UNIVERSIDAD DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO

TESIS DOCTORAL

Selection, Maintenance, and Removal of Working Memory Contents: A Behavioral and Electrophysiological Approach

Selección, Mantenimiento y Eliminación de Contenidos de la Memoria de Trabajo: Un Enfoque Conductual y Electrofisiológico

D.ª Miriam Tortajada Gomariz

2024

UNIVERSIDAD DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO

TESIS DOCTORAL

Selection, Maintenance, and Removal of Working Memory Contents: A Behavioral and Electrophysiological Approach

Selección, Mantenimiento y Eliminación de Contenidos de la Memoria de Trabajo: Un Enfoque Conductual y Electrofisiológico

Autor: D.ª Miriam Tortajada Gomariz

Director/es: D. Guillermo Campoy Menéndez y
D. Luis José Fuentes Melero

UNIVERSIDAD DE
MURCIA

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA EN MODALIDAD DE COMPENDIO O ARTÍCULOS PARA OBTENER EL TÍTULO DE DOCTOR

Aprobado por la Comisión General de Doctorado el 19-10-2022

D./Dña. Miriam Tortajada Gomariz

doctorando del Programa de Doctorado en

Psicología

de la Escuela Internacional de Doctorado de la Universidad Murcia, como autor/a de la tesis presentada para la obtención del título de Doctor y titulada:

Selection, Maintenance, and Removal of Working Memory Contents: A Behavioral and Electrophysiological Approach

y dirigida por,

D./Dña. Guillermo Campoy Menéndez

D./Dña. Luis José Fuentes Melero

D./Dña.

DECLARO QUE:

La tesis es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, de acuerdo con el ordenamiento jurídico vigente, en particular, la Ley de Propiedad Intelectual (R.D. legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, modificado por la Ley 2/2019, de 1 de marzo, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), en particular, las disposiciones referidas al derecho de cita, cuando se han utilizado sus resultados o publicaciones.

Además, al haber sido autorizada como compendio de publicaciones o, tal y como prevé el artículo 29.8 del reglamento, cuenta con:

- *La aceptación por escrito de los coautores de las publicaciones de que el doctorando las presente como parte de la tesis.*
- *En su caso, la renuncia por escrito de los coautores no doctores de dichos trabajos a presentarlos como parte de otras tesis doctorales en la Universidad de Murcia o en cualquier otra universidad.*

Del mismo modo, asumo ante la Universidad cualquier responsabilidad que pudiera derivarse de la autoría o falta de originalidad del contenido de la tesis presentada, en caso de plagio, de conformidad con el ordenamiento jurídico vigente.

En Murcia, a 19 de Septiembre de 2024

Fdo.:

Esta DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD debe ser insertada en la primera página de la tesis presentada para la obtención del título de Doctor.

The present work has been supported by the *Ministerio de Ciencia e Innovación* (grant nº FPU19/06016) and by the *Agencia Estatal de investigación* and “ERDF A way of making Europe” (grant PID2021-125408NB-I00).

AGRADECIMIENTOS

“*Happiness is only real when shared*” - Alexander Supertramp

Christopher Johnson McCandless, conocido también como Alexander Supertramp, fue un joven californiano recién graduado que decidió donar todas sus pertenencias para emprender una vida nómada. Después de varios años viviendo en soledad, y poco antes de morir trágicamente tras ingerir una planta venenosa, escribió en su ejemplar de *Doctor Zhivago*: "La felicidad solo es real cuando se comparte". Durante la etapa predoctoral, uno inevitablemente atraviesa momentos de soledad: días sin apartar la vista y la mente de artículos y documentos de Word, horas frente a líneas de código rebeldes, y semanas que se encadenan sin salir del laboratorio. Sin embargo, al igual que Alexander, si algo he aprendido es que esos momentos encuentran su sentido cuando el resto del tiempo es compartido.

Las dificultades más complejas se empequeñecen con los cafés y palabras de ánimo de tus directores. Las complicaciones de organizar una estancia en el extranjero se olvidan cuando tus supervisores te acogen y enseñan como si fueras su propio doctorando. La desmotivación intermitente se emborrona al compartirla, a veces entre risas y a veces entre lágrimas, con tus compañeros predoctorales. Y, muy importante, todo el esfuerzo y los obstáculos se alejan gracias a nuestra familia y amigos que nos mantienen los pies en la tierra. Por estos y otros muchos motivos, gracias de corazón a todos los que me habéis empujado hasta el final.

Title

Selection, Maintenance, and Removal of Working Memory Contents: A Behavioral and Electrophysiological Approach

Abstract

Despite the long-standing interest in working memory (WM), its functional mechanisms remain a major focus of research in psychology. Recent advancements in techniques and analytical methods have enhanced our understanding of the essential subprocesses that compose WM functioning. This dissertation aims to understand better three of these subprocesses: selection, maintenance, and removal. To achieve this, three studies are presented across three chapters, each addressing different questions related to WM functioning.

The first study examines the removal of contents from WM using classifiers on electroencephalography (EEG) data to decode variations in load following removal. The results show that attentional selection processes are not easy to separate from retention and that the EEG signal may be difficult to decode under certain conditions. The second study shifts the focus to the maintenance component of WM and the role of oscillations in its correct functioning. Specifically, the theta-gamma coupling model is tested using transcranial alternating stimulation to modulate WM capacity and the EEG consequences of this entrainment are evaluated. Lastly, the third study continues the exploration of content removal by examining the differences between temporary and permanent removal of information from WM, this time using a behavioral design that allows for the complete dissociation of these two types of removal. The results reveal that while the processes may appear similar, there are deeper, significant differences between them.

Keywords

Working memory; electroencephalography; removal; multivariate pattern analysis; oscillations; theta-gamma coupling

Table of Contents

CHAPTER 1: GENERAL INTRODUCTION 1

1.1. The Definition of Working Memory 1

1.2. Working Memory as Different Activation States 3

1.3. Selection and the Role of Attention in Working Memory 7

1.4. Maintenance or Temporary Bindings 9

1.4.1. The Role of Oscillations in the Maintenance of Information 12

1.5. Removal of Information from Working Memory 14

1.6. Aims and Hypotheses of the Dissertation 17

References 20

CHAPTER 2: DECODING LOAD OR SELECTION IN VISUOSPATIAL WORKING MEMORY? 35

Abstract 36

Introduction 37

Method 41

Participants 42

Apparatus and Stimuli 42

General Procedure 45

Data Recording and Preprocessing 46

Behavioral Data Analysis 47

Multivariate Pattern Analysis (MVPA) 48

Results 52

Behavioral Time Course (Session 1) 52

Behavioral Retro-cue Effect (Session 2) 55

Can the Change in VWM Load be Decoded from the EEG Data? 56

Can we Decode the Drop in Load After the Retro Cue? 63

Correlation of Classifier Performance and VWM Capacity.....66

Discussion..... 68

Load Effects: Retention versus Selection.....71

Conclusion..... 76

References..... 77

Supplementary Materials: Tortajada et al. (2024)..... 92

Behavioral time course for all participants who completed the first session.....92

Empirical chance.....94

Robustness analyses.....95

Testing with False Discovery Rate (FDR) as multiple comparisons correction method.....97

References.....98

CHAPTER 3: QUESTIONING THETA-GAMMA COUPLING PREDICTIONS: INEFFECTIVENESS OF 4 HZ TACS ON WORKING MEMORY MAINTENANCE AND THETA FREQUENCY MODULATION.....99

Abstract..... 99

Introduction.....101

Method.....105

Participants.....105

Tasks and stimuli.....106

Procedure.....108

tACS and EEG protocol.....110

EEG Preprocessing and Analyses.....111

Statistical analyses.....114

Results.....115

tACS effects on WM performance.....115

tACS Effects on Theta Power.....118

Mean theta variations.....119

Phase Locking Value between Stimulation Electrodes 121

Discussion **122**

References **128**

CHAPTER 4: DISTINGUISHING BETWEEN TEMPORARY AND PERMANENT REMOVAL IN VERBAL WORKING MEMORY **143**

Abstract **143**

Introduction **145**

Method **151**

Participants 151

Materials 151

Procedure 151

Results **154**

Discussion **162**

References **167**

CHAPTER 5: GENERAL DISCUSSION AND CONCLUSIONS **175**

RESUMEN EN ESPAÑOL **179**

Chapter 1: General Introduction

1.1. The Definition of Working Memory

The study of memory has traditionally undergone a fundamental division, which we now study as long-term and short-term memory (LTM and STM, respectively). The philosopher John Locke (1690) had already reflected on these two types of retention in mind: the "storehouse of ideas," where ideas are permanently stored and can be revived after having been set aside, and "contemplation," which would consist of a temporary state in which ideas would be kept in view. Perhaps better known is the distinction between primary and secondary memory proposed by William James (1890). Primary memory reflects the contents actively present in consciousness, whereas secondary memory stores items that are not currently in consciousness and require a retrieval process to recall them. Despite the temporal gap between the two authors, both “contemplation” and “primary memory” seem to describe a store of information ready for immediate use, similar to what we now refer to as Short Term Memory (Logie, 1996).

However, the term “working memory” was not used until the 60s when Miller et al. (1960) stated:

> “Something important does happen to a Plan when the decision is made to execute it. It is taken out of dead storage and (...) It is brought into the focus of attention (...) The parts of a Plan that is being executed have special access to consciousness and special ways of being remembered (...) When we have decided to

> execute some particular Plan, it is probably put into some special state or place where it can be remembered while it is being executed (...). Therefore, we should like to speak of the memory we use for the execution of our Plans as a kind of quick-access, ‘working memory’. (p.65)”

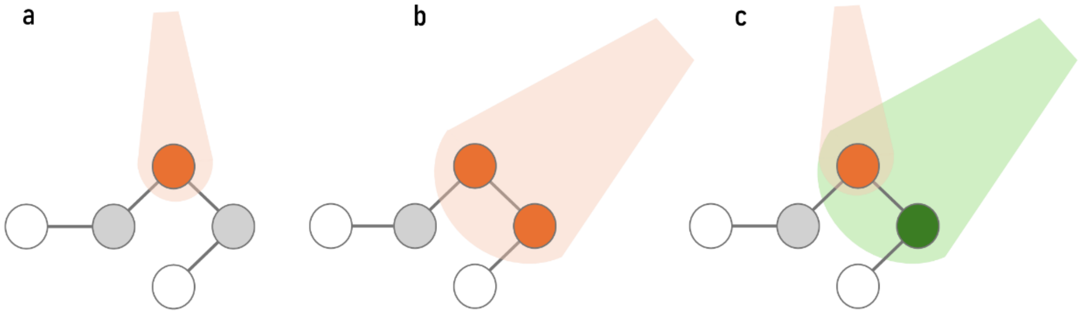
In this way, 'working memory' referred to this temporary memory concept intuitively discussed by Locke and James. However, for the first time, it introduced the significant nuance that it serves as a temporary store for 'plans' or actions that need to be remembered and executed shortly, not only for the maintenance of data or items consciously. Similarly, Atkinson and Shiffrin (1968) proposed that both the conscious information (or representations in STM) and the control of this information, which is intrinsic to human cognition, are encompassed within “working memory”, thereby emphasizing the functional role of its short-term storage. In this scenario, and also attributing WM a functional role, Baddeley and Hitch (1974) proposed that working memory constitutes a multicomponent system, as opposed to the single-store short-term storage view dominant at the time. Consequently, with the focus on understanding the differences between LTM and STM, these authors significantly pushed the study of WM to its current relevance (Baddeley, 2007).

There are currently different theories and conceptualizations about WM (Logie et al., 2020). However, for all of them, the essence of WM is that it is a system (or set of processes) holding mental representations temporarily available for use in thought and action (Cowan, 2017). Therefore, the two fundamental and universally accepted characteristics of WM include this temporary storage of representations and their functional availability for use. Recently, Oberauer (2020) stated that “the function of WM is to provide a medium for representations that we can manipulate and that we can use to guide our thoughts and actions in ways that are novel and potentially depart from what we have learned” (p. 117). This statement builds upon the previous definition by emphasizing the temporary storage and manipulation of representations, but it also adds the role of WM in guiding novel thoughts and actions that may diverge from what has been learned. This highlights the dynamic interaction and relative independence between LTM and WM. In this sense, LTM serves to provide a stable framework of representations, while WM enables us to temporarily manipulate these representations and modify them as necessary to suit the situation. Thus, the “Plans” previously proposed by Miller et al. (1960) would require information from LTM, as well as mechanisms for selecting and maintaining relevant information or filtering out irrelevant information, which would be provided by WM (Postle & Oberauer, 2022). This close link between LTM and WM has received much interest recently, and it underlies some of the main WM frameworks explained in more depth in the next section.

1.2. Working Memory as Different Activation States

Different conceptualizations agree that working memory is not a homogeneous mental device (LaRocque et al., 2014; Rerko & Oberauer, 2013). Rather, information in WM can be presented in different activation states based on the allocation of attention to the information stored in LTM. In other words, WM can be conceptualized as an interaction of LTM and attention, where stable stored information from LTM can enter more activated states when this information is attended. There are three predominant frameworks supporting this idea (see Figure 1), which mainly differ in two aspects: the number of activation states and the number of representations that can be kept in the most active state, that is, in the focus of attention (FoA).

Figure 1. Activation states frameworks of WM. (a) McElree's conceptualization, (b) Cowan's model, and (c) Oberauer's framework. Orange nodes represent the items in the FoA. Green nodes represent the items in the broad focus or region of direct access.



The first conceptualization (McElree, 1998, 2006; McElree & Dosher, 1993) posits the existence of a unique WM component: the FoA. From this framework, the last presented item to retain is supposed to have a priority access status as long as it is attended to and kept in the FoA, while the rest of the items would remain in LTM. Importantly, it goes against the existence of a broader intermediate activation state and suggests that WM and LTM are governed by the same principles—contrary to the conceptualizations stated below. Support for this framework comes from behavioral experiments using stimuli serial presentation tasks, in which performance is enhanced for the last encoded item (McElree, 2006; McElree & Dosher, 1993; Wickelgren et al., 1980). Additional support also comes from functional magnetic resonance imaging (fMRI) studies (Öztekin et al., 2010). Under the assumption that the hippocampus is required for retrieval of relevant information (Mayes et al., 2007), Öztekin et al. (2010) showed that hippocampal activity supports all items but the last presented one, which was proposed to be the only attended item and, consequently, the only item that did not need to be encoded or retrieved. However, other research also found evidence for the distinction of an additional intermediate level of activation of contents in WM—broader than the FoA—supported by different brain regions (Nee & Jonides, 2011), which more closely aligns with the following exposed frameworks.

The second model is Cowan's embedded-processes approach (Cowan, 1988, 1999). This model considers that information in WM can be in two states: in a central component (i.e. FoA), which owns the most accessible and activated part of WM, and its capacity is limited to four chunks of information, or in a peripheral component (the *activated part of long-term memory,* or aLTM) which is a brief sensory store that activates information from LTM, has no capacity limit and is susceptible to decay and interference. The embedded approach emerged from the observations that WM and LTM relation is highly interrelated: information processed in WM includes sensory input but also meaning taken from LTM, and information must be processed in WM before it becomes a long-term representation. This framework serves to accommodate different findings from the literature. For instance, Cowan et al. (2005) showed very little proactive interference when the set size to remember was three or four items, so within the FoA capacity, opposite to larger set sizes, which can be explained by the fact that the FoA protects items from proactive interference (but see also Oberauer et al., 2017). Also, the embedded-processes approach is congruent with the fact that more than one item can be maintained in the FoA simultaneously (Gilchrist & Cowan, 2011).

The three embedded-processes framework advocated by Oberauer (2002, 2013) arose as a modification of the previous model. It proposes a first wide component, the aLTM, which contains potentially relevant contextual information. Then, the next activation step would be occupied by the *region of direct access* (DAR) or *broad focus* (Oberauer & Hein, 2012), which is the equivalence of Cowan's FoA, holding the information that is relevant to the task. Then, the FoA from Oberauer's framework mostly holds a single relevant item from the region of direct access at each moment. The FoA in this framework does not have a limited capacity of a single item per se because it simply serves the role of "selecting" the most relevant item for the current task, which is usually only one item (together with its context). However, when multiple items are equally relevant to the task, they can also be superimposed at the FoA (Oberauer, 2020).

Importantly, Oberauer's framework proposes that WM representations are organized in a network of temporary bindings, where contents (i.e. numbers, words, colors, etc.) are bound to contextual representations (locations in a mental coordinate system). Hebbian learning quickly creates these bindings when a context and a representation are activated simultaneously. Subsequently, when these bindings become irrelevant, they are gradually removed by a Hebbian anti-learning process (Koch et al., 2013; Lewis-Peacock et al., 2018; Oberauer et al., 2012). This process would account for the rapid removal of WM contents once they become irrelevant (see section *Removal of Information from Working Memory*).

Moreover, based on this framework, it has been possible to develop computational models to characterize specific mechanisms deployed in different WM tasks (Oberauer et al., 2012, 2013; Oberauer & Lin, 2017).

In summary, although different predictions can be derived from these frameworks, they all support the notion that information in WM can be presented in different activation states. However, terminology for these states varies across models. Oberauer's FoA aligns closely with McElree's proposal, while the DAR is closer to Cowan's FoA. To minimize ambiguity, Oberauer's framework will be adopted in this dissertation. This choice is given because it successfully accounts for the key processes involved in WM functioning relevant to the present work, namely internal selective attention, information maintenance, and removal of irrelevant information. Subsequent sections will address these specific subprocesses in greater detail.

1.3. Selection and the Role of Attention in Working Memory

Until recently, the study of attention focused on the prioritization of sensory information from our environment. However, in recent years, research has expanded to include the study of attention towards our mental representations (Chun et al., 2011; Ede & Nobre, 2023). Numerous studies have demonstrated similarities between external and internal attention (i.e., towards the environment and the contents of working memory). For instance, research indicates that the brain regions involved are similar in both modalities (Nobre et al., 2004; Panichello & Buschman, 2021; Zhou et al., 2022; but see Jungerius et al., 2024; Myers et al., 2017). When discussing attention toward contents in WM, "internal selective attention” is currently used as an umbrella term that encompasses several mechanisms crucial for information processing in WM, including orientation, selection, consolidation, and prioritization, among others (Ede & Nobre, 2023; Olivers & Roelfsema, 2020).

Traditionally, there has been a tendency to equate internal selective attention solely with initial transient and time-limited processes where prioritization of information that is necessary for the tasks to be performed occurs (Nobre & Kastner, 2014). This is often observed in Event-Related Potentials (ERP) studies through neural markers such as N2PC (Woodman & Luck, 1999, 2003). However, several findings indicate that internal selective attention is not limited to initial transient processes but also includes operations that transform an item into another state for future processing (Ede & Nobre, 2023; Olivers & Roelfsema, 2020). In this sense, these attentional processes play a key role in linking sensory representations with actions (i.e., the 'items' with the 'plans'), consistent with studies showing that sensory contents and plans for actions involving these contents are processed in parallel rather than independently (van Ede et al., 2019).

Since internal attentional selection processes, such as prioritization or consolidation, are not time-constrained, the measures of internal selective attention also need not be. For example, it has been argued that consolidation processes can last between 500 and 2000 ms (Ricker et al., 2018). Also, Quentin et al. (2019) showed sustained decoding for a selection component in a VWM task (i.e., focusing on the orientation or frequency of a grating pattern) that remained stable for approximately 1.5 seconds. However, the dynamics of processes involving some form of internal selective attention are not yet well defined. Chapter 2 of this thesis sheds some light on this. Specifically, Multivariate Pattern Analysis (MVPA) of EEG and EOG data suggests that internal selective attention plays a crucial role in flexibly updating the relevance of VWM content during a prolonged delay.

1.4. Maintenance or Temporary Bindings

How information is maintained in WM is a matter of debate. One of the most common approaches to study maintenance involves manipulating the amount of information retained in WM (i.e., the load) and examining how this information is maintained during a delay following its encoding. During these delays, variations in different electrophysiological markers are typically observed as a function of load. For example, with higher loads, sustained negative activity is seen in the hemisphere contralateral to the stimulus presentation, compared to the ipsilateral hemisphere. This activity becomes more negative in amplitude as the load increases, until it reaches an asymptote that corresponds with the limit of WM capacity (Adam et al., 2018; Vogel & Machizawa, 2004). However, although we can identify markers to estimate load, the neural mechanisms underlying the maintenance of these items remain unclear. To properly explain the possible hypotheses currently being considered, we must revisit the idea that information in WM can be found in different activation states.

It is well known that information in WM can exist in two states depending on its relevance to current goals. These states have been labeled differently in the literature: classical models of WM distinguish between attended and unattended items (Cowan, 2001; Oberauer, 2002), while others refer to items that guide attention versus those that do not (Olivers et al., 2011), or prioritized versus (temporally) unprioritized items (Lewis-Peacock et al., 2012), or currently versus prospectively relevant items (van Loon et al., 2018). Importantly, regardless of the terminology used to describe these two states in the literature, the neural mechanisms supporting maintenance in one or the other state differ.

One of the most influential current proposals suggests that items in an attended state in WM are maintained through persistent neural spiking, whereas unattended items remain in a silent or hidden state of activity (LaRocque et al., 2013, 2014; Lewis-Peacock et al., 2012; Olivers et al., 2011; Stokes, 2015). Supporting evidence for this proposal comes from studies demonstrating that only contents prioritized in the current moment can be decoded, whereas unprioritized representations are not decodable unless they become relevant later (e.g. LaRocque et al., 2013, 2017; Sprague et al., 2016). However, this notion has also been challenged and is not absent from debate.

On one hand, it has been argued that persistent spiking is not a requirement for information maintenance in WM (Miller et al., 2018). Different findings support this idea. For instance, Contralateral Delay Activity (CDA) is attenuated when the same information is maintained across multiple trials, i.e., when it has already involved some learning, which may indicate that persistent activity is necessary to prioritize information initially but not to keep it active later (Gunseli et al., 2014). Furthermore, different results with electrophysiological data suggest persistent activity due to the averaging of trials for statistical analysis. However, closer examination at the single-trial level reveals that activation occurs sparsely, with maintenance achieved through bursts of activity at rapid frequencies (Lundqvist et al., 2016; Miller et al., 2018; Stokes & Spaak, 2016).

On the other hand, how information would be maintained in a passive state is also unclear. Stokes (2015) proposed that the initial activity following the encoding of a stimulus creates a functional connectivity network (probably through a process of short-term synaptic plasticity) unique to this stimulus. After encoding, this network is maintained in a state of "silent activity" that may briefly manifest during any spontaneous activity within the network and could be fully reactivated when the representation becomes relevant again, e.g., if required to provide a behavioral response. A major challenge of this conceptualization is how to measure this passive activation state. One proposed method to detect it involves perturbing these networks so that functional connectivity becomes decodable. This has been achieved through techniques such as using bursts with Transcranial Magnetic Stimulation (Rose et al., 2016) or presenting task-irrelevant high-contrast visual stimuli (Wolff et al., 2015, 2017). The results indicate that activity related to unattended information can be transiently decoded, thereby supporting Stokes's proposal (2015).

