusion of the NoThink associate from working memory. Additionally, we hypothesized that prior suppression episodes would affect later retrieval efforts, demonstrated by a reduction in NSW amplitude for NoThink contexts that were later presented with a retrieval cue.

Methods

Participants

We collected scalp EEG data for 23 participants. Three participants were excluded due to excessive noise in the EEG data, resulting in 20 participants in the final analysis (15 women, 5 men; right-handed; mean age= 19.15 yr, SD= 1.79 yr). Participants received $20/hr compensation for their time. All participants provided written informed consent according to a protocol approved by the University of California San Diego (UCSD).

Task Design

The task used features from both the standard TNT task (Anderson & Green, 2001) and from a a recent study that used the NSW to monitor the retrieval of objects from long term memory into working memory (Itthipuripat & Woodman, 2018). Because our primary goal was to track intruding information during the Think/NoThink retrieval window, the task consisted of only two phases: a learning phase and the TNT phase. **Phase 1 Learning**: Participants learned to associate unique objects (e.g. a shoe) with a specific context (i.e. a large colored background). Each trial started with a display of the context and a fixation dot in the center of the screen. The associated object then appeared in one of 6 possible locations on the context screen and remained on screen for 500ms. The location was randomized for each presentation of the object-context pair and was not part of the memory association. The object was then removed from the screen for a 1500ms memory delay. Participants were instructed to try to remember the association between the object and the colored context. A subsequent "Search Screen" with 6 objects appeared, prompting the participant to report whether or not the associated object was present. Participants pressed the "J" key to report "yes" or the "K" key to report "no". Feedback was provided after each response with a "$" indicating a correct response and a "0" indicating an

error. Participants learned a total of six different object-context associations during Phase 1. Each object-context association appeared 40 times for a total of 240 trials split into 5 blocks (48 trials per block). In Phase2, two of these contexts were assigned to the "NoThink" condition, two were assigned to the "Think" condition and the remaining two contexts served as "Baseline" contexts which will be described in more detail below. **Phase2 TNT**: One of the previously learned contexts appeared with either an "X" ("NoThink" instruction) or a "+" ("Think" instruction) in the center of the context. If shown a "+", the participant tried to immediately retrieve the object associated with that context. If shown an "X", the participant tried to immediately prevent the associated object from coming to mind. The memory cue (i.e. the colored context) and the Think(+)/NoThink (X) cue remained on the screen for two seconds. The participant then saw one of two screens: the context with six objects (Search Screen) or the context with 6 question marks (Intrusion Rating Screen). If shown the Search Screen, participants reported whether or not the associated object was present, "yes" or "no". If shown the intrusion rating screen, participants reported whether or not the object came to mind (0= "not at all", 1= "kind of", 2= "definitely"). To test whether repeated efforts to block retrieval affects future retrieval attempts, 50% of the NoThink context trials were actually paired with a retrieval cue (+), which prompted participants to try to retrieve the associated object. A Search Screen with 6 objects then appeared and the participants reported whether or not the associated object was present. Each Think and NoThink context was displayed a total of 160 times during Phase2 with 80 intrusion rating trials and 80 search screen trials per context. The two Baseline contexts always occurred with a retrieval cue (+) followed by a search screen. Importantly, however, there were only 80 trials per Baseline context to match the number of retrieval trials presented with the NoThink contexts (see Figure 2.1 for task design).

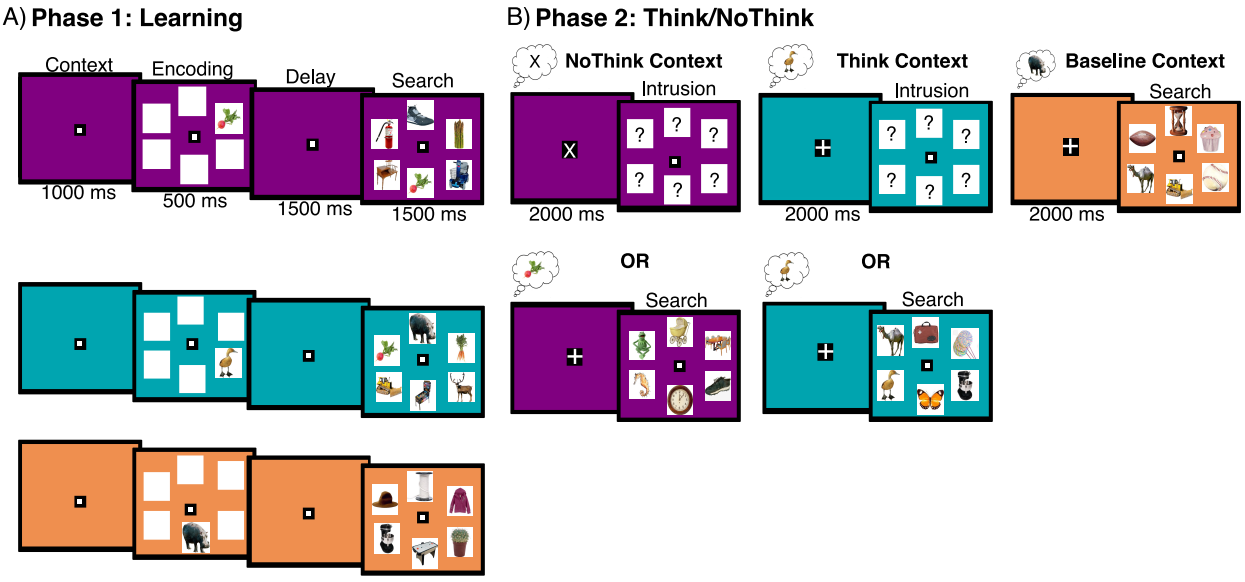

Figure 2.1. A) Participants learned 6 object-context associations. Two of these contexts became "Think" contexts, two became "NoThink", and two became "Baseline" contexts in Phase2. B) In Phase2 participants were given a memory cue (i.e. one of the colored contexts) and were prompted to "think" (+) or "not think" (X) of the associate. After a Think cue, participants were shown one of two screens: a search test screen or an intrusion rating screen where they indicated if the associate came to mind "not at all", "kind of", or "definitely". Trials with a NoThink cue (X) were always followed by an intrusion rating. A proportion of NoThink context trials (50%) were presented with a retrieval cue (+), prompting the participant to try to retrieve the associate. Participants then responded to a search array. Baseline contexts were always presented with a retrieval cue (+) but were only presented 80 times each to match the number of NoThink context trials that were paired with a retrieval cue.

EEG Recording and Preprocessing

We recorded scalp EEG using a 64 electrode ActiveTwo system with an electrode montage in accordance with the 5% 10/20 system (Biosemi Instrumentation, The Netherlands). Additional electrodes were placed bilaterally on the mastoids, the canthi and below and above each eye. Data were sampled at 1024Hz and on-line referenced to the BioSemi CMS-DRL reference. We kept all offsets from the reference below 25 µV.

The EEG data were preprocessed using custom MATLAB scripts and EEGLAB 14_1_1b (Delorme & Makeig, 2004). During preprocessing, we filtered the data using a .1hz high-pass and 80hz low-pass Butterworth filter (3rd order). We re-referenced the data to the average mastoids and then identified any channels with excessive noise via visual inspection. We used interpolation to mitigate the effect of any channel identified as excessively noisy. Channel

interpolation was only required for one channel in one subject. Next, we segmented the continuous EEG data into epochs from 2000ms prior to the context memory cue to 5000ms after context onset. We then conducted independent component analysis to remove prominent eye artifacts (Makeig, Bell, Jung, & Sejnowski, 1996) and we used threshold rejection and visual inspection to remove any trials with excessive residual artifact (due to muscle tension, sweat etc.). The artifact rejection procedure resulted in the removal of 19.36% +/- 8.74% SD of trials across the 20 participants.

NSW pre- intrusion rating

We first selected all of the trials that preceded an intrusion rating screen. In this subset of trials, NoThink contexts were always paired with the NoThink instruction (i.e. an "X" in the center of the screen) while Think contexts were always paired with a retrieval cue ("+"). We sorted trials into their respective condition, Think or NoThink contexts. We then computed the average NSW amplitude at electrodes Fz and Cpz during the 400-900ms period after memory cue onset for each condition. Although the scalp distribution for the NSW has been shown to vary according to the type of memory information, object-related memories typically focus over mid-frontal sites (Mecklinger & Pfeifer, 1996). Further, Hellerstedt et al. (2016) reported the greatest NoThink vs. Think effect on NSW at electrode Fz. We included Cpz in the analysis because prior work using a very similar object-context task reported more posterior effects on NSW amplitude (Itthipuripat & Woodman, 2018). The 400-900ms time window was also selected according to these studies (Hellerstedt et al., 2016; Itthipuripat & Woodman, 2018). We conducted a paired t-test for each electrode to test for changes in NSW amplitude across conditions (Think, NoThink). We expected a reduction in NSW amplitude for NoThink trials compared to Think trials.

NSW pre- search

We next selected all of the trials that preceded a Search Screen. The procedure for computing NSW amplitude on search trials was identical to the above "pre-intrusion rating" analysis except that we now included three conditions: Think contexts, NoThink contexts (now shown with a "+" retrieval cue), and Baseline contexts (contexts always shown with a "+" retrieval cue but with the same trial count as the NoThink contexts shown with a retrieval cue). We then used a one-way repeated measures ANOVA to test for an effect of condition on NSW amplitude. We performed this analysis for both electrodes. If repeated suppression attempts affects future retrieval efforts, we again expected a reduction in NSW amplitude for NoThink contexts despite the cue to retrieve.

NSW amplitude vs. behavior

Lastly, we examined the relationship between NSW amplitude and each participant's frequency of reported intrusions for NoThink trials. We expected that people who reported fewer NoThink intrusions would have a smaller NSW amplitude for NoThink trials, reflecting less information retrieved into working memory. We conducted a Pearson's correlation to examine the relationship between each participant's average NSW amplitude for NoThink contexts that preceded an intrusion rating (i.e. when trying to prevent retrieval) with the percentage of reported intrusions for NoThink trials. We computed this correlation for both electrode sites.

Behavioral Analysis

Phase1 Learning

We first examined memory performance for each context-object pair in Phase1. This analysis provided a sanity check and was important for demonstrating that there were roughly equivalent learning rates for each object-context pair prior to Phase2. We calculated the average

hit rate (i.e. the probability of correctly responding "yes" when the associate was present in the search array) and the average false alarm rate (i.e. incorrectly responding "yes" when the associate was absent). We conducted two repeated measure ANOVAs to test whether there was 1) a difference in the average hit rate and 2) a difference in the false alarm rate for each object-context pair during Phase1.

Phase2

Intrusion Ratings. Consistent with prior work, we predicted a decline in reported intrusions for NoThink trials as the task endured. Specifically, we examined the change in reported intrusions across the 10 blocks. We ran a two-way repeated measures ANOVA with block (1-10) and condition (Think, NoThink) as the factors and percentage of reported intrusions as the dependent measure. In accordance with prior work, ratings were coded as an intrusion whenever participants reported that the associate either "kind of" or "definitely" came to mind.

Search Performance. Our task included a proportion (50%) of trials in which a NoThink context was paired with a retrieval cue and followed by a search array. We aimed to test whether repeated efforts to suppress retrieval affected later efforts to retrieve the previously "inhibited" memory. We conducted three one-way repeated measures ANOVAs to examine whether memory performance suffered during retrieval attempts for NoThink contexts paired with a retrieval cue compared to Think and Baseline contexts. These ANOVAs tested for changes in 1) the rate of false alarms, 2) the rate of correct hits, and 3) the speed of responding to a hit in the search array (i.e. the response time).

Results

Behavior

Phase1:

Search Performance. There was no effect of object-context pair for either the probability of hits, $F(5,95)= 0.226$, $p=.950$ nor the probability of false alarms $F(5,95)= 2.24$, $p=.057$ during Phase1 learning. This result demonstrates that there was no initial difference in memory accuracy for each object-context pair prior to advancing to the Phase2 Think/NoThink manipulation.

Phase2:

Intrusion Ratings. We expected participants to report fewer intrusions for NoThink trials compared to Think. We also predicted that, consistent with prior work, participants would report fewer intrusions for NoThink trials across the duration of the task. We found a main effect of Block $F(9,171)=2.59$, $p=.038$, a main effect of Condition $F(1,19)=47.45$, $p<.001$, and a Condition x Block interaction $F(9,171)=3.42$, $p<.001$ for the frequency of reported intrusions (Figure 2.2A). The main effect of Condition showed that people reported fewer overall intrusions for NoThink trials compared to Think. Further, these findings replicate prior reports of a decrease in reported intrusions for NoThink trials across the duration of the task. In contrast to past work, however, we did also observe a slight reduction in reported intrusions for Think trials later in the task, as indicated by the main effect of Block. Importantly, however, the decline was greater for NoThink compared to Think trials. This was confirmed by conducting a paired t-test that compared the change in reported intrusions for Think (Block10-Block1, mean delta=-2.50, sd=19.81) versus the change in reported intrusions for NoThink (Block10-Block1, mean delta=-17.5, sd= 27.92), $t(19)=3.54$, $p=.002$. This pattern of intrusion ratings is consistent with prior reports and suggests that participants were performing our novel version of the TNT task in a similar manner to previous studies.

Search Performance. Our secondary hypothesis was that prior suppression episodes might impair later attempts to retrieve NoThink associates. We examined performance on search screens (i.e. trials in which all contexts were presented with a "+" retrieval cue) to test whether participants were less accurate or slower to identify memory hits when retrieving an associate that had been previously paired with a NoThink instruction. There was a main effect of condition on hit rate during Phase2 search trials, $F(2,19)=3.485$, $p=0.041$. We conducted three paired samples t-tests to examine how hit accuracy changed across conditions, with a Bonferroni-adjusted alpha level of 0.17 (.05/3). Participants had fewer correct hits on NoThink trials compared to Baseline, $t(19)=3.2781$, $p=.004$. Although the pattern of data suggests a similar reduction in accuracy for NoThink contexts compared to Think contexts, the effect was not significant, $t(19)=1.537$, $p=.141$. There was also no significant difference between Think and Baseline contexts, $t(19)=0.907$, $p=.376$ (Figure 2.2B). A similar pattern emerged when we examined the speed of responding to successful "hit" search trials (i.e. speed of correctly identifying the presence of the associated object). Again, there was a main effect of condition, $F(2,19)= 4.996$, $p=.012$. Participants were slower to correctly respond to the associated object for NoThink contexts compared to Baseline contexts, $t(19)=-3.6885$, $p=.002$. Again, NoThink contexts appeared to be slower compared to Think contexts but the effect did not reach significance, $t(19)=-1.870$, $p=.077$. There was again no significant difference between response times for Baseline and Think contexts $t(19)=-1.284$, $p=.215$ (Figure 2.2C). There was also no effect of condition on false alarm rate, $F(2,19)=1.569$, $p=0.221$.Together, these results show a slight decline in accuracy and a slowing of responses for contexts that were previously presented with a NoThink instruction. The effect, however, was only reliable when comparing NoThink and Baseline contexts.

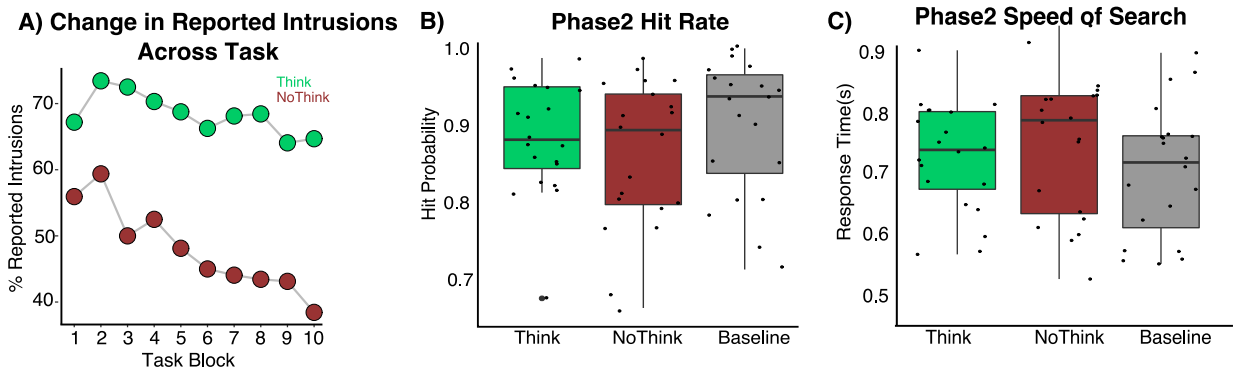

Figure 2.2. A) Consistent with prior studies, participants reported fewer intrusions for NoThink trials later in the task. B) The probability of a successful hit during Phase2 search arrays (i.e. accurately reporting "yes" when the associated object was present) varied across conditions. Post-hoc t-tests, revealed a decrease in accuracy for NoThink contexts compared to Baseline. The decrease in accuracy for NoThink vs. Think contexts, however, did not reach significance. C) The speed of correctly identifying a hit also varied across conditions. Participants were slower to identify a hit for NoThink contexts compared to Baseline contexts. A similar pattern was observed for the NoThink vs. Think comparison but it, again, did not reach significance.

NSW amplitude

NSW pre- intrusion rating:

We first tested for modulations of NSW amplitude prior to intrusion ratings, when NoThink contexts were paired with a NoThink instruction and Think contexts were paired with a Think instruction. Thus, this comparison reflects the standard comparison of interest between when individuals are actively attempting to retrieve an associate versus intending to prevent retrieval. We predicted a reduction in NSW amplitude for NoThink trials, presumably reflecting less information retrieved into working memory as participants sought to prevent retrieval. Consistent with this hypothesis, both paired t-tests (for electrode Fz and electrode Cpz) revealed a reduction in NSW amplitude for NoThink trials compared to Think trials, $t(19)=-4.253$, $p<.001$ for electrode Fz (Figure 2.3A) and $t(19)=-7.43$, $p<.001$ for electrode Cpz (Figure 2.3C).

NSW pre- search:

Our secondary hypothesis was that previously "suppressing" retrieval of an associate would hinder later efforts to retrieve the item and that this would be reflected in the NSW

amplitude. To test this, we examined NSW amplitude on the trials that preceded a search screen. For this set of trials, all context types (NoThink, Think and Baseline) were paired with a retrieval cue (+). We expected a reduction in NSW amplitude for NoThink contexts (i.e. a context that had been previously paired with a NoThink instruction) even as participants were prompted to now retrieve. Contrary to this hypothesis, however, we did not find a reliable effect of Condition on NSW amplitude for trials preceding search screens, $F(2,38)=1.69$, $p=.197$ for electrode Fz (Figure 2.3B) and $F(2,38)=0.899$, $p=.415$ for electrode Cpz (Figure 2.3D). To assess support for the null hypothesis we also computed bayes statistics for the Think versus NoThink comparison at both electrodes. We found anecdotal evidence in support of the null hypothesis at both electrodes, $B_{10}=0.795$ for electrode Cpz and $B_{10}=0.274$ at electrode Fz (Kass & Raftery, 1995).