However, various alternative explanations have been proposed to describe the active and passive states of WM. For instance, it has been suggested that these two states can be represented in distinct distributed networks (Christophel et al., 2017, 2018), or within the same networks but characterized by different patterns of neuronal activation for each item (Druckmann & Chklovskii, 2012; Spaak et al., 2017), or within the same neural patterns but differing in their levels of activation (e.g., Schneegans & Bays, 2017). Therefore, taking into account these other conceptualizations, there is also no consensus in stating that the active state is equivalent to persistent or sparse activity and that the passive state is equivalent to silent or hidden activity (Stokes et al., 2020).

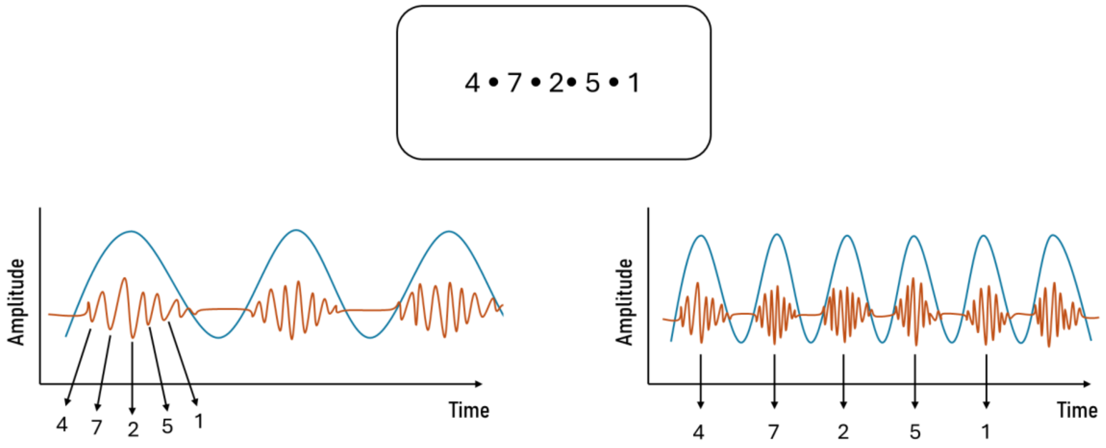
Given the lack of consensus on how information is maintained in WM, studying its characteristics requires techniques sensitive to any of the possibilities discussed so far. Techniques such as Multivariate Pattern Analysis to decode EEG or fMRI measures appears to be a promising approach (Adam et al., 2020; Hakim et al., 2020; Postle, 2020). However, this technique is not without potential confounds (Mostert et al., 2018; Quax et al., 2019). Chapter 2 of the present dissertation focuses on decoding WM content held in an active state and examines how this decoding may be influenced by other factors, particularly when part of this information becomes irrelevant.

1.4.1. The Role of Oscillations in the Maintenance of Information

On the other hand, oscillations also play a significant role in maintaining information in WM (Miller et al., 2018; Roux & Uhlhaas, 2014; Sauseng et al., 2019; Vries et al., 2020). Specifically, evidence suggests that the interaction between theta and gamma frequency bands is likely underlying the maintenance of WM contents. However, the exact way in which the WM representations relate to these frequencies is still under debate (Sauseng et al., 2019; see Figure 2).

The predominant theory is the theta-gamma coupling theory (Jensen & Lisman, 1996; Lisman & Jensen, 2013), which suggests that maintenance occurs through the coupling of fast gamma oscillations—representing individual items—with slower theta waves that organize these fast oscillations by nesting them (see left panel of Figure 2). According to this view, the number of gamma oscillations nested within a single theta cycle would determine the capacity limits of WM. In contrast, a more recent view (Herman et al., 2013) proposes that each item is represented by a complete gamma burst, with only one gamma burst coupled to each theta cycle. Therefore, the capacity limit of WM is determined by the time required to refresh several theta cycles. Although both theories have some supporting evidence, it remains unclear which aspects of each theory best accommodate the scientific evidence so far (Sauseng et al., 2019).

Figure 2. Schematic representation of the two predominant oscillation models about WM maintenance: the theta-gamma coupling theory (left panel), in which all the items are represented in a gamma burst (orange), nested in a theta cycle (blue), and another view that proposes that each WM representation is refreshed in an individual theta cycle (right panel).



One way to study the specific role of oscillations in cognitive processes is by attempting to manipulate the brain rhythms. The main method for manipulating brain oscillations is transcranial Alternating Current Stimulation (tACS), a non-invasive stimulation technique that consists of applying weak currents through electrodes in the scalp, allowing the current to penetrate the cortex. tACS facilitates the synchronization of different brain regions at the frequency of stimulation and can influence cognitive functions dependent on the targeted brain networks and oscillations (Antal et al., 2008, 2022). In the context of WM, tACS has been used to apply theta, gamma, or coupled theta-gamma rhythms to different brain regions, with the primary aim of enhancing WM performance (see Booth et al., 2022 for a recent review). However, the results regarding the technique's efficacy are mixed, with several factors that can influence the cognitive impact of tACS (Krause & Cohen Kadosh, 2014). In Chapter 3 of this dissertation, tACS is employed to modulate WM performance to test various WM maintenance models involving oscillations, while also considering factors that may influence the stimulation outcomes, in this case, task difficulty and baseline WM capacity.

1.5. Removal of Information from Working Memory

The third subprocess relevant to the present work is the removal of WM contents when they become irrelevant. Given the limited capacity of WM (Cowan, 2010), it is crucial to have a mechanism to get rid of information. Using different paradigms, such as retro-cue or directed forgetting (Chiu et al., 2021; Dames & Oberauer, 2022; Lintz & Johnson, 2021), different ways by which information stops loading the WM have been proposed (Barrouillet et al., 2012; Berman et al., 2009; Lewis-Peacock et al., 2018), among which the so-called *removal* stands out. Removal is defined as an agile and fast mechanism for actively target-directed removal of information from WM to achieve a current goal (Lewis-Peacock et al., 2018).

Removal support comes from studies using the retro-cue paradigm that have shown how flagging some WM information as irrelevant improves the performance of subsequent WM operations. For example, it frees up working memory capacity, facilitating the encoding of new information (Popov et al., 2019; Souza et al., 2014), and responses related to the relevant information are faster when part of the information has been removed (Dames & Oberauer, 2022; Oberauer, 2018; Souza et al., 2014). From a neural point of view, several EEG and fMRI studies have also shown how neural traces for information cued as irrelevant are reduced during the delay after the retro-cue (LaRocque et al., 2013; Lewis-Peacock et al., 2012).

There are increasing efforts to understand how removal functions. Specifically, a recent study experimentally examined the boundaries and conditions under which removal occurs, as well as the time course of removal with verbal information (Oberauer, 2018). It was observed that it takes approximately 1 second to completely remove irrelevant verbal information from WM. However, this aspect has not been studied extensively in visuospatial working memory until now. The main objective of Chapter 2 of this dissertation was to assess the temporal dynamics of permanent removal in visuospatial WM, both behaviorally and in EEG data.

Importantly, the results exposed so far were obtained based on the assumption that uncued information is deemed permanently irrelevant. However, sometimes information is flagged as temporarily irrelevant and cannot be completely removed because it may be needed later, but the WM must be left unloaded for some time. This concept is referred to as temporary removal. For instance, during a conference, if two questions are posed and the second one needs to be set aside temporarily while addressing the first, it cannot be entirely forgotten. Although such scenarios are common, the extent to which temporary removal mirrors the impact of permanent removal on WM remains unexplored.

One way to investigate temporary removal is by following the first target by a second target. In this paradigm, the response to this second target consistently revolves around the initially uncued information, ensuring that it was not permanently removed to provide the first response. However, experiments featuring both types of situations (i.e., where information could be permanently or temporarily removed) adopted a design in which both types of trials were randomized (LaRocque et al., 2013; Oberauer, 2005). This setup prevented participants from knowing when to implement one type of removal over the other, making the direct comparison of temporary and permanent removal impossible. To our knowledge, only one study has directly used a retro-cue paradigm to address the comparison of temporary and permanent removal and it observed strong differences between both conditions (Oberauer, 2001, Experiment 2). However, rejected items could appear as distractors during the temporary removal task, which increased the familiarity elicited by these items for this condition. This produced an uneven familiarity between both conditions, which could potentially account for the observed differences. Thus, the main aim of Chapter 4 of the present dissertation was to study the differential impact of temporary and permanent removal on the retrieval of the remaining information in WM and how this unfolded over time.

1.6. Aims and Hypotheses of the Dissertation

The general objective of the present dissertation was to better characterize the mechanisms underlying the selection, maintenance, and removal of information from WM. To that aim, three primary studies were conducted, using both behavioral and electrophysiological methods to address different questions related to these processes. The studies were structured around Oberauer's approach, which served as the central framework for understanding WM from a cognitive perspective. The studies are developed in depth in the following three chapters.

Chapter 2 mainly aimed to investigate the temporal dynamics of content removal from visuospatial WM. The interest was focused on whether changes in load could be monitored following WM removal. To achieve this, MVPA was used to track the electrophysiological signals associated with load, in conditions where load reductions were anticipated after removal. Thus, Chapter 2 explores how removal (or selection, as detailed below) impacts load over time in a visuospatial WM task and the electrophysiological basis of these changes. It was hypothesized that:

- With a longer time to remove, performance on the condition with an informative retro-cue would improve. It was expected that performance would be comparable to a high-load baseline level when no time to remove was available; however, as more time was provided (thus, as more items were removed), performance should become comparable to a low-load baseline level.

- In the EEG, low and high load conditions were expected to be decoded before any removal was depleted.
- Also, mirroring the behavioral performance, it was expected to decode the retro-cue condition as low-load baseline at the beginning of the delay but as high-load baseline by the end of the delay.

In **Chapter 3**, the interest was in studying the role of oscillations in the maintenance of representations in WM. Given the attention that has been given to theta-gamma coupling theory as an electrophysiological substrate for this process, this study builds on this conceptualization to test whether applying extrinsic low-frequency theta electrical stimulation (4 Hz) can enhance performance on a task that primarily requires the maintenance of information in WM. Additionally, factors known to be relevant in previous studies for enhancing tACS effects were controlled. Therefore, this study has two aims: first, to test the prediction derived from theta-gamma coupling theory that a lower theta frequency facilitates the maintenance of more items in WM; second, to examine the influence of individual differences and task difficulty on the effects of tACS. The main two specific hypotheses were:

- Participants with lower working memory capacity (WMC) will benefit more from 4Hz theta transcranial alternating current stimulation (tACS) compared to those with higher WMC. These benefits are expected to be particularly evident during the high-load condition.

- After the 4Hz theta tACS, there will be observable EEG modulations, including a reduction in the predominant average theta frequency and an increase in power at the target theta frequency, especially in the low WMC group.

Lastly, in **Chapter 4** the primary aim was to examine the distinct impact of permanent and temporary removal of verbal representations from WM. Specifically, the study focused on the effects that these two types of removal had on the retrieval of the remaining contents in verbal WM, a comparison that had not been directly examined until now. Besides, by manipulating the time available between removal and retrieval, the temporal dynamics of these effects were evaluated. Additionally, the influence of the order of information encoding on subsequent retrieval of representations and its interaction with each type of removal was explored.

- Both permanent and temporary removal would lead to performance improvements with longer time available to remove the irrelevant items. This improvement was expected to be greater in the permanent removal condition.
- When temporarily removed information had to be retrieved to provide a second response, performance was expected to deteriorate when there was more time to remove the information for the first response, compared to when the removal time was short.

References

Adam, K. C. S., Robison, M. K., & Vogel, E. K. (2018). Contralateral Delay Activity Tracks Fluctuations in Working Memory Performance. *Journal of Cognitive Neuroscience*, *30*(9), 1229-1240. https://doi.org/10.1162/jocn_a_01233

Adam, K. C. S., Vogel, E. K., & Awh, E. (2020). Multivariate analysis reveals a generalizable human electrophysiological signature of working memory load. *Psychophysiology*, *57*(12), e13691. https://doi.org/10.1111/psyp.13691

Antal, A., Boros, K., Poreisz, C., Chaieb, L., Terney, D., & Paulus, W. (2008). Comparatively weak after-effects of transcranial alternating current stimulation (tACS) on cortical excitability in humans. *Brain Stimulation*, *1*(2), 97-105. https://doi.org/10.1016/j.brs.2007.10.001

Antal, A., Luber, B., Brem, A., Bikson, M., Brunoni, A., Kadosh, R., Dubljevic, V., Fecteau, S., Ferreri, F., Floel, A., Hallett, M., Hamilton, R., Herrmann, C., Lavidor, M., Loo, C., Lustenberger, C., Machado, S., Miniussi, C., Moliadze, V., … Paulus, W. (2022). Non-invasive brain stimulation and neuroenhancement. *CLINICAL NEUROPHYSIOLOGY PRACTICE*, *7*, 146-165. https://doi.org/10.1016/j.cnp.2022.05.002

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes. En K. W. Spence & J. T. Spence (Eds.), *Psychology of Learning and Motivation* (Vol. 2, pp. 89-195). Academic Press. https://doi.org/10.1016/S0079-7421(08)60422-3

Baddeley, A. (2007). *Working Memory, Thought, and Action*. OUP Oxford.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. En *Psychology of Learning and Motivation* (Vol. 8, pp. 47-89). Elsevier. https://doi.org/10.1016/S0079-7421(08)60452-1

Barrouillet, P., De Paepe, A., & Langerock, N. (2012). Time causes forgetting from working memory. *Psychonomic Bulletin & Review*, *19*(1), 87-92. https://doi.org/10.3758/s13423-011-0192-8

Berman, M. G., Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 317.

Booth, S., Taylor, J., Brown, L., & Pobric, G. (2022). The effects of transcranial alternating current stimulation on memory performance in healthy adults: A systematic review. *CORTEX*, *147*, 112-139. https://doi.org/10.1016/j.cortex.2021.12.001

Chiu, Y.-C., Wang, T. H., Beck, D. M., Lewis-Peacock, J. A., & Sahakyan, L. (2021). Separation of item and context in item-method directed forgetting. *NeuroImage*, *235*. https://doi.org/10.1016/j.neuroimage.2021.117983

Christophel, T. B., Iamshchinina, P., Yan, C., Allefeld, C., & Haynes, J.-D. (2018). Cortical specialization for attended versus unattended working memory. *Nature Neuroscience*, *21*(4), Article 4. https://doi.org/10.1038/s41593-018-0094-4

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J.-D. (2017). The Distributed Nature of Working Memory. *Trends in Cognitive Sciences*, *21*(2), 111-124. https://doi.org/10.1016/j.tics.2016.12.007

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A Taxonomy of External and Internal Attention. *Annual Review of Psychology*, *62*(Volume 62, 2011), 73-101. https://doi.org/10.1146/annurev.psych.093008.100427

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, *104*(2), 163-191. https://doi.org/10.1037/0033-2909.104.2.163

Cowan, N. (1999). An Embedded-Processes Model of working memory. En A. Miyake & P. Shan (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62-101). Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.006

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87-114. https://doi.org/10.1017/S0140525X01003922

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current directions in psychological science*, *19*(1), 51-57. https://doi.org/10.1177/0963721409359277

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, *24*(4), 1158-1170. https://doi.org/10.3758/s13423-016-1191-6

Cowan, N., Johnson, T. D., & Scott Saults, J. (2005). Capacity limits in list item recognition: Evidence from proactive interference. *Memory*, *13*(3-4), 293-299. https://doi.org/10.1080/09658210344000206

Dames, H., & Oberauer, K. (2022). Directed forgetting in working memory. *Journal of Experimental Psychology. General*, *151*(12), 2990-3008. https://doi.org/10.1037/xge0001256

Druckmann, S., & Chklovskii, D. B. (2012). Neuronal Circuits Underlying Persistent Representations Despite Time Varying Activity. *Current Biology*, *22*(22), 2095-2103. https://doi.org/10.1016/j.cub.2012.08.058

Ede, F. van, & Nobre, A. C. (2023). Turning Attention Inside Out: How Working Memory Serves Behavior. *Annual Review of Psychology*, *74*(Volume 74, 2023), 137-165. https://doi.org/10.1146/annurev-psych-021422-041757

Gilchrist, A. L., & Cowan, N. (2011). Can the focus of attention accommodate multiple, separate items? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(6), 1484-1502. https://doi.org/10.1037/a0024352

Gunseli, E., Olivers, C. N. L., & Meeter, M. (2014). Effects of Search Difficulty on the Selection, Maintenance, and Learning of Attentional Templates. *Journal of Cognitive Neuroscience*, *26*(9), 2042-2054. https://doi.org/10.1162/jocn_a_00600

Hakim, N., Awh, E., & Vogel, E. K. (2020). Manifold Visual Working Memory. En R. Logie, V. Camos, & N. Cowan (Eds.), *Working Memory: The state of the science* (p. 0). Oxford University Press. https://doi.org/10.1093/oso/9780198842286.003.0011

Herman, P. A., Lundqvist, M., & Lansner, A. (2013). Nested theta to gamma oscillations and precise spatiotemporal firing during memory retrieval in a simulated attractor network. *Brain Research*, *1536*, 68-87. https://doi.org/10.1016/j.brainres.2013.08.002

James, W. (1890). *Principles of Psychology: Vol. I*. Methuen.

Jensen, O., & Lisman, J. E. (1996). Novel lists of 7 +/- 2 known items can be reliably stored in an oscillatory short-term memory network: Interaction with long-term memory. *Learning & Memory*, *3*(2-3), 257-263.

Jungerius, C., Ede, F. van, & Slagter, H. A. (2024). *No evidence for attentional prioritization in visual working memory through statistical learning*. OSF. https://doi.org/10.31234/osf.io/ksqb4

Kiyonaga, A., & Egner, T. (2013). Working memory as internal attention: Toward an integrative account of internal and external selection processes. *Psychonomic Bulletin & Review*, *20*(2), 228-242. https://doi.org/10.3758/s13423-012-0359-y

Koch, G., Ponzo, V., Lorenzo, F. D., Caltagirone, C., & Veniero, D. (2013). Hebbian and Anti-Hebbian Spike-Timing-Dependent Plasticity of Human Cortico-Cortical Connections. *Journal of Neuroscience*, *33*(23), 9725-9733. https://doi.org/10.1523/JNEUROSCI.4988-12.2013

Krause, B., & Cohen Kadosh, R. (2014). Not all brains are created equal: The relevance of individual differences in responsiveness to transcranial electrical stimulation. *Frontiers in Systems Neuroscience*, *8*, 25. https://doi.org/10.3389/fnsys.2014.00025

LaRocque, J. J., Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2013). Decoding Attended Information in Short-term Memory: An EEG Study. *Journal of Cognitive Neuroscience*, *25*(1), 127-142. https://doi.org/10.1162/jocn_a_00305

LaRocque, J. J., Lewis-Peacock, J. A., & Postle, B. R. (2014). Multiple neural states of representation in short-term memory? It'sa matter of attention. *Frontiers in human neuroscience*, *8*, 5.

LaRocque, J. J., Riggall, A. C., Emrich, S. M., & Postle, B. R. (2017). Within-Category Decoding of Information in Different Attentional States in Short-Term Memory. *Cerebral Cortex*, *27*(10), 4881-4890. https://doi.org/10.1093/cercor/bhw283

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural Evidence for a Distinction between Short-term Memory and the Focus of Attention. *Journal of Cognitive Neuroscience*, *24*(1), 61-79. https://doi.org/10.1162/jocn_a_00140

Lewis-Peacock, J. A., Kessler, Y., & Oberauer, K. (2018). The removal of information from working memory. *Annals of the New York Academy of Sciences*, *1424*(1), 33-44. https://doi.org/10.1111/nyas.13714

Lintz, E. N., & Johnson, M. R. (2021). Refreshing and removing items in working memory: Different approaches to equivalent processes? *Cognition*, *211*, 104655. https://doi.org/10.1016/j.cognition.2021.104655

Lisman, J. E., & Jensen, O. (2013). The Theta-Gamma Neural Code. *Neuron*, *77*(6), 1002-1016. https://doi.org/10.1016/j.neuron.2013.03.007

Locke, J. (1690). *An essay concerning humane understanding*. Thomas Bassett.

Logie, R., Camos, V., & Cowan, N. (Eds.). (2020). *Working Memory: The state of the science*. Oxford University Press. https://doi.org/10.1093/oso/9780198842286.001.0001

Logie, R. H. (1996). The Seven Ages of Working Memory. En J. T. E. Richardson, R. W. Engle, L. Hasher, R. H. Logie, E. R. Stoltzfus, & R. T. Zacks (Eds.), *Working Memory and Human Cognition* (p. 0). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195100990.003.0002

Lundqvist, M., Rose, J., Herman, P., Brincat, S. L., Buschman, T. J., & Miller, E. K. (2016). Gamma and Beta Bursts Underlie Working Memory. *Neuron*, *90*(1), 152-164. Scopus. https://doi.org/10.1016/j.neuron.2016.02.028

Mayes, A., Montaldi, D., & Migo, E. (2007). Associative memory and the medial temporal lobes. *Trends in Cognitive Sciences*, *11*(3), 126-135. https://doi.org/10.1016/j.tics.2006.12.003

McElree, B. (1998). Attended and non-attended states in working memory: Accessing categorized structures. *Journal of Memory and Language*, *38*(2), 225-252.

McElree, B. (2006). Accessing recent events. *Psychology of learning and motivation*, *46*, 155-200.

McElree, B., & Dosher, B. A. (1993). Serial Retrieval Processes in the Recovery of Order Information. *Journal of Experimental Psychology: General*, *122*(3), 291. https://doi.org/10.1037/0096-3445.122.3.291

Miller, E. K., Lundqvist, M., & Bastos, A. M. (2018). Working Memory 2.0. *Neuron*, *100*(2), 463-475. https://doi.org/10.1016/j.neuron.2018.09.023

Miller, G. A., Eugene, G., & Pribram, K. H. (1960). Plans and the Structure of Behaviour. En *Systems Research for Behavioral Science*. Henry Holt and Company.

Mostert, P., Albers, A. M., Brinkman, L., Todorova, L., Kok, P., & de Lange, F. P. (2018). Eye Movement-Related Confounds in Neural Decoding of Visual Working Memory Representations. *eNeuro*, *5*(4). https://doi.org/10.1523/ENEURO.0401-17.2018

Myers, N. E., Stokes, M. G., & Nobre, A. C. (2017). Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. *Trends in Cognitive Sciences*, *21*(6), 449-461. https://doi.org/10.1016/j.tics.2017.03.010

Nee, D. E., & Jonides, J. (2011). Dissociable contributions of prefrontal cortex and the hippocampus to short-term memory: Evidence for a 3-state model of memory. *NeuroImage*, *54*(2), 1540-1548. https://doi.org/10.1016/j.neuroimage.2010.09.002

Nobre, K., Coull, J. T., Maquet, P., Frith, C. D., Vandenberghe, R., & Mesulam, M. M. (2004). Orienting attention to locations in perceptual versus mental representations. *Journal of Cognitive Neuroscience*, *16*(3), 363-373. Scopus. https://doi.org/10.1162/089892904322926700

Nobre, K., & Kastner, S. (2014). *The Oxford Handbook of Attention*. OUP Oxford.

Oberauer, K. (2001). Removing irrelevant information from working memory: A cognitive aging study with the modified Sternberg task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(4), 948.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 411.

Oberauer, K. (2005). Control of the contents of working memory—A comparison of two paradigms and two age groups. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(4), 714.

Oberauer, K. (2013). The focus of attention in working memory—From metaphors to mechanisms. *Frontiers in human neuroscience*, *7*, 673.

Oberauer, K. (2018). Removal of irrelevant information from working memory: Sometimes fast, sometimes slow, and sometimes not at all. *Annals of the New York Academy of Sciences*, *1424*(1), 239-255.

Oberauer, K. (2020). Towards a Theory of Working Memory: From Metaphors to Mechanisms. En R. Logie, V. Camos, & N. Cowan (Eds.), *Working Memory: The state of the science*. Oxford University Press. https://doi.org/10.1093/oso/9780198842286.003.0005

Oberauer, K., Awh, E., & Sutterer, D. W. (2017). The Role of Long-Term Memory in a Test of Visual Working Memory: Proactive Facilitation but no Proactive Interference. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(1), 1. https://doi.org/10.1037/xlm0000302

Oberauer, K., & Hein, L. (2012). Attention to Information in Working Memory. *Current Directions in Psychological Science*, *21*(3), 164-169. https://doi.org/10.1177/0963721412444727

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, *19*(5), 779-819. https://doi.org/10.3758/s13423-012-0272-4

Oberauer, K., & Lin, H.-Y. (2017). An Interference Model of Visual Working Memory. *Psychological Review*, *124*(1), Article 1. https://doi.org/10.1037/rev0000044

Oberauer, K., Souza, A. S., Druey, M. D., & Gade, M. (2013). Analogous mechanisms of selection and updating in declarative and procedural working memory: Experiments and a computational model. *Cognitive Psychology*, *66*(2), 157-211. https://doi.org/10.1016/j.cogpsych.2012.11.001

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, *15*(7), 327-334. https://doi.org/10.1016/j.tics.2011.05.004

Olivers, C. N. L., & Roelfsema, P. R. (2020). Attention for action in visual working memory. *Cortex*, *131*, 179-194. https://doi.org/10.1016/j.cortex.2020.07.011

Öztekin, I., Davachi, L., & McElree, B. (2010). Are Representations in Working Memory Distinct From Representations in Long-Term Memory?: Neural Evidence in Support of a Single Store. *Psychological Science*, *21*(8), 1123-1133. https://doi.org/10.1177/0956797610376651

Panichello, M. F., & Buschman, T. J. (2021). Shared mechanisms underlie the control of working memory and attention. *Nature*, *592*(7855), 601-605. https://doi.org/10.1038/s41586-021-03390-w

Popov, V., Marevic, I., Rummel, J., & Reder, L. M. (2019). Forgetting Is a Feature, Not a Bug: Intentionally Forgetting Some Things Helps Us Remember Others by Freeing Up Working Memory Resources. *Psychological Science*, *30*(9), 1303-1317. https://doi.org/10.1177/0956797619859531

Postle, B. R. (2020). Cognitive Neuroscience of Visual Working Memory. En R. Logie, V. Camos, & N. Cowan (Eds.), *Working Memory: The state of the science* (p. 0). Oxford University Press. https://doi.org/10.1093/oso/9780198842286.003.0012

Postle, B. R., & Oberauer, K. (2022). Working Memory: Theoretical, Computational, and Neural Considerations. En *The Oxford Handbook of Human Memory*. Oxford University Press. https://osf.io/963kf

Quax, S. C., Dijkstra, N., van Staveren, M. J., Bosch, S. E., & van Gerven, M. A. J. (2019). Eye movements explain decodability during perception and cued attention in MEG. *NeuroImage*, *195*, 444-453. https://doi.org/10.1016/j.neuroimage.2019.03.069

Quentin, R., King, J.-R., Sallard, E., Fishman, N., Thompson, R., Buch, E. R., & Cohen, L. G. (2019). Differential Brain Mechanisms of Selection and Maintenance of Information during Working Memory. *Journal of Neuroscience*, *39*(19), 3728-3740. https://doi.org/10.1523/JNEUROSCI.2764-18.2019

Rerko, L., & Oberauer, K. (2013). Focused, unfocused, and defocused information in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1075.

Ricker, T. J., Nieuwenstein, M. R., Bayliss, D. M., & Barrouillet, P. (2018). Working memory consolidation: Insights from studies on attention and working memory. *Annals of the New York Academy of Sciences*, *1424*(1), 8-18. https://doi.org/10.1111/nyas.13633

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science*, *354*(6316), 1136-1139. https://doi.org/10.1126/science.aah7011

Roux, F., & Uhlhaas, P. J. (2014). Working memory and neural oscillations: Alpha–gamma versus theta–gamma codes for distinct WM information? *Trends in Cognitive Sciences*, *18*(1), 16-25. https://doi.org/10.1016/j.tics.2013.10.010

Sauseng, P., Peylo, C., Biel, A. L., Friedrich, E. V. C., & Romberg-Taylor, C. (2019). Does cross-frequency phase coupling of oscillatory brain activity contribute to a better understanding of visual working memory? *British Journal of Psychology*, *110*(2), 245-255. https://doi.org/10.1111/bjop.12340

Schneegans, S., & Bays, P. M. (2017). Restoration of fMRI Decodability Does Not Imply Latent Working Memory States. *Journal of Cognitive Neuroscience*, *29*(12), 1977-1994. https://doi.org/10.1162/jocn_a_01180

Souza, A. S., Rerko, L., & Oberauer, K. (2014). Unloading and reloading working memory: Attending to one item frees capacity. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 1237-1256. https://doi.org/10.1037/a0036331

Spaak, E., Watanabe, K., Funahashi, S., & Stokes, M. G. (2017). Stable and Dynamic Coding for Working Memory in Primate Prefrontal Cortex. *Journal of Neuroscience*, *37*(27), 6503-6516. https://doi.org/10.1523/JNEUROSCI.3364-16.2017

Sprague, T. C., Ester, E. F., & Serences, J. T. (2016). Restoring Latent Visual Working Memory Representations in Human Cortex. *Neuron*, *91*(3), 694-707. https://doi.org/10.1016/j.neuron.2016.07.006

Stokes, M. (2015). 'Activity-silent' working memory in prefrontal cortex: A dynamic coding framework. *Trends in Cognitive Sciences*, *19*(7), 394-405. https://doi.org/10.1016/j.tics.2015.05.004

Stokes, M. G., Muhle-Karbe, P. S., & Myers, N. E. (2020). Theoretical distinction between functional states in working memory and their corresponding neural states. *Visual Cognition*, *28*(5-8), 420-432. https://doi.org/10.1080/13506285.2020.1825141

Stokes, M., & Spaak, E. (2016). The Importance of Single-Trial Analyses in Cognitive Neuroscience. *Trends in Cognitive Sciences*, *20*(7), 483-486. https://doi.org/10.1016/j.tics.2016.05.008

van Ede, F., Chekroud, S. R., Stokes, M. G., & Nobre, A. C. (2019). Concurrent visual and motor selection during visual working memory guided action. *Nature Neuroscience*, *22*(3), 477-483. https://doi.org/10.1038/s41593-018-0335-6

van Loon, A. M., Olmos-Solis, K., Fahrenfort, J. J., & Olivers, C. N. (2018). Current and future goals are represented in opposite patterns in object-selective cortex. *eLife*, *7*, e38677. https://doi.org/10.7554/eLife.38677

Verschooren, S., Schindler, S., De Raedt, R., & Pourtois, G. (2019). Switching attention from internal to external information processing: A review of the literature and empirical support of the resource sharing account. *Psychonomic Bulletin & Review*, *26*(2), 468-490. https://doi.org/10.3758/s13423-019-01568-y

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748-751. https://doi.org/10.1038/nature02447

Vries, I. E. J. de, Slagter, H. A., & Olivers, C. N. L. (2020). Oscillatory Control over Representational States in Working Memory. *Trends in Cognitive Sciences*, *24*(2), 150-162. https://doi.org/10.1016/j.tics.2019.11.006

Wickelgren, W. A., Corbett, A. T., & Dosher, B. A. (1980). Priming and retrieval from short-term memory: A speed accuracy trade-off analysis. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 387-404. https://doi.org/10.1016/S0022-5371(80)90276-5

Wolff, M. J., Ding, J., Myers, N., & Stokes, M. (2015). Revealing hidden states in visual working memory using electroencephalography. *Frontiers in Systems Neuroscience*, *9*. https://www.frontiersin.org/articles/10.3389/fnsys.2015.00123

Wolff, M. J., Jochim, J., Akyürek, E. G., & Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience*, *20*(6), 864-871. https://doi.org/10.1038/nn.4546

Woodman, G. F., & Luck, S. J. (1999). Electrophysiological measurement of rapid shifts of attention during visual search. *Nature*, *400*(6747), 867-869. https://doi.org/10.1038/23698

Woodman, G. F., & Luck, S. J. (2003). Serial deployment of attention during visual search. *Journal of Experimental Psychology. Human Perception and Performance*, *29*(1), 121-138. https://doi.org/10.1037//0096-1523.29.1.121

Zhou, Y., Curtis, C. E., Sreenivasan, K. K., & Fougnie, D. (2022). Common Neural Mechanisms Control Attention and Working Memory. *Journal of Neuroscience*, *42*(37), 7110-7120. https://doi.org/10.1523/JNEUROSCI.0443-22.2022

Chapter 2: Decoding Load or Selection in Visuospatial Working Memory?

The present chapter is currently published in the scientific journal *Psychophysiology:*

Tortajada, M., Fahrenfort, J. J., Sandoval-Lentisco, A., Martínez-Pérez, V., Palmero, L. B., Castillo, A., Fuentes, L.J., Campoy, G., & Olivers, C. N. (2024). Decoding load or selection in visuospatial working memory? *Psychophysiology*, e14636. https://doi.org/10.1111/psyp.14636

Abstract

Flexible updating of information in Visual Working Memory (VWM) is crucial to deal with its limited capacity. Previous research has shown that the removal of no longer relevant information takes some time to complete. Here, we sought to study the time course of such removal by tracking the accompanying drop in load through behavioral and neurophysiological measures. In the first experimental session, participants completed a visuospatial retro-cue task in which the Cue-Target Interval (CTI) was manipulated. The performance revealed that it takes about half a second to make full use of the retro-cue. In a second session, we sought to study the dynamics of load-related electroencephalographic (EEG) signals to track the removal of information. We applied Multivariate Pattern Analysis (MVPA) to EEG data from the same task. Right after encoding, results replicated previous research using MVPA to decode load. However, especially after the retro-cue, results suggested that classifiers were mainly sensitive to a selection component, and not so much to load per se. Additionally, visual cue variations, as well as eye movements that accompany load manipulations can also contribute to decoding. These findings advise caution when using MVPA to decode VWM load, as classifiers may be sensitive to confounding operations.

Keywords: Visual working memory, electroencephalography, multivariate pattern analysis, working memory load, attentional selection

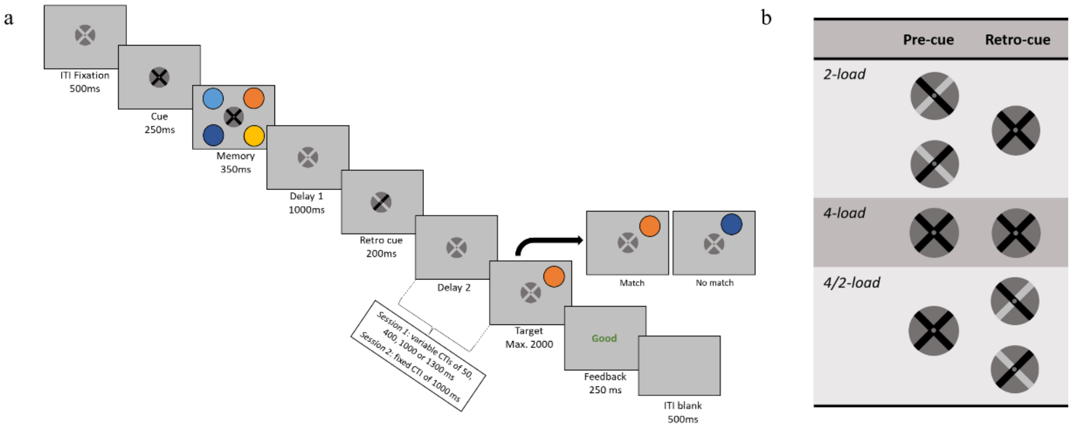
Introduction

Visual working memory (VWM) is the ability to maintain relevant visual information for an ongoing task. The amount of information that we can keep active in VWM is generally thought to be limited to about three or four items (Cowan, 2010; Vogel et al., 2001), but even below this capacity limit, performance tends to decline as load increases. Importantly, retrospectively cueing an item within VWM as relevant can at least partially save it from such detrimental effects, as compared to information that is not cued (Astle et al., 2012; Günseli et al., 2015; Kuo et al., 2012; Shepherdson et al., 2018; Souza & Oberauer, 2016; van Moorselaar et al., 2015). While there is no clear consensus yet about the exact underlying mechanisms of these retro-cueing benefits, most accounts assume some form of attentional selection process within VWM that increases the robustness of the selected item against decay or interference (Lepsien & Nobre, 2006; Myers et al., 2017; Souza et al., 2014; Souza & Oberauer, 2016). Here we were interested in the dynamics of this process: How long does it take for cue-induced benefits to occur in VWM? And can we then track its underlying mechanisms using electrophysiological measures (specifically EEG)?

One way to explain retro-cueing benefits is that the cue typically allows for irrelevant items to be removed from WM, thus reducing any interference such items may have on, and freeing memory capacity for, the target information (Lewis-Peacock et al., 2018; Makovski, 2012; Williams et al., 2013). In other words, cueing part of the information in VWM as relevant (and as a consequence the rest as irrelevant) effectively allows for a reduction of memory load, where load refers to the number of items to be retained. The aim of the current study was to track this load reduction across time. To this end, participants completed a visuospatial probe recognition task in which they were asked to remember a number of colored disks, one of which would be probed at the end of the trial. There were three main conditions (see Figure 1). In the 2- and 4-load baseline conditions, respectively two and four colored items had to be maintained during the entire trial, as was signaled by a pre-cue appearing prior to the stimulus display. These two conditions served as a comparison for the third condition, in which a retro-cue was introduced during the maintenance of the items in VWM, and which signaled which half of the information would be relevant – as the other half would not be tested. We refer to this as the 4/2-load condition because four items were to be initially encoded and afterwards two remained relevant and two could be dropped. The 4/2-load condition was thus meant to induce removal of half of the information from VWM.

The study consisted of two components. First, in a behavioral version of the task, we systematically varied what can been referred to as the Cue-Test Interval (CTI), which is the interval between the retro-cue and the memory probe. With time after cue, we would then expect performance to improve, with performance initially being comparable to high load baselines, while with the removal of items from memory it should become comparable to low load baseline levels. Earlier work by Oberauer (2018) using this procedure showed that indeed, for short CTIs, cued performance still behaved like the high load condition, but it resembled low load performance as CTI increased. These and other results have suggested that it takes at least 500 to 1000 ms for the retro-cue to have full effect (see Souza & Oberauer, 2016, for a review). However, so far, these studies have used verbal stimuli. There have been studies that have looked at the time course of retro-cueing effects in VWM (Gressmann & Janczyk, 2016; Pertzov et al., 2013; Shepherdson et al., 2018; Souza et al., 2014; van Moorselaar et al., 2015), but these did not assess how performance transitions from resembling high to resembling low load baselines, e.g. because such conditions were not included. Here we were specifically interested in how selection within visual working memory reduces load across time.

Figure 1. (a) Task procedure for a 4/2-load trial. The task was a visuospatial working memory task that required delayed colored item recognition. First, participants saw a pre-cue, followed by four colored disks in four positions. The pre-cue pointed to the disks that had to be encoded. In this present example, all disks had to be encoded and maintained in memory. In the case of the 2-load condition, the uncued disks could be ignored. Then, after delay 1 (1000 ms) a retro-cue signaled the disks that remained relevant. This retro- cue was irrelevant for the 2- and 4-load conditions, but was informative for the 4/2-load condition. After a second delay (variable in the first session; fixed to 1000 ms in second session), a target disk appeared in one of the relevant locations and participants had to press a button reporting whether it matched or not the retained items. Inter Trial Interval (ITI), Cue Target Interval (CTI). (b) Summary of possible cue combinations for each trial type.



Second, we sought to uncover the electrophysiological counterpart of these dynamics, using EEG. There have been electrophysiological studies of the time course of the retro-cue effect in VWM (Kuo et al., 2012; Schneider et al., 2016), but these made use of univariate measures involving event-related components that required averaging across extended time windows, precluding a detailed time course assessment. In the current experiment, we wanted to track the removal of contents from VWM using highly time-resolved multivariate decoding methods. Recent studies have suggested that multivariate pattern analyses (MVPA) of the EEG signal provides an index of visual working memory load (Adam et al., 2020; Thyer et al., 2022). Specifically, these studies showed that MVPA-based decoding analyses could distinguish between levels of memory load, capturing differences as little as one item. Compared to univariate VWM load measures, such as contralateral delay activity (McCollough et al., 2007; Vogel & Machizawa, 2004) or negative slow wave (Fukuda et al., 2015), MVPA provides the advantages of being more sensitive and temporarily precise as it does not necessarily rely on the timing of a specific event-related potential measured from at most a few electrodes. In addition, load decoding with MVPA promises to be more universally applicable, as it does not depend on lateralized stimuli. Moreover, it has shown generalization across several factors, including item complexity and the type of information retained (Adam et al., 2020; Thyer et al., 2022). Thus the method promises to be both more sensitive and more versatile. We therefore applied this method in a second session, where we did not manipulate CTI, but provided sufficient post-cue time and then instead assessed whether multivariate signal traces the assumed reduction in VWM load. Specifically, we hypothesized that if the uncued information was indeed removed from VWM, the load-related multivariate signal should first resemble the high-load baseline, but over time start to resemble more and more the low-load baseline.

To foreshadow the findings, behaviorally we found a clear increase in the retro-cueing benefit with time, suggesting a relatively rapid shedding of load within about half a second. This extends earlier findings showing a similar time course to that observed for verbal information (Oberauer, 2018). However, the EEG findings indicated that the MVPA method was sensitive, not only to a load component but also to selection mechanisms. That is, whenever observers had to consolidate (after a pre-cue) or keep (after a retro-cue) a certain number of items in VWM, this entailed both a retention component (i.e., the more items selected, the more items to be maintained), and a selection component (i.e., select the items to consolidate, or select the items to keep in memory). We argue that it makes MVPA analyses susceptible to confounding interpretations, especially given the frequent co-occurrence of selection and maintenance requirements in standard VWM tasks. This is especially the case after the retro-cue, where the selection component precluded successful tracking of load reduction in VWM. Given that these components tend to correlate in VWM experiments in general, our study serves as a warning that it is important to clearly define what is meant by load. We will return to this in the General Discussion, but in the meantime we will use the term load as referring to the number of items retained in memory.

Method

Data and scripts from the two sessions are available at OSF (https://osf.io/fh3wa/).

Participants

Fifty-six undergraduate students (mean age = 20.5, sd = 3.6, 50 females) from the Faculty of Psychology of the University of Murcia completed the first experimental session and were informed that they would be contacted again to complete a second experimental session. From these participants, those who performed close to or below chance in any of the three experimental conditions (i.e., under .60) were excluded from the analyses and were not contacted for the second session. From the remaining fifty-two, ten participants did not respond to the invitation for the second session. Forty-two participants assisted in the EEG session. Three of them could not complete the task due to technical issues. Thus, a final sample of thirty-nine participants completed the EEG session. All participants reported normal or corrected-to-normal vision and signed a written consent at the beginning of every session. Participants received course credits for their participation. The study was approved by the Ethics Committee of the University of Murcia and was conducted according to the ethical standards of the 1964 Declaration of Helsinki.

Apparatus and Stimuli

The tasks were programmed in E-Prime 3 and were performed in individual sound-attenuated booths where participants were seated and responded using a five buttons Chronos® device (Psychology Software Tools).

Figure 1 illustrates the task procedure. The task began with a fixation point of 500 ms which consisted of a dark grey colored circle with a dark grey point in the middle, inside an outline cross. The cross was designed to minimize eye movements (the ABC shape in Thaler et al., 2013) and was rotated 45 degrees from the original to accommodate the purpose of this experiment. The fixation point was followed by a cue of 250 ms duration in black. The cue consisted of the previous fixation point plus the filling of one or two of the arms of the cross Next, it appeared together with four colored disks for 350 ms. The cue indicated the disks that had to be encoded into memory, with a left diagonal cue indicating that the top left and bottom right items had to be encoded, and a right diagonal cue indicating the top right and bottom left items (2-load condition), and a fully filled cue indicated that all four items had to be encoded (4-load and 4/2-load conditions). This way we fully matched the initial perceptual information in all load conditions except for the cue. Furthermore, in the 2-load condition, we intentionally had participants recall items arranged diagonally, i.e., both to the left and right of fixation, thus avoiding any lateralization effects on EEG and eye movements. Next, the fixation point was presented during a first delay of 1000 ms while participants had to keep in memory the indicated colored disks. Delay 1 was followed by a 200 ms retro-cue that differed between conditions. In the 4/2-load condition, the retro-cue was one of the diagonals, indicating that only two disks of the previously presented colored items should be retained, while the remaining items could be forgotten, as they would never be tested. In both the 2- and 4-load conditions, a non-informative fully filled cross was presented to indicate that what had to be encoded also had to be retain. Looking at the results, we observe a larger area under the ROC curve, A, and faster RTs in the 2-load condition (see Figure 2). This shows that participants correctly understood these instructions.

After the retro-cue there was a second delay after which a target probe was presented until the response, with a maximum of 2000 ms. In the behavioral session, the CTI varied between 50 and 1300 ms, while in the EEG session it lasted 1000 ms (see General procedure). The target consisted of a colored disk placed in one of two (2-load) or four (4 and 4/2-load) possible locations. Participants had to decide whether the previously memorized colored item at that position matched the target or not. On mismatch trials, the target disk was filled in with one color of the other item(s) that should have also been encoded, thus preventing participants from responding simply on the basis of familiarity.