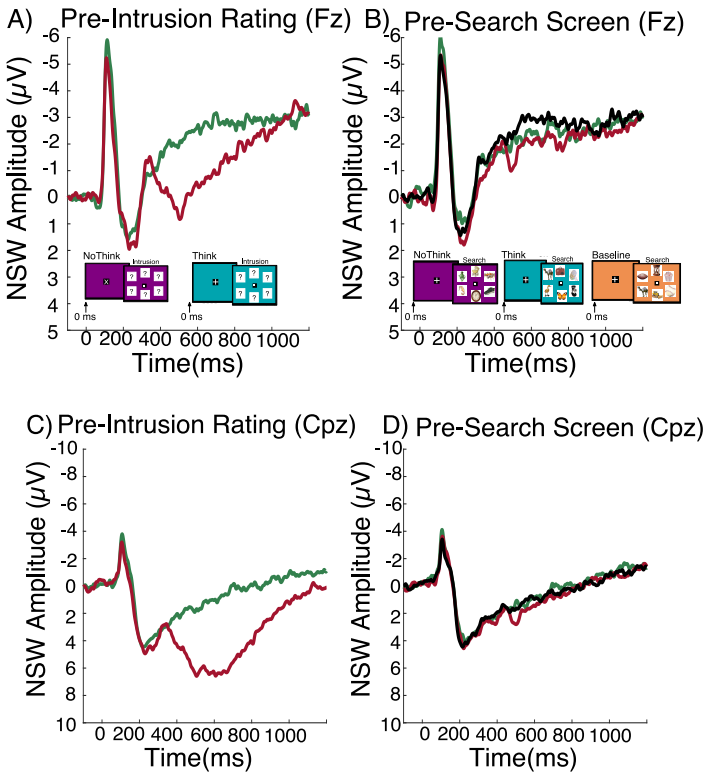

Figure 2.3. A) NSW amplitude at electrode Fz was significantly reduced during efforts to prevent retrieval (NoThink) vs. retrieve (Think), suggesting that participants were largely successful in excluding the NoThink associate from working memory. B) We did not find a significant reduction in NSW amplitude at electrode Fz when NoThink contexts were paired with a retrieval cue. C) Electrode Cpz also showed a reduced NSW amplitude for NoThink trials compared to Think trials. D) Electrode Cpz showed no effect on NSW amplitude across conditions for trials that preceded a search screen and when participants were instructed to retrieve for NoThink contexts.

Relationship between NoThink NSW amplitude and Reported Intrusions:

We observed a negative correlation between participants' percentage of reported intrusions for NoThink trials and the pre-intrusion NSW amplitude for NoThink trials from electrode Fz, R=-0.455, p=0.044 (Figure 2.4A). Specifically, participants who reported more intrusions showed a large NSW amplitude which may reflect more information coming to mind. There was no significant relationship for NSW amplitude at electrode Cpz and reported intrusions, R=-0.104, p=0.663. Together these data suggest that the NSW amplitude recorded at electrode Fz is more relevant to NoThink performance (Figure 2.4).

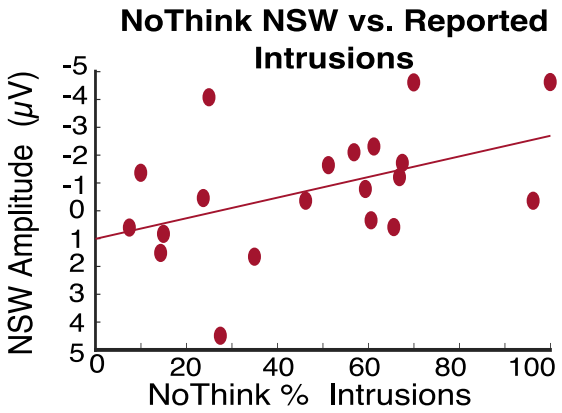

Figure 2.4. A) Participants who reported more NoThink intrusions had larger average NSW amplitudes at electrode Fz on NoThink trials.

Discussion

We measured NSW amplitude as participants attempted to either retrieve or prevent the retrieval of a cued memory. As predicted, we found a reduction in NSW amplitude for NoThink trials compared to Think trials. We interpret this as evidence that participants were largely successful in preventing the associate from coming to mind when instructed to cancel retrieval. Additionally, participants who reported greater success in preventing retrieval (i.e. fewer NoThink intrusions), had a smaller NSW amplitude measured from a frontal electrode site. We did not, however, observe a reduction in NSW amplitude when participants were prompted to try to retrieve in a NoThink context. Together these results provide objective evidence that

individuals can voluntarily prevent retrieval of an associated memory yet still maintain access to the memory if later required to retrieve it.

Our key finding was that NSW amplitude was reduced when participants sought to prevent retrieval. We demonstrate this result in a novel version of the TNT task using context-object pairs rather than arbitrary word pairs. Assuming NSW tracks the amount of information in working memory, the reduction in amplitude may reflect less retrieved information. Further, because NSW amplitude has been shown to scale with the quantity of information, the reduction in NSW amplitude for NoThink trials is consistent with the inhibition account. Alternative strategies for avoiding the NoThink associate, including thought substitution, would likely yield the same amount or an even greater quantity of items in working memory compared to Think trials (Benoit & Anderson, 2012). Additionally, NSW amplitude was smaller during NoThink trials for participants who reported fewer intrusions. This result suggests that NSW amplitude may specifically relate to the amount of information that we consciously retrieve and our mnemonic awareness.

Our secondary hypothesis was that repeated efforts to prevent retrieval would interfere with future retrieval attempts. This prediction was based on prior reports that use the standard TNT task and show a decline in final recall for NoThink items, an observation termed "suppression induced forgetting" (e.g., Benoit & Anderson, 2012; Bergström et al., 2009b; Castiglione et al., 2019; Brendan E Depue et al., 2007; Detre et al., 2013; Levy & Anderson, 2012). As the name implies, suppression induced forgetting is thought to result from repeated suppression episodes that erode the memory trace and impede later access (Anderson & Green, 2001). Contrary to our hypothesis, we did not find a reliable effect on NSW amplitude when participants were given a NoThink memory context and were instructed to actively retrieve the

associate. Behavioral metrics of memory accuracy, however, did show subtle effects. Specifically, participants were less accurate and slower to respond to hit trials for NoThink contexts that were paired with a retrieval cue compared to Baseline contexts. Although, the pattern of data was similar (i.e. reduced accuracy and slower response times) for NoThink retrieval trials compared to Think trials, these behavioral effects were not significant. Thus, our data suggest that despite subtle effects on behavioral metrics of retrieval, participants maintain access to NoThink associates and are largely successful in retrieving these items when prompted to do so. The current study, however, is the first to our knowledge to examine retrieval for NoThink items during the Think/NoThink phase itself. We speculate that inserting retrieval prompts for NoThink items may have counteracted the standard effect on memory. This result may have important implications as it suggests that repeated *and* consistent efforts to inhibit retrieval may be necessary to induce long-lasting changes on intrusive memories.

One limitation of the NSW approach is that the ERP is thought to track the amount of information in working memory rather than the specific content. Due to the lack of memory-specific information, we are unable to entirely rule out other more task-general processes that could also contribute to changes in the NSW amplitude (e.g. increased task difficulty for NoThink trials). We aim to address this shortcoming in the subsequent chapter.

Overall, these data support theories of executive control over retrieval with evidence of reduced intruding information during NoThink trials in an entirely novel version of the TNT task. This study highlights the need for a more objective method for measuring intrusions and identifies the NSW as a useful tool for probing theories of executive control over long term memory retrieval.

Chapter 2, in full, is currently being prepared for submission for the publication of the material. Sundby Kelsey; Itthipuripat, Sirawaj; Aron, Adam. The dissertation author was the primary investigator and author of this paper.

Chapter 3:

Using multivariate methods to track memory retrieval in a Think/NoThink Task

Abstract

Items in our environment often elicit the retrieval of associated memories. This retrieved information can confer behavioral benefits by guiding our actions and drawing our attention to features that proved important in the past. On the other hand, cues can also bring to mind unwanted memories. A growing body of literature suggests that individuals are able to voluntarily cancel retrieval via a right prefrontal frontal (PFC) network. According to the direct suppression account, the right PFC actively inhibits the retrieval process and thereby erodes the associate memory. The extant data, however, lack objective evidence that the memory representation itself is modulated during retrieval. The current study seeks to develop methods that can both objectively and continuously track memory reactivation during efforts to suppress retrieval. In Chapter 2 we observed a marked reduction in the the negative slow wave potential (NSW), an event related potential that tracks the amount of information in working memory, as participants sought to prevent retrieval (NoThink). NSW amplitude, however, does not provide information about the specific content retrieved into working memory, making it challenging to rule out alternative explanations for the NoThink effect. We now apply an EEG-based decoding approach to more accurately determine whether the *specific* prohibited memory intrudes during NoThink efforts. These data reveal reliable decoding of the associate memory during both Think and NoThink trials, suggesting that information may frequently intrude despite efforts to prevent retrieval. Importantly, however, we observed a later reduction in decoding accuracy for the NoThink memory representations relative to Think. This suggests that memory content may still be susceptible to control even once retrieval has begun. This study demonstrates the utility of multivariate decoding methods as a novel approach to objectively and continuously monitoring

the retrieval of memories during Think/NoThink tasks, and thus, provides a critical tool for testing theories of executive control over cued retrieval.

Introduction

We rely on on our past experiences to point us to relevant information and to guide our behavior. Reminders of the past, however, can become detrimental when they elicit unwelcomed memories. Research suggests that individuals can voluntarily "stop" cued retrieval via a right prefrontal network (PFC) (Brendan E Depue et al., 2016; Guo et al., 2018). There is currently, however, little direct evidence that recruitment of the right PFC prevents or suppresses the reactivation of a memory trace. The current evidence for retrieval suppression relies largely on self report and later deficits in memory that are thought to be the consequence of suppressing retrieval (Anderson & Bell, 2001; Levy & Anderson, 2012). These behavioral data are linked to both increased activation of the right PFC network and reduced activity in the downstream hippocampus (Anderson et al., 2004; Benoit & Anderson, 2012; Benoit et al., 2015; Gagnepain et al., 2014; Schmitz et al., 2017). None of these data, however, provide an objective measure of the memory information itself as individuals are trying to prevent retrieval. An objective measure of the memory representation is important for both validating claims of active inhibition and for resolving ongoing debates as to whether the right PFC exerts control proactively, in response to the cue (Castiglione et al., 2019), or reactively, in response to an intruding memory (Castiglione & Aron, 2021; Levy & Anderson, 2012). The current study addresses this limitation by adopting methods that use neural actively to better track the reactivation of a memory across the retrieval window and as participants seek to suppress.

In Chapter 2, we used the negative slow wave potential (NSW), to track the amount of information entering working memory during a Think/NoThink (TNT) task. The standard TNT

task instructs participants to either retrieve (Think) or prevent retrieval (NoThink) of a cued memory (see General Introduction Chapter for a detailed description) (Anderson & Bell, 2001; Levy & Anderson, 2012). We observed a reduction in NSW amplitude for NoThink trials which may reflect less information retrieved into working memory. NSW amplitude, however, does not provide a specific metric of memory content, making it challenging to rule out other potential explanations for the observed effect. The current study expands on Chapter 2 by using a multivariate approach to more accurately determine whether the *specific* prohibited memory intrudes during NoThink efforts.

Multivariate methods have been used to track a range of information but several methods have proven particularly successful in tracking spatial information (e.g., Bae & Luck, 2018; Foster, Sutterer, Serences, Vogel, & Awh, 2017; Sutterer, Foster, Serences, Vogel, & Awh, 2019). Thus, we designed a novel object-location version of the TNT task and used a support vector machine (SVM) decoding approach. We applied this method to scalp EEG to monitor the emergence of retrieved memories reflected in neural activity. We first expected SVM decoding to reliably track retrieval during a learning Phase in which participants were prompted to retrieve on every trial. This result would simply help to validate our application of the SVM methods by replicating prior work (Bae & Luck, 2018). Our core prediction, however, focused on the Think/NoThink phase. We predicted that SVM decoding would succeed in tracking Think memories, i.e. when trying to retrieve, yet fail to reliably track NoThink memories, i.e. when participants sought to prevent retrieval. By applying SVM decoding to each timepoint, this method also provides a continuous metric of the memory representation across the retrieval window and may yield unique insight into when memory information intrudes and/or is successfully modulated.

Methods

Participants

We collected data for 26 participants. Five participants were excluded due to excessive artifact in the EEG data (i.e. >60% trials rejected due to artifact) and one participant was excluded due to abnormally poor task performance[10]. One additional participant withdrew before completing the visit. Thus, 19 participants were included in the final analysis (mean age=20.21, SD=2.57; right handed; 16 female). Participants received $20/hour for their time. All participants provided written informed consent according to a protocol approved by the University of California San Diego (UCSD).

Object-Location Think/NoThink Task

We developed a novel object-location version of the Think/NoThink task to use with the SVM analysis technique (Figure 3.1). This task was adapted from Sutterer et al. (2019). The task consisted of two phases: a two-part learning phase and the Think/NoThink phase. **Phase1 Part1**: Phase1 was split into two parts. Part1 consisted of a brief learning period to ensure that participants had acquired some degree of learning prior to collecting the scalp EEG data that would be used for training the SVMs. Each trial began with an object presented in a specific location on the computer screen. Participants viewed the object-location display for 1000ms and were instructed to try to remember the association between the presented object and its designated location. The memory cue (i.e. the object) then disappeared as participants sought to maintain their memory for the association. After a 1250ms memory delay, the memory cue reappeared in the center of the screen and participants had to click the remembered location on a response circle. After responding, participants received feedback with the object displayed in the correct location and their response error for that trial. The response error reflected the difference

[10] Specifically, this participant failed to respond to trials prior to the response deadline on every trial excluding one.

between the participant's response and the correct location ranging from -180 (a counterclockwise error) to 180 degrees (a clockwise error). Participants learned a total of 12 object-location pairs and each pair was presented 5 times during Part1of learning, resulting in a total of 60 trials. **Phase1 Part2 (EEG training dataset):** During Part2 of learning, one of the previously learned objects appeared in the center of the screen as the memory cue and remained on screen for 2000ms. We instructed participants to try to immediately retrieve and think of the associated location for the duration of the memory cue. We used the scalp EEG data collected during this 2000ms retrieval window to train the SVM model (described in more detail below). After the retrieval period, a response circle appeared prompting participants to click the remembered location (unspeeded response). Participants again received feedback with the object displayed in the correct location and their response error. Part2 of learning included 600 total trials (50 trials per memory cue) divided into three blocks with 200 trials per block. We encouraged participants to briefly rest between blocks. **Phase2 Think/NoThink (EEG testing dataset):** Phase 2 introduced the Think/NoThink manipulation. Eight of the 12 pairs were assigned to the Think and NoThink conditions (i.e. 4 per condition)**.** The remaining 4 "Baseline" pairs did not reappear in the task but allowed greater coverage of the response circle and prevented the memory task from being too easy. On each trial, one of the Think/NoThink objects appeared in the center of the screen as the memory cue. A green memory cue served as a "Think" cue and prompted participants to immediately retrieve the associated location from long term memory (LTM). Alternatively, a red memory cue indicated that the participant must try to *not* think of the associated location. The memory cue remained on the screen for 2000ms. After Think trials, participants saw one of two screens: 1) a test screen with a response circle that prompted them to click the associated location or 2) an intrusion rating screen where they

indicated if the location came to mind “never”, “briefly”, or “often”. Participants were informed that there was a time limit for their response on the test screen and that responding quickly was important. We included this instruction to encourage participants to quickly retrieve the remembered location immediately upon seeing the memory cue. The screen advanced if participants failed to respond within 2000ms. After NoThink trials, participants again saw one of two screens: 1) a speeded (2000ms for response) visual perception test in which they clicked a dot that appeared randomly on the response circle or 2) an intrusion rating screen where they indicated if the associated location came to mind “never”, “briefly”, or “often”. The visual perception test was designed to match the behavior required from the test screen on Think trials without ever requesting that participants respond with the true associated NoThink location. We never tested participants for the associated location on NoThink trials because this may have increased the likelihood that some participants would ignore the NoThink instruction in anticipation of a potential test. Phase2 included 640 total trials divided into 4 blocks (160 trials per block) with each of the 8 memory cues appearing a total of 80 times. Participants responded to intrusion ratings on 75% of the trials for each condition (240 intrusion rating trials per Think/NoThink condition, 60 per memory cue) and responded to the memory/perception test screen on the remaining 25% of the trials in each condition (80 test screen trials per Think/NoThink condition, 20 per memory cue). We used the EEG data collected as participants sought to either “Think” or “Not Think” of the associated locations as the testing dataset for the SVM analysis.

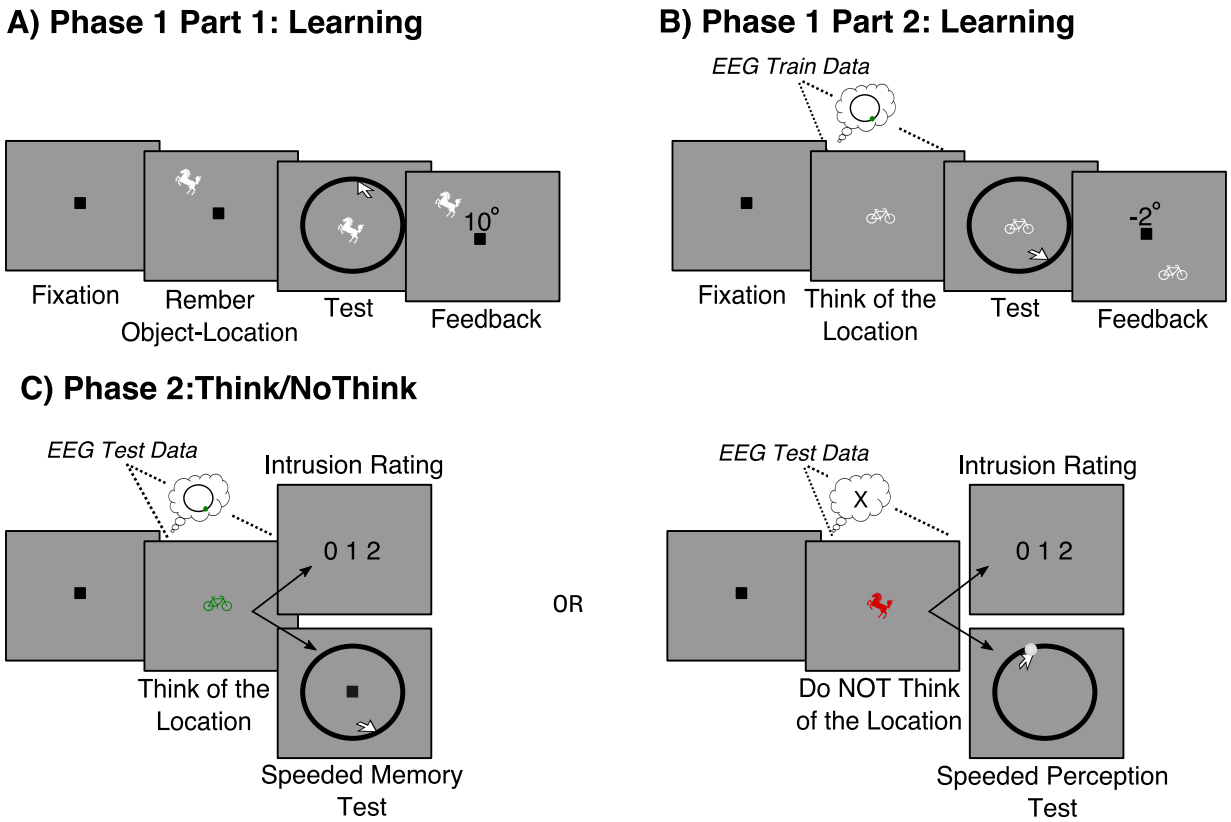

Figure 3.1 A) In Phase1Part1, prior to collecting EEG data to train the SVM models, participants completed a brief learning period in which they learned to associate objects with specific locations on the computer screen. B) In Phase1Part2, participants were shown one of the previously learned objects as a memory cue, and were instructed to try to retrieve the associated location. After the retrieval period, they clicked the remembered location on a response circle and received feedback. C) In Phase2 participants were given one of the previously learned objects either in green (Think) or in red (NoThink). If the memory cue appeared in green, the participant tried to immediately retrieve and think of the associated location. They were then shown one of two screens: 1) a test screen (25% of trials) for which they clicked the remembered location or 2) an intrusion rating screen (75% of trials) for which they reported whether or not the location came to mind never (0), briefly (1), or often (2). If the memory cue appeared in red, the participant instead tried to block retrieval of the associated location. After each NoThink trial, participants again saw one of two screens: 1) a speeded perception test (25% of trials) in which they had to click the location where a dot randomly appeared or 2) an intrusion rating screen (75% of trials). In our primary analysis, we used Phase1Part2 data to train the SVMs and Phase2 data to test the models.

Stimuli

We drew the memory cue objects from the Sutterer and Awh (2016) clip art library. We randomly assigned the selected objects to unique angular locations (0-360 degrees within 30 degree steps) for each participant. Thus, every quadrant of the response circle had three assigned objects that were randomly assigned to one of the three conditions (Baseline, Think, NoThink).

EEG Acquisition and Preprocessing

We recorded scalp EEG using a 64 electrode ActiveTwo system (BioSmei Instrumentation, Netherlands) with a sampling rate of 1024Hz. External electrodes were placed under and above each eye, on the bilateral mastoids and on each canthi. The data were on-line referenced to the BioSemi CMS-DRL reference and we kept all offsets below 25μV during recordings.

We conducted all EEG analyses using Matlab2016b (The Math Works, United States) with the EEGLAB (Delorme & Makeig, 2004) and ERPLAB toolboxes (Lopez-Calderon & Luck, 2014). Data were first downsampled to 512Hz, referenced to the average mastoids and filtered (0.01-80hz). We identified channels with excessive noise via visual inspection and used interpolation to mitigate the effect of bad channels. The data were then epoched from -500 ms prior to the memory cue to 2000ms post memory cue. We then performed baseline correction with the 200ms preceding the memory cue (i.e. during fixation). The epoched data were sorted into bins for each spatial memory included in the analyses, resulting in eight bins. We then used the ERPlab function pop_artstep to identify epochs with artifact exceeding a threshold of 75 μV. These epochs were excluded from the subsequent analyses. Lastly, we performed independent component analysis and removed independent components reflecting eye artifacts (i.e. blinks and saccades).

Analysis of Task Behavior

Phase1

We first computed each participant's average response error for both Part1 and Part2 of Phase1 Learning. Participants learned a total of 12 object-location pairs with only one object

assigned to each 30degree segment of the response circle. Thus, an average response error <30degrees indicates that participants were within the appropriate segment of the circle.

Phase2

Intrusion Ratings. We calculated the frequency of reported intrusions for both Think and NoThink conditions. The frequency was calculated by dividing the number of trials with a reported intrusion in one condition (i.e. Think or NoThink) by the overall number of trials that had an intrusion rating screen for that condition. This value was then converted to a percentage by multiplying by 100. Consistent with prior work we coded both "briefly" and "often" as intrusions. We used a paired t-test to examine differences in reported intrusions for Think and NoThink conditions. We also calculated the frequency of reported intrusions for Think and NoThink trials in each Phase2 block to test for changes in reported intrusions across the duration of the task. We conducted a two way repeated measures ANOVA with condition (Think, NoThink) and Block (1 to 4) as the factors and percentage of reported intrusions as the dependent measure. Consistent with prior studies, we expected a decline in reported intrusions for NoThink trials across the duration of the task.

Response Accuracy. We expected accurate responding on both types of test screens in Phase2 (memory and perception). We calculated each participant's average response error on the memory tests that occurred on 25% of the Think trials. Again, we assumed that an average error of ~30degrees was indicative of successful retrieval. We next calculated average response error for the perception tests that occurred on 25% of NoThink trials. High accuracy (i.e. within a few degrees) on the perception test was important for demonstrating that participants remained vigilant and attentive on NoThink trials.

SVM Overview

EEG-based decoding, also referred to as stimulus classification, uses patterns of neural activity to predict which stimulus or task condition produced the observed neural response. A model is first trained to distinguish between different stimulus classes according to differences in neural activity. The algorithm used to train the model can vary but the current study focuses on the use of support vector machines (see Appendix for inverted encoding model approach). Support vector machines (SVMs) are a supervised machine learning algorithm frequently used for classification problems. The method classifies data by estimating a hyperplane or decision boundary from a training dataset that maximizes the distance between the hyperplane and the nearest datapoint of each class. The model is then able to classify a novel datapoint (from a testing dataset) according to which side of the hyperplane the datapoint falls.

In the case of TNT, we assume that if a memory can be accurately classified into its correct location bin based on neural data, this implies that memory information was retrieved and present in the neural signal. Our decoding approach was motivated by a recent study that used the distribution of phase-locked ERP voltages to decode precise spatial locations during a working memory task (Bae & Luck, 2018). We applied these methods to the object-location version of TNT to examine our ability to decode spatial memories from whole brain patterns of ERPs during Think trials and to test for reductions in decoding accuracy during NoThink trials.

SVM Analysis

Phase1 Training/Phase1 Testing

We first applied the ERP decoding methods to only Phase1Part2 data (split into training and testing sets) to confirm that we could reliably decode retrieved spatial memories when participants were instructed to retrieve on every trial. The model was trained and tested using the

scalp distribution of phase-locked ERP voltage. The raw voltage was filtered (0.1-80hz) during preprocessing and thus, was not specific to any frequency band. The procedure, described below, closely followed the methods from Bae and Luck (2018) and was conducted with adapted versions of their analysis scripts.

The analyses were conducted on the epoched data (-500 to 2000ms relative to the memory cue) and applied to each participant's data separately. We first partitioned the Phase1 data into three independent sets to be used as either training or testing data (see Trial Assignments for more details). The data were then resampled at 50Hz (i.e. ~1 data point per 20ms) to improve computing efficiency. Thus, each dataset initially contained a 4-dimensional matrix for each participant with the dimensions of time (126 timepoints), spatial bin (8 unique remembered locations-i.e. those that were later assigned to be Think or NoThink), trials, and electrode site (64 electrodes). We then computed the average ERP for each of the 8 spatial bins at each electrode and at every timepoint in all three of the sets. Thus, after averaging, each of the 3 datasets (per participant) contained one averaged ERP distribution (across the 64 electrodes) for each spatial bin and at every point in time (i.e. the dataset dimensions were 8 spatial bins x 64 electrodes x 126 timepoints).

We first trained the classifier using SVMs to define the decision boundaries between classes. Because SVMs are typically used for binary classification problems (i.e. discriminating between only two classes), we trained 8 separate SVMs to account for all 8 remembered locations. Using a "one vs. all" approach, each SVM learned to implement a binary classification that distinguished one of the 8 locations from the other 7 locations. This procedure was performed for every timepoint.

Next, during the testing stage, we used the 8 trained SVMs combined with error-correcting output codes (ECOC) to classify each of the retrieved locations in the independent test dataset (Dietterich & Bakiri, 1994). ECOC models combine the results from multiple binary classifiers (e.g. multiple SVMs) to solve multiclass categorization problems. Specifically, unique 8 bit codes were used to denote each class. The eight numbers reflect the targets for each of the eight trained SVMs (Dietterich & Bakiri, 1994). For example, (1, 0, 0, 0, 0, 0, 0, 0, 0) and, (0, 1, 0, 0, 0, 0, 0, 0, 0) represent class one and class two respectively. When given a new datapoint to classify, if all 8 SVMs accurately classify it, then the ECOC reports no error. However, if at least one of the classifiers fails, then the class with its code closest in hamming distance to the given output code (i.e. the closest match) will be assigned as the label. Thus, these 8 trained ECOC models were together used to predict the location label for each item in the test dataset. Using the MATLAB predict() function, a single spatial bin label was assigned to each observation in the testing dataset by minimizing the average binary loss over the 8 SVMs. We then computed the decoding accuracy by comparing the true spatial bin label (from the test dataset) with the predicted label. Decoding was deemed correct if the classifier correctly determined which of the 8 locations was being retrieved from long term memory.

Phase1 Training/Phase2 Testing

Our primary aim was to test whether memory information was present (i.e. above chance decoding) when participants were instructed to retrieve on Think trials versus prevent retrieval on NoThink trials. To test this, we applied the ERP decoding methods to the Phase2 Think/NoThink data. The procedure followed the exact steps described above except that we used Phase1 data for training and Phase2 data for testing (see below Trial Assignment).

Trial Assignment

Phase1 Training/Phase1 Testing

We first applied the SVM procedure to Phase1 data. As described above, the Phase1 data were divided into three datasets while ensuring that each set had an equal number of trials per spatial bin. Using a cross-validation procedure, we used two of the datasets to train the SVM classifier and the remaining dataset to test the classifier. We repeated this procedure three times such that each set served as the testing set once. We computed the average decoding performance across each test set at each timepoint. Chance performance was 12.5% (1/8).

Phase1 Training/Phase2 Testing

To examine decoding during the Phase2 Think/NoThink manipulation, we used the Phase1 data for training and the Phase2 data for testing. We first partitioned the Phase1 data into two independent training sets and the Phase2 data into two independent testing sets. The number of trials for each location bin and each condition were balanced across both training sets and across both testing sets. We used both training sets to train the models and then conducted the testing stage with each of the independent test datasets. The decoding accuracies were then averaged across each test set and calculated separately for Think spatial bins and NoThink spatial bins.

Resampling

Phase1 Training/Phase1 Testing

For the Phase1 only analysis, we iterated the entire SVM procedure, including random trial assignment, ten times to avoid spurious results driven by idiosyncrasies from trial assignment. We averaged decoding accuracy for each of the 8 spatial locations across the ten iterations. Thus, this produced a single estimate of each participant's decoding accuracy for each spatial memory at each timepoint during retrieval in Phase1. We then averaged across the 8

spatial memories to achieve each participant's average decoding accuracy for spatial memories at each timepoint during Phase1 retrieval. In line with methods by Bae and Luck (2018), we applied a five-point moving window (reflecting a +/- 40ms period) to smooth decoding accuracies across timepoints and reduce noise. Based on estimates from Bae and Luck (2018), we expect the final temporal precision from the full decoding pipeline to be ~50ms.

Phase1 Training/Phase2 Testing

For the Think/NoThink analysis, we again repeated the entire decoding procedure ten times with new random trial assignments for each of the training and testing datasets. We then averaged decoding accuracy for the four spatial bins in each condition (Think and NoThink) across each iteration. Thus, this produced single estimates of each participant's decoding accuracy for either Think spatial memories or for NoThink spatial memories during retrieval in Phase2. We again applied a five-point moving window to smooth decoding accuracies across timepoints and mitigate noise.

Decoding accuracy and statistics

Phase1 Training/Phase1 Testing

We first tested whether we could reliably decode retrieved spatial memories during Phase1 only, i.e. as participants sought to retrieve on every trial. Decoding accuracy should exceed chance levels (i.e. 1/8 with 8 spatial bins) if the retrieved spatial memory was accurately represented in the ERP scalp distribution during Phase1 Learning. We used a 3 step nonparametric cluster-based Monte Carlo simulation technique to statistically compare decoding accuracy to chance levels at each time point while controlling for multiple comparisons (Bae & Luck, 2018). We restricted our statistical analyses to a time window in which retrieval was expected to occur (500-2000ms post memory cue). We selected this time window according to

hippocampal data that estimates pattern completion to occur around 500ms post memory cue (Staresina & Wimber, 2019).

Step1. We first used one-tailed one-sample t-tests at each timepoint to identify times at which the mean decoding accuracy, averaged across all participants, exceeded chance levels in the retrieval window. We then identified clusters of contiguous time points for which the single-point t-tests were significant ($p<.05$). We summed the t-scores in each cluster to obtain a cluster-level t-mass that was then compared against a cluster level t-mass calculated from a null distribution. This analysis assessed whether a cluster of contiguous significant t-values was larger than a mass that would occur by chance alone. The procedure for producing the null distribution is described below in step2.

Step2. We produced the Monte Carlo null distribution and calculated cluster level t-mass values from the null. The null distribution was created by simulating the decoding accuracy when the decoder randomly guessed the retrieved spatial location while naïve to the true location. Specifically, on each simulated trial, we randomly sampled an integer from 1 to 8 to serve as the decoder prediction label for a given target. The prediction was deemed correct if it matched the target value. Consistent with the procedure run on the true data, we repeated this routine 240 times (8 locations x 3 cross-validations x 10 iterations) and then averaged across the 240 repetitions to obtain the mean simulated decoding accuracy for each time point. Using the same procedure applied to the true data, the simulated decoding accuracies were then smoothed with a five-point running average filter. Step 2 was repeated 19 times to reflect each of the 19 participants. We then calculated the cluster-level t-mass from these simulated data using the same procedure that was applied in Step1 (i.e. focusing on the window of retrieval 500-2000ms

post cue). In cases with more than one significant cluster, we selected the t-mass from the largest cluster.

Step3. We produced a null distribution for the cluster mass. Specifically, we simulated 10,000 experiments (i.e. 10,000 iterations of step2) in which the decoder randomly guessed the retrieved location. We then created a distribution with the maximum t-mass value from each of the 10,000 iterations. We next calculated the p-value of each cluster in the true data by testing where each observed t-mass from the true data fell within the simulated null distribution. We considered the observed cluster-level t-mass to be significant if it fell above the top 95% of the null distribution.

Phase1 Training/Phase2 Testing

Our primary hypothesis what that we would be able to reliably decode spatial memories for Think trials and that decoding accuracy would be reduced during NoThink trials. To test this, we applied the same statistical procedure as we described above to the decoding accuracies that we obtained when we trained the model with Phase1 data and tested the model with the Phase2 Think/NoThink data. The cluster-based Monte Carlo simulation technique (described above) was used to identify time points with above chance decoding for the Think and for the NoThink conditions separately. To compare the two conditions and to identify the timepoints at which Think and NoThink differed, we used a non-parametric bootstrap method (with 5000 iterations), followed by FDR to correct for multiple comparisons.

SVM Predictions

Phase1 Training/ Phase1 Testing

We first performed the decoding analysis on only the Phase1 data (split into separate training and testing datasets). We expected above-chance decoding of the retrieved spatial

memories during Phase1. We also expected participants with higher decoding accuracy to perform better on the memory tests given after each trial.

Phase1 Training/Phase2 Testing

Our primary analysis examined decoding accuracy for Phase2 Think/NoThink. We expected reliable decoding for Think trials and reduced decoding accuracy for NoThink trials as participants sought to prevent retrieval.

Results

Behavior

Phase1

The group average for response error was within 30degrees of the correct location for both Part1 and Part2 of Phase1 learning. This demonstrates adequate learning of the object-location pairs prior to advancing to Phase2 (Part1 mean error=4.947, SD=1.369; Part2 mean error=13.732, SD=8.615) (Figure 3.2A).

Phase2

Intrusion Ratings. Participants reported more intrusions on average for Think trials compared to NoThink trials, $t(18)=8.632$, $p<.001$. A repeated measures ANOVA with Condition (Think, NoThink) and Block(1 to 4) as factors also revealed changes in reported intrusions as the task progressed. There was a main effect of Block ($F(3,54)=9.622$, $p<.001$), a main effect of Condition ($F(1,18)=73.38$, $p<.001$), and a Condition x Block interaction ($F(3,54)=14.33$, $<.001$). Thus, consistent with the prior NSW study, we observed a decline in reported intrusions from Block1 for both conditions. Importantly, however, we again found that the decline in reported intrusions was significantly greater for the NoThink condition. This was confirmed by a paired t-test that compared the change in reported intrusions for Think (Block4-Block1, mean delta=-

0.439) versus the change in reported intrusions for NoThink (mean delta=-18.772), t(18)= 4.792, p<.001 (Figure 3.2C).

Response Accuracy. The group average for response error was within 30degrees of the correct location on the post Think test screens that occurred on 25% of the Think trials (mean response error=10.962, SD=7.891). Participants were also accurate on the perception tests that occurred after 25% of NoThink Trials (mean response error= 2.470, SD=2.267) (Figure 3.2B).

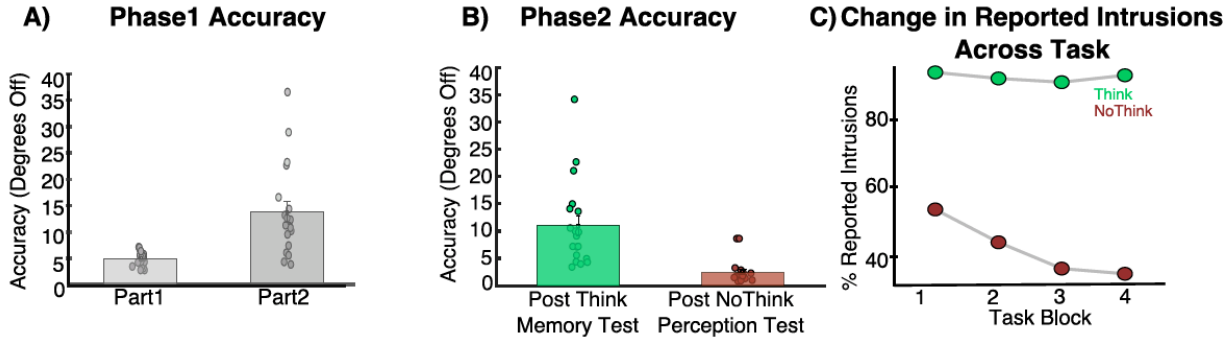

Figure 3.2. A) Participants showed accurate retrieval of spatial memories during both parts of Phase1 Learning. B) Participants remained accurate in responding to test screens that occurred after a proportion of Think trials. Participants were also very accurate on the perception tests that occurred after a proportion of the NoThink trials, helping to confirm that participants remained vigilant and attentive on NoThink trials. C) Consistent with prior work, we observed a reduction in the frequency of reported intrusions for NoThink trials across the task duration.