Stimuli were presented on a 23-inch flatscreen LED monitor (LG 23MP68VQ-P) with a resolution of 1920 x 1080 pixels on a grey background, RGB (192, 192, 192). Participants were sitting 70 cm away from the screen. The fixation point was made out of a 1.23º diameter circle, a cross inside the circle (arms were 0.18º thick) and an additional circle (0.12º of diameter) in the intersection of the cross (see Figure 1), based on Thaler et al. (2013) recommendations to avoid eye movements during fixation. The circle of the fixation point was dark grey, and the arrows were the same color as the background. The distance of the memory disks from the fixation point was 1.23º. Disks had a radius of 0.6º and were equally distanced from each other. The color pool comprised 12 colors and it was obtained from Vries et al. (2017). Colors were firstly determined in DKL color space and later converted to RGB. They were chosen to have the same contrast and luminance, differing only in hue. The 12 colors were discrete colors from an imaginary wheel in which consecutive colors were more similar than colors at the other extreme of the wheel (see methods in de

Vries et al., 2017, for a detailed description of color extraction). Four non-consecutive colors were then randomly chosen for the memory set of each trial.

General Procedure

All participants completed two experimental sessions. At the beginning of the first session, they received instructions and practiced each condition separately for 17 trials. Additionally, they performed one extra practice block of 18 trials with all the three conditions mixed. Subsequently, they completed 17 experimental blocks of 18 trials each with a total of 306 trials, 102 in each condition. For the experimental blocks, the three conditions were always mixed. Importantly, in this first session, the CTI was manipulated to examine the time course of performance as a function of load reduction during the second delay. Four CTIs were chosen: 50, 400, 1000 and 1300 ms, with a mean of 25.5 and a minimum of 21 experimental trials for each combination of CTI and condition per participant.

For the second session (i.e., the EEG session), participants were given the same instructions and administered three short practice blocks of 10 trials, one for each condition. In this session, they completed 900 experimental trials, 300 of each condition, presented in 45 experimental blocks of 20 randomly shuffled trials. The CTI had a fixed duration of 1000 ms. Before beginning the practice trials, participants were explicitly encouraged to look at the fixation point, to avoid eye movements or blinking, and to keep their eyes as still as possible during the experiment. At the end of the practice, the light was dimmed, the door was closed, and EEG was recorded while they completed the experimental trials.

Data Recording and Preprocessing

EEG data were recorded using BrainVision Recorder (Brain Products, 2020b) with a 64-channel ActiCap (Brain Products, 2020a) setup at 1000 Hz following the 10-20 system with active online referencing to the right mastoid. Impedances of the Ag/AgCl electrodes were kept below 5 kΩ. Bilateral horizontal EOG electrodes were placed on the outer canthi, together with two vertical EOG electrodes above and below the left eye. All preprocessing steps and analyses were performed in MATLAB, version R2020a (MATLAB, 2020). Preprocessing was performed using custom code in conjunction with EEGLAB (Delorme & Makeig, 2004). For the MVPA analysis, the Amsterdam Decoding and Modeling toolbox (ADAM; Fahrenfort et al., 2018) was used.

EEG raw data were first imported into EEGLAB (v2021.0) using the NE EEGLAB NIC plugin and re-referenced to the average of the left and right mastoids. The data were then downsampled to 200 Hz, followed by high-pass filter of 0.01 Hz using the pop_eegfiltnew function. This filter was chosen because of the artifacts that can occur when less conservative filters are applied to the data when performing multivariate classification, especially when trial durations are long as in this case (van Driel et al., 2021). Epochs from -0.45 to 2.8 s were extracted, locked at the beginning of the first cue until the end of the second delay. A baseline correction was applied, which spanned the 250 ms prior to the first cue. Independent Component Analysis (ICA) was then run with the compute_ICs_new function of the ADAM toolbox using the pop_runica function of EEGLAB. Eye blink components were removed using the ADJUST plugin and components were visually inspected to ensure that only the blink components were removed from the data. Errors and no responses were excluded from the analyses.

Behavioral Data Analysis

Behavioral data were preprocessed in R-studio (RStudio Team, 2022) and analyzed with JASP version 0.16.2.0 (JASP Team, 2022) for Reaction Times (RT) and A, a non-parametric measure of sensitivity that takes into account hits and false alarms (Zhang & Mueller, 2005). This measure is derived from signal detection theory (Green & Swets, 1966), and resolves some of the problems associated with d' and its non-parametric counterpart A' (Pollack & Norman, 1964). The R code to compute A is in OSF, and the formula is:

$$A = \begin{cases} \frac{3}{4} + \frac{H - F}{4} - F\,(1 - H) & if \quad F \leq 0.5 \leq H\,; \\ \frac{3}{4} + \frac{H - F}{4} - \frac{F}{4\,H} & if \quad F \leq H < 0.5\,; \\ \frac{3}{4} + \frac{H - F}{4} - \frac{1 - H}{4\,(1 - F)} & if \quad 0.5 < F \leq H\,, \end{cases} \tag{1}$$

where H is the hit rate and F is the false alarm rate for a given participant and condition (Zhang & Mueller, 2005). Participants with an accuracy of less than 0.6 (the proportion of trials in which a correct response was given) in any of the three main conditions in the first session were excluded from all analyses and were not contacted for the second session. When Mauchly's sphericity test reached statistical significance, the Greenhouse-Geisser correction was applied. For post-hoc tests, the Holm-Bonferroni correction was used. For the data from both sessions, trials with errors and non-responses were removed from the RT analysis. This accounted for 13.25% and 14.86% of the data from the first and second session, respectively. We also excluded trials with responses below 250 ms, above

1900 ms, or 3 SD from the participant's mean. Based on these criteria, 0.27% and 0.25% of the data from the first and second session, respectively, were removed.

For the behavioral analysis of the first session, we analyzed the time course of the different conditions during the second delay. Data from the participants who completed both sessions were analyzed. For the final sample (N= 39), two repeated measures ANOVAs were conducted with Condition (3 levels: retro-cue, low load, and high load) and CTI (4 levels: 50, 400, 1000, and 1300 ms) as within-participant factors and RTs and A as dependent variables. In Supplementary Materials we present analyses including all participants who successfully completed the first experimental session (52 participants). The main results were replicated for the total sample.

For the second session, the CTI was set at 1000 ms. Here, RTs and A were entered into one-way ANOVAs with Condition (3 levels: retro-cue, low load, and high load) as the main factor.

All plots were made in R (R Core Team, 2021) with RStudio (RStudio Team, 2022) using the ggplot2 package (Wickham, 2016) and the Rainclouds tool (Allen et al., 2021).

Multivariate Pattern Analysis (MVPA)

We analyzed data from the 39 participants who completed the EEG session. To assess the effects of the different conditions on eye movements, we performed separate analyses of the fifty-nine EEG electrodes (excluding EOG) and the four EOG electrodes. Analyses on EOG data were performed on VEOG and HEOG channels, after subtraction of vertical (i.e., lower minus upper ocular electrode) and horizontal (i.e., right minus left ocular electrode) activity. MVPA analyses were performed with the Amsterdam Decoding and Modeling toolbox (ADAM; Fahrenfort et al., 2018). All analyses used Linear Discriminant Analysis (LDA) to predict the different conditions from the EEG data. Analyses were performed from the first cue until the target was presented, including the first delay (600 ms – 1600 ms epoch) and the second delay (1800 ms – 2800 ms epoch). Therefore, the entire epoch lasted 2.8 seconds. The data were downsampled to 40 Hz, resulting in 70 time points of interest. In addition, between-class and within-class balancing was applied prior to the analyses. Within-class balancing was applied by ensuring that the cue directionality of each condition was equally represented within each stimulus class. Small between-class imbalances were corrected to ensure that all classes were equally represented (e.g., the same number of trials in the 2- and 4-load conditions). In the current data, the mean percentage of oversampled trials was 6.03%, 0.36%, and 17% for the 4/2-load, 2-load, and 4-load conditions, respectively. Main analyses were repeated downsampling the number of trials used for decoding in each condition to that of the condition with the lowest number of trials, for each participant. Results are similar to the oversampling correction (see Figures S3 and S4).

In analyses where two classes were compared (e.g., 2-load versus 4-load), a 10-fold cross-validation method was used, where all trials were randomly distributed in the 10 folds. Within each participant, the model was trained on 90% of the trials at a specific time point and then tested on the remaining 10% of the trials (fold) at the same time point. This procedure was repeated 10 times, until each fold was tested once. The classifier performance was obtained by averaging the performance of all folds. This process was repeated for each time point in the epoch and for each participant. To estimate performance, we used the Area Under the Curve (AUC) of a Receiver Operating Characteristic (ROC), which is the area covered when plotting the cumulative probabilities of a class being classified as the class it belongs to (i.e., the true positive rate) against the cumulative probabilities of being classified as another class (i.e., the false positive rate). The AUC shows how well the classes are distinguished by the model and it goes from 0.5 (classification at theoretical chance level) to 1 (perfect classification), regardless of the number of classes in the analysis. Unlike binary classification accuracy, the AUC takes into account the confidence to classify each individual case, i.e., the distance from the decision boundary. This allows each case to be weighted according to the confidence with which it has been classified. To check that that theoretical chance level was appropriate, we calculated empirical chance level for the 2-load versus 4-load classification analysis. The AUC derived from this analysis provides an estimation of a sample-size dependent chance-level classification AUC (Combrisson & Jerbi, 2015). This was calculated by decoding two classes but shuffling the labels of the conditions. We performed ten iterations of this analysis and, on average, it performed almost exactly at the theoretical chance level (0.5010 vs 0.5). The variation in empirical chance across participants for the peak point of decoding based on these iterations was relatively low (min: 0.481; max: 0.531; M: 0.503; SD: 0.014). Thus, to reduce computational costs, the theoretical chance level was chosen for the analyses (see Figure S2).

The AUC was computed for each time point for each participant. To obtain group-level results, a t-test was performed at each epoch time point between each participant's AUC values and the chance level (0.5). To deal with multiple comparisons in EEG data (large number of time points), the cluster-based random permutation testing (Maris & Oostenveld, 2007) was used. This method computes the probability of finding an observed cluster size (i.e., significant contiguous t-tests) under random permutation. The size of a cluster is determined by the sum of the t-values of that cluster. The significance threshold we used was $p<0.05$, both for the individual t-tests and for the cluster-based analysis (for a detailed description of cluster-based random permutation testing see Maris & Oostenveld, 2007). Three types of analyses were conducted. First, several two-classes decoding analyses were performed. These analyses aim to show whether two classes can be distinguished based on the multivariate EEG data at different time points. Second, we also performed temporal generalization analyses, in which training and testing were performed on all possible combinations of epoch time points, generating what is known as a temporal generalization matrix. This matrix contains the performance for all possible combinations of training and test times and provides information about the stability/dynamics of brain activity over time (King & Dehaene, 2014). Finally, to assess the time course of the retro-cue effect on load, we trained the LDA algorithm in the two load conditions (2-load versus 4-load) and tested this classifier in the third condition (4/2-load). This analysis was used to determine whether the 4/2-load condition could be classified differently as one or the other condition at the time points of the delays. All significant latencies were relative to the onset of the initial cue.

Results

Our predictions were as follows. At the behavioral level we expected performance in the 4/2 load condition to, over time, transition from resembling the 4-load baseline to the 2-load baseline. At the electrophysiological level we expected something similar. First, when applying MVPA to the maintenance delays of the *2-* and *4-load* baseline conditions, we expected to replicate Adam et al. (2020) findings, with VWM load being decodable from the multivariate EEG signal. Second, to investigate the time course of the removal stage, we then used these baselines as a comparison for the 4/2-load condition, where a retro-cue signaled only half of the information as relevant. We hypothesized that if the uncued information was indeed removed from VWM, the load-related signal should first resemble the high-load baseline, but over time start to resemble more and more the low-load baseline.

Behavioral Time Course (Session 1)

Table 1 shows the descriptive statistics for the RTs and A (Zhang & Mueller, 2005) in the first experimental session including only participants who completed both sessions (N = 39)[1]. During this session, the time course of performance was assessed as a function of CTI. For the RTs, we observed a main effect of CTI, $F_{(2.25, 85.63)} = 58.745$, $p < .001$, $\eta_p^2 = .607$, with overall RTs being faster with longer CTIs and stabilizing at the 1000 ms CTI. We also observed a main effect of Condition $F_{(2, 76)} = 48.846$, $p < .001$, $\eta_p^2 = .562$, where the RTs were faster for the 2-load condition (M =

---

[1] Analyses with the whole sample that completed the first experimental (N=52) session are in Supplementary Materials and the main interactions and post-hoc tests show the same results.

746, *SD* = 130) than for the 4/2-load and the 4-load conditions (*M* = 758, *SD* = 130 and *M* = 762, *SD* = 127, respectively). As expected, the Condition x CTI interaction reached statistical significance, $F_{(6, 228)} = 3.564$, $p =.002$, $\eta_p^2 = .086$, showing different time courses for the three conditions. In the short 50 ms CTI, post-hoc tests showed that the 4/2-load condition differed significantly from the 2-load condition, $t_{(38)} = 3.904$, $p = .004$, $d =.366$, but not from the 4-load condition, $t_{(38)} = 2.222$, $p = .464$, $d = .208$. However, from the second CTI onwards, the tendency reversed until the longest CTI, with statistically significant differences between the 4/2-load condition and the 4-load condition (CTI 2: $t_{(38)} = 4.657$, $p < .001$, $d= .436$; CTI 3: $t_{(38)} = 6.033$, $p < .001$, $d= .565$; CTI 4: $t_{(38)} = 6.174$, $p < .001$, $d= .579$) , but not the 2-load condition, (CTI 2: $t_{(38)} = 1.301$, $p = 1$, $d = .122$; CTI 3: $t_{(38)} = 1.572$, $p = 1$, $d= .147$; CTI 4: $t_{(38)} = 0.823$, $p = 1$, $d= .077$). Thus, the development of RTs during delay 2 showed how the 4/2-load condition gradually distanced itself from the 4-load condition and ended up being comparable to the 2-load condition. This has been interpreted as the removal of information from WM (e.g. Souza et al., 2014), and the results here suggest it takes about half second.

For A values (sensitivity measure), a repeated-measures ANOVA showed a main effect of Condition, $F_{(1.68, 63.77)} = 47.427$, $p < .001$, $\eta_p^2 = .555$, with the highest mean A for the 2-load condition (*M* = .96, *SD* = 0.05), lower for the 4/2-load condition (*M* = .91, *SD* = 0.08), and the lowest performance for the 4-load condition (*M* = .87, *SD* = 0.1). However, the main effect of CTI did not reach significance, showing that overall performance did not change across the second delay, $F_{(3, 114)} = 1.287$, $p = .282$, $\eta_p^2 = .033$. Similar to the RT analysis and in line with what was expected, a significant CTI x Condition interaction was observed, $F_{(6, 228)} = 3.214$, $p = .005$, $\eta_p^2 =$

.078. Post-hoc comparisons showed a similar pattern to the one observed with RTs: the 4/2-load condition and the 4-load condition were equivalent at the shortest CTI ($t_{(38)} = .375$, $p = 1$, $d = .067$) and differed in the second and third CTIs ($t_{(38)} = 4.517$, $p < .001$, $d = .804$ and $t_{(38)} = 4.588$, $p < .001$, $d = .817$, respectively) but not in the last CTI ($t_{(38)} = 2.960$, $p = .102$, $d = .527$), and the reverse was true for the 4/2-load condition and the 2-load condition, which differed at the first delay ($t_{(38)} = 2.960$, $p < .001$, $d = 1.033$), became equal in the second and third interval ($t_{(38)} = 3.144$, $p = .062$, $d = .560$ and $t_{(38)} = 2.248$, $p = .536$, $d = .400$, respectively) and differed in the last one ($t_{(38)} = 3.578$, $p < .016$, $d = .637$). Thus, we observed a similar pattern of results, that is, a general improvement in performance for the 4/2-load condition with longer CTIs (see time course of RTs and A in Figure 2).

Figure 2. Reaction Times (RT) and mean A for the time course during the second delay for participants who completed both sessions, N=39. Small colored points represent the mean of each participant in each condition. Points with black border represent the mean of each trial type for each interval. Error bars represent standard error of the mean. The individual distributions for all the trial types x CTI combinations are represented in the split violin plots.

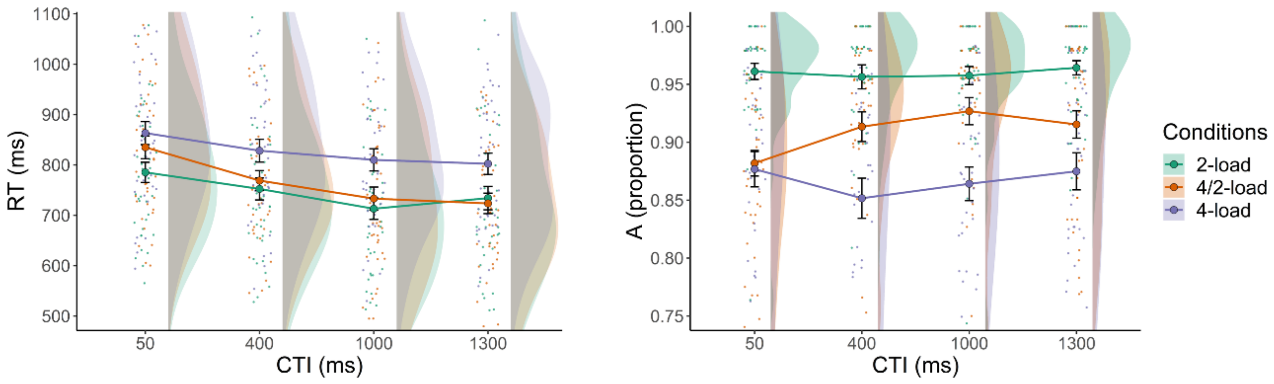


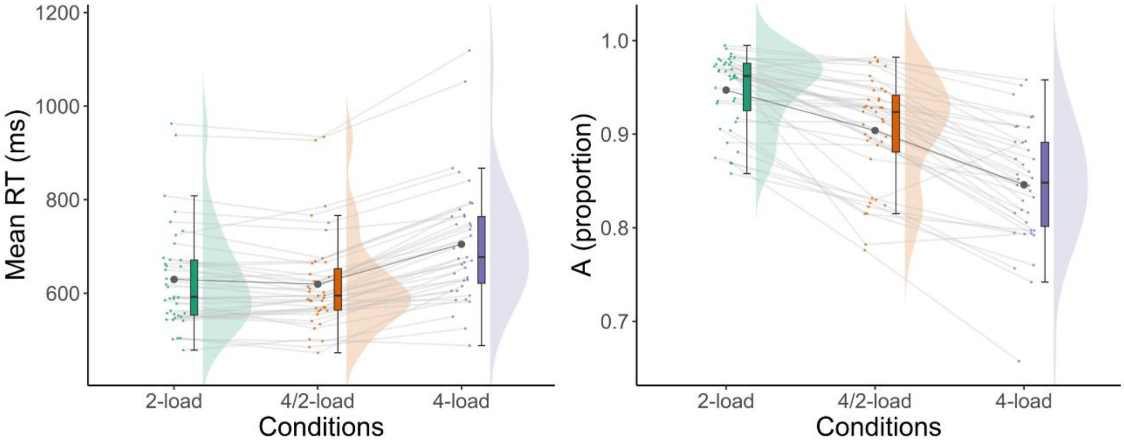
Table 1. Descriptive statistics of behavioral data from participants who completed both experimental sessions, N=39. Reaction Times (RT) and mean A (A) split by trial type (Condition) and Cue-Target Interval (CTI).

		RT		**A**	
Condition	**CTI (ms)**	**Mean**	**SD**	**Mean**	**SD**
2-load	50	785	125.1	0.961	0.044
	400	752	136.8	0.956	0.065
	1000	713	133.0	0.958	0.049
	1300	734	145.0	0.964	0.038
4/2-load	50	835	141.4	0.882	0.069
	400	769	123.2	0.913	0.080
	1000	733	143.4	0.927	0.073
	1300	724	126.4	0.915	0.074
4-load	50	863	143.2	0.877	0.095
	400	828	141.8	0.852	0.108
	1000	810	137.5	0.864	0.090
	1300	802	132.4	0.875	0.100

Behavioral Retro-cue Effect (Session 2)

In the second session, participants completed the same task but with a fixed 1000 ms CTI. Mirroring the first session results, the main effect of Condition was significant for the RTs, $F_{(1.64,\ 62.39)} = 108.072$, $p < .001$, $n2p = .740$, and A, $F_{(2,\ 76)} = 117.744$, $p < .001$, $n2p = .756$. For the RTs, post-hoc comparisons showed that the 4-load condition ($M = 704.75$, $SD = 127.32$) was significantly slower than both the 2-load condition ($M = 629.68$, $SD = 109.17$), $t_{(38)} = 11.848$, $p < .001$, $d = .661$, and the 4/2-load condition ($M = 619.44$, $SD = 103.05$), $t_{(38)} = 13.463$, $p < .001$, $d = .751$, but the last two were equivalent, $t_{(38)}$ = 1.615, *p* = .111, *d* = .090. For A values, post-hoc comparisons showed a higher A for the 2-load condition (*M* = .95, *SD* = .04) than for the 4/2-load condition (*M* = .90, *SD* = .06), $t_{(38)}$ = 6.527, *p* < .001, *d* = .803, and higher for the 4/2-load condition than for the 4-load condition (*M* = .85, *SD* = .06), $t_{(38)}$ = 8.764, *p* < .001, *d* = 1.078 (see *Figure 3*).

Figure 3. Mean Reaction Times (RT) and mean A in EEG session. Small colored points represent the mean of each participant in each condition, joined by a light grey line. Black points represent the mean of each trial type (conditions). Box plots of each condition are represented. The individual distributions for all the trial types x CTI combinations are represented in the split violin plots.

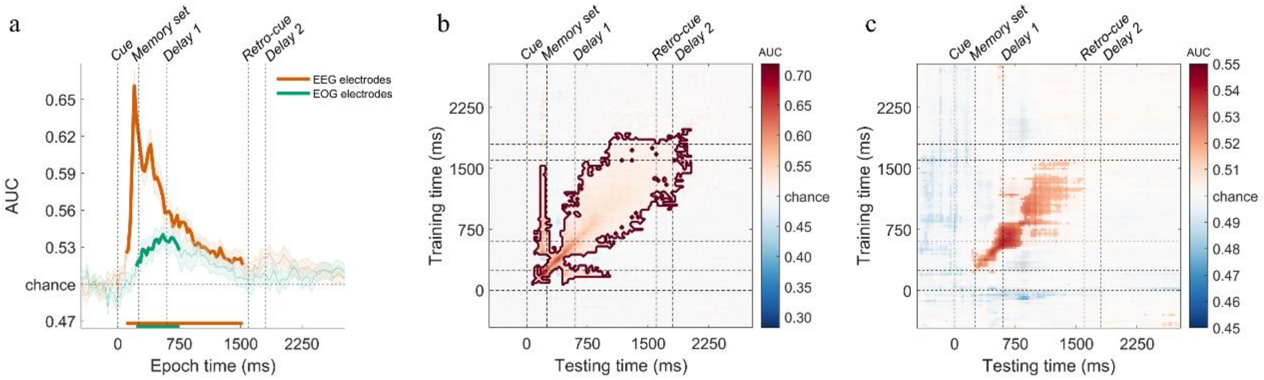


Can the Change in VWM Load be Decoded from the EEG Data?

First, to determine whether different levels of VWM load could be distinguished in the EEG data, MVPA was performed on two classes: 2-load and 4-load. The classifier was trained at each epoch time point to dissociate 4- from 2-load and then it was tested at the same time points, using k-folding (see Method section). The EOG electrodes were excluded from this analysis and the raw data from the remaining 59 EEG electrodes were used. This analysis showed significant decoding from the beginning of the epoch to the end of the first delay (85 – 1535 ms). These results replicate the results of Adam et al. (2020). However, the two conditions became indistinguishable from the appearance of the second cue (the retro-cue) until the end of the epoch (see Figure 4a). Since the second cue was uninformative in these conditions (it should not alter the load), the pattern for this second delay was unexpected, as one would expect to still be able to decode the load even at the second delay. To further study the dynamics of these conditions, a temporal generalization analysis was performed. In this analysis, the algorithm was trained at each time point and tested at all other time points of the epoch. This method was then repeated for all the possible train-test combinations of time points. This allowed us to assess the stability of the pattern. In addition, it allowed us to see whether training at a fixed time point (e.g., at peak activity) improved classification for the remaining time windows. Figure 4b showed a moderate generalization of decoding within the first delay period. However, at the time of the second cue and within the second subsequent delay period, the signal remained weak regardless of the training time point.

Figure 4. Decoding of load from electroencephalography (EEG) and electrooculogram (EOG) signal. AUC = Area Under the Curve. Dotted lines correspond with the onset of the following task moments, from left to right: initial cue, memory set, first delay, retro-cue, and second delay. (a) Diagonal decoding of 2-load versus 4-load conditions from EEG (orange line) and EOG (green line) electrodes. Bold lines show p-values that survived multiple comparisons corrections and shaded area surrounding the line is the standard error. (b) Temporal generalization matrix for 2-load vs 4-load decoding on EEG electrodes. (c) Temporal generalization matrix for 2-load vs 4-load decoding on EOG electrodes. In both (b) and (c), saturated colors reflect uncorrected $p<0.05$ decoding, whereas areas

circumscribed by a dark bold line highlight clusters that survive cluster-based permutation testing at $p<0.05$. Above chance decoding is colored in red color and below chance decoding in blue.



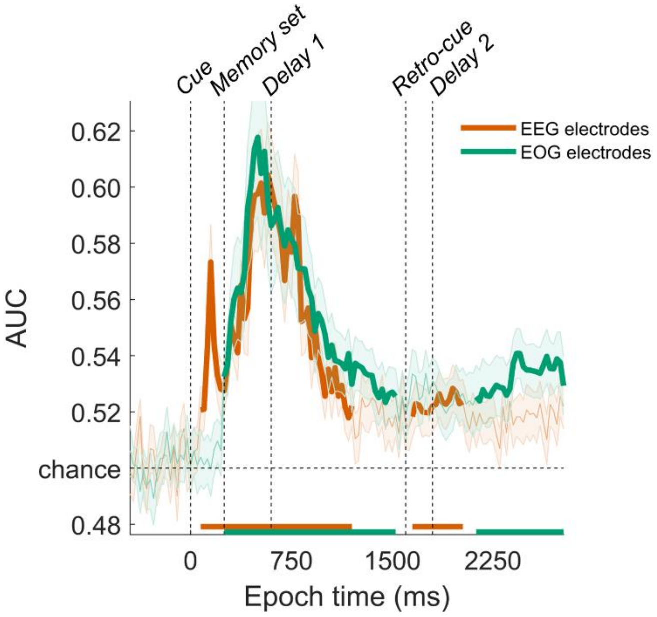
The results observed so far allow for multiple interpretations. On the one hand, the visuospatial information in WM may simply decay, and more so for the 4-load condition, which would then reduce the distinction between the two load conditions (Ricker et al., 2014, 2016). However, behavioral performance renders this unlikely, as a clear difference in performance between the 2-load and 4-load conditions was observed in both the behavioral and the EEG sessions. Another possibility is that decoding was decreased due to the greater distance from the pre-trial baselining period, which could increase noise in the data. However, further analyses showed significant decoding later in the epoch (see Figures 5, 6, and 7), making this unlikely. A third possibility is that some processes triggered by the retro-cue—and other than maintenance—mask the load manipulation decoding. In the present design, trials were randomly presented. In the 4/2-load condition, participants had to select some disks as relevant and drop the rest. In contrast, in the 2- and 4-load conditions—the ones of the present analysis—the second cue instruction was the same in both cases (i.e., a full cross), and indicated that nothing had changed (observers simply had to retain what they already had in memory). Then, it might be that a selection (vs non-selection) component is prevailing over the maintenance at the time of the retro-cue. If it was the case that the retro-cue was driving an attentional selection component that was superimposed on the load decoding, a similar case could apply to the pre-cue. In other words, in addition to load, the selection processes linked to the load manipulation, and triggered by the pre-cue, could in principle also contribute to the decoding during the first delay.

To test this possibility further, and given the close relationship between attentional selection and eye movements (van Ede et al., 2019; Zhao et al., 2012), we repeated the same analyses on the EOG electrodes. To isolate eye movements from brain activity, subtraction for VEOG (i.e., right minus left ocular electrode) and HEOG (i.e., lower minus upper ocular electrode) were computed. Thus, all the EOG analyses were performed on the obtained HEOG and VEOG subtracted channels. The idea behind these analyses was that, if 2- and 4-load conditions could be decoded from the signal coming from the ocular electrodes after the first cue, this would support the idea that attentional selection was also playing a role in the previous decoding analyses. The EOG analysis showed a significant decoding of the 2- and 4-load conditions during the memory set and the beginning of the first delay (0.235 – 0.760 ms; see Figure 4a). The temporal generalization was then also repeated in EOG. Although we observed a similar descriptive pattern as for the 59 EEG electrodes, there was no significant decoding after cluster correction (Figure 4c). This lack of significant decoding after cluster-based correction and the lower AUC values of the present analyses could be explained by the smaller number of variables included (4 EOG channels versus 59 EEG channels above).

To further test the role of selection, we additionally decoded the *directionality* of the initial cue of the 2-load condition. For this, two classes were made based on the diagonal direction of the encoding cue, that is, whether it was tilted to the left or to the right. Figure 5 shows diagonal decoding performed on EEG and EOG electrodes. In the EEG electrodes, the analysis showed significant decoding until the end of the first delay (85 – 1210 ms) and again during the second cue and the beginning of the second delay (1660 – 2035 ms). When the classification was performed at the EOG electrodes, significant decoding extended from memory encoding to the end of the trial (260 – 2785 ms), with a gap around the retro-cue. Again, these analyses showed a strong contribution of eye movements, closely linked to the selection required in the 2-load condition.

Finally, using the *adam_correlate_CONF_stats* function, we computed Spearman correlations at each time point between EEG and EOG classifier confidence scores (i.e. resulting from the distance to the decision boundary) of 2- vs 4-load results across trials for each participant, and tested the Fisher-transformed correlations against zero at the group level. The confidence scores in this analysis represent the single trial evidential support for the classification that the classifier provides at any given time point (see Figure 6). Thus, this analysis tested whether both decoding analyses captured similar information at any given time point.

Figure 5. Diagonal decoding of the direction of the initial cue in the 2-load condition for EEG (orange) and EOG electrodes (green). AUC for Area Under the Curve. Bold lines show p-values that survived multiple comparisons corrections and shaded area surrounding the line is the standard error.

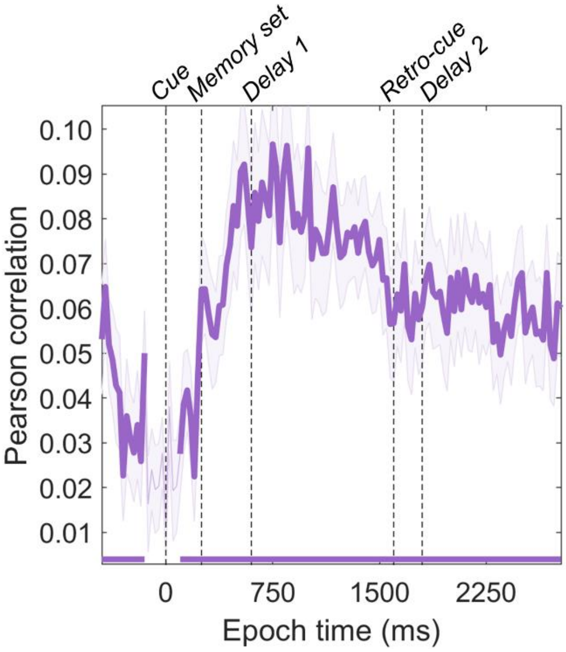


Results show a significant positive correlation during the complete epoch. However, the correlation is numerically small, capturing relatively little shared variance. If we take the EOG analyses as evidence of selection, this analysis supports a contribution of selection to the 2- vs 4-load decoding. The correlation is higher during the first delay, congruent with the decoding results of Figure 4a, and where we expect differences in selection in these conditions. However, given the weak correlation, we can conclude that the decoding analysis cannot be driven by this selection component alone and that, in line with previous experiments, load is also being decoded (Adam et al., 2020; Thyer et al., 2022).

Apart from load and selection, there are two additional factors that could be partly contributing to the present significant decoding in the EEG channels: eye movements and the visual differences in the cue. Regarding eye movements, we interpreted here the EOG decoding as evidence for selection. However, we did not expect eye movements to be the only contribution to the significant decoding of the analyses run with the EEG channels because, in some cases, they may not co-occur with selection (Liu et al., 2022). Robustness analyses removing eye movements showed that decoding remained significant when we removed them before running the 4- vs 2-load decoding analysis (see Figure S3). Thus, we can confirm that eye movements information cannot explain by itself the present analysis results. This is also congruent with the previous correlational analyses (Figure 6), that show a small shared variance between EEG and EOG electrodes. Regarding any visual differences between different pre-cues, earlier work showed that any sensory modulations would occur early, during the first 300 ms after the stimuli presentation (e.g. Jongen et al., 2007; Luck, 2006; Quentin et al., 2019). However, other decoding work showed that differential signals can be observed beyond that (Dijkstra et al., 2018; Noah et al., 2023). Thus, we cannot discard the possibility that these perceptual differences of the pre-cues (and not only the attentional modulation produced by them) are partially driving the delay decoding.

Overall, we replicated previous work decoding visual WM load. However, we argue that differences in attentional selection of relevant items (as partially reflected in the overt eye movements) may also partially contribute to the observed decoding.

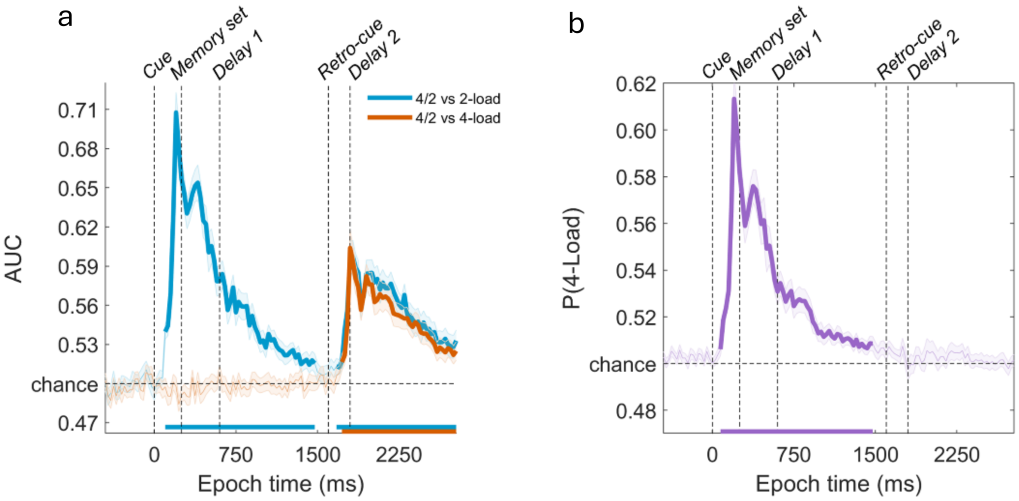
Figure 6. Correlation of EEG and EOG decoding for 2- vs 4-load. Bold lines show p<.05 values for Pearson correlations after cluster correction. Shaded area surrounding the line is the standard error.



Can we Decode the Drop in Load After the Retro Cue?

The second and main goal of the present work was to test the time course with which item information was removed from the VWM. For this purpose, we wanted to compared the 4/2-load condition with, on the one hand, the 2-load condition and, on the other hand, the 4-load condition (Figure 7a). If from the current data the load had been decoded, we would expect (1) the 4/2-load condition would not be distinguishable from the 4-load condition during the first delay, prior to the cue, while it would be distinguishable from the 2-load condition; (2) the 4/2-load would be similar to the 4-load condition and would be distinguishable from the 2-load condition at the beginning of the second delay and after the cue. As the second delay develops, this pattern should reverse, as the 4/2 load condition should begin to resemble more like the 2-load and less to the 4-load condition.

Figure 7. (a) Diagonal decoding 4/2-load vs 2- and 4-load conditions (blue and orange, respectively). AUC = Area Under the Curve. Bold lines show p-values that survived multiple comparisons corrections and shaded area surrounding the line is the standard error. (b) Training the 4- vs 2-load conditions and testing the 4/2- load condition. When classification is above chance, the 4/2-load condition is classified as 4-load and when it is below chance, the 4/2-load condition is classified as 2-load.

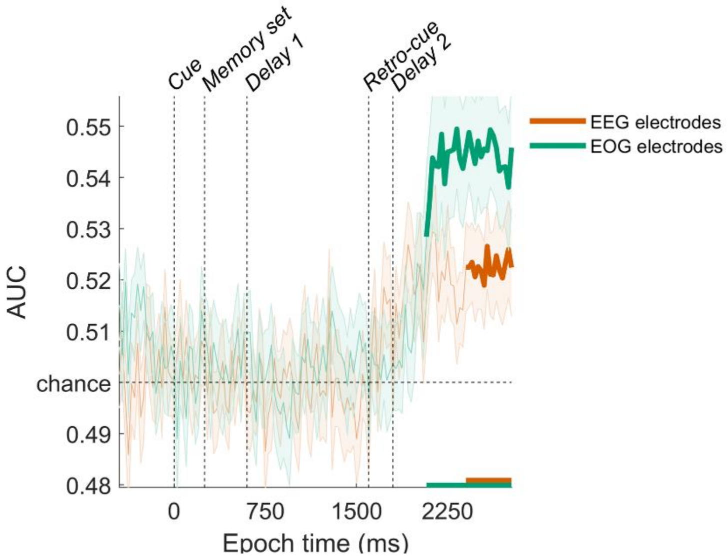


Thus, in a first step, we trained two linear discriminant classifiers at each epoch time point: one classifier was trained to discriminate between the 4/2-load and 2-load conditions, while the other was trained to discriminate between the 4/2-load and 4-load condition. These multivariate analyses showed significant decoding during the first delay between the 4/2-load condition and the 2-load condition (110 – 1485 ms), but not between the 4/2-load and the 4-load conditions. In other words, prior to the second cue, the 4/2-load condition resembled the 4-load condition more than the 2-load condition, as would be expected. Analyses of the second delay revealed above-chance decoding between the 4/2-load and 2-load conditions

(1685 – 2785 ms), as well as between the 4/2-load and 4-load conditions (1735 – 2785 ms) during virtually the same time windows. That is, the 4/2-load condition could be distinguished from both baseline conditions during the whole second delay and there was no sign of a transition from high to low load. This is congruent with previous studies that showed sustained decoding of an attentional selection component after a retro-cue (Quentin et al., 2019). For the sake of completeness, we performed an additional analysis in which the algorithm was trained on the baseline conditions (4-load vs. 2-load), and then tested in the 4/2-load condition at each time point, the result of which is fully in line with what would be predicted given the results in Figure 4a. This analysis allowed us to check at each time point whether the 4/2-load trials were classified as 4-load or 2-load. As in previous analyses, the 4/2-load condition could be classified as the 4-load condition during the first delay (85 – 1485 ms), but could not be classified as either category during the second delay (Figure 7b). These results confirm that the load reduction hypothesis could not be tested, given the lack of decoding of the 2-load and 4-load conditions during the second delay. Although we cannot discard the possibility that load was partly contributing to this decoding, the present results for the second delay are congruent with the multivariate analysis capturing the effects of cued-induced selection, as only the retro-cue condition required further selection for the second maintenance period.

To further support the conclusion that the second delay results were driven by the selection of information (this time, within the VWM), we trained linear classifiers to decode the directionality of the retro-cue (i.e., right- or left-tilted cue). This analysis was performed independently on EEG and EOG electrodes. The classifier found significant information about the directionality of the retro-cue at EEG electrodes (2410 – 2785 ms) and also at EOG electrodes (1985 – 2785 ms) (Figure 8). Since the retro-cue disappeared from the screen at 1800 ms, the fact that significant decoding was detected especially towards the end of the delay showed that participants could use eye movements to support the maintenance of the relevant items active in VWM (van Ede & Nobre, 2022). The present results confirm that it is not a reduction of the VWM load, but the selection of relevant information within the VWM that underlies the decoding of the second delay results.

Figure 8. Diagonal decoding of the directionality of the retro- cue in the 4/2-load condition for EEG (orange) and EOG electrodes (green). AUC for Area Under the Curve. Bold lines show p-values that survived multiple comparisons corrections and shaded area surrounding the line is the standard error.

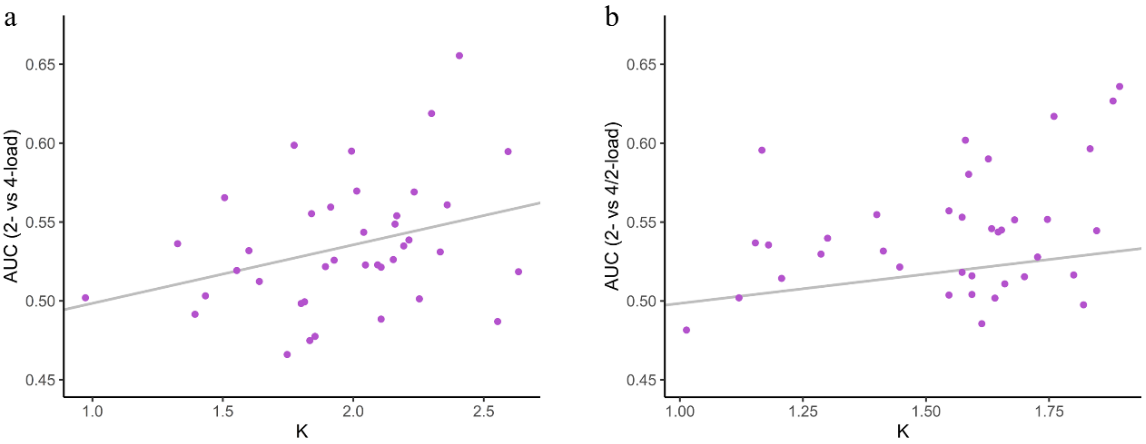


Correlation of Classifier Performance and VWM Capacity

To test whether participants with higher working memory capacity might show both better maintenance and better selection of items, we ran two additional correlational analyses. First, we correlated a measure of VWM capacity with the AUC from the MVPA analysis to classify 2- and 4-load trials (see Figure 4a). We know from previous work that decoding of 2- vs 4-load reflects load decoding (Adam et al., 2020; Thyer et al., 2022), and we also know from previous analyses that it might partially reflect the selection of items driven by the initial cue (see *Results* section). To provide a measure of visuospatial working memory capacity, we calculated the estimated number of items remembered for each participant using the Cowan's K formula (Cowan, 2001; Rouder et al., 2008): $K = (H - FA) * N$, where K is the number of items remembered, H and FA are the hit and false alarm rates, and N is the number of item presents to be remembered. As for the AUC measure, the first delay decoding was chosen because (1) it was the only time at which the classifier could differentiate conditions, and (2) to be as equivalent as possible to Adam et al. (2020). The Pearson correlation of the 39 participants who completed both experimental sessions showed a significant positive trend between the two variables, $r = .329$, $p = 0.041$, 95% CI = [.584, .015], depicted in Figure 9a.

Additionally, we also performed a correlation where we explicitly tested whether decoding of selection would also be correlated with working memory capacity (K). For that, we used the second delay AUC of decoding of 2- and 4/2-load conditions and mean K of these same two conditions. The reason for choosing the second delay of these two conditions is because the same amount of load is relevant by the end of the delay (seen in K), but one of them has required a selection within WM of this information, the 4/2-load condition. Therefore, this analysis should largely restrict to selection. The Pearson correlation showed a significant positive trend, $r = .420$, $p = 0.008$, 95% CI = [.649, .120], depicted in Figure *9b*. This results show that working memory capacity does not correlate only with load maintenance, but also with successful items selection.

Figure 9. (a) Correlation of decoding for 2- vs 4-load and K during the first delay. (b) Correlation of decoding 2- vs 4/2-load and K during the second delay. Each point represents a single participant. The grey line represents the regression line of the correlation.



Discussion

During the past decade, the cognitive neurosciences have seen an upsurge in the application of Multivariate Pattern Analysis (MVPA) to data from electro- and magnetoencephalography (EEG and MEG, respectively). Contrary to univariate methods, MVPA allows for the decoding or identification of different representations or states by considering multidimensional patterns of sensor activity, which makes it a highly sensitive technique (Grootswagers et al., 2017; Haxby et al., 2014). This has made MVPA a particularly popular tool in the field of working memory research, where it is used to track item-related representations and different memory states during delays, when the stimulus is absent (e.g. Bae & Luck, 2018; Bocincova & Johnson, 2019; King et al., 2016; LaRocque et al., 2013; Rose et al., 2016;

Trübutschek et al., 2017; Wolff et al., 2015, 2017). However, the increased sensitivity and complexity comes with a price tag, as it is not always evident what the source is of the information that is being used to successfully decode. For instance, it has been shown that artifacts introduced by regular cleaning steps such as high-pass filtering can lead to spurious decoding (van Driel et al., 2021), as can unintended eye movements (Mostert et al., 2018; Quax et al., 2019). This may lead to potential confounds that can hinder or, in the worst case, invalidate the conclusions of a study, in that the multivariate patterns may not capture the mental state that they were intended to capture. The present study serves as another case in point.