SVM Phase1 Training/Phase1 Testing

To validate that we could use the SVM approach to decode the retrieval of spatial memories, we first applied the decoding regime to only the Phase1 data (split into training and testing sets) during which participants sought to retrieve on every trial. We found above chance decoding in the window 700ms –1980ms post memory cue (Figure 3.3A). We also observed a negative correlation between response accuracy in Phase1 (i.e. average degrees off from the correct location) and decoding accuracy across participants. Specifically, participants with better behavioral accuracy had higher EEG decoding accuracy of the retrieved location, R=-0.473, p=.041 (Figure 3.3D). This correlation helps to validate that the quality of decoding from the neural representation relates to behavioral measures of memory accuracy.

SVM Phase1 Training/Phase2 Testing

Our primary analysis applied the decoding methods to Phase2 data to test whether decoding accuracy was reduced during NoThink trials when participants sought to prevent retrieval. We found above-chance decoding for both Think and NoThink trials in the window 700ms –1920ms post memory cue. We also observed a later reduction in NoThink decoding relative to Think from 1060ms-1140ms post memory cue (Figure 3.3B).

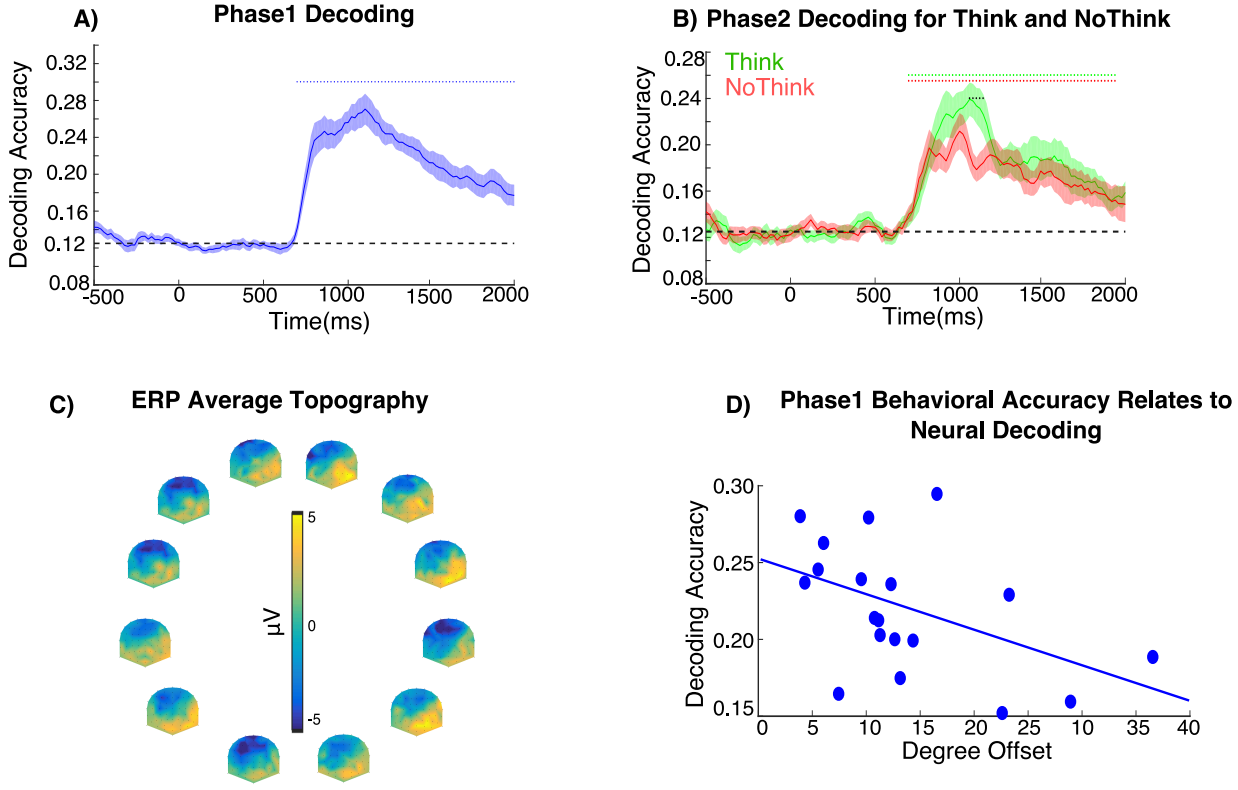

Figure 3.3. A) We could reliably decode retrieved spatial memories in Phase1 data, when participants retrieved on every trial (significant period 700ms –1980ms post memory cue, indicated by blue dotted bar). B) There was above-chance decoding for both Think and NoThink spatial memories (significant period 700-1920ms post memory cue, indicated by green (Think) and red (NoThink) dotted bars) during Phase2. Later during retrieval, decoding was reduced for NoThink memories relative to Think (significant period 1060-1140ms post memory cue, indicated by black dotted bar). C) Scalp topography of ERPs for each remembered location in the Phase1 training set, averaged across 700-2000ms post memory cue. The decoding analyses included only the 8 object-location pairs (from the 12 learned) for each participant that were subsequently assigned to Think or NoThink conditions in Phase2. This topography appears to show a negative voltage over anterior sites which may relate to the NSW ERP measured in Chapter 2. D) Participants with better neural decoding also performed better on the memory tests in Phase1.

Post-hoc Analysis

Decoding vs. self-reported intrusions

Consistent with prior work, we observed a reduction in self-reported NoThink intrusions across the duration of the task (Figure 3.2C). We next conducted an exploratory analysis to test whether decoding accuracy showed a similar pattern. Specifically, we predicted that decoding in the second half of the task (blocks 3-4) would be reduced for NoThink memories compared to the first half of the task (blocks 1-2). Current theories contend that the decline in self-reported intrusions is the consequence of executive control (Levy & Anderson, 2012). Thus, we did not expect decoding accuracy to decline later in the task for the Think condition. Indeed, we found that the reduction in NoThink decoding accuracy was amplified in the second half of the task (i.e. the last two blocks) versus the first half (Figure 3.4A). There was no difference between task halves in the Think condition (Figure 3.4B). This result is important for showing that decoding accuracy at least partially reflects the participants' subjective experience.

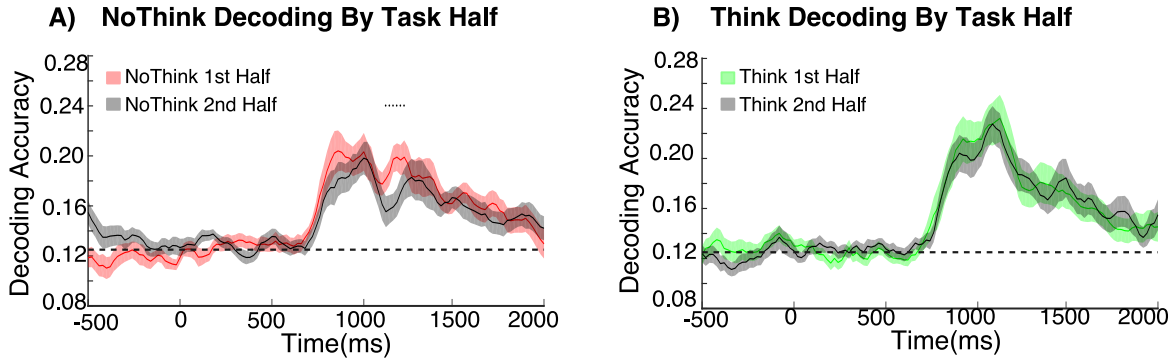

Figure 3.4. A) The decoding accuracy for NoThink memories is further reduced for the second half of the task compared to the first (significant divergence 1120-1220ms post cue). This pattern is consistent with the tendency for participants to report fewer NoThink intrusions later in the task. B) There was no significant difference between decoding for Think trials when comparing the first vs. second half of the task.

Discussion

The current study presents a novel approach for tracking the retrieval of memories during a Think/NoThink task. Using SVM decoding methods, we observed above-chance decoding for both Think and NoThink spatial memories. We interpret this as evidence that memory information frequently intrudes in response to an associated cue despite efforts to suppress retrieval. NoThink decoding accuracy was, however, reduced later in the retrieval period. This later reduction may reflect the consequence of inhibitory control or other strategies aimed at preventing retrieval of the NoThink item. Further, decoding accuracy decreased across the duration of the task for NoThink memories. This pattern mirrors the tendency for people to report fewer intrusions during NoThink trials later in the task and suggests that decoding accuracy is at least partially related to one's subjective experience. Together these data contribute objective support for the theory that individuals can voluntarily modulate the retrieval process when confronted with an associated cue.

Although the TNT task has greatly advanced the field's understanding of executive control over LTM retrieval, the outcome of NoThink trials has proven challenging to measure. Evidence for success on NoThink trials is often inferred from post trial self-reports and impaired recall in subsequent memory tests. Thus, the current study makes an important contribution by revealing measureable changes in the neural representation of the NoThink memories as participants attempt to suppress retrieval.

A major advantage of the SVM approach is that it can be applied to each timepoint during retrieval, providing a continuous measure of the neural representation across the critical period. Having a continuous metric is important for examining when information intrudes and when the putative executive process intervenes. As a result of this, the current dataset revealed

an interesting pattern in which NoThink memory information appears to initially intrude and is then subsequently reduced relative to intentional retrieval (i.e. Think trials). This later dip in decoding for NoThink memories suggests that memory information may still be susceptible to modulation even after the initiation of retrieval. This finding aligns with current explanations for suppression induced forgetting which posit that a memory trace must be at least partially active in order for it to be susceptible to suppression (Detre et al., 2013; Dudai, 2004; Lee, 2009; Levy & Anderson, 2012; Nader et al., 2000). Our observation is also consistent with prior work that reports increased activity in the right PFC network, the proposed source of control, specifically on trials with reported intrusions. This has been interpreted as evidence that inhibitory control intervenes to suppress or clear the intruding memory (Levy & Anderson, 2012).

The current study is limited in several ways. First, it is unclear how decoding accuracy relates to one's subjective experience of a memory. That is, the ability to decode a memory from neural activity does not necessitate that the individual is overtly aware of the intruding content. Although, the present data shows some parallels between self-reported intrusions and decoding (i.e. both are reduced later in the task), decoding may also reflect memory information that is present in neural activity yet not conscious. The relationship between decoding accuracy and one's subjective experience requires further investigation. Yet it is important to note that latent or subconscious memory intrusions may still be detrimental. Although the data are mixed, a large body of literature suggests that subconscious information may sometimes effect decisions and behavior (for review, Kouider & Dehaene, 2007; Newell & Shanks, 2014). As a result, measuring information that is retrieved yet does not reach conscious awareness may still be relevant to understanding executive control of LTM.

Second, it is important to acknowledge, that the decoding analysis was conducted on data averaged across trials. This is because raw voltage from scalp EEG is very noisy and difficult to measure on a single trial level. Based on these averaged data, the current results do not appear to support a "gating" theory in which rapid control preemptively cancels retrieval upon seeing an associated cue. However, because the current approach does not provide a single trial metric, it is still certainly possible that an earlier executive process may successfully prevent retrieval on some trials and that these trials are simply masked when examining the averaged result.

Lastly, the current analysis is unable to determine what causes the observed changes in decoding accuracy during NoThink trials. In addition to active inhibition, there are a number of non-inhibitory explanations that could produce reductions in decoding accuracy (e.g. diverting attention to another thought). Linking reductions in decoding to inhibition will require testing the relationship between the right PFC network and changes in decoding accuracy.

In summary, the present study provides objective evidence that individuals are able to intentionally modulate memory retrieval even once the process has begun. These data go beyond our prior NSW work, presented in Chapter 2, by tracking the specific NoThink memory representation across the retrieval window. Although the gating and clearing accounts of inhibitory control are not necessarily mutually exclusive, the extant literature has lacked the tools to adequately test either theory. The current study helps to address this limitation by providing an objective and continuous metric of memory information that greatly improves our ability to test these theories and to gain a deeper understanding of when retrieval becomes susceptible to control.

Chapter 3 is coauthored with David W. Sutterer, Sirawaj Itthipuripat and Adam Aron. The dissertation author was the primary investigator and author of this paper.

Chapter 4:

Examining the relationship between right frontal beta and memory representations during a Think/NoThink task

Abstract

A recent theory posits that action-stopping and "canceling" memory retrieval requires a common inhibitory process that is implemented via a right prefrontal (PFC) network. According to this theory, recruitment of the right PFC network should result in less intruding memory content when provoked by an associated cued. There is currently, however, little direct evidence linking the right PFC network to the active clearing or exclusion of memory information during efforts to prevent retrieval. Testing this relationship has been hindered by the absence of an objective and continuous metric of the associate memory across the retrieval window. In Chapter 3, we addressed this problem by using support vector (SVM) decoding methods applied to EEG data to track the retrieval of memories as participants sought to either retrieve (Think) or prevent retrieval (NoThink). In the current chapter, we conducted an exploratory analysis that examines the relationship between right frontal beta, a marker of the right PFC network, and SVM decoding accuracy. We hypothesized that right frontal beta on NoThink trials would correspond to less intruding memory information and that this would be reflected in the decoding accuracy. Consistent with this prediction, we found that participants with a higher probability of right frontal beta bursts, within 500ms of the memory cue, exhibited lower overall decoding accuracy for NoThink memories. Since decoding accuracy is thought to reflect the quality or amount of memory specific content present in the neural signal, this relationship between right frontal beta and decoding accuracy is consistent with inhibitory accounts of the right PFC network. This preliminary evidence thus links a right prefrontal executive system with an objective measure of the associate memory, and helps to validate the theory that this network contributes to executive control over cued retrieval.

Introduction

Inhibition is often considered essential for maintaining control over our actions and thoughts. A substantial body of research implicates a right prefrontal (PFC) network in both the inhibition of actions and in the "inhibition" of cued memory retrieval (Benoit & Anderson, 2012; Benoit et al., 2015; Brendan E Depue et al., 2016; Guo et al., 2018; Levy & Anderson, 2002, 2012). To date, however, there is little direct evidence that this right PFC network actively suppresses or prevents the retrieval of memories. The present dissertation contributes two tools that may aid in testing the relationship between right PFC recruitment and its effect on memory retrieval. First, in Chapter 1, we helped to validate right frontal beta oscillations as an EEG signature of the right PFC network (Sundby, Jana, & Aron, 2021). Next, in Chapter 3, we identified support vector machine (SVM) decoding methods applied to EEG data as a reliable tool for tracking the neural representation of the associate memory during a Think/NoThink task. In the current chapter, using the same participants from the SVM study of Chapter 3, we conducted an exploratory analysis that combined these tools to test whether heightened right frontal beta relates to reduced decoding accuracy of NoThink memories. This would provide a more direct link between recruitment of the right PFC network and disruption of the retrieval process.

Support for right frontal beta as a signature of the right PFC network comes primarily from studies that examine its role in action-stopping (Enz et al., 2021; Hannah et al., 2020; Schaum et al., 2021; Sundby et al., 2021; Swann et al., 2009; Wagner et al., 2018). Specifically, right frontal beta power increases during successful stop trials and in a time window that is critical for stopping, i.e. after the stop signal but before stop signal reaction time (SSRT). Although prior work has shown similar increases in right frontal beta during NoThink trials that

relate to fewer self-reported intrusions, this relationship has not yet been tested using an objective measure of the associate memory (Castiglione et al., 2019).

A growing body of research on beta oscillations and inhibitory control indicate that these increases in beta power are the result of brief bursting events within the beta band rather than a prolonged modulation (Enz et al., 2021; Errington, Woodman, & Schall, 2020; Jana, Hannah, Muralidharan, & Aron, 2020; Little, Bonaiuto, Barnes, & Bestmann, 2018; Wessel, 2020). Analyses of beta bursts provide a rich set of features including burst probability and burst timing relative to an event. These features have provided new insight into the relationship between beta and action-stopping. Studies using burst timing, for example, reveal a tight relationship between beta burst timing and stop signal reaction time (SSRT), a behavioral "timestamp" that is thought to mark the end of the stopping process (Jana et al., 2020). One advantage of beta burst analyses rather than beta power, is that meaningful changes in bursts do not require beta modulations to consistently occur at the same moment across trials. This is particularly pertinent for tasks that study executive control over long term memory retrieval, for which the timing of intrusions and implementing control may vary widely both within and across participants. The current study tests the relationship between right frontal beta bursts and SVM decoding, our objective metric of memory retrieval, during a TNT task.

As described in Chapter 3, we used SVM decoding applied to whole brain patterns of ERP voltages to track the retrieval of spatial memories during a novel object-location version of the TNT Task. This approach revealed above-chance decoding for both Think and NoThink memory representations during the retrieval period. Importantly, however, we observed a later reduction in NoThink decoding accuracy which may reflect successful efforts to modulate the retrieval process. In the current chapter we tested whether decoding accuracy of the NoThink

memories was related to the probability of beta bursts during NoThink trials. Unlike action-stopping and SSRT, however, it is currently unclear which precise timepoint during the decoding window would reflect the "timestamp" for when the putative control process affects retrieval. Thus, the current study instead focused on beta burst probability in a time window that is critical for action-stopping (i.e. within 500ms of the cue) and the decoding accuracy for NoThink memories averaged across the full window of above-chance decoding (700-1920ms post memory cue). Therefore, this exploratory analysis simply tested whether a higher probability of beta bursts soon after the cue corresponds to less memory specific information being retrieved and represented in neural activity. We predicted that early recruitment of the right PFC network, reflected in increased beta bursts (within 500ms of the cue), would relate to lower overall decoding accuracy of the NoThink memories. To measure right frontal beta, we used generalized eigenvalue decomposition (GED), a guided multivariate source separation technique, to select a right frontal filter from each participant's EEG data (Cohen, 2017). We then used this spatial filter to examine the relationship between beta busts and decoding accuracy for NoThink memories.

Methods

Participants

This analysis included the same 19 participants that were used in the SVM analysis described in Chapter 3 (mean age=20.21, SD=2.57; right handed; 16 female).

Behavior and EEG

The beta burst analysis was applied to the same dataset used for the SVM analysis. These data were collected during the object-location version of the TNT task. We did not conduct any

new behavioral analyses nor change any steps in the preprocessing of the EEG data (see Chapter 3 for details on task, behavioral data, and EEG preprocessing).

Identifying a right frontal filter

Our method for identifying a right frontal spatial filter and examining beta bursts was modeled after methods used to study beta bursts in action-stopping. As noted above, we used GED (Generalized Eigen Decomposition), a guided multivariate source separation technique, to identify a right frontal filter (i.e. a weighting of the electrodes) for each participant. GED identifies a spatial filter, in other words a subspace, which maximally separates two conditions of interest. Moreover, GED is guided because it uses information about features of interest including the frequency, topography and time window in which you expect the underlying activity of two different conditions to diverge. We used GED rather than selecting specific electrodes because this method has been shown to substantially improve the signal-to-noise ratio in the analysis (Cohen, 2017). Selecting the filter for each participant involved the following steps:

1. Epoching the EEG data from -1000 to 3000ms, time-locked to the memory cue (we selected a large epoch to avoid edge artifacts from filtering)
2. Narrow band filtering the data at each frequency in the beta band (12-24 Hz) using a Gaussian window with full-width half-maximum of 5Hz. We used 12-24Hz as the beta band range (rather than 12-30Hz) because this reduced the number of components to manually review and because the majority of work on right frontal beta and inhibitory control shows the relevant changes in lower beta (for review see, Schmidt et al., 2019; Wagner et al., 2018).