We attempted to unravel the time course of the reduction in VWM load after observers have been cued which information needs to be retained, while other information can be dropped. To this aim, participants completed a visuospatial working memory task involving delayed recognition of colored stimuli in which VWM load was manipulated in three conditions: two or four items maintained throughout the trial (low and high load conditions, respectively), or two retro-cued items out of four initially encoded items, allowing a transition from high to low load. Two main results were obtained in the current study. First, the time course of behavioral performance (during the pre-EEG session) suggests that, following the cue, it takes about half a second to exclude irrelevant information that may affect behavior. Specifically, we varied the CTI from 50 to 1300 ms, and found that a retro-cue that told participants that only two of the four items would need to be retained caused RTs to drop from close to load 4 levels to close to load 2 levels within about 500 ms. To our knowledge, this is the first time that the time course of visual working memory has been tracked. A similar pattern was previously reported by Oberauer (2018) using verbal stimuli, who observed a time course suggesting that participants needed around one second to drop the irrelevant information. Such timing differences could reflect not only the type of information (verbal versus visuospatial), but also the amount of information selected/removed, which was three out of six items in Oberauer (2018) and two out of four items in the present design. Future studies should take into account the variations in this time course by modifying the amount of information maintained and selected.

In a subsequent step, we sought to track the same time course of load reduction using EEG measurements. First, we replicated previous results that found that the MVPA analyses were sensitive to load (Adam et al., 2020). However, we found that decoding can also be sensitive to selection processes during the same epochs where one might expect load to change. Several characteristics of our results support this idea. First, during the time windows in which we observed a reliable load effect in the EEG, we also observed a reliable load effect in the EOG electrodes, suggesting the involvement of eye movements. Moreover, these measures were positively correlated, albeit weakly. This is consistent with the close link between selection of items within VWM and eye movements (van Ede et al., 2019). Additionally, the pre-cue directionality was also decoded from both EEG and EOG electrodes, supporting the enrollment of selection processes that were to some extent captured by the classifier. Second, during the second delay we predicted a transition in the 4/2-load retro-cue condition from being classified as more similar to the high-load condition (load four) at the beginning of the delay, to being classified as more similar to the low-load condition (load two) towards the end of the delay. In contrast, the 4/2-load retro-cue condition remained distinct from both baseline conditions throughout the delay and, if anything, the two baseline conditions were more similar, as indicated by weak classification performance. What distinguishes the retro-cue condition from the two baseline conditions is the selection component. Third, during the same time window we were able to distinguish the directionality of the retro-cue in both EEG and EOG electrodes, which allowed us to confirm that there is a general selection mechanism behind both cues. This is congruent with the conceptualization that proposes that selecting content maintained in working memory relies on processes similar to those involved when attending to perceptual information (Gazzaley & Nobre, 2012; Kiyonaga & Egner, 2013; Panichello & Buschman, 2021; van Ede & Nobre, 2022). Below we elaborate on this argument.

Load Effects: Retention versus Selection

For our argument, it is important to clearly define what we mean by load and selection. Note that load as such is an independent variable, which is then assumed to affect certain mental processes. It is probably fair to say that when they think of load, researchers tend to think of retention (or maintenance), specifically the number of items that are held in some active state in working memory. Indeed, studies that have used MVPA to decode load also appear to interpret load this way (Adam et al., 2020; Feldmann-Wüstefeld, 2021; Thyer et al., 2022). This is also the definition of load that we adopt in the current work. From this conceptualization, retention is a stable and sustained process, which is reflected in neural indicators such as CDA (Vogel & Machizawa, 2004), negative slow wave (Fukuda et al., 2015), or load decoding temporal generalization (Adam et al., 2020). Note that more dynamic maintenance processes have also been proposed (Miller et al., 2018; Stokes, 2015). However, even those dynamic trajectories are stable and reproducible.

On the other side, we define selection as a relatively transient, item-specific operation that transforms an item into another state for further processing. Examples of transition operations are separating items from distractors (i.e., input gating; Chatham & Badre, 2015; Frank et al., 2001; Nir-Cohen et al., 2020), encoding or consolidating information (i.e., transforming them into a maintenance state; Ricker et al., 2018; Woodman & Vogel, 2005), refreshing or updating relevant items (e.g. Kessler & Meiran, 2008; Lemaire et al., 2018), dropping items after a retro-cue raising or lowering the representational state (e.g., Kruijne et al., 2021; Oberauer, 2018), and retrieving items for response (i.e., output gating; Chatham et al., 2014). Many of these operations are assumed to involve some form of attention, in the sense of prioritizing or enhancing some representations over others. While these operations—and therefore also the selection process—are assumed to be transient (Woodman & Luck, 1999, 2003), this does not imply that they would not occur and be measurable throughout longer periods into the delay. First, such selection processes may last longer than typically assumed especially when they concern “internal selection”, within memory (e.g., Oberauer, 2018; Quentin et al., 2019; Ricker et al., 2018; see also the time course after the retro-cue in our behavioral results). Second, even if the processes are transient and short-lived, they may occur at a delay, or occur repeatedly within working memory, especially when multiple items need to be selected for the next operation (e.g., Lemaire et al., 2018; Vogel et al., 2006; Vogel & Luck, 2002).

As a last remark, it is noteworthy that some researchers have argued that maintenance itself is a repeated attentional serial selection process to refresh items and raise their strength (Camos et al., 2018; Kiyonaga & Egner, 2013; Lepsien & Nobre, 2006; Olivers, 2008; Rac-Lubashevsky & Frank, 2021; Souza et al., 2015; Souza & Oberauer, 2017). If so, then any measured load effects would essentially be selection effects and vice versa. However, we would then have expected decoding during the second delay period in our study to be sensitive to the remaining load, which is not what we observed. Thus, to what extent selection and maintenance overlap or are independent processes is still a matter of debate.

We note that positive correlations between load and selection exist for most if not all of the experiments reanalyzed in Adam et al. (2020). Many of these experiments involved an initial cue pointing to the half of the display containing the set to be memorized. The size of this set was then varied to induce different loads. This then also involves different numbers of items to be selected for encoding and memory consolidation. Some other recent works trying to decode load also used experimental designs where load depended on selection operations, since only some colors or geometric figures needed to be maintained among other stimuli that had to be ignored (Feldmann-Wüstefeld, 2021; Thyer et al., 2022). Even when displays do not include spatial or feature cues, simply varying set sizes still implies varying the number of items that need to be selected and consolidated from a display. The same argument also goes for another piece of evidence reported by Adam et al. (2020) in support of load decoding, which was a positive correlation between classifier performance and individual working memory capacity. However, this correlation could also reflect more effective selection mechanisms, as previous work has shown that VWM capacity may actually reflect attentional filtering efficiency (Fukuda & Vogel, 2009; Vogel et al., 2005). We performed two additional analyses of our data where we correlated, for each participant, a measure of the mean number of items in memory (K) with the mean AUC of two decoding analyses that reflect the initial load manipulation (with the potential of perceptual selection) or selection from VWM (i.e., first delay 2- versus 4-load and second delay 2- versus 4/2-load, respectively). We too observed significant positive correlations for both analyses (see Figure 9), supporting the idea that high-capacity individuals may also be better selectors in addition to better loaders. Our point that decoding may at least in part reflect selection mechanisms does not in any way exclude the possibility that retention load was being decoded from those experiments. What we call for is for future experimental designs to focus on the dissociation between these two processes.

One might argue that Adam et al. (2020) already controlled for attention effects in their Experiment 3, where they compared a VWM task in which a number of lateralized colored squares had to be remembered to a sustained spatial attention task, in which the locations indicated by the same number of lateralized colored squares had to be monitored for the appearance of an unrelated visual target (a small line segment). Adam and colleagues found an effect of the number of colored squares in the VWM condition that lasted throughout the 1300 ms delay period. In the attention condition, the number of squares could also be decoded, but for about half that time period, around 700 ms. Adam et al. therefore concluded that load decoding is caused by more than attention alone. However, this still leaves a decent, relatively long-lasting attention effect. Finally, it is also noteworthy that

Adam et al. also observed strong cross-task decoding between the VWM and the attention task for those same 700 ms, suggesting shared mechanisms for at least the first half of the delay period.

A number of additional issues remain. First, it is possible that we failed to decode retention load because VWM activity fell "silent" during the first delay (Oberauer & Awh, 2022; Stokes, 2015). We cannot exclude this possibility, but we note that this delay was not that long (1000 ms), and others showed significant decoding with similar or longer delays (e.g. Feldmann-Wüstefeld, 2021; Thyer et al., 2022). Second, within the current design the 4/2 retro-cue may create an additional overall task demand, and the classifier may pick up on that rather than the selection operation per se. This could be addressed in future work by parametrically varying the number of items to be retained as indicated by the retro-cue. This would also help to further dissociate the relative contributions of selection versus retention processes. Third, we observed a contribution of eye movements to the decoding see Figure S3. While such eye movements are consistent with selective attention operations (but may also play a role in maintenance), future studies should include additional eye-tracking measures to remove eye movements more accurately (Quax et al., 2019). Fourth, although we tried to minimize the visual differences of the cues, we observed an early peak in decoding with the appearance of the pre-cue, before the memory delay was presented (see Figures 5 and 7), and thus suggestive of differential perceptual signals. While for univariate EEG studies, these perceptual signals tend to emerge early (< 300 ms; Jongen et al., 2007; Luck, 2006), the case might be different for multivariate analyses, where decoding of such signals may last further into the epoch (e.g., Dijkstra et al., 2018; Noah et al., 2023; but see

Goddard et al., 2022; Quentin et al., 2019 for shorter perceptual decoding). Also using the retro-cue paradigm, Quentin et al. (2019) compared the decoding of two cues: one that was simply perceived with one that required selecting certain stimuli for posterior retrieval. They found that the cue that implied selection showed a longer significant decoding (~ 1.5 s) than the cue that was merely perceived (~ 0.5 s). Even though we have no “cue perception” condition to make this direct comparison in the present data, our sustained decoding results would be in line with the sustained decoding of selection after the cue observed by Quentin et al. (2019). However, as stated before, the contribution of perceptual differences to decoding may vary between experiments. This will need to be taken into account in future studies attempting to differentiate between working memory load and working memory operations.

Conclusion

The original goal of the present study was to track the time course of changes in visual working memory (VWM) load when part of the stored information in VWM is no longer relevant. While the behavioral performance suggested a gradual drop in load, we failed to observe an EEG correlate of this load reduction after the retro-cue. Instead, the multivariate analyses of EEG and EOG electrodes suggested an important role of attentional selection to flexibly update the relevance of VWM content. The present results are in line with previous work supporting the role of attention as a mechanism for information selection and prioritization within VWM (e.g. Astle et al., 2012; Murray et al., 2013; Panichello & Buschman, 2021; Serin & Günseli, 2022). Finally, the current results advise caution when using MVPA to track

VWM retention per se, as other mechanisms may also contribute to the classification outcome.

References

Adam, K. C. S., Vogel, E. K., & Awh, E. (2020). Multivariate analysis reveals a generalizable human electrophysiological signature of working memory load. Psychophysiology, 57(12), e13691. https://doi.org/10.1111/psyp.13691

Allen, M., Poggiali, D., Whitaker, K., Marshall, T., van Langen, J., & Kievit, R. (2021). Raincloud plots: A multi-platform tool for robust data visualization. Wellcome Open Research, 4(63). https://doi.org/10.12688/wellcomeopenres.15191.2

Astle, D. E., Summerfield, J., Griffin, I., & Nobre, A. C. (2012). Orienting attention to locations in mental representations. Attention, Perception, & Psychophysics, 74(1), 146-162. https://doi.org/10.3758/s13414-011-0218-3

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. https://doi.org/10.1523/JNEUROSCI.2860-17.2017

Bocincova, A., & Johnson, J. S. (2019). The time course of encoding and maintenance of task-relevant versus irrelevant object features in working memory. Cortex, 111, 196-209. https://doi.org/10.1016/j.cortex.2018.10.013

Brain Products. (2020a). actiCAP snap [Apparatus] [Software]. Brain Products GmbH.

Brain Products. (2020b). BrainVision Recorder (Version 1.23.0003) [Software] [Software]. Brain Products GmbH.

Camos, V., Johnson, M., Loaiza, V., Portrat, S., Souza, A., & Vergauwe, E. (2018). What is attentional refreshing in working memory? Annals of the New York Academy of Sciences, 1424(1), 19-32. https://doi.org/10.1111/nyas.13616

Chatham, C. H., & Badre, D. (2015). Multiple gates on working memory. Current Opinion in Behavioral Sciences, 1, 23-31. https://doi.org/10.1016/j.cobeha.2014.08.001

Chatham, C. H., Frank, M. J., & Badre, D. (2014). Corticostriatal output gating during selection from working memory. Neuron, 81(4), 930-942. https://doi.org/10.1016/j.neuron.2014.01.002

Combrisson, E., & Jerbi, K. (2015). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. Journal of Neuroscience Methods, 250, 126-136. https://doi.org/10.1016/j.jneumeth.2015.01.010

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences, 24(1), 87-114. https://doi.org/10.1017/S0140525X01003922

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? Current directions in psychological science, 19(1), 51-57. https://doi.org/10.1177/0963721409359277 de Vries, I. E. J., van Driel, J., & Olivers, C. N. L. (2017). Posterior α EEG Dynamics Dissociate Current from Future Goals in Working Memory-Guided Visual Search. Journal of Neuroscience, 37(6), 1591-1603. https://doi.org/10.1523/JNEUROSCI.2945-16.2016

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. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dijkstra, N., Mostert, P., Lange, F. P. de, Bosch, S., & van Gerven, M. A. (2018). Differential temporal dynamics during visual imagery and perception. eLife, 7, e33904. https://doi.org/10.7554/eLife.33904

Fahrenfort, J. J., van Driel, J., van Gaal, S., & Olivers, C. N. L. (2018). From ERPs to MVPA Using the Amsterdam Decoding and Modeling Toolbox (ADAM). Frontiers in Neuroscience, 12. https://doi.org/10.3389/fnins.2018.00368

Feldmann-Wüstefeld, T. (2021). Neural measures of working memory in a bilateral change detection task. Psychophysiology, 58(1), e13683. https://doi.org/10.1111/psyp.13683

Frank, M. J., Loughry, B., & O’Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: A computational model. Cognitive, Affective, & Behavioral Neuroscience, 1(2), 137-160. https://doi.org/10.3758/CABN.1.2.137

Fukuda, K., Mance, I., & Vogel, E. K. (2015). α Power Modulation and Event-Related Slow Wave Provide Dissociable Correlates of Visual Working Memory. The Journal of Neuroscience, 35(41), 14009-14016. https://doi.org/10.1523/JNEUROSCI.5003-14.2015

Fukuda, K., & Vogel, E. K. (2009). Human Variation in Overriding Attentional Capture. Journal of Neuroscience, 29(27), 8726-8733. https://doi.org/10.1523/JNEUROSCI.2145-09.2009

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: Bridging selective attention and working memory. Trends in Cognitive Sciences, 16(2), 129-135. https://doi.org/10.1016/j.tics.2011.11.014

Goddard, E., Carlson, T. A., & Woolgar, A. (2022). Spatial and Feature-selective Attention Have Distinct, Interacting Effects on Population-level Tuning. Journal of cognitive neuroscience, 34(2), 290-312. https://doi.org/10.1162/jocn_a_01796

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics (pp. xi, 455). John Wiley.

Gressmann, M., & Janczyk, M. (2016). The (Un)Clear Effects of Invalid Retro-Cues. Frontiers in Psychology, 7. https://doi.org/10.3389/fpsyg.2016.00244

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. Journal of Cognitive Neuroscience, 29(4), 677-697. https://doi.org/10.1162/jocn_a_01068

Günseli, E., van Moorselaar, D., Meeter, M., & Olivers, C. N. L. (2015). The reliability of retro-cues determines the fate of noncued visual working memory representations. Psychonomic Bulletin & Review, 22(5), 1334-1341. https://doi.org/10.3758/s13423-014-0796-x

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. Annual Review of Neuroscience, 37, 435-456. https://doi.org/10.1146/annurev-neuro-062012-170325

JASP Team. (2022). JASP (Version 0.16.2.0) [Computer software] [Software]. https://jasp-stats.org/

Jongen, E. M. M., Smulders, F. T. Y., & Van der Heiden, J. S. H. (2007). Lateralized ERP components related to spatial orienting: Discriminating the direction of attention from processing sensory aspects of the cue. Psychophysiology, 44(6), 968-986. https://doi.org/10.1111/j.1469-8986.2007.00557.x

Kessler, Y., & Meiran, N. (2008). Two dissociable updating processes in working memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(6), 1339-1348. https://doi.org/10.1037/a0013078

King, J. R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. Trends in Cognitive Sciences, 18(4), 203-210. https://doi.org/10.1016/j.tics.2014.01.002

King, J. R., Pescetelli, N., & Dehaene, S. (2016). Brain Mechanisms Underlying the Brief Maintenance of Seen and Unseen Sensory Information. Neuron, 92(5), 1122-1134. https://doi.org/10.1016/j.neuron.2016.10.051

Kiyonaga, A., & Egner, T. (2013). Working memory as internal attention: Toward an integrative account of internal and external selection processes. Psychonomic Bulletin & Review, 20(2), 228-242. https://doi.org/10.3758/s13423-012-0359-y

Kruijne, W., Bohte, S. M., Roelfsema, P. R., & Olivers, C. N. L. (2021). Flexible Working Memory Through Selective Gating and Attentional Tagging. Neural Computation, 33(1), 1-40. https://doi.org/10.1162/neco_a_01339

Kuo, B. C., Stokes, M. G., & Nobre, A. C. (2012). Attention Modulates Maintenance of Representations in Visual Short-term Memory. Journal of Cognitive Neuroscience, 24(1), 51-60. https://doi.org/10.1162/jocn_a_00087

LaRocque, J. J., Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2013). Decoding Attended Information in Short-term Memory: An EEG Study. Journal of Cognitive Neuroscience, 25(1), 127-142. https://doi.org/10.1162/jocn_a_00305

Lemaire, B., Pageot, A., Plancher, G., & Portrat, S. (2018). What is the time course of working memory attentional refreshing? Psychonomic Bulletin & Review, 25(1), 370-385. https://doi.org/10.3758/s13423-017-1282-z

Lepsien, J., & Nobre, A. C. (2006). Cognitive control of attention in the human brain: Insights from orienting attention to mental representations. Brain Research, 1105(1), 20-31. https://doi.org/10.1016/j.brainres.2006.03.033

Lewis-Peacock, J. A., Kessler, Y., & Oberauer, K. (2018). The removal of information from working memory. Annals of the New York Academy of Sciences, 1424(1), 33-44. https://doi.org/10.1111/nyas.13714

Liu, B., Nobre, A. C., & van Ede, F. (2022). Functional but not obligatory link between microsaccades and neural modulation by covert spatial attention. Nature Communications, 13(1), 3503. https://doi.org/10.1038/s41467-022-31217-3

Luck, S. J. (2006). The Operation of Attention—Millisecond by Millisecond—Over the First Half Second. En The first half second: The microgenesis and temporal dynamics of unconscious and conscious visual processes (pp. 187-206). MIT Press.

Makovski, T. (2012). Are multiple visual short-term memory storages necessary to explain the retro-cue effect? Psychonomic Bulletin & Review, 19(3), 470-476. https://doi.org/10.3758/s13423-012-0235-9

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. Journal of Neuroscience Methods, 164(1), 177-190. https://doi.org/10.1016/j.jneumeth.2007.03.024

MATLAB. (2020). 9.8.0.1323502 (R2020a). The MathWorks Inc.

McCollough, A. W., Machizawa, M. G., & Vogel, E. K. (2007). Electrophysiological Measures of Maintaining Representations in Visual Working Memory. Cortex, 43(1), 77-94. https://doi.org/10.1016/S0010-9452(08)70447-7

Miller, E. K., Lundqvist, M., & Bastos, A. M. (2018). Working Memory 2.0. Neuron, 100(2), 463-475. https://doi.org/10.1016/j.neuron.2018.09.023

Mostert, P., Albers, A. M., Brinkman, L., Todorova, L., Kok, P., & de Lange, F. P. (2018). Eye Movement-Related Confounds in Neural Decoding of Visual Working Memory Representations. eNeuro, 5(4). https://doi.org/10.1523/ENEURO.0401-17.2018

Murray, A. M., Nobre, A. C., Clark, I. A., Cravo, A. M., & Stokes, M. G. (2013). Attention Restores Discrete Items to Visual Short-Term Memory. Psychological Science, 24(4), 550-556. https://doi.org/10.1177/0956797612457782

Myers, N. E., Stokes, M. G., & Nobre, A. C. (2017). Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. Trends in Cognitive Sciences, 21(6), 449-461. https://doi.org/10.1016/j.tics.2017.03.010

Nir-Cohen, G., Kessler, Y., & Egner, T. (2020). Neural Substrates of Working Memory Updating. Journal of Cognitive Neuroscience, 32(12), 2285-2302. https://doi.org/10.1162/jocn_a_01625

Noah, S., Meyyappan, S., Ding, M., & Mangun, G. R. (2023). Time courses of attended and ignored object representations. Journal of cognitive neuroscience, 35(4), 645-658. https://doi.org/10.1162/jocn_a_01972

Oberauer, K. (2018). Removal of irrelevant information from working memory: Sometimes fast, sometimes slow, and sometimes not at all. Annals of the New York Academy of Sciences, 1424(1), 239-255.

Oberauer, K., & Awh, E. (2022). Is There an Activity-silent Working Memory? Journal of Cognitive Neuroscience, 34(12), 2360-2374. https://doi.org/10.1162/jocn_a_01917

Olivers, C. N. L. (2008). Interactions between visual working memory and visual attention. Frontiers in Bioscience: A Journal and Virtual Library, 13, 1182-1191. https://doi.org/10.2741/2754

Panichello, M. F., & Buschman, T. J. (2021). Shared mechanisms underlie the control of working memory and attention. Nature, 592(7855), 601-605. https://doi.org/10.1038/s41586-021-03390-w

Pertzov, Y., Bays, P. M., Joseph, S., & Husain, M. (2013). Rapid forgetting prevented by retrospective attention cues. Journal of Experimental Psychology: Human Perception and Performance, 39, 1224-1231. https://doi.org/10.1037/a0030947

Pollack, I., & Norman, D. A. (1964). A non-parametric analysis of recognition experiments. Psychonomic Science, 1(1), 125-126. https://doi.org/10.3758/BF03342823

Psychology Software Tools, Inc. (2016). E-Prime 3.0 (Retrieved from https://support.pstnet.com/) [Software].

Quax, S. C., Dijkstra, N., van Staveren, M. J., Bosch, S. E., & van Gerven, M. A. J. (2019). Eye movements explain decodability during perception and cued attention in MEG. NeuroImage, 195, 444-453. https://doi.org/10.1016/j.neuroimage.2019.03.069

Quentin, R., King, J.-R., Sallard, E., Fishman, N., Thompson, R., Buch, E. R., & Cohen, L. G. (2019). Differential Brain Mechanisms of Selection and

Maintenance of Information during Working Memory. Journal of Neuroscience, 39(19), 3728-3740. https://doi.org/10.1523/JNEUROSCI.2764-18.2019

R Core Team. (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/

Rac-Lubashevsky, R., & Frank, M. J. (2021). Analogous computations in working memory input, output and motor gating: Electrophysiological and computational modeling evidence. PLOS Computational Biology, 17(6), e1008971. https://doi.org/10.1371/journal.pcbi.1008971

Ricker, T. J., Nieuwenstein, M. R., Bayliss, D. M., & Barrouillet, P. (2018). Working memory consolidation: Insights from studies on attention and working memory. Annals of the New York Academy of Sciences, 1424(1), 8-18. https://doi.org/10.1111/nyas.13633

Ricker, T. J., Spiegel, L. R., & Cowan, N. (2014). Time-Based Loss in Visual Short-Term Memory is from Trace Decay, not Temporal Distinctiveness. Journal of experimental psychology. Learning, memory, and cognition, 40(6), 1510-1523. https://doi.org/10.1037/xlm0000018

Ricker, T. J., Vergauwe, E., & Cowan, N. (2016). Decay theory of immediate memory: From Brown (1958) to today (2014). Quarterly Journal of Experimental Psychology (2006), 69(10), 1969-1995. https://doi.org/10.1080/17470218.2014.914546

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. Science, 354(6316), 1136-1139. https://doi.org/10.1126/science.aah7011

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. Proceedings of the National Academy of Sciences, 105(16), 5975-5979. https://doi.org/10.1073/pnas.0711295105

RStudio Team. (2022). RStudio: Integrated Development Environment for R (2022.02.3) [Software]. RStudio, Inc. http://www.rstudio.com/

Schneider, D., Mertes, C., & Wascher, E. (2016). The time course of visuo-spatial working memory updating revealed by a retro-cuing paradigm. Scientific Reports, 6(1), Article 1. https://doi.org/10.1038/srep21442

Serin, F., & Günseli, E. (2022). Internal attention is the only retroactive mechanism for controlling precision in working memory. Attention, Perception, & Psychophysics. https://doi.org/10.3758/s13414-022-02628-7

Shepherdson, P., Oberauer, K., & Souza, A. S. (2018). Working memory load and the retro-cue effect: A diffusion model account. Journal of Experimental Psychology: Human Perception and Performance, 44, 286-310. https://doi.org/10.1037/xhp0000448

Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory: 13 years of the retro-cue effect. Attention, Perception, & Psychophysics, 78(7), 1839-1860. https://doi.org/10.3758/s13414-016-1108-5

Souza, A. S., & Oberauer, K. (2017). The contributions of visual and central attention to visual working memory. Attention, Perception, & Psychophysics, 79(7), 1897-1916. https://doi.org/10.3758/s13414-017-1357-y

Souza, A. S., Rerko, L., & Oberauer, K. (2014). Unloading and reloading working memory: Attending to one item frees capacity. Journal of Experimental Psychology: Human Perception and Performance, 40(3), 1237-1256. https://doi.org/10.1037/a0036331

Souza, A. S., Rerko, L., & Oberauer, K. (2015). Refreshing memory traces: Thinking of an item improves retrieval from visual working memory. Annals of the New York Academy of Sciences, 1339(1), 20-31. https://doi.org/10.1111/nyas.12603

Stokes, M. G. (2015). 'Activity-silent' working memory in prefrontal cortex: A dynamic coding framework. Trends in Cognitive Sciences, 19(7), 394-405. https://doi.org/10.1016/j.tics.2015.05.004

Thaler, L., Schütz, A. C., Goodale, M. A., & Gegenfurtner, K. R. (2013). What is the best fixation target? The effect of target shape on stability of fixational eye movements. Vision Research, 76, 31-42. https://doi.org/10.1016/j.visres.2012.10.012

Thyer, W., Adam, K. C. S., Diaz, G. K., Velázquez Sánchez, I. N., Vogel, E. K., & Awh, E. (2022). Storage in Visual Working Memory Recruits a Content-Independent Pointer System. Psychological Science, 33(10), 1680-1694. https://doi.org/10.1177/09567976221090923

Trübutschek, D., Marti, S., Ojeda, A., King, J.-R., Mi, Y., Tsodyks, M., & Dehaene, S. (2017). A theory of working memory without consciousness or sustained activity. eLife, 6, e23871. https://doi.org/10.7554/eLife.23871

van Driel, J., Olivers, C. N. L., & Fahrenfort, J. J. (2021). High-pass filtering artifacts in multivariate classification of neural time series data. Journal of Neuroscience Methods, 352, 109080. https://doi.org/10.1016/j.jneumeth.2021.109080

van Ede, F., Chekroud, S. R., & Nobre, A. C. (2019). Human gaze tracks attentional focusing in memorized visual space. Nature Human Behaviour, 3(5), 462-470. https://doi.org/10.1038/s41562-019-0549-y

van Ede, F., & Nobre, A. C. (2022). Turning Attention Inside Out: How Working Memory Serves Behavior. Annual Review of Psychology, Forthcoming. https://doi.org/10.2139/ssrn.4082572

van Moorselaar, D., Gunseli, E., Theeuwes, J., & Olivers, C. N. L. (2015). The time course of protecting a visual memory representation from perceptual interference. Frontiers in Human Neuroscience, 8. https://www.frontiersin.org/articles/10.3389/fnhum.2014.01053

Vogel, E. K., & Luck, S. J. (2002). Delayed working memory consolidation during the attentional blink. Psychonomic Bulletin & Review, 9(4), 739-743. https://doi.org/10.3758/BF03196329

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. Nature, 428(6984), 748-751. https://doi.org/10.1038/nature02447

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. Nature, 438(7067), Article 7067. https://doi.org/10.1038/nature04171

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions and objects in visual working memory. Journal of Experimental Psychology Human Perception and Performance, 27(1), 92-114. https://doi.org/10.1037//0096-1523.27.1.92

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. Journal of Experimental Psychology: Human Perception and Performance, 32(6), 1436-1451. https://doi.org/10.1037/0096-1523.32.6.1436

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. https://ggplot2.tidyverse.org

Williams, M., Hong, S. W., Kang, M.-S., Carlisle, N. B., & Woodman, G. F. (2013). The benefit of forgetting. Psychonomic Bulletin & Review, 20(2), 348-355. https://doi.org/10.3758/s13423-012-0354-3

Wolff, M. J., Ding, J., Myers, N., & Stokes, M. (2015). Revealing hidden states in visual working memory using electroencephalography. Frontiers in Systems Neuroscience, 9. https://www.frontiersin.org/articles/10.3389/fnsys.2015.00123

Wolff, M. J., Jochim, J., Akyürek, E. G., & Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. Nature Neuroscience, 20(6), 864-871. https://doi.org/10.1038/nn.4546

Woodman, G. F., & Luck, S. J. (1999). Electrophysiological measurement of rapid shifts of attention during visual search. Nature, 400(6747), 867-869. https://doi.org/10.1038/23698

Woodman, G. F., & Luck, S. J. (2003). Serial deployment of attention during visual search. Journal of Experimental Psychology. Human Perception and Performance, 29(1), 121-138. https://doi.org/10.1037//0096-1523.29.1.121

Woodman, G. F., & Vogel, E. K. (2005). Fractionating working memory: Consolidation and maintenance are independent processes. Psychological Science, 16(2), 106-113. https://doi.org/10.1111/j.0956-7976.2005.00790.x

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. Psychometrika, 70(1), 203-212. https://www.doi.org/10.1007/s11336-003-1119-8

Zhao, M., Gersch, T. M., Schnitzer, B. S., Dosher, B. A., & Kowler, E. (2012). Eye movements and attention: The role of pre-saccadic shifts of attention in perception, memory and the control of saccades. Vision Research, 74, 40-60. https://doi.org/10.1016/j.visres.2012.06.017

Supplementary Materials: Tortajada et al. (2024)

Behavioral time course for all participants who completed the first session

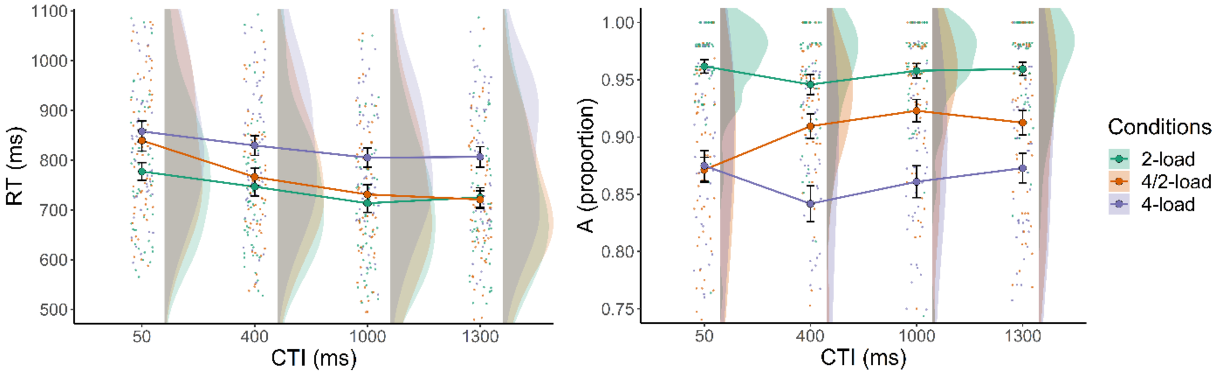
Descriptive statistics for the full sample that completed the first session are available in Table S1 for reaction times (RTs) and mean A. Figure S1 shows the time course of the three conditions for RTs and A during the second delay, separately for each CTI. For RTs, the main effect of CTI was significant, $F_{(2.28, 116.48)} = 87.888$, $p < .001$, $\eta^2_p = .633$, as was the main effect of Condition, $F_{(1.78, 90.77)} = 73.993$, $p < .001$, $\eta^2_p = .592$. RTs were slower for the shortest CTIs and became faster with longer CTIs and stabilized at 1200 ms CTI. In general, RTs were faster for the 2-load condition than for the 4-load condition and mostly also faster than the 4/2-load condition (except for the shortest CTI). In this line, the Condition x CTI interaction reached statistically significant, $F_{(6, 306)} = 6.130$, $p < .001$, $\eta^2_p = .107$, showing different time courses for the different conditions. For post-hoc analyses, the Holm-Bonferroni correction was applied. Post-hoc comparisons showed that the 4/2-load condition differed statistically from the 2-load condition, $t_{(51)} = 5.710$, $p < .001$, $d = .447$ in the short 250 ms CTI but not from the 4-load condition, $t_{(51)} = 1.677$, $p = 1$, $d = .131$. As in the subsample reported in the main text, the trend reverted from the second CTI. From here, the 4/2-load condition differed from the 4-load condition, $t_{(51)} = 5.808$, $p < .001$, $d = .454$, and at the same time became similar to the 2-load condition, $t_{(51)} = 1.787$, $p = .90$, $d = .140$, which remained so for the rest of the CTIs.

Table S1. Descriptive statistics of behavioral data from all participants who completed the first session, N=52. Reaction Times (RT) and mean A (A) split by trial type (Condition) and Cue-Target Interval (CTI).

		RT		**A**	
Condition	**CTI (ms)**	**Mean**	**SD**	**Mean**	**SD**
2-load	50	777	129.7	0.962	0.042
	400	747	136.3	0.946	0.063
	1000	713	133.0	0.958	0.046
	1300	725	139.7	0.959	0.042
4/2-load	50	840	154.1	0.871	0.079
	400	766	126.2	0.909	0.077
	1000	731	141.3	0.923	0.071
	1300	721	127.3	0.913	0.076
4-load	50	858	150.5	0.875	0.096
	400	830	142.9	0.842	0.112
	1000	805	140.8	0.861	0.100
	1300	807	149.2	0.873	0.093

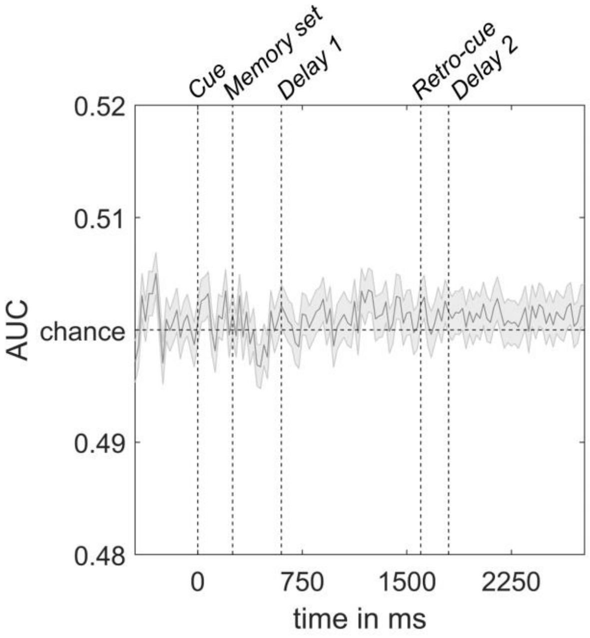
On the other side, the main effect of Condition was significant for A, $F_{(1.76,\ 89.75)} = 64.388$, $p < .001$, $\eta_p^2 = .558$. Contrary to the analyses in the results section, this ANOVA also showed a main effect of CTI, $F_{(3,\ 153)} = 3.134$, $p = .027$, $\eta_p^2 = .058$. Again, post-hoc comparisons on A showed a similar pattern of results to that observed for RTs. The retro-cue condition differed from the 2-load condition in the shortest CTI, $t_{(51)} = .7.60$, $p < .001$, $d = 1.16$ but became equivalent for the remaining three CTIs. The opposite pattern was observed for the differences between the 4/2- and the 4-load conditions, where no differences were detected for the first CTI ($t < 1$) but showed statistically significant differences from the second CTI onwards ($p < .03$).

Figure S1. Reaction Times (RT) and mean A for the time course during the second delay for all participants who completed the first session, N=52. Small colored points represent the mean of each participant in each condition. Points with black border represent the mean of each trial type for each interval. Error bars represent standard error of the mean. The individual distributions for all the trial types x CTI combinations are represented in the split violin plots.



Empirical chance

Figure S2. Classifier decoding performance randomizing labels, showing empirical chance level. AUC = Area Under the Curve. Bold lines show p-values that survived multiple comparison corrections and the shaded area surrounding the line is the standard error.



Robustness analyses

To discard the possibility that load decoding was not detected in the second delay due to pre-processing and/or analytic settings, we re-run the 2- versus 4-load decoding analysis with a series of modifications. First, we removed activity identified as eye movements (activity higher than 50 uV) from the subtracted VEOG and HEOG channels (Thyer et al., 2022). Second, to ensure the same number of trials per condition, we downsampled trials for unbalanced classes instead of oversampling. Lastly, instead of resampling to 40 Hz by discarding the unselected time points, we averaged data within each time bin to 40 Hz. Results show similar decoding after removing eye movements and applying the rest of the analytical modifications but significant decoding slightly increased for EEG electrodes during the second delay (until 2.16 s in the epoch; Figure S3).

To test whether this modest increase in decoding would allowed us to test the main hypothesis of this study (i.e., is the 4/2-load condition first classified as 4-load and then as 2-load?), analyses from Figure 7 were repeated after applying the new pre-processing steps. However, again, results do not show this transition in decoding after the retro-cue and the 4/2-load condition is decoded from both 2- and 4-load conditions during the second delay (Figure S4).

Figure S3. Reanalysis after different EEG preprocessing of decoding of load from electroencephalography (EEG) and electrooculogram (EOG) signal. AUC = Area Under the Curve. (a) Diagonal decoding of 2-load versus 4-load conditions from EEG (orange line) and EOG (green line) electrodes. Bold lines show p-values that survived multiple comparisons corrections and shaded area surrounding the line is the standard error. (b) Temporal generalization matrix for 2-load vs 4-load decoding on EEG electrodes. (c) Temporal generalization matrix for 2-load vs 4-load decoding on

EOG electrodes. In both (b) and (c), saturated colors reflect uncorrected p<0.05 decoding, whereas areas circumscribed by a dark bold line highlight clusters that survive cluster-based permutation testing at p<0.05. Above chance decoding is colored in red color and below chance decoding in blue.

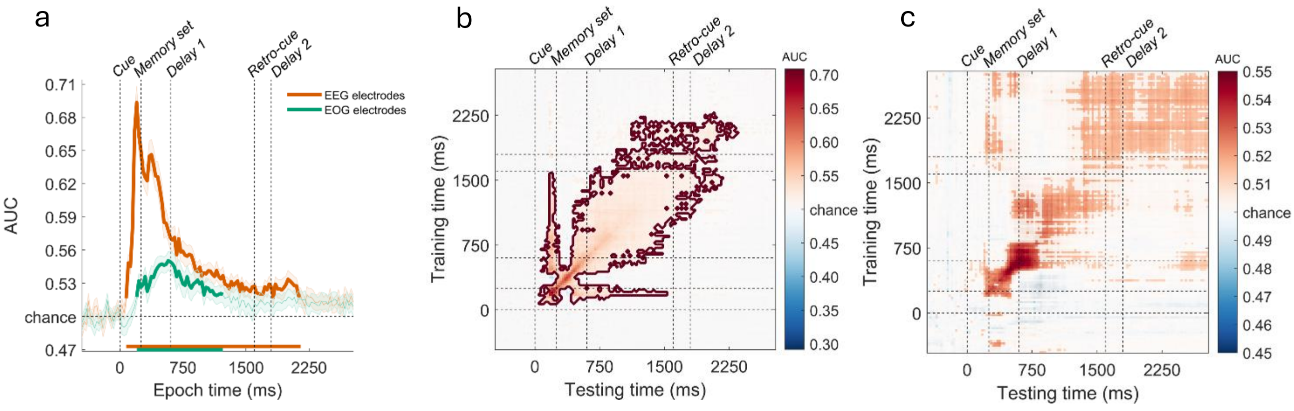
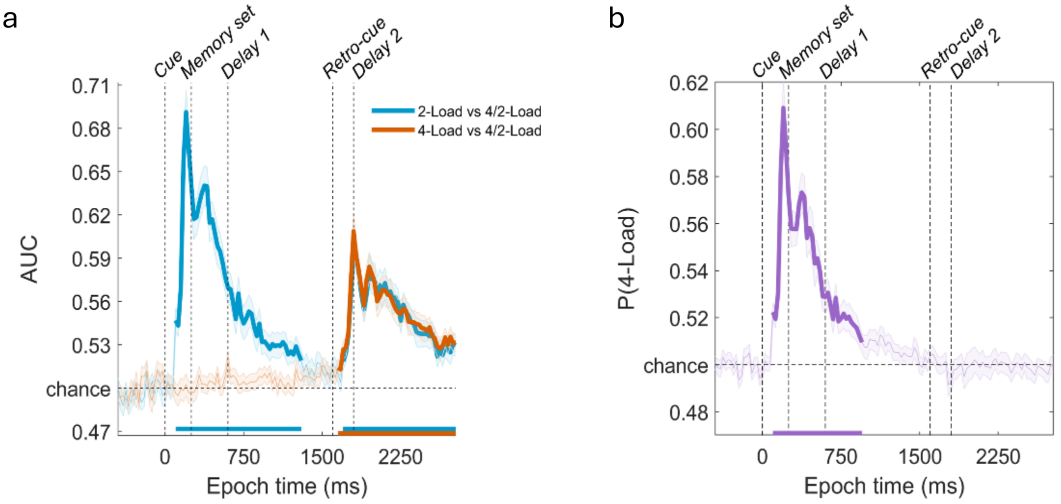


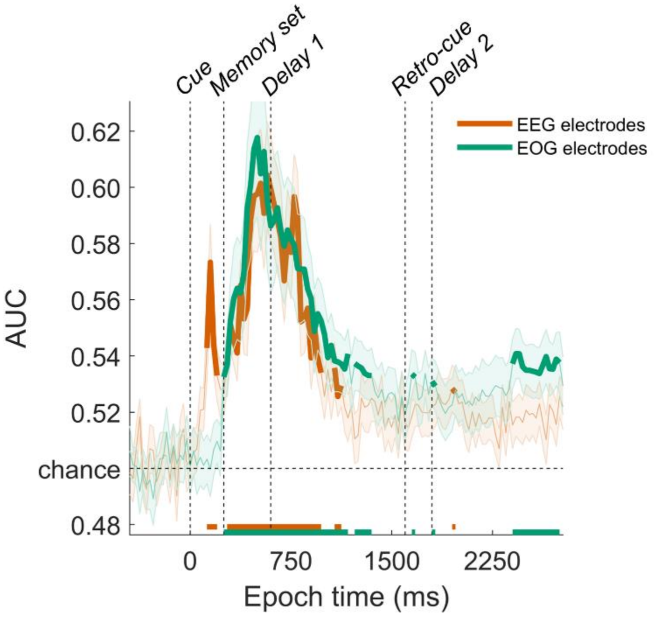
Figure S4. Reanalysis after preprocessing modifications. (a) Diagonal decoding 4/2-load vs 2- and 4-load conditions (blue and orange, respectively). AUC = Area Under the Curve. Bold lines show p-values that survived multiple comparisons corrections and shaded area surrounding the line is the standard error. (b) Training the 4- vs 2-load conditions and testing the 4/2-load condition. When classification is above chance, the 4/2-load condition is classified as 4-load and when it is below chance, the 4/2-load condition is classified as 2-load.



Testing with False Discovery Rate (FDR) as multiple comparisons correction method

To test if the cluster correction method we used for multiple comparisons was insensitive to transient responses, we repeated the analyses where we decoded the directionality of the pre-cue (Figure 5) because there seems to be a small rebound at the height of the retro-cue which, perhaps, could be extended during the second delay. Then, we used the False Discovery Rate (FDR) to deal with the multiple comparisons problem (Benjamini & Yekutieli, 2001). The FDR limits the false positive rate in a way that type 1 errors are not expected to be more than 5% of the tests. Results do not show any unseen transient response, on the contrary, the significantly decoded time points were reduced with this method (see Figure S5).

Figure S5. Diagonal decoding of the direction of the initial cue in the 2-load condition for EEG (orange) and EOG electrodes (green) with False Discovery Rate as multiple comparisons correction. AUC for Area Under the Curve. Bold lines show p-values that survived multiple comparisons corrections and shaded area surrounding the line is the standard error.



References

Benjamini, Y., & Yekutieli, D. (2001). The Control of the False Discovery Rate in Multiple Testing under Dependency. *The Annals of Statistics*, 29(4), 1165-1188.