3. Computing the covariance matrices at each frequency (12-24Hz) for each condition (Think trials and NoThink trials) in the window of interest (0-500ms post cue) and then running GED. This requires solving the generalized eigen value equation (see below) to factorize the covariance matrices of both conditions (i.e. the covariance matrix for Think and the covariance for NoThink) to its respective eigen values (λ_D) and eigen vectors (W_D). A small 1% regularization was added to the reference matrix (i.e. the Think covariance matrix) to improve the quality of the decomposition (Cohen, 2021). The generalized eigen value equation was as follows:

$$C^{-1}_{Think} C_{NoThink} W_D = W_D \lambda$$

4. The eigenvectors (W_D) represent the spatial filters which maximally separate Think from the NoThink condition and the corresponding eigenvalues (λ) represent the amount of variance explained by each eigenvector. Selection of the filter/component for further analysis is done by looking at the spatial topography, the % of variance explained and the signal (NoThink) to reference (Think) power ratio of each component. The spatial topography of a filter is estimated by computing the forward activation, i.e. projecting the selected eigenvector onto the corresponding covariance matrix (in this case the NoThink covariance matrix). The signal to reference ratio of the filter, or the power ratio is estimated by projecting the power time series (i.e. power at a specific frequency) of the respective conditions (NoThink and Think) on the selected component and then computing the ratio of mean power within the window of interest (in this case 0-500ms).
5. We manually reviewed the top 6 components from the GED output, based on the % variance explained, for each beta frequency. Among these, we then selected the filter

with features that best matched those identified in action-stopping. It is important to note that we did not have a separate action-stopping task to help localize a right frontal beta filter that could then be applied to the TNT data. Thus, we selected the component according to the following characteristics typical of the beta signature for action-stopping: a power ratio greater than 1 (i.e. greater power for NoThink vs. Think within 500ms of the memory cue) and a right frontal topography (Figure 4.1).

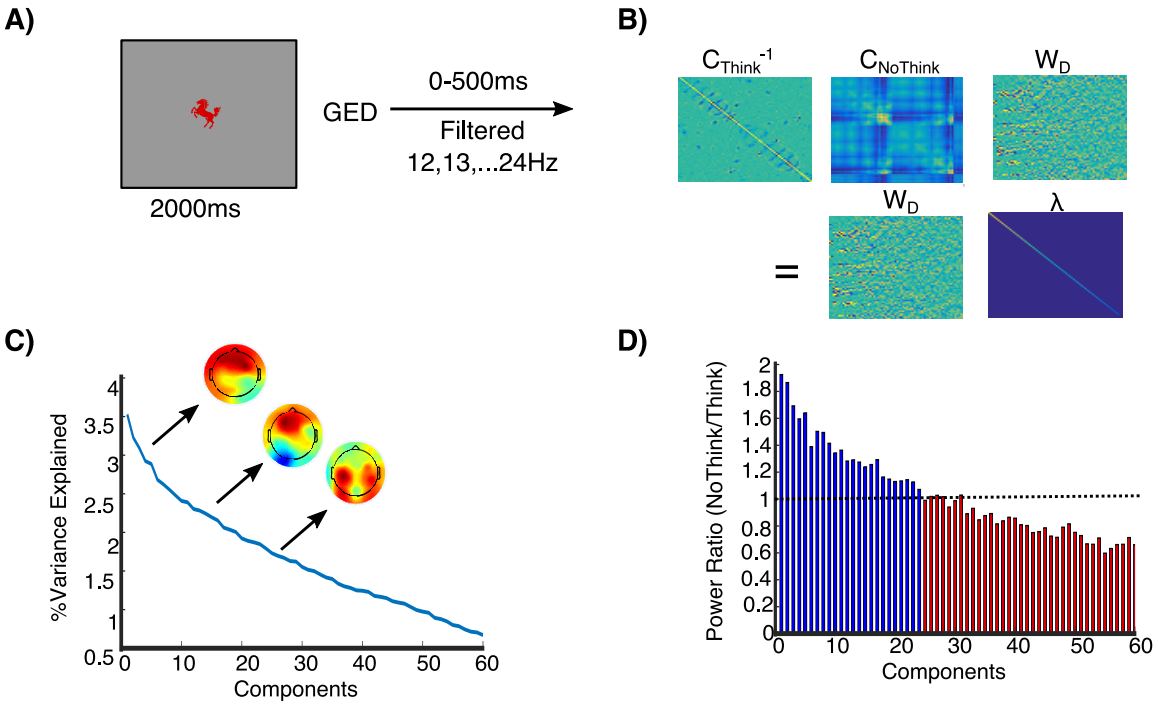

Figure 4.1 A) The time window for GED was from 0 to 500ms after memory cue onset. B) An example illustration of the GED implementation in one participant: GED was performed for NoThink versus Think trials with C^{-1}_{Think} and $C_{NoThink}$ representing the covariance matrices (0-500ms) for each condition. W_D denotes the eigenvector matrix corresponding to the matrix of eigenvalues (λ). C) The eigenvalue spectrum for the same participant with three example components. The component topography reflects the filter forward model (activation= $C_{NoThink}$ W_D). The first component in panel C shows an example of a right frontal topography. D) The power ratio between the mean power in NoThink vs. Think in the 0-500ms window. Only components that had a power ratio >1 were considered for selection.

Extracting Beta Bursts from the right frontal filter

Our procedure for identifying beta bursts, described below, again closely follows methods used in studies on action-stopping (Figure 4.2). After identifying the right frontal filter,

we projected the weights onto the epoched data. We then computed the event-related spectral perturbations (ERSP) of NoThink and Think trials. The NoThink ERSP was used to identify each participant's peak beta band, i.e. the the beta frequency that showed the greatest power increase in the window of interest (0-500ms post memory cue). We then used that peak frequency to extract the beta amplitude time series in each condition (Think and NoThink). This required first filtering the data at each participant's peak beta band using a frequency domain Gaussian window with full-width half-maximum of 5Hz. We then computed beta amplitude by using a Hilbert transform to derive the complex analytic time series and then taking the absolute of the analytic signal. The amplitude was computed for each timepoint within a trial and for the baseline period (i.e. the 500ms preceding the memory cue during fixation). We calculated both the median and standard deviation of the baseline period after pooling across all trial types (i.e. Think and NoThink). Bursts were defined as any period within a trial in which the beta amplitude exceeded the median baseline amplitude by + 1.5 SD. We identified each trial timepoint that exceeded this burst threshold. This analysis provided us with a measure of burst probability, i.e. the likelihood of a beta burst at each timepoint within 500ms of the memory cue. This time window was selected based on when right frontal beta is typically observed during action-stopping (Jana et al., 2020) and during retrieval suppression according to one study that used the standard TNT task (Castiglione et al., 2019). Thus, burst probability reflects the percentage of timepoints with beta bursts in the window of interest, i.e. the first 500ms of encountering the Think/NoThink memory cue, averaged across trials.

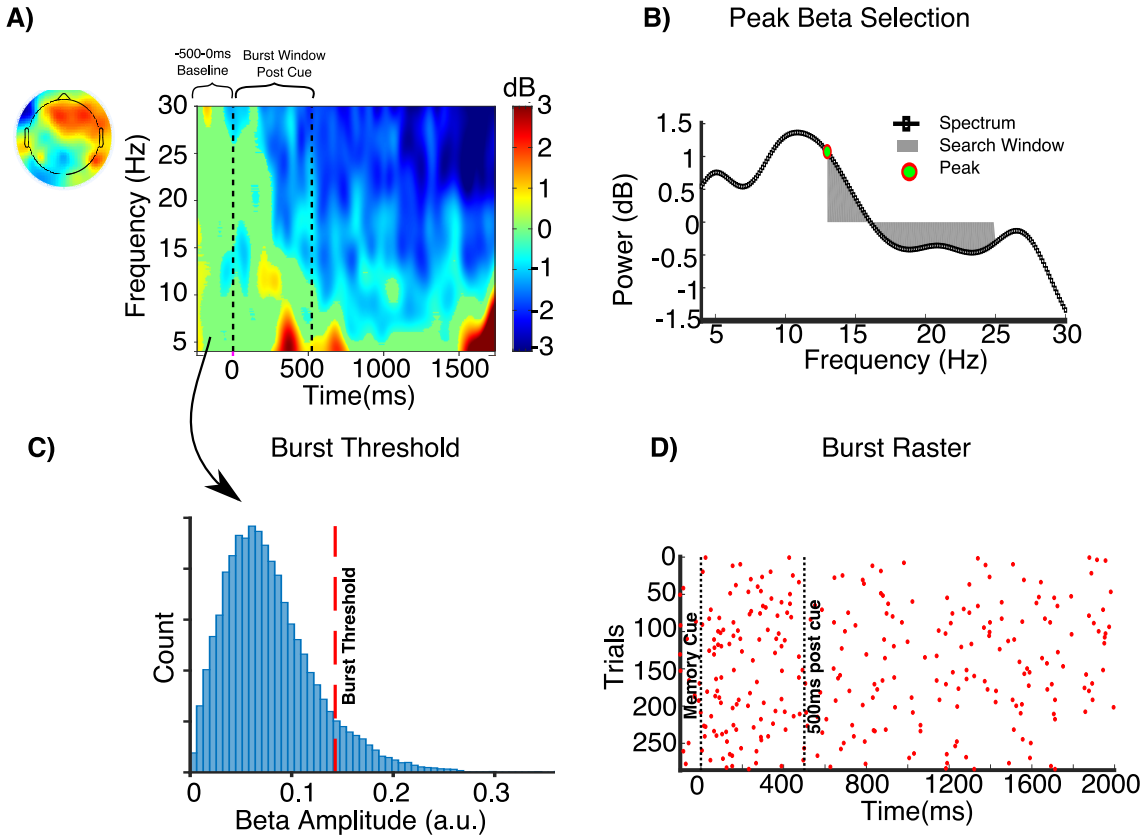

Figure 4.2. A) An example ERSP from one subject with the neural activity during NoThink trials after applying the right frontal filter. B) The ERSP was used to identify the subject specific beta band, i.e. the band that showed a peak increase in the window of interest (0-500ms post cue). The data were filtered at each participant's peak beta band prior to computing beta amplitude and extracting bursts. C) The histogram of beta amplitude for all trials in a participant during the baseline window (-500-0ms preceding the cue). This period was used to define the burst threshold, i.e. +1.5 SD of the median beta amplitude from for this baseline period. D) The burst raster for all NoThink trials across time in an example participant. Each red dot represents a burst in that trial. Burst probability was estimated as the % of timepoints with a burst within 500ms of the memory cue, averaged across trials.

The relationship between beta bursts and decoding accuracy

We next tested whether the probability of beta bursts within 500ms of the memory cue corresponded to reduced decoding accuracy for NoThink memories. We computed 1) the probability of beta bursts within 500ms of the NoThink memory cue and 2) decoding accuracy for NoThink memories averaged across the retrieval period that showed above-chance decoding for both conditions (i.e. 700ms-1920ms post memory cue). We conducted a Pearson's correlation to quantify the relationship between burst probability and NoThink decoding accuracy across participants.

The relationship between beta bursts and self-reported intrusions

We conducted the same analysis for self-reported measures of memory intrusions. Specifically, we performed a Pearson's correlation to examine the relationship between burst probability and the percentage of reported NoThink intrusions for each participant.

Results

The relationship between beta and decoding accuracy

We first tested the hypothesis that increases in right frontal beta bursts would relate to lower overall decoding of the NoThink memories. Consistent with this prediction, we found a significant relationship such that participants with a higher probability of beta bursts showed lower overall decoding accuracy for NoThink trials, which may reflect less intruding memory content, R=-.471, p=.042 (Figure 4.3C). To test whether this relationship was specific for NoThink memories, we also tested the correlation between beta burst probability on Think trials and decoding accuracy for Think memories. We did not find a significant relationship, R=0.147, p=0.548.

The relationship between beta and self-reported intrusions

We next tested the hypothesis that increases in right frontal beta bursts would relate to a lower frequency of self-reported intrusions. In contrast to this prediction, we did not find a significant relationship between the percentage of self reported intrusions and burst probability on NoThink trials, R=.428, p=.067 (Figure 4.3D).

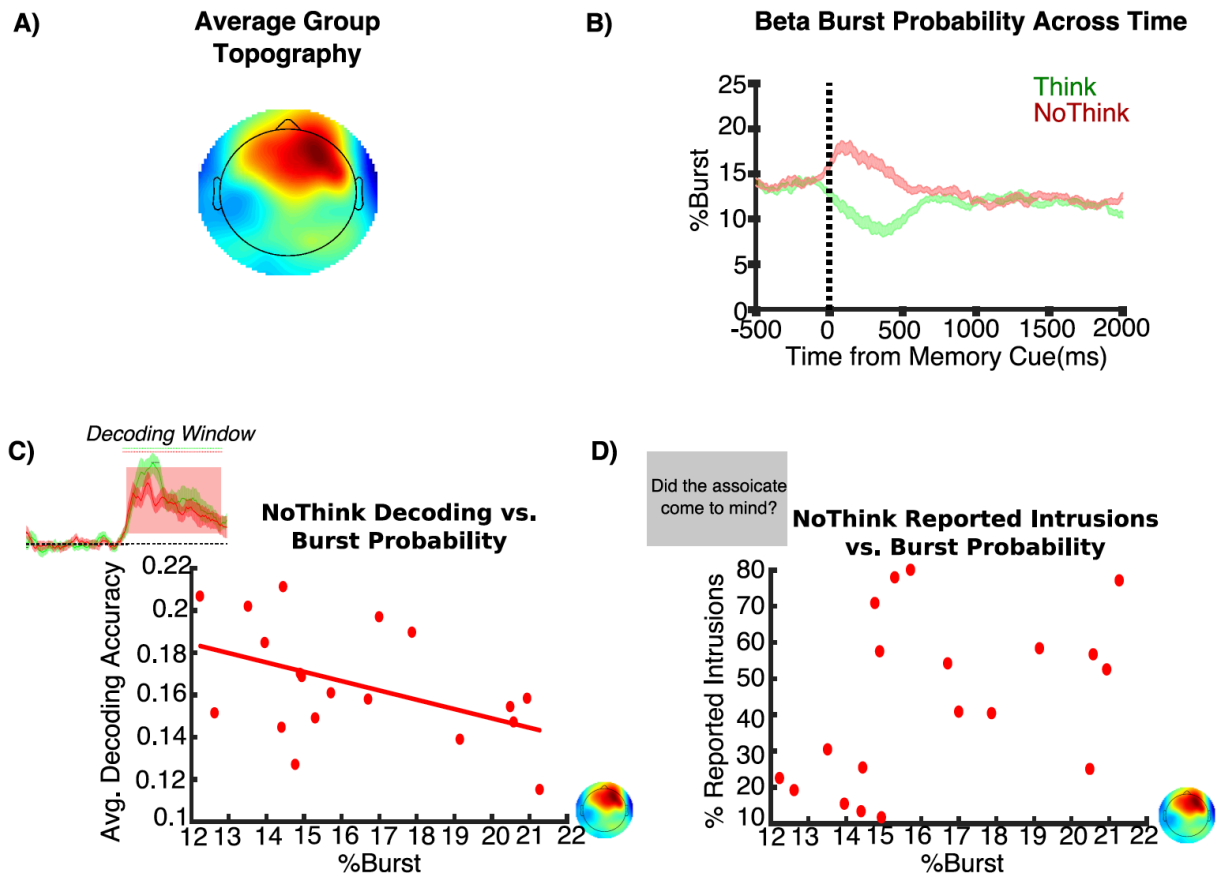

Figure 4.3. A) Average right frontal topography of the filters selected for each participant using GED. B) An illustration of the probability of right frontal beta bursts across time for Think (green) and NoThink (red) averaged across all participants. C) Participants with a higher probability of beta bursts showed reduced decoding accuracy (averaged across the 700-1920ms above-chance decoding window) for NoThink trials. D) We did not observe a reliable relationship between beta bursts and the frequency of self-reported intrusions.

Discussion

The current chapter examined the relationship between right frontal beta bursts and SVM decoding accuracy of NoThink memories. We found that participants with a greater probability of bursts, in the first 500ms of encountering a NoThink cue, showed reduced overall decoding accuracy for the NoThink memories (an objective measure of intrusions), although there was no such relationship with a behavioral measure of self-reported intrusions. According to prior work, we assume that right frontal beta bursts reflect the recruitment of the underlying right PFC network (Sundby et al., 2021). Together we interpret these results as evidence that early

recruitment of the right PFC network relates to one's ability to modulate retrieval and the presence of an intruding memory. These data are the first, to our knowledge, to reveal a relationship between an objective metric of the associate memory and the right PFC network during a TNT task. Thus, this study helps to validate the theory that recruitment of the right PFC network corresponds to changes in retrieval and less intruding information.

The current results are consistent with a number of fMRI studies that report right PFC activation during NoThink trials. These fMRI studies posit that the right PFC network acts to inhibit retrieval and any intruding information. This is inferred from the observation that right PFC activity corresponds to reductions in downstream hippocampal activity, fewer self-reported intrusions, and reduced memory in a final recall test (Anderson et al., 2004; Benoit & Anderson, 2012; Benoit et al., 2015; Brendan E Depue et al., 2007; Gagnepain et al., 2014; Schmitz et al., 2017). These data, however, provide only indirect or subjective evidence linking the right PFC to inhibitory control over LTM retrieval. By using SVM decoding, the current study contributes more specific evidence that recruitment of the right PFC network, reflected in beta bursts, corresponds to observable changes in the neural representation of the associate memory. Further, these changes in decoding were measured across the retrieval period (i.e. when participants must actively prevent retrieval), a critical window that is largely inaccessible with the existing methods (e.g. post trial self-report and final recall tests).

It is important to note, however, that the current analysis was exploratory and is limited in several ways. First, future work should adopt a dual task approach in which the participants perform both the spatial TNT task (or another version of TNT that is amenable to the SVM approach) and an action-stopping task. The action-stopping task could then help to localize the

right frontal filter and to validate that the selected filter shows the standard result for action-stopping prior to testing it in the TNT task.

Second, although we selected the filter according to the established features observed in action-stopping (i.e. right frontal topography, beta within 500ms of the cue etc.), we are unable to validate that the observed beta modulations reflect the right PFC network specifically and an active inhibitory process. Indeed, another line of research more broadly links changes in frontal beta (without a right specific topography) to the gating of working memory and decisions about which information to encode versus ignore (Lundqvist, Herman, Warden, Brincat, & Miller, 2018; Zavala, Jang, & Zaghloul, 2017). Thus, the relationship between right frontal beta bursts and decoding accuracy may instead reflect a broader function of beta oscillations relevant to controlling the contents of working memory but that are not necessarily specific to active inhibition of retrieval or the right PFC network (for review see, Schmidt et al., 2019). We did however, find that the relationship between beta bursts and decoding accuracy was unique to the NoThink condition. We speculate that if beta modulations were instead related to a non-specific gating function (in which, for example, decreased beta permits information and increased beta precludes information) then we might also expect beta bursts to relate to decoding accuracy for the Think condition. Specifically, a lower probability of beta bursts may have corresponded to improved decoding accuracy when actively retrieving on Think trials. The present data did not show this relationship for the Think condition.