Thyer, W., Adam, K. C. S., Diaz, G. K., Velázquez Sánchez, I. N., Vogel, E. K., & Awh, E. (2022). Storage in Visual Working Memory Recruits a Content-Independent Pointer System. *Psychological Science*, 33(10), 1680-1694. https://doi.org/10.1177/09567976221090923

Chapter 3: Questioning Theta-Gamma Coupling Predictions: Ineffectiveness of 4 Hz tACS on Working Memory Maintenance and Theta Frequency Modulation

Abstract

Recent findings have shown the influential role of theta-gamma cross-frequency coupling in the maintenance of items in working memory (WM). Specifically, it has been proposed that individual items are represented within gamma oscillations that are nested within slower theta waves. Thus, longer theta cycles would theoretically allow the retention of more items. Based on this theory, the present study aimed to test whether slowing theta rhythms via 4 Hz transcranial alternating current stimulation (tACS) in the fronto-parietal network improves WM capacity. Given the potential role of individual differences in the effects of stimulation, participants were divided into high and low-WM capacity groups based on initial screening. Besides, task difficulty was also manipulated to maximize the likelihood of obtaining a performance improvement. Contrary to our hypothesis, tACS did not improve behavioral outcomes, such as response time or accuracy, regardless of task difficulty or baseline WM capacity. EEG analysis revealed no significant increase in theta power post-stimulation compared to a sham condition, and the expected reduction in the mean theta frequency was neither observed. Importantly, mean theta frequency was higher in participants with better WM performance, contrary to the theta-gamma coupling theory hypothesis. Besides, no mean theta frequency variations were observed based on task difficulty. Together these results raise questions about this theory that may require further consideration.

Keywords: verbal working memory, theta-gamma coupling, Transcranial alternating current stimulation, electroencephalography.

Introduction

Maintaining information in an available state so that it can be used for task objectives is a process supported by working memory (WM). Given the importance of this process in daily life, many efforts have been made to understand the neural mechanisms underlying it. In this regard, there is no evidence for the existence of a single brain region responsible for this maintenance (Christophel et al., 2017). Instead, the activity is distributed among various regions, with the fronto-parietal cortex being particularly prominent (Owen et al., 2005). Due to the distributed nature of this process, it is important to study brain oscillations as a form of long-range connectivity to coordinate the different brain regions (Buzsáki & Draguhn, 2004; D'Esposito & Postle, 2015; Gazzaley et al., 2004; Sauseng et al., 2005).

The influence of several oscillatory rhythms has been involved in different WM processes (Lisman & Jensen, 2013; Miller et al., 2018; Roux & Uhlhaas, 2014; Sauseng et al., 2019). For maintenance, an important role of theta and gamma oscillations has been proposed. The theta-gamma coupling theory (Jensen & Lisman, 1996; Lisman & Jensen, 2013) proposes that maintenance occurs by coupling slow and fast brain waves. According to this theory, fast oscillations represent individual items, while slow waves, such as theta waves, organize these fast oscillations by coupling them to different phases of the theta cycle. This theta-gamma coupling has been observed both in the hippocampus (Axmacher et al., 2010; Chaieb et al., 2015) and cortical regions (Brooks et al., 2020; Canolty et al., 2006; Fernández et al., 2021; Holz et al., 2010; Köster et al., 2014).

A clear prediction derived from this theory is that lower-frequency theta waves (i.e., longer theta cycles) would allow faster bursts to be nested, enabling more information to be retained in WM. One approach that can be used to test this hypothesis is to modulate the oscillations to slow down the theta rhythms and observe whether this increases WM capacity. One way to extrinsically modulate endogenous oscillations is using transcranial Alternating Current Stimulation (tACS). tACS is a widely used non-invasive brain stimulation (NIBS) technique (Antal et al., 2008, 2022; Elyamany et al., 2021; Fröhlich et al., 2015; Helfrich et al., 2014) that applies weak electrical currents (typically sinusoidal) to the scalp using two or more electrodes, simulating natural brain oscillations. Part of this current reaches the cortex by traveling through the skull, allowing the entrainment of distant brain regions to the stimulation frequency, which can improve or hinder connectivity between these areas (Alekseichuk et al., 2019; Schwab et al., 2019; Weinrich et al., 2017). In this way, different cognitive abilities can be targeted by adjusting the frequency and the regions to stimulate.

Given the role of theta oscillations in WM functioning, several studies have applied theta tACS with the aim of modulating endogenous brain oscillations. Polania et al. (2012) were the first to apply theta tACS to the fronto-parietal network finding better performance in a Sternberg task when 6Hz stimulation was applied in phase (i.e. 0º lag between stimulation electrodes). However, studies attempting to replicate this effect showed mixed results (Abellaneda-Perez et al., 2020; Alekseichuk et al., 2016; Bender et al., 2019; Biel et al., 2022; Chander et al., 2016; Jausovec et al., 2014; Jausovec & Jausovec, 2014; Jones et al., 2019; Kleinert et al., 2017; Reinhart & Nguyen, 2019; Tseng et al., 2016; Violante et al., 2017; Vosskuhl et al., 2015; Wolinski et al., 2018; for a recent review, see Booth et al., 2022). Biel et al. (2022) proposed that one reason for the inconsistency of results could be how stimulation electrodes were referenced, and they decided to try to replicate Polania et al. (2012) but using a ring montage, which allows for a better current focality (Alekseichuk et al., 2019; Bortoletto et al., 2016; Saturnino et al., 2017). Thus, they observed a WM improvement when stimulation was applied in phase with a ring montage, but only when participants performed a more demanding task than the one of Polania et al. (2012).

In this line, previous results have also reported higher stimulation benefits with more difficult tasks (e.g. Bender et al., 2019; Hoy et al., 2015; Violante et al., 2017; Wu et al., 2014). Violante et al. (2017) applied in-phase tACS and observed better performance and higher fronto-parietal activity when the cognitive demands of the task were high. Bender et al. (2019) also applied parietal tACS while performing a visual WM task. They found an improvement when the stimuli were presented in the hemifield contralateral to the stimulation site, but only when the task was difficult (i.e. six items to retain). This improvement is consistent with research that enhanced frontal midline theta with more demanding verbal WM tasks (Pavlov & Kotchoubey, 2022). However, Jausovec et al. (2014) only found tACS effects in easier 1- and 2-back tasks, but not in a more demanding 3-back. Given this scenario, the present work used a task similar to Polania et al. (2012) and manipulated task difficulty to test whether this could explain the mixed results in the literature.

Furthermore, this variability in stimulation efficacy is not unique to tACS but occurs with all types of stimulation. This has led researchers to emphasize the study of individual differences in this area (Krause & Cohen Kadosh, 2014), showing that factors such as chronotype (Salehinejad et al., 2021) or the menstrual cycle (Rudroff et al., 2020) may determine its efficacy. Among works that use NIBS to improve WM, several have shown that the benefits of stimulation are conditioned to participants having low baseline WM capacity (WMC; e.g., Arciniega et al., 2018; Hsu et al., 2016; Sahu & Tseng, 2021; Tseng et al., 2018; but see Feurra et al., 2016, for the opposite pattern), suggesting that low performers have more room for improvement. Sahu and Tseng (2021) applied 6 Hz tACS to the fronto-parietal network and observed a facilitatory effect of tACS driven by low WM capacity participants in a visual working memory task. In two experiments, Tseng et al. (2018) also applied 6Hz tACS while low and high performers completed a visual WM task, observing that in-phase stimulation improved low-performers' execution while anti-phase stimulation only impaired high-performers execution. These findings highlight the influence of individual differences on the efficacy of tACS.

However, despite the apparent importance of these factors—task difficulty and participant WMC baseline—for the efficacy of stimulation, only one study has explicitly taken both of them into account using tDCS (Hsu et al., 2016). Hsu et al. (2016) manipulated baseline WM and task difficulty to see if they modulated the benefits of tDCS in a visuospatial task (the corsi block tapping task). They found a modulation of anodal stimulation only for low performers and only when the task difficulty was high. However, no study has systematically modulated both factors when applying tACS, similar to Biel et al. (2022) or Polania et al. (2012). Thus, the present study aims to apply tACS while controlling these factors to maximize the effectiveness of stimulation, in order to test the prediction from the theta-gamma coupling theory, i.e., lower theta frequency allows the maintenance of more items in WM. Since the main goal was to lower the participants' theta frequency, 4Hz stimulation was applied instead of the 6Hz used in previous research. It was hypothesized that participants with lower WMC would benefit more from 4Hz theta tACS and that these benefits should emerge especially during the high-load task condition. Additionally, EEG modulations would also be expected after stimulation, including a reduction in the predominant average theta frequency and an increase in the target theta frequency power, especially in the low WMC group.

Method

The methods and hypotheses were preregistered in Open Science Framework before data collection (https://osf.io/ghqk6).

Participants

To select participants with high and low WMC, 352 undergraduate students from the University of Murcia were screened in exchange for a credit course. From these, 220 students were selected, 99 for the high WMC group and 121 for the low WMC group (see below for the description of the screening task and the participant classification procedure). The remaining participants were contacted to attend the lab. Of the participants that replied to the email or phone, 108 attended the experimental session after exclusion for meeting any exclusion criteria. The exclusion criteria were pregnancy, personal or close family history of neurological or psychiatric disorders, metal implants, migraines, epilepsy, cardiac pacemaker, or infusion pumps. After rejecting data from eleven participants due to device issues, the final sample was 97 participants: 49 high WMC (36 females; M_{age} = 19.82, SD_{age} = 4.12; M_{span} = 6.80), 48 low WMC (42 females; M_{age} = 19.48, SD_{age} = 1.22; M_{span} = 4.75).

The sample size for the tACS session was previously determined using G*Power 3.1 software to detect a medium effect size (partial $\eta2$ = 0.06) with an alpha level set at .05 for the most stringent interaction term of the design to obtain .80 power. A minimum sample size of 92 participants was required, thus, a sufficient sample was obtained. All participants indicated that they had normal or corrected vision. Approval for this study was obtained from the Ethics Committee at the University of Murcia, and the research was conducted according to the ethical standards of the 1964 Declaration of Helsinki.

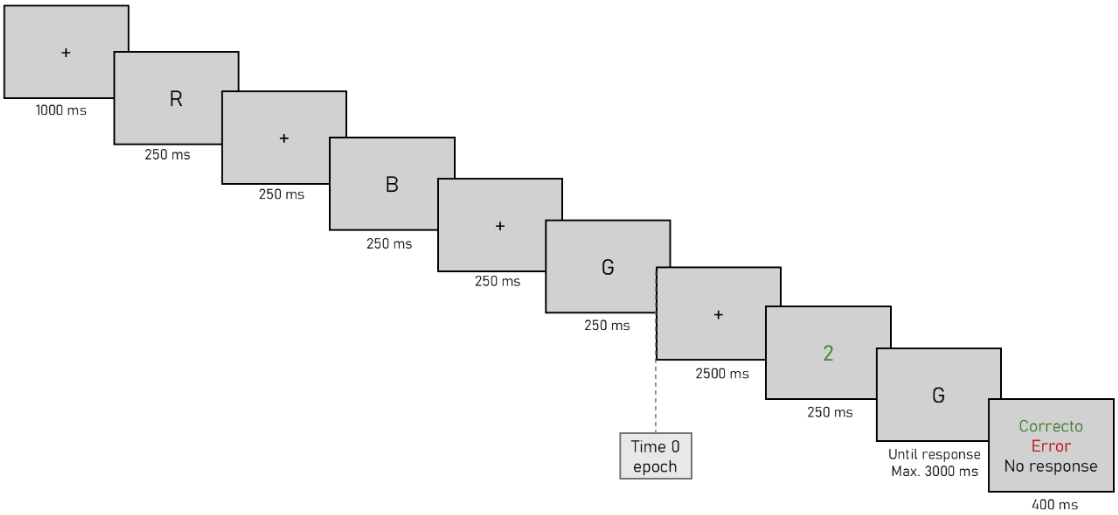
Tasks and stimuli

An online version of the Digit Span Task was used for the participants' WMC screening. The task duration was about ten minutes. The task consisted of sequences of random digits that participants had to remember and recall in the same order after the presentation of the last digit. A total of 18 sequences were presented: three sequences of three digits, three sequences of four digits, and so on until the last three sequences of eight digits. Three measurements could be obtained from this task: the total number of digits correctly recalled, the number of complete lists correctly recalled, and the participant’s span (defined as the length of the longest lists for which the participant correctly recalled at least two lists). The criteria to classify participants in the low, intermediate, or high WMC groups were as follows. Participants were classified as low WMC if they could recall less than 70 digits, less than eight lists, and their span was lower than six. Participants were classified as high WMC if they recalled 75 or more digits, nine or more complete lists, and they had a span of six or more. The rest of the participants were classified as intermediate and were not allowed to participate in the stimulation experimental session. By this procedure, 34.38 % of the participants were classified as low and 28.13 % as high in capacity.

In the tACS session, participants performed a modified version of the delayed letter discrimination task used by Polania et al. (2012). In Figure 1, there is a schematic representation of the WM task used in the experiment. Each trial began with a fixation point (+) for 1000 ms followed by the sequential presentation of three or five (low or high load trials, respectively) capital letters. Each letter had a duration of 250 ms followed by a fixation cross for another 250 ms. After the last letter, another fixation point was presented for 2500 ms (usually referred to as "delay period"). Then, a number in green color was presented in the center of the screen for 250 ms. This number would be from one to three in the low-load trials and from one to five in the high-load trials, and it referred to the position that occupied one of the previously presented letters. Then, another capital letter was presented for a maximum time of 3000 ms and participants had to judge whether or not this letter corresponded to the position indicated by the green number. The probability of the letter being correct was 50%, and for the incorrect trials, a letter to be maintained in another position was displayed to avoid answering merely by familiarity. The button responses (i.e., right or left) were counterbalanced across participants. Then, a feedback display was presented for 400 ms followed by a 1000-ms blank screen that preceded the presentation of the next trial. The stimuli were randomly selected from a pool of 17 possible letters (B, C, D, F, G, H, J, K, L, M, P, Q, R, S, T, V, X, and Z). All letters had the same likelihood of being presented, and all of them were presented once before being repeated. Besides, a letter would never repeat itself before the following six letters were presented.

This task was programmed in E-Prime 3 (Psychology Software Tools, 2016) and they used a five-button Chronos device to respond (Psychology Software Tools, 2016). The task was performed in an individual sound-attenuated booth where participants were seated around 70 cm from the screen. Stimuli were presented on a 23-inch flatscreen LED monitor (LG 23MP68VQ-P) with 1920 × 1080 pixels on a gray background RGB (192,192,192).

Figure 1. Schematic synthesis of a load 3 trial from the delayed letter discrimination task used during the stimulation session.



Procedure

Participants completed two experimental sessions. In the first session, the screening was performed in groups of around thirty participants. After explaining the instructions of the experiment, they completed an online version of the Digit Span Task on their phone devices. Then, participants classified as high or low WMC were contacted to attend the laboratory in a second session. Menstruating participants who were not using hormonal contraceptives were dated outside of days 13-15 of their menstrual cycle to avoid ovulatory days (Rudroff et al., 2020). It was calculated by asking about their last four periods to calculate the average of their cycle length and subtracting fourteen days (average luteal phase duration). Then, this was added to the onset day of their last period. If the experiment date matched this estimated ovulatory day, the day before or the day after, it was rescheduled to avoid ovulation.

Then, participants attended the laboratory for the second session (i.e., stimulation session). Once they arrived at the laboratory, they were randomly assigned to active or sham stimulation. Both the participant and the experimenter were blinded to both the stimulation condition and the WMC group they belonged to. Then, they entered the individual isolated room, the tES/EEG electrodes were placed, and the task was explained to them. First, they remained with their eyes open and completed three minutes of resting state EEG recording. Then, the instructions for the task were displayed, and they completed five practice trials of each type (i.e., low and high load). After practicing, they completed thirty randomly presented trials of each type while EEG was being recorded. Then, the message "PLEASE, CALL THE RESEARCHER" was displayed on the screen and the researcher would change from EEG to tACS. After that, they received either sham or active tACS while performing 120 randomly ordered trials, 60 of each type. Following stimulation, the same message was displayed and the researcher changed again to EEG recording while the participants performed 60 more trials, 30 of each type. Then, the participants remained still with their eyes open to complete the last three minutes of resting state recording EEG. After finishing the experiment, they completed a survey about the stimulation sensations (Fertonani et al., 2015; Hemmerich et al., 2023).

tACS and EEG protocol

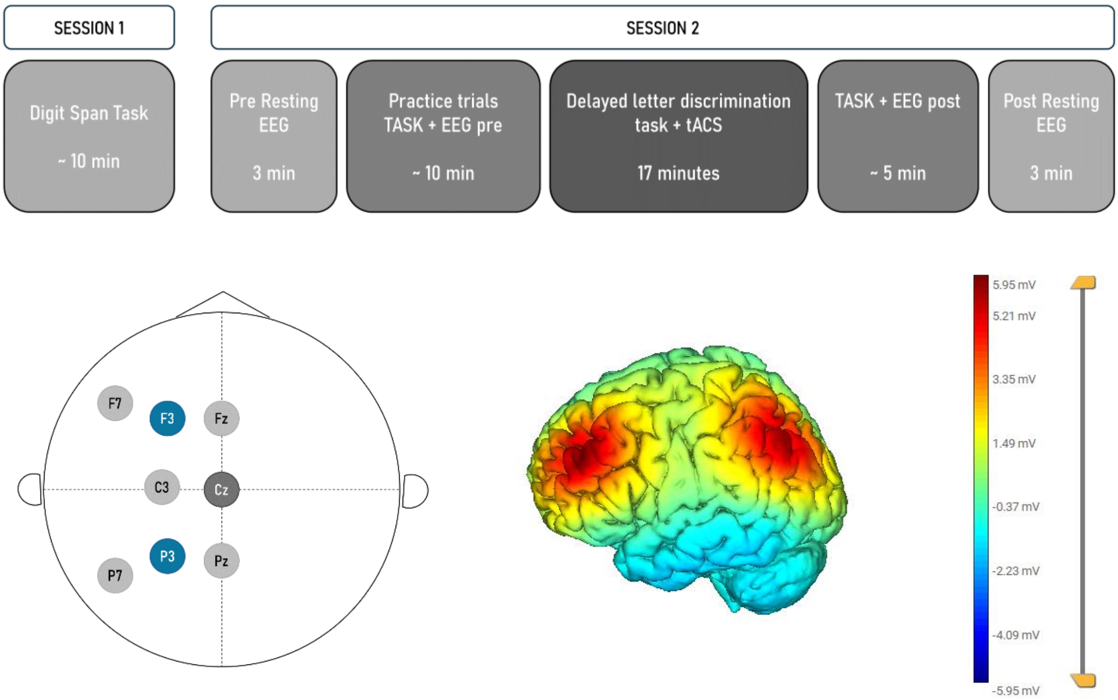
Transcranial alternating current stimulation (tACS) was administered with a StarStim 8® wireless system (Neuroelectrics, Barcelona, Spain) connected to hybrid tCS/EEG NG Pistim circular electrodes (pi cm^2). Stimulation electrodes followed the 10-20 system and were located at the left dorsolateral prefrontal cortex (DLPFC) and left posterior parietal cortex (PPC), respectively, at the F3 and P3 positions. The stimulation was administered in phase, 0°[2], between the F3 and P3 electrodes, and return electrodes surrounding stimulation electrodes were located on positions F7, Fz, C3, P7, and Pz, in a triangular focal shape. Stimulation at F3 and P3 was set at 4 Hz with an intensity of 1500 μA, zero-to-peak, and return electrodes at 600 μA (following the montage from Biel et al., 2022). The duration of the stimulation was fifteen minutes with one-minute ramp-up and down, and was administered online.

EEG data was also collected using the Starstim 8 system. Apart from the stimulation electrodes, the additional Cz electrode was recorded. The reference electrode was placed on the right earlobe using an EarClip electrode. Data were recorded with a sampling frequency of 500 Hz, a bandwidth of 0–250 Hz. Electrode impedances were kept below 5 kΩ. EEG data was collected both during resting and while performing the task, before and after stimulation (see Figure 2). During resting state recording, participants were instructed to remain with their eyes open, looking at the fixation point in the center of the screen.

---

[2] The preregistered 180° referred to the timing between the active and the reference electrodes.

Figure 2. Experimental procedure of the two session (upper panel), stimulation montage (lower left panel) and electric field (lower right panel). In the stimulation montage panel, the blue electrodes represent the active stimulation electrodes, the light grey electrodes are the reference electrodes and the dark grey electrode was only used during EEG recording.



EEG Preprocessing and Analyses

Preprocessing was conducted in MATLAB using the EEGLAB toolbox v2024.1 (Delorme & Makeig, 2004). The raw datasets were imported to EEGLAB using the NE EEGLAB NIC plugin. EEG data was downsampled to 250 Hz, high pass filtered to 1 Hz, and low pass filtered to 120 Hz using *pop_eegfiltnew*. Then, the *pop_cleanline* function was used to clean the 50-line noise and the 100 Hz harmonic, and data was re-referenced to the average of all the electrodes. Afterward, the resting epochs and the delay epochs were extracted. The delay epoch was from -800 ms to 2499 ms, locked to the beginning of the delay after the last presented letter. The baseline correction was from -600 ms to -100 ms, following (Cohen, 2014) recommendations of avoiding the 0 time in the baseline. No baseline correction was applied to the resting state epochs. Then, Independent Component Analysis (ICA) was run with the *pop_runica* function of EEGLAB. Artifacts were removed using the *pop_iclabel* function to classify the components. Then, those components classified with a probability higher than 0.7 as eye, heart activity, or channel noise, or higher than 0.5 as line noise were flagged with the function *pop_icflag*, and later removed from the data. Trials with no responses and errors were excluded from the EEG analyses.

Power spectral density (PSD) was calculated in resting and during the first and last block of the task (i.e. before and after stimulation). Analyses were performed with custom code using the Fast Fourier Transform (FFT) following the recommendations of Cohen (2014). Resting-state data were segmented into 2-second epochs. The frequencies to calculate the FFT were determined with the function *linspace* with the maximum frequency in the Nyquist number (125 Hz), and the number was half the number of time points in an epoch plus one. A Hanning window was applied to each epoch. To determine the effects of stimulation in PSD, the normalized percentage change was calculated for the EEG signal after stimulation/sham compared to the baseline (i.e. resting state or task signal average before receiving tACS). This normalization has several advantages over non-normalized data, such as making the data more easily interpretable, standardizing it, and reducing differences that could be due to interindividual factors at baseline

(Clayton et al., 2019; Cohen, 2014). Where X is the mean individual PSD, the percentage change formula was:

$$Percent\ change\ = \frac{(X_{post} - X_{pre})}{X_{pre}} * 100 \tag{1}$$

Additionally, the functional connectivity between the two main stimulation electrodes (i.e., F3 and P3) was explored before and after stimulation to assess variations in long-range synchronization using a method called Phase Locking Value (PLV, Lachaux et al., 1999). To calculate