Third, the current approach does not address the temporal relationship between beta bursts and the precise time at which a memory representation is "suppressed." This relationship is important for better understanding the functional relevance of right frontal beta and whether it is involved in either the prevention or clearing of memory intrusions. Interestingly, the

divergence in decoding accuracy between Think and NoThink conditions occurs far later (~1000ms) than the measured beta bursts. This raises interesting questions as to why very early modulations in beta relate to changes in NoThink decoding. We speculate that early modulations of beta may reflect an initial process aimed at controlling retrieval that may also prepare the system to respond to later intrusions. Whether or not this proposed secondary process is related to beta or the right PFC network is an open question.

Lastly, we did not find a reliable relationship between beta bursts and self-reported intrusions. This is inconsistent with prior work that found greater right frontal beta power for trials with no reported intrusion (Castiglione et al., 2019). The relationship between beta and self-reported intrusions should be investigated further. However, the fact that we observe a relationship to an "objective" metric of memory intrusions yet not subjective self-report may underscore the importance of using methods beyond self-report to more reliably determine whether or not memory information is retrieved during efforts to suppress.

In conclusion, by combining two novel tools, the current study provides important evidence linking a marker of the right PFC network to a measureable decline in the neural representation of the associate memory. This work demonstrates the utility of developing more objective methods for examining the proposed role of the right PFC network. Further, it provides support for the broader theory that individuals can voluntarily exert control over cued retrieval, which has profound implications for both several clinical populations and for our everyday experience.

GENERAL DISCUSSION

We encounter an abundance of cues in our daily lives that prompt actions and sometimes even memories. Controlling these intrusions, whether an inappropriate action or an unwanted thought, is essential for our productivity and mental well-being. A compelling theory in cognitive psychology asserts that, much like stopping actions, individuals are able to actively inhibit the retrieval of an unwanted memory and that this process relies on a shared right PFC network. Yet our ability to test theories of inhibitory control over LTM retrieval is limited in several critical ways: 1) the majority of neuroimaging studies use methods with limited temporal resolution and are unable to determine *when* the putative inhibitory network intervenes, and 2) memory intrusions are typically measured using post trial self-report rather than an objective metric. The studies in this dissertation sought to overcome these problems in order to develop a deeper understanding of executive control over cued retrieval. The discussion below reviews the results from Chapters 1-4 and examines how they contribute to the existing literature.

In Chapter 1, we addressed the first limitation, i.e. the lack of a temporally precise method for determining when the proposed inhibitory process interrupts retrieval. Specifically, we sought to validate right frontal beta oscillations as an EEG marker of the right PFC network and inhibitory control. Scalp EEG has a temporal resolution in the order of milliseconds, making it a valuable method for examining the timing of neural events like network recruitment. Evidence linking right frontal beta to the right PFC network comes primarily from studies on action-stopping that report a robust increase in right frontal beta when participants successfully cancel an action (Enz et al., 2021; Hannah et al., 2020; Schaum et al., 2021; Swann et al., 2009; Wagner et al., 2018; Wessel et al., 2016). To use this as a reliable signature of the right PFC network, however, requires first demonstrating that these modulations of beta depend on the

integrity of the underlying network. Thus, to causally test the relationship between beta and the right PFC network, we used rTMS to disrupt a key node in the network and to examine how this affected beta during an action-stopping task. Consistent with our hypothesis, disruption of the right PFC network resulted in reduced right frontal beta power during the time window that is critical for stopping. These results support right frontal beta as a reliable signature of the right PFC network. By validating this EEG signature in the more established domain of action stopping, future work can more confidentially use it as a tool to examine how and when the network intervenes during efforts to prevent retrieval.

A temporally precise marker of the network is particularly important for distinguishing between two theories that make distinct predictions about when the right PFC intervenes: the "gating" account which predicts early network recruitment in response to the memory cue and the "clearing" account which predicts later recruitment in response to an intruding memory. Although prior fMRI studies were instrumental in identifying the right PFC network and its potential role in controlling retrieval, these Chapter 1 results provide a method that is better suited for examining the temporal dynamics of network recruitment (Benoit & Anderson, 2012; Benoit et al., 2015; Brendan E Depue et al., 2016; Guo et al., 2018; Levy & Anderson, 2012).

In Chapter 2 we moved on to address the second major limitation, i.e. the lack of an objective method for tracking memory retrieval during efforts to suppress. In the case of action-stopping, the absence of a motor response provides overt evidence of successful inhibition. Evidence for active inhibition of a cognitive process like retrieval is far more difficult to establish. Prior work points to reductions in final recall, declines in self-reported intrusions, and reduced hippocampal activity as evidence of successful retrieval suppression (e.g., Anderson & Bell, 2001; Anderson et al., 2004; Benoit et al., 2015; Levy & Anderson, 2012). Yet none of

these methods measure the memory representation itself during the critical retrieval window, i.e. when an individual must implement control to avoid an intrusion.

Thus, in Chapter2, we use an event-related potential (ERP) to more reliably track the presence of intruding information during the NoThink instruction. Specifically, we tested whether the negative slow wave (NSW), an ERP that has been shown to track the number of items held in working memory, was reduced when participants sought to prevent the retrieval of a cued memory. We conducted this study using a novel version of the TNT task in which participants learned to associate a context in the form of a large colored background, with a specific object. Participants were then given a context as a memory cue and were instructed to either retrieve or prevent retrieval of the associated object. We found that NSW amplitude was reduced for NoThink items compared to Think items. Further, participants who reported fewer intrusions, had a smaller NSW amplitude for NoThink trials. We interpreted these results as evidence that participants were largely successful in preventing the associate from coming to mind when instructed to stop retrieval. The NSW approach, however, is limited in a number of ways. Based on prior literature, NSW amplitude appears to reflect the amount of information entering conscious working memory, but it does reflect the specific content of the items in working memory (Drew et al., 2006; Fukuda et al., 2015; Itthipuripat & Woodman, 2018). Due to this lack of memory-specific information, we are unable to fully dismiss other factors that could contribute to the observed reduction in amplitude (e.g. task difficulty). We addressed this limitation in the subsequent chapter.

In Chapter 3, we sought to expand on the NSW results by using multivariate methods to track the specific prohibited NoThink memories rather than the mere quantity of information retrieved during a Think/NoThink task. In a novel object-location version of the TNT task, we

used EEG-based decoding methods that have previously proven successful in tracking precise spatial memories (Bae & Luck, 2018). We discovered that we could reliably decode both Think and NoThink memories from neural activity during the retrieval period. Importantly, however, we observed a subsequent reduction (~1000ms post cue) in decoding accuracy for NoThink memories compared to intentional retrieval episodes (i.e. Think trials). We assume that the reduction in decoding accuracy reflects successful efforts to modulate retrieval.

Consistent with the NSW study, these data provide additional objective evidence that individuals are able to voluntarily modulate the retrieval of information. However, using this more specific metric, the decoding results reveal that, on average, modulation occurs after the initiation of retrieval. Accordingly, the decoding data appear to align with the "clearing" account in which control strategies are implemented in response to intruding information. This observation also fits with current explanations of suppression induced forgetting. Suppression induced forgetting refers to impaired memory for NoThink items and is thought to result from a weakening of the memory trace due to suppression (e.g., Anderson & Bell, 2001; Benoit & Anderson, 2012; Bergström et al., 2009b). Explanations for suppression induced forgetting point to reconsolidation theory and the notion that reactivating a memory pushes it into a labile state that can then be actively suppressed (Detre et al., 2013; Dudai, 2004; Nader et al., 2000). By this logic, an intrusion and partial reactivation of a memory is in fact necessary to produce final forgetting effects. The EEG decoding approach, presented in this chapter, may prove useful for testing this assertion. Specifically, future studies could expand our work by adding a final recall phase and testing whether NoThink decoding accuracy relates to the degree of final forgetting. Indeed, one study that used decoding methods applied to fMRI data suggests that the strength of reactivation may be a critical yet largely overlooked factor in the TNT literature. Specifically,

Detre et al. (2013) found that only moderate levels of memory activation were amenable to suppression and final forgetting. This may be due to the fact that particularly weak reactivation does not elicit an active inhibitory process while particularly strong reactivation may be resistant to inhibitory control. The decoding approach applied to EEG (i.e. rather than fMRI) has the added advantage of examining the quality of the neural representation at each timepoint and thus, could provide a more detailed understanding of how memory reactivation (e.g. the strength or duration of reactivation) relates to subsequent forgetting.

An overarching goal of this dissertation was to develop tools that, when combined, might provide deeper insight into whether and how the right PFC exerts control over retrieval. In the final Chapter 4, we provide one example of how these tools can be coupled to test whether recruitment of the PFC network corresponds to measureable changes in the neural representation of the associate memory. Specifically, we conducted an exploratory analysis to test whether increases in right frontal beta, presumably reflecting the right PFC network, relates to poorer decoding of the NoThink memories. Our analysis of beta focused, on an early time window (within 500ms of the memory cue) when right frontal beta has been shown to correspond to inhibitory control in the context of action-stopping (Castiglione et al., 2019; Hannah et al., 2020; Jana et al., 2020; Wagner et al., 2018). Indeed, we found that a greater probability of beta bursts on NoThink trials corresponded to lower overall decoding accuracy of the NoThink memory. We interpret this as preliminary evidence that early recruitment of the right PFC network relates to one's ability to modulate retrieval and the degree to which a memory intrudes.

A Hypothetical Two-Phase Model

Taken together, our data support a more complex view of executive control of LTM retrieval that likely involves both early and late phases of control. We first observed early

reductions in NSW amplitude (400-900ms), an ERP that is thought to reflect the amount of information present in working memory. A more specific measure of the memory information -- the EEG decoding– suggests, however, that some memory content is retrieved (on average) regardless of trial type and that modulation of the retrieved content occurs later during retrieval. Interestingly, the overall decoding accuracy across retrieval was related to very early modulations of right frontal beta bursts, a putative marker of inhibitory control via the right PFC network. The variable timing observed across these studies should be investigated further and could reflect a number of factors. We posit, however, that these results may in part reflect two phases of control and two types of "stopping".

Even in the far "simpler" context of action-stopping, the timing of control has been shown to vary. Electromyography (EMG) recordings from the muscle being stopped, reveal two types of successful stop trials: trials in which muscle activity is canceled before movement initiation and trials in which muscle activity begins and is subsequently shutdown (Jana et al., 2020). A similar two-phase model may apply to executive control over cued retrieval (Figure 5.1). Specifically, an early gate-like process may interject rapidly upon seeing a cue and act to weaken or, in some cases, even prevent retrieval of the associated memory. If this process fails or if the memory later reaches a conscious threshold, a secondary process may be required to cope with the intruding memory. This dual phase model also accommodates the extant literature that shows variable results as to whether the right PFC is recruited on trials with or without reported intrusions (Castiglione & Aron, 2021; Castiglione et al., 2019; Levy & Anderson, 2012).

The demand for each of these proposed processes may depend on a number of factors including the strength or saliency of the associate memory. Future studies could, for example,

examine whether the saliency of a memory determines whether or not a secondary phase is necessary. The tasks used in the NSW and SVM studies varied in several ways that may affect the frequency of intruding information. It may be, for example, that spatial memories are particularly prepotent (e.g. more so than objects) and produce a higher rate of intrusions that require later subsequent control. This type of circumstance is likely more akin to what we experience with intrusive memories in the real world which tend to be highly salient thoughts.

A few studies propose similar two-phase hypotheses (Castiglione & Aron, 2021; Brendan E Depue et al., 2007). In all of these cases, there are several outstanding questions about the specifics of the two-phase model. Here we focus on just two. First, while our work provides a more objective method for tracking memories and thus, a tool for examining whether the content is reduced, more work is needed to confirm that changes in the neural representation are the result of an active inhibitory process (see below Combing Tools and Future Applications). Indeed, other strategies including diverting attention away from the intruding memory, could also reduce decoding accuracy for the ignored item. Second, future studies should examine whether each phase or type of "stopping" is implemented by distinct nodes within the right PFC network. The majority of our work focuses on the right inferior frontal gyrus (rIFG), a locus in the right PFC network that is more tightly linked to the right frontal beta signature and early control processes. Retrieval suppression, however, recruits several regions of the right lateral PFC, including the right medial frontal gyrus (rMFG), that may uniquely contribute to executive control over LTM.

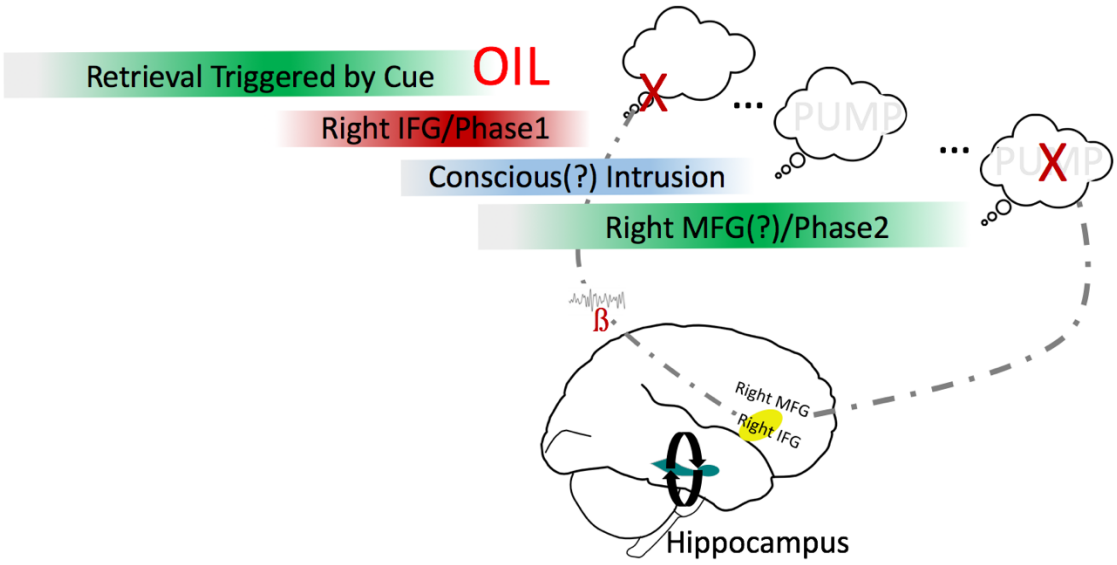

Figure 5.1 A hypothetical two-phase model of executive control over LTM retrieval. An early process intervenes rapidly to prevent or weaken the retrieved associate. This early process is likely mediated by right frontal beta and the rIFG. A secondary process may often be required to cope with later intrusions. This process may rely on other nodes in the right PFC network, including regions like the right MFG.

Stopping Cued Retrieval in the World Outside the Laboratory

An important question is whether retrieval suppression is an effective strategy for avoiding unwanted memories outside of a laboratory setting. Thought substitution (i.e. diverting attention to a different thought), may be more ecologically valid and it produces similar benefits, including impaired recall for the avoided item (Hertel & Calcaterra, 2005). Further, retrieval suppression has also been shown to be less effective when the situation requires suppressing a memory across an extended period of time (e.g. 5 seconds vs. 2.5 seconds) (van Schie & Anderson, 2017). This presents a problem for everyday living where memory cues are likely to persist longer than a few short seconds.

This raises the question as to why a suppression strategy would be used at all. We posit that suppression may be most useful when paired with thought substitution. Specifically, enacting a rapid "stop" may facilitate the transition to a substitute thought. This hypothesis is motivated by studies that use the stop-change paradigm. The stop-change paradigm requires

participants to stop one motor response and switch to a new motor response. Several studies find that inhibiting the original response is important for reducing interference and efficiently changing to the new response (for review see, Boecker, Gauggel, & Drueke, 2013). A similar principle may apply to memory retrieval, in which actively stopping the unwanted memory may enable a more efficient switch to a new thought. This idea is also consistent with more standard switching tasks in which lingering activity related to a prior task set interferes with performance of a new task set (Yeung, Nystrom, Aronson, & Cohen, 2006). Importantly, this dissertation provides tools that would allow future studies to test this hypothesis. For instance, by using the SVM decoding approach, one could test whether increased right frontal beta precedes thought substitution and whether the magnitude or timing of beta relates to how quickly an individual switches. If carefully designed, this type of study could track both the prohibited NoThink memory *and* the substituted item.

It is also unclear how individuals know to apply inhibitory control upon encountering a memory cue outside of the lab. This point is particularly relevant to the "gating" account and the notion that retrieval can be canceled proactively in response to an associated cue. In the lab setting, the memory cue is paired with an explicit instruction to prevent retrieval (i.e. the word written in red). Yet, in the real world, the memory cue itself does not signal any need to stop retrieval. That is, an innocuous cue for one individual may trigger a harmful memory for another. One possibility is that individuals can learn to associate a cue with a need to prevent retrieval. A number of studies show that, in the case of action-stopping, inhibition via the right PFC network can become automatic when people learn to associate certain stimuli with stopping (Lenartowicz, Verbruggen, Logan, & Poldrack, 2011; Verbruggen & Logan, 2008). This raises interesting

questions as to whether individuals could learn to suppress retrieval in response to a provocative memory cue without an explicit NoThink instruction.

In summary, the Think/NoThink task, like many paradigms, suffers from the scalability problem (Badre, 2020). The task creates a simplified and controlled version of a very complex real world problem (i.e. the need to prevent retrieval when provoked by a reminder). This simplification allows us to study complex questions in a controlled lab setting yet may fail to accurately capture the nuances of our day-to-day experiences. Although future work should certainly develop more realistic models of cued memory suppression, it is important to note that the Think/NoThink task may still provide valuable insights. For instance, with the proper tools, the task can help to address a basic cognitive neuroscience question: does the right PFC network serve a "supramodal" inhibitory function or is it specific to stopping actions? Additionally, if further evidence demonstrates that recruiting the right PFC network helps to preclude unwanted memory information, people may be able to receive training to effectively use this strategy even if it is not the default in the real world. If training proves successful, this could have important implications for individuals who are particularly susceptible to intrusive thoughts.

Combining Tools and Future Applications

The utility of an EEG signature of the right PFC network extends beyond examining correlations between the timing of network recruitment and evidence of "suppression" in the neural signal during retrieval. An EEG signature may also allow for future studies to causally test the relationship between right frontal beta and changes in memory representations. Specifically, TMS could be used to entrain right frontal beta oscillations and test how this affects an intruding memory representation. This would be a powerful demonstration that recruitment of the right PFC network may actively suppress the intruding memory trace.

Conclusion

The experiments detailed in this dissertation contribute several tools that are necessary for gaining a deeper understanding of executive control over cued retrieval. Specifically, these studies provide a high temporal resolution marker of the candidate right PFC network and methods for objectively and continuously tracking the reactivation of memories during efforts to suppress. Using these metrics, we find evidence of both early and late modulations of memory retrieval. We suggest that this may warrant a two-phase model of control including both an early gate-like process and a later process that may function to clear intruding information. Our hope is that these tools lay the groundwork for future studies to rigorously test these models of executive control over LTM and to continue to make progress towards understanding a truly profound problem in cognitive and clinical psychology: How can we prevent the intrusion of unwanted memories and maintain control of our thoughts?

Table 5.1 Summary of Chapters

Chapter	Task	Methods	Aim Addressed	Key Takeaway
1	Stop Signal	• fMRI • rTMS • EEG • EMG	Temporally precise marker of right PFC network	• Right frontal beta provides marker of right PFC network
2	Object-Context TNT	• EEG • NSW	Objective metric of memory	• Reduced *amount* of intruding information during NoThink • Aligns with "gating" account
3	Object-Location TNT	• EEG • SVM decoding	Objective *(specific & continuous)* metric of memory	• Retrieval initiated regardless of condition & NoThink modulation occurs later • Aligns with "clearing" account
4	Object-Location TNT	• EEG • SVM decoding • Beta Bursts	Relationship between right PFC network & objective metric of memory	• Early recruitment of a right PFC network (within 500ms of cue) relates to poorer neural representation of NoThink memory

Table 5.2 Questions for future research

Questions for future research
Are the observed modulations in the neural representation of memories indicative of active suppression or an alternative strategy?
Are both proposed phases of control (early and late) reliant on the right PFC network and active inhibition? Or do distinct network nodes control different aspects?
Does the strength of an associated memory determine how susceptible it is to executive control?

APPENDIX

Inverted Encoding Model (IEM) Approach

We initially applied an inverted encoding model to the EEG data collected during the object-location version of TNT (Chapter 3 data). Inverted encoding models (IEMs) are often used to reconstruct stimulus information (e.g. location, color, orientation, etc.) from patterns of neural activity. The below analysis was motivated by a recent study that used an IEM applied to alpha power to track the retrieval of precise spatial memories from LTM (Sutterer et al., 2019). We applied these methods to the object-location version of TNT to estimate spatially selective channel tuning functions (CTFs) from whole brain alpha topography for 1) Phase1 Learning data, i.e. when actively retrieving on each trial and 2) Phase2 TNT data, when trying to either retrieve (Think) or prevent retrieval (NoThink) of a spatial memory.

Analysis

1) Phase1 Training/Phase1 Testing

As a validation of our approach, we first conducted the IEM analysis on Phase1 data only (split into separate training and testing datasets). Phase1 was part of learning and instructed participants to actively retrieve the cued spatial memory on every trial, a procedure that closely matched the task used in prior work (Sutterer et al., 2019). Consistent with Sutterer et al. (2019), the IEM was applied to the multivariate distribution of alpha power across all electrodes. We modeled the alpha power at each electrode as the weighted sum of eight hypothetical channels (i.e. neuronal populations), each tuned to a unique angular location. The response profile of each channel was modeled as a half sinusoid raised to the seventh power (i.e. the basis set). We circularly shifted the response profile so that each channel's peak response was centered over one of the eight location bins. This design assumed that each of our hypothetical channels was

tuned to optimally respond to one of the eight location bins. The location bins were centered at 0°, 45°, 90° etc.

The Phase1 data were first partitioned into three independent sets that then served as training or testing datasets (see Trial assignments for more details). During step1(training), we used the neural activity from a training set (B_1) to solve for a set of weights (W). The weights approximate the degree to which each of the eight channels contributes to the power at each electrode. A general linear model was used to describe the relationship between the training data (B_1), the predicted channel responses to each training stimulus (C_1 estimated from the basis functions) and the weights (W):

$$B_1=WC_1$$

We computed the weight matrix using a least-squares estimation procedure:

$$\widehat{W}=B_1C_1^T(C_1C_1^T)^{-1}$$

During the testing stage, we then inverted the model and used the weight-matrix to estimate the channel responses (C_2) for the test data (B_2).

$$\hat{C}_2=(\widehat{W}^T\widehat{W})^{-1}\widehat{W}^TB_2$$

This provided an 8-point channel response function for each location bin. The channel response functions were then circularly shifted such that the channel matching the remembered location was situated in the center of the tuning curve. This aligned each channel response profile to a common center (i.e. 0 degrees). We then computed the average channel tuning function across the eight remembered locations. The above IEM routine (training and testing) was applied to every timepoint to produce CTFs for each point during retrieval.

Phase1 Training/Phase2 Testing

Our core analysis aimed to use an IEM to track the retrieval of spatial memories during the Think/NoThink manipulation in Phase2. We applied the same methods as described above. The only difference was that we now trained the model using only Phase1 data and tested the model using only Phase2 Think/NoThink data. Thus, we computed an average CTF at each timepoint for each condition (Think and NoThink) to examine how the retrieval of memory information changed when trying to prevent retrieval vs. actively retrieve. We used Phase1 data for training because it was important to train the model with a dataset that included neural activity when participants were actively retrieving each spatial memory.

Trial Assignment

Phase1 Training/Phase1 Testing

For the Phase1 validation analysis, we divided the Phase1 data into three independent sets to use as either training or testing data in the IEM routine. During trial assignment, we ensured that each set contained an equal number of trials per location bin. Within each set we averaged across trials to compute the average alpha power at each electrode for each spatial bin. Thus, each set contained one average scalp topography of alpha power for each of the eight spatial bins. We used a cross-validation procedure in which two sets were used for training and the remaining set was used for testing. We repeated the procedure three times such that each set served as the testing set once. The resulting CTFs were then averaged across each test set, resulting in an averaged CTF for each timepoint.

Phase1 Training/Phase 2 Testing

To monitor the retrieval of memories during the Think/NoThink phase, we used the Phase1 data for training and the Phase2 data for testing. We partitioned the Phase1 data into two independent training sets and the Phase2 data into two independent testing sets. The number of

trials for each location bin and each condition were balanced across both training sets and across both testing sets. The data were then averaged across trials for each spatial bin in each set resulting in one average scalp topography of alpha power for each of the eight spatial bins in each set. We repeated the procedure four times such that each training set was used to train the model with each testing set. The CTFs were computed separately for the Think and NoThink conditions. We averaged the CTFs across each test set, resulting in an averaged CTF for Think and an averaged CTF for NoThink at each timepoint.

Resampling

Phase1 Training/Phase1 Testing

We repeated the Phase1 validation analysis (training and testing with only Phase1 data) 10 times with a different random assignment of trials to account for spurious results that could arise from idiosyncrasies during trial assignments.

Phase1 Training/Phase2 Testing

When training with Phase1 data and testing with Phase2, we again conducted the full analysis 10 times to avoid spurious results due to trial assignments.

Calculating CTF Selectivity and Identifying significant timepoints

We quantified spatial selectivity (i.e. the quality of the retrieved memory) by calculating the slope of the CTF at each timepoint. A larger CTF slope reflects greater spatial selectivity. Thus, accurately retrieving a specific spatial memory should yield a higher CTF slope. To calculate the CTF slope, we averaged the response from channels of equidistance (e.g. +/- 2 bins) and used linear regression. To identify the timepoints at which the CTF slopes were significantly above chance, we applied a nonparametric cluster approach that corrects for multiple comparisons (for detailed description see, Sutterer et al., 2019).

Results

IEM

Phase1 Training/Phase1 Testing

We observed reliable reconstructions of spatial memories during Phase1 when participants were prompted to retrieve the remembered location on each trial. CTF slopes were reliably above chance in the periods 572ms-916ms and 1364-1578ms post memory cue (FigureA.1 A). This confirmed that we were able to use an IEM applied to alpha power to track the retrieval of spatial memories during a task Phase in which participants were retrieving on every trial. We also observed a negative correlation such that participants who were more accurate in their remembered response (according to post trial test performance) had a higher CTF slope i.e. greater spatial selectivity in their neural data, R=-.628, p=.004. (Figure A.1 B).

Phase1 Training/Phase2 Testing

Our primary prediction was that we would observe reliable CTF slopes for Think trials, reflecting accurate retrieval, and that CTF slopes would be reduced for NoThink trials when trying to prevent retrieval. The data showed periods in which NoThink slopes were reduced below Think (710ms-1038ms, 1126ms-1384ms, 1640ms-1730ms). There were not, however, any time periods in which either Think or NoThink slopes were reliably above chance (Figure A.1 C). Thus, although these data show modulations of NoThink memory representations relative to Think, it is unclear how to interpret this result since we failed to produce reliable memory reconstructions during Think trials, i.e. trials in which participants were instructed to retrieve.

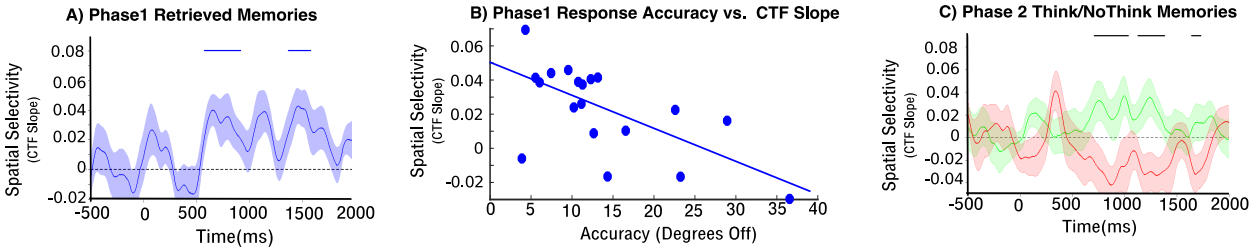

Figure A.1. A) We observed reliable reconstructions of retrieved spatial memories during Phase1. The blue bars indicate significant CTF slopes. B) Participants who were more accurate in their remembered responses in Phase1 also showed higher quality reconstructions of each spatial memory (i.e. larger CTF slopes). C) We were unable to reliably reconstruct spatial memories above chance for either Think or NoThink trials in Phase2. The black bars indicate periods with a significant difference between Think and NoThink CTF slopes.

Summary of results and discussion

We used an IEM applied to alpha power to test whether we could reliably track the retrieval of spatial memories in a novel object-location version of the Think/NoThink task. We first showed that we could replicate past work and reconstruct retrieved spatial memories during Phase1 of the task, when participants were actively trying to retrieve on each trial. Additionally, participants that showed greater spatial selectivity in the neural data during Phase1, performed better on the memory tests. This suggests that the memory information reflected in the CTF slope was related to memory accuracy.

Our primary goal, however, was to use the IEM to examine the retrieval of memories during the Think/NoThink phase. To do this we trained the IEM with Phase1 data (i.e. retrieving on each trial) and tested the model with Phase2 data (i.e. retrieving on Think trials and "stopping" retrieval on NoThink trials). Our core question was whether an IEM approach would provide more objective evidence that memory representations were precluded or suppressed as participants sought to avoid retrieval. We did observe a reduction in memory reconstructions for NoThink trials compared to Think. However, it is unclear how to interpret this result because we failed to produce reliable reconstructions of the spatial memories on Think trials, i.e. when participants were instructed to retrieve.

Because our primary goal for this chapter was to find methods that could objectively track the emergence of an associate memory during the TNT phase, we next tried a different multivariate approach. As described in Chapter3, we applied an EEG-based decoding method (i.e. SVM decoding) to the raw voltage in the EEG signal rather than alpha power. This method was applied to the same dataset and thus, shared several of the qualities that could be contributing to poor IEM reconstructions (e.g. increased task difficulty in Phase2).

It is important to note that we did not observe consistent results using the IEM and SVM decoding methods. We propose two potential explanations for why these methods may yield different results when using the TNT task. First, we speculate that neural representations may become nosier even when information is accurately retrieved in Phase2 (due to increased task difficulty, switching between retrieving and not retrieving etc.) and that different types of measures may suffer more as a result of the added noise. CTF slope, the dependent measure from the IEM, is a very precise metric that quantifies the quality of a stimulus representation. As a result of its precision, a measure like CTF slope may be more affected by noisy neural representations compared to classification models that are forced to assign a label that is either correct or incorrect.

Another, perhaps more likely, possibility is that raw voltage captures aspects of memory representations that are not specific to the alpha band. Maps of the ERP topography from the Phase1 data that were used to train the SVMs appear to show a negative voltage over anterior electrodes which may relate to the negative slow wave, an ERP that we have previously shown to be modulated during the Think/NoThink task (see Figure 3.3C). It remains unclear, however, why alpha oscillations, a frequency that is consistently linked to spatial information, did not reliably track spatial memories during Think trials (Foster, Sutterer, Serences, Vogel, & Awh,

2016; Foster et al., 2017; Stokes, Atherton, Patai, & Nobre, 2012). It is worth noting, that alpha oscillations have not been studied in the context of the Think/NoThink task. We speculate that these results may relate to an ongoing debate about whether alpha oscillations track the specific memory representation or covert shifts in attention to a remembered location (Carlisle & Woodman, 2011; Foster et al., 2017; Hakim, Adam, Gunseli, Awh, & Vogel, 2019; Wang, Rajsic, & Woodman, 2019). If tracking spatial attention, it is possible that attention is not automatically deployed to the remembered location in the TNT context, i.e. when participants are required to switch between retrieving and preventing retrieval. Though currently purely speculative, restraining attentional shifts may be an adaptive strategy for a setting in which you must frequently try to prevent the intrusion of a cued memory. Future work could address these questions by manipulating the percentage of NoThink trials and testing whether attention is more readily pulled to remembered locations in task settings that require less cognitive control over memories.

References

Anderson, M. C. (2005). The role of inhibitory control in forgetting unwanted memories: A consideration of three methods. *Dynamic cognitive processes*, 159-190.

Anderson, M. C., & Bell, T. (2001). Forgetting our facts: the role of inhibitory processes in the loss of propositional knowledge. *Journal of Experimental Psychology: General, 130*(3), 544.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(5), 1063.

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.

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

Anderson, M. C., & Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. *Trends in cognitive sciences, 18*(6), 279-292.

Anderson, M. C., & Hulbert, J. C. (2021). Active forgetting: Adaptation of memory by prefrontal control. *Annual Review of Psychology*.

Anderson, M. C., Ochsner, K. N., Kuhl, B., Cooper, J., Robertson, E., Gabrieli, S. W., . . . Gabrieli, J. D. (2004). Neural systems underlying the suppression of unwanted memories. *Science, 303*(5655), 232-235.

Apšvalka, D., Ferreira, C. S., Schmitz, T. W., Rowe, J. B., & Anderson, M. C. (2020). Dynamic targeting enables domain-general inhibitory control over action and thought by the prefrontal cortex. *bioRxiv*.

Aron, A. R. (2007). The neural basis of inhibition in cognitive control. *The neuroscientist, 13*(3), 214-228.

Aron, A. R. (2011). From reactive to proactive and selective control: developing a richer model for stopping inappropriate responses. *Biological psychiatry, 69*(12), e55-e68.

Aron, A. R., Behrens, T. E., Smith, S., Frank, M. J., & Poldrack, R. A. (2007). Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *Journal of Neuroscience, 27*(14), 3743-3752.

Aron, A. R., Fletcher, P. C., Bullmore, E. T., Sahakian, B. J., & Robbins, T. W. (2003). Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nature neuroscience, 6*(2), 115-116.

Aron, A. R., & Poldrack, R. A. (2006). Cortical and subcortical contributions to stop signal response inhibition: role of the subthalamic nucleus. *Journal of Neuroscience, 26*(9), 2424-2433.

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.

Badre, D. (2020). What Lies in the Gap between Knowledge and Action? *On Task* (pp. 1-31): Princeton University Press.

Badry, R., Mima, T., Aso, T., Nakatsuka, M., Abe, M., Fathi, D., . . . Fukuyama, H. (2009). Suppression of human cortico-motoneuronal excitability during the Stop-signal task. *Clinical Neurophysiology, 120*(9), 1717-1723.

Bae, G.-Y., & Luck, S. J. (2018). Dissociable decoding of spatial attention and working memory from EEG oscillations and sustained potentials. *Journal of Neuroscience, 38*(2), 409-422.

Bari, A., & Robbins, T. W. (2013). Inhibition and impulsivity: behavioral and neural basis of response control. *Progress in neurobiology, 108*, 44-79.

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(3), 603.

Benoit, R. G., & Anderson, M. C. (2012). Opposing mechanisms support the voluntary forgetting of unwanted memories. *Neuron, 76*(2), 450-460.

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

Bergström, Z. M., de Fockert, J., & Richardson-Klavehn, A. (2009a). Event-related potential evidence that automatic recollection can be voluntarily avoided. *Journal of Cognitive Neuroscience, 21*(7), 1280-1301.

Bergström, Z. M., de Fockert, J. W., & Richardson-Klavehn, A. (2009b). ERP and behavioural evidence for direct suppression of unwanted memories. *NeuroImage, 48*(4), 726-737.

Bergström, Z. M., Velmans, M., de Fockert, J., & Richardson-Klavehn, A. (2007). ERP evidence for successful voluntary avoidance of conscious recollection. *Brain research, 1151*, 119-133.

Boecker, M., Gauggel, S., & Drueke, B. (2013). Stop or stop-change—Does it make any difference for the inhibition process? *International Journal of Psychophysiology, 87*(3), 234-243.

Bulevich, J. B., Roediger, H. L., Balota, D. A., & Butler, A. C. (2006). Failures to find suppression of episodic memories in the think/no-think paradigm. *Memory & cognition, 34*(8), 1569-1577.

Cai, W., Oldenkamp, C. L., & Aron, A. R. (2012). Stopping speech suppresses the task-irrelevant hand. *Brain and language, 120*(3), 412-415.

Cai, W., Ryali, S., Chen, T., Li, C.-S. R., & Menon, V. (2014). Dissociable roles of right inferior frontal cortex and anterior insula in inhibitory control: evidence from intrinsic and task-related functional parcellation, connectivity, and response profile analyses across multiple datasets. *Journal of Neuroscience, 34*(44), 14652-14667.

Carlisle, N. B., & Woodman, G. F. (2011). Automatic and strategic effects in the guidance of attention by working memory representations. *Acta psychologica, 137*(2), 217-225.

Castiglione, A., & Aron, A. R. (2021). Unwanted Memory Intrusions Recruit Broad Motor Suppression. *Journal of cognitive neuroscience, 33*(1), 119-128.

Castiglione, A., Wagner, J., Anderson, M., & Aron, A. R. (2019). Preventing a thought from coming to mind elicits increased right frontal beta just as stopping action does. *Cerebral Cortex, 29*(5), 2160-2172.

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

Chambers, C. D., Bellgrove, M. A., Stokes, M. G., Henderson, T. R., Garavan, H., Robertson, I. H., . . . Mattingley, J. B. (2006). Executive "brake failure" following deactivation of human frontal lobe. *Journal of cognitive neuroscience, 18*(3), 444-455.

Chang, C., Crottaz-Herbette, S., & Menon, V. (2007). Temporal dynamics of basal ganglia response and connectivity during verbal working memory. *NeuroImage, 34*(3), 1253-1269.

Chatham, C. H., & Badre, D. (2015). Multiple gates on working memory. *Current opinion in behavioral sciences, 1*, 23-31.

Chen, C., Liu, C., Huang, R., Cheng, D., Wu, H., Xu, P., . . . Luo, Y.-J. (2012). Suppression of aversive memories associates with changes in early and late stages of neurocognitive processing. *Neuropsychologia, 50*(12), 2839-2848.

Chen, W., de Hemptinne, C., Miller, A. M., Leibbrand, M., Little, S. J., Lim, D. A., . . . Starr, P. A. (2020). Prefrontal-subthalamic hyperdirect pathway modulates movement inhibition in humans. *Neuron, 106*(4), 579-588. e573.

Cinel, C., Cortis Mack, C., & Ward, G. (2018). Towards augmented human memory: Retrieval-induced forgetting and retrieval practice in an interactive, end-of-day review. *Journal of Experimental Psychology: General, 147*(5), 632.

Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(6), 1403.

Cohen, M. X. (2017). Comparison of linear spatial filters for identifying oscillatory activity in multichannel data. *Journal of neuroscience methods, 278*, 1-12.

Cohen, M. X. (2021). A data-driven method to identify frequency boundaries in multichannel electrophysiology data. *Journal of neuroscience methods, 347*, 108949.

Coxon, J. P., Stinear, C. M., & Byblow, W. D. (2006). Intracortical inhibition during volitional inhibition of prepared action. *Journal of neurophysiology, 95*(6), 3371-3383.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods, 134*(1), 9-21.

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.

Depue, B. E., Burgess, G. C., Willcutt, E. G., Ruzic, L., & Banich, M. T. (2010). Inhibitory control of memory retrieval and motor processing associated with the right lateral prefrontal cortex: evidence from deficits in individuals with ADHD. *Neuropsychologia, 48*(13), 3909-3917.

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.

Depue, B. E., Ketz, N., Mollison, M. V., Nyhus, E., Banich, M. T., & Curran, T. (2013). ERPs and neural oscillations during volitional suppression of memory retrieval. *Journal of cognitive neuroscience, 25*(10), 1624-1633.

Depue, B. E., Orr, J., Smolker, H., Naaz, F., & Banich, M. (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.

Detre, G. J., Natarajan, A., Gershman, S. J., & Norman, K. A. (2013). Moderate levels of activation lead to forgetting in the think/no-think paradigm. *Neuropsychologia, 51*(12), 2371-2388.

Dietterich, T. G., & Bakiri, G. (1994). Solving multiclass learning problems via error-correcting output codes. *Journal of artificial intelligence research, 2*, 263-286.

Drew, T. W., McCollough, A. W., & Vogel, E. K. (2006). Event-related potential measures of visual working memory. *Clinical EEG and Neuroscience, 37*(4), 286-291.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annu. Rev. Psychol., 55*, 51-86.

Duque, J., Lew, D., Mazzocchio, R., Olivier, E., & Ivry, R. B. (2010). Evidence for two concurrent inhibitory mechanisms during response preparation. *Journal of Neuroscience, 30*(10), 3793-3802.

Enz, N., Ruddy, K. L., Rueda-Delgado, L. M., & Whelan, R. (2021). Volume of beta bursts, but not their rate, predicts successful response inhibition. *Journal of Neuroscience*.

Errington, S. P., Woodman, G. F., & Schall, J. D. (2020). Dissociation of medial frontal β-bursts and executive control. *Journal of Neuroscience, 40*(48), 9272-9282.

Fawcett, J. M., Benoit, R. G., Gagnepain, P., Salman, A., Bartholdy, S., Bradley, C., . . . Anderson, M. C. (2015). The origins of repetitive thought in rumination: Separating cognitive style from deficits in inhibitory control over memory. *Journal of Behavior Therapy and Experimental Psychiatry, 47*, 1-8.

Foster, J. J., Sutterer, D. W., Serences, J. T., Vogel, E. K., & Awh, E. (2016). The topography of alpha-band activity tracks the content of spatial working memory. *Journal of neurophysiology, 115*(1), 168-177.

Foster, J. J., Sutterer, D. W., Serences, J. T., Vogel, E. K., & Awh, E. (2017). Alpha-band oscillations enable spatially and temporally resolved tracking of covert spatial attention. *Psychological science, 28*(7), 929-941.

Friedman, D., & Johnson Jr, R. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. *Microscopy research and technique, 51*(1), 6-28.

Fukuda, K., Mance, I., & Vogel, E. K. (2015). α power modulation and event-related slow wave provide dissociable correlates of visual working memory. *Journal of Neuroscience, 35*(41), 14009-14016.

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

Gagnepain, P., Hulbert, J., & Anderson, M. C. (2017). Parallel regulation of memory and emotion supports the suppression of intrusive memories. *Journal of Neuroscience, 37*(27), 6423-6441.

Garavan, H., Ross, T., & Stein, E. (1999). Right hemispheric dominance of inhibitory control: an event-related functional MRI study. *Proceedings of the National Academy of Sciences, 96*(14), 8301-8306.

Glynn, R., Salmon, K., & Low, J. (2019). Short-and longer-term effects of selective discussion of adolescents' autobiographical memories. *Journal of experimental child psychology, 184*, 232-240.

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

Hakim, N., Adam, K. C., Gunseli, E., Awh, E., & Vogel, E. K. (2019). Dissecting the neural focus of attention reveals distinct processes for spatial attention and object-based storage in visual working memory. *Psychological science, 30*(4), 526-540.

Hannah, R., Muralidharan, V., Sundby, K. K., & Aron, A. R. (2020). Temporally-precise disruption of prefrontal cortex informed by the timing of beta bursts impairs human action-stopping. *NeuroImage, 222*, 117222.

Hanslmayr, S., Leipold, P., Pastötter, B., & Bäuml, K.-H. (2009). Anticipatory signatures of voluntary memory suppression. *Journal of Neuroscience, 29*(9), 2742-2747.

Harrington, M. O., Ashton, J. E., Sankarasubramanian, S., Anderson, M. C., & Cairney, S. A. (2021). Losing control: Sleep deprivation impairs the suppression of unwanted thoughts. *Clinical Psychological Science, 9*(1), 97-113.

Hasher, L., Lustig, C., & Zacks, R. (2007). Inhibitory mechanisms and the control of attention.

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

Hertel, P. T., & Calcaterra, G. (2005). Intentional forgetting benefits from thought substitution. *Psychonomic Bulletin & Review, 12*(3), 484-489.

Hertel, P. T., & Gerstle, M. (2003). Depressive deficits in forgetting. *Psychological science, 14*(6), 573-578.

Hertel, P. T., Maydon, A., Ogilvie, A., & Mor, N. (2018). Ruminators (unlike others) fail to show suppression-induced forgetting on indirect measures of memory. *Clinical Psychological Science, 6*(6), 872-881.

Itthipuripat, S., & Woodman, G. F. (2018). Context triggers the retrieval of long-term memories into working memory to guide attention. *bioRxiv*, 401216.

Jahanshahi, M., Obeso, I., Rothwell, J. C., & Obeso, J. A. (2015). A fronto–striato–subthalamic–pallidal network for goal-directed and habitual inhibition. *Nature Reviews Neuroscience, 16*(12), 719-732.

Jana, S., Hannah, R., Muralidharan, V., & Aron, A. R. (2020). Temporal cascade of frontal, motor and muscle processes underlying human action-stopping. *Elife, 9*, e50371.

Johnson, S. K., & Anderson, M. C. (2004). The role of inhibitory control in forgetting semantic knowledge. *Psychological science, 15*(7), 448-453.

Joormann, J., Hertel, P. T., LeMoult, J., & Gotlib, I. H. (2009). Training forgetting of negative material in depression. *Journal of abnormal psychology, 118*(1), 34.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association, 90*(430), 773-795.

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: a critical review of visual masking. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1481), 857-875.

Küpper, C. S., Benoit, R. G., Dalgleish, T., & Anderson, M. C. (2014). Direct suppression as a mechanism for controlling unpleasant memories in daily life. *Journal of Experimental Psychology: General, 143*(4), 1443.

Lambert, A. J., Good, K. S., & Kirk, I. J. (2010). Testing the repression hypothesis: Effects of emotional valence on memory suppression in the think–no think task. *Consciousness and cognition, 19*(1), 281-293.

Lee, J. L. (2009). Reconsolidation: maintaining memory relevance. *Trends in neurosciences, 32*(8), 413-420.

Legrand, N., Etard, O., Vandevelde, A., Pierre, M., Viader, F., Clochon, P., . . . Gagnepain, P. (2020). Long-term modulation of cardiac activity induced by inhibitory control over emotional memories. *Scientific reports, 10*(1), 1-19.

Lenartowicz, A., Verbruggen, F., Logan, G. D., & Poldrack, R. A. (2011). Inhibition-related activation in the right inferior frontal gyrus in the absence of inhibitory cues. *Journal of cognitive neuroscience, 23*(11), 3388-3399.

Levy, B. J., & Anderson, M. C. (2002). Inhibitory processes and the control of memory retrieval. *Trends in cognitive sciences, 6*(7), 299-305.

Levy, B. J., & Anderson, M. C. (2008). Individual differences in the suppression of unwanted memories: The executive deficit hypothesis. *Acta psychologica, 127*(3), 623-635.

Levy, B. J., & Anderson, M. C. (2012). Purging of memories from conscious awareness tracked in the human brain. *Journal of Neuroscience, 32*(47), 16785-16794.

Lewis-Peacock, J. A., & Norman, K. A. (2014). Competition between items in working memory leads to forgetting. *Nature Communications, 5*(1), 1-10.

Little, S., Bonaiuto, J., Barnes, G., & Bestmann, S. (2018). Motor cortical beta transients delay movement initiation and track errors. *bioRxiv*, 384370.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological review, 91*(3), 295.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in human neuroscience, 8*, 213.

Lundqvist, M., Herman, P., Warden, M. R., Brincat, S. L., & Miller, E. K. (2018). Gamma and beta bursts during working memory readout suggest roles in its volitional control. *Nature Communications, 9*(1), 1-12.

Majid, D. A., Cai, W., George, J. S., Verbruggen, F., & Aron, A. R. (2012). Transcranial magnetic stimulation reveals dissociable mechanisms for global versus selective corticomotor suppression underlying the stopping of action. *Cerebral Cortex, 22*(2), 363-371.

Makeig, S., Bell, A. J., Jung, T.-P., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. *Advances in neural information processing systems*, 145-151.

Mary, A., Dayan, J., Leone, G., Postel, C., Fraisse, F., Malle, C., . . . De la Sayette, V. (2020). Resilience after trauma: The role of memory suppression. *Science, 367*(6479).

Marzi, T., Regina, A., & Righi, S. (2014). Emotions shape memory suppression in trait anxiety. *Frontiers in psychology, 4*, 1001.

Mecklinger, A., Parra, M., & Waldhauser, G. T. (2009). ERP correlates of intentional forgetting. *Brain research, 1255*, 132-147.

Mecklinger, A., & Pfeifer, E. (1996). Event-related potentials reveal topographical and temporal distinct neuronal activation patterns for spatial and object working memory. *Cognitive Brain Research, 4*(3), 211-224.

Melton, A. W., & Irwin, J. M. (1940). The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *The American Journal of Psychology, 53*(2), 173-203.

Mink, J. W. (1996). The basal ganglia: focused selection and inhibition of competing motor programs. *Progress in neurobiology, 50*(4), 381-425.

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

Nambu, A., Tokuno, H., & Takada, M. (2002). Functional significance of the cortico–subthalamo–pallidal 'hyperdirect'pathway. *Neuroscience research, 43*(2), 111-117.

Nemeth, V. L., Kurgyis, E., Csifcsak, G., Maraz, A., Almasi, D. A., Drotos, G., . . . Must, A. (2014). The impact of intermediate-term alcohol abstinence on memory retrieval and suppression. *Frontiers in psychology, 5*, 1396.

Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and brain sciences, 37*(1), 1-19.

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.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in cognitive sciences, 10*(2), 59-63.

Raaijmakers, J. G., & Jakab, E. (2013). Rethinking inhibition theory: On the problematic status of the inhibition theory for forgetting. *Journal of memory and language, 68*(2), 98-122.

Racsmány, M., Conway, M. A., Keresztes, A., & Krajcsi, A. (2012). Inhibition and interference in the think/no-think task. *Memory & cognition, 40*(2), 168-176.

Reas, E. T., & Brewer, J. B. (2013). Effortful retrieval reduces hippocampal activity and impairs incidental encoding. *Hippocampus, 23*(5), 367-379.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in cognitive sciences, 11*(6), 251-257.

Schaum, M., Pinzuti, E., Sebastian, A., Lieb, K., Fries, P., Mobascher, A., . . . Tüscher, O. (2021). Right inferior frontal gyrus implements motor inhibitory control via beta-band oscillations in humans. *Elife, 10*, e61679.

Schmidt, R., Ruiz, M. H., Kilavik, B. E., Lundqvist, M., Starr, P. A., & Aron, A. R. (2019). Beta oscillations in working memory, executive control of movement and thought, and sensorimotor function. *Journal of Neuroscience, 39*(42), 8231-8238.

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), 1-12.

Scimeca, J. M., & Badre, D. (2012). Striatal contributions to declarative memory retrieval. *Neuron, 75*(3), 380-392.

Simpson, G. B., & Kang, H. (1994). Inhibitory processes in the recognition of homograph meanings.

Staresina, B. P., Reber, T. P., Niediek, J., Boström, J., Elger, C. E., & Mormann, F. (2019). Recollection in the human hippocampal-entorhinal cell circuitry. *Nature Communications, 10*(1), 1-11.

Staresina, B. P., & Wimber, M. (2019). A neural chronometry of memory recall. *Trends in cognitive sciences, 23*(12), 1071-1085.

Stephens, E., Braid, A., & Hertel, P. T. (2013). Suppression-induced reduction in the specificity of autobiographical memories. *Clinical Psychological Science, 1*(2), 163-169.

Stokes, M. G., Atherton, K., Patai, E. Z., & Nobre, A. C. (2012). Long-term memory prepares neural activity for perception. *Proceedings of the National Academy of Sciences, 109*(6), E360-E367.

Stramaccia, D. F., Meyer, A.-K., Rischer, K. M., Fawcett, J. M., & Benoit, R. G. (2020). Memory suppression and its deficiency in psychological disorders: A focused meta-analysis. *Journal of Experimental Psychology: General*.

Sullivan, D. R., Marx, B., Chen, M. S., Depue, B. E., Hayes, S. M., & Hayes, J. P. (2019). Behavioral and neural correlates of memory suppression in PTSD. *Journal of psychiatric research, 112*, 30-37.

Sundby, K. K., Jana, S., & Aron, A. R. (2021). Double-blind disruption of right inferior frontal cortex with TMS reduces right frontal beta power for action stopping. *Journal of neurophysiology, 125*(1), 140-153.

Sutterer, D. W., & Awh, E. (2016). Retrieval practice enhances the accessibility but not the quality of memory. *Psychonomic Bulletin & Review, 23*(3), 831-841.

Sutterer, D. W., Foster, J. J., Serences, J. T., Vogel, E. K., & Awh, E. (2019). Alpha-band oscillations track the retrieval of precise spatial representations from long-term memory. *Journal of neurophysiology, 122*(2), 539-551.

Swann, N., Tandon, N., Canolty, R., Ellmore, T. M., McEvoy, L. K., Dreyer, S., . . . Aron, A. R. (2009). Intracranial EEG reveals a time-and frequency-specific role for the right inferior frontal gyrus and primary motor cortex in stopping initiated responses. *Journal of Neuroscience, 29*(40), 12675-12685.

Teyler, T. J., & Rudy, J. W. (2007). The hippocampal indexing theory and episodic memory: updating the index. *Hippocampus, 17*(12), 1158-1169.

Tomlinson, T. D., Huber, D. E., Rieth, C. A., & Davelaar, E. J. (2009). An interference account of cue-independent forgetting in the no-think paradigm. *Proceedings of the National Academy of Sciences, 106*(37), 15588-15593.

Tonegawa, S., Liu, X., Ramirez, S., & Redondo, R. (2015). Memory engram cells have come of age. *Neuron, 87*(5), 918-931.

van den Wildenberg, W. P., Burle, B., Vidal, F., van der Molen, M. W., Ridderinkhof, K. R., & Hasbroucq, T. (2010). Mechanisms and dynamics of cortical motor inhibition in the stop-signal paradigm: a TMS study. *Journal of cognitive neuroscience, 22*(2), 225-239.

van Schie, K., & Anderson, M. C. (2017). Successfully controlling intrusive memories is harder when control must be sustained. *Memory, 25*(9), 1201-1216.

van Schie, K., Geraerts, E., & Anderson, M. C. (2013). Emotional and non-emotional memories are suppressible under direct suppression instructions. *Cognition & Emotion, 27*(6), 1122-1131.

Verbruggen, F., & Logan, G. D. (2008). Automatic and controlled response inhibition: associative learning in the go/no-go and stop-signal paradigms. *Journal of Experimental Psychology: General, 137*(4), 649.

Visser, R., Anderson, M., Aron, A., Banich, M., Brady, K., Huys, Q., . . . Schooler, J. (2020). Neuropsychological Mechanisms of Intrusive Thinking.

Wagner, J., Wessel, J. R., Ghahremani, A., & Aron, A. R. (2018). Establishing a right frontal beta signature for stopping action in scalp EEG: implications for testing inhibitory control in other task contexts. *Journal of cognitive neuroscience, 30*(1), 107-118.

Waldhauser, G. T., Dahl, M. J., Ruf-Leuschner, M., Müller-Bamouh, V., Schauer, M., Axmacher, N., . . . Hanslmayr, S. (2018). The neural dynamics of deficient memory control in heavily traumatized refugees. *Scientific reports, 8*(1), 1-12.

Waldhauser, G. T., Lindgren, M., & Johansson, M. (2012). Intentional suppression can lead to a reduction of memory strength: behavioral and electrophysiological findings. *Frontiers in psychology, 3*, 401.

Wang, S., Rajsic, J., & Woodman, G. F. (2019). The contralateral delay activity tracks the sequential loading of objects into visual working memory, unlike lateralized alpha oscillations. *Journal of cognitive neuroscience, 31*(11), 1689-1698.

Wegner, D. M., Schneider, D. J., Carter, S. R., & White, T. L. (1987). Paradoxical effects of thought suppression. *Journal of personality and social psychology, 53*(1), 5.

Wessel, J. R. (2020). β-bursts reveal the trial-to-trial dynamics of movement initiation and cancellation. *Journal of Neuroscience, 40*(2), 411-423.

Wessel, J. R., & Aron, A. R. (2017). On the globality of motor suppression: unexpected events and their influence on behavior and cognition. *Neuron, 93*(2), 259-280.

Wessel, J. R., Ghahremani, A., Udupa, K., Saha, U., Kalia, S. K., Hodaie, M., . . . Chen, R. (2016). Stop-related subthalamic beta activity indexes global motor suppression in Parkinson's disease. *Movement Disorders, 31*(12), 1846-1853.

Wessel, J. R., Reynoso, H. S., & Aron, A. R. (2013). Saccade suppression exerts global effects on the motor system. *Journal of neurophysiology, 110*(4), 883-890.

Wiecki, T. V., & Frank, M. J. (2013). A computational model of inhibitory control in frontal cortex and basal ganglia. *Psychological review, 120*(2), 329.

Yeung, N., Nystrom, L. E., Aronson, J. A., & Cohen, J. D. (2006). Between-task competition and cognitive control in task switching. *Journal of Neuroscience, 26*(5), 1429-1438.

Zacks, R. T., Radvansky, G., & Hasher, L. (1996). Studies of directed forgetting in older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(1), 143.

Zavala, B. A., Jang, A. I., & Zaghloul, K. A. (2017). Human subthalamic nucleus activity during non-motor decision making. *Elife, 6*, e31007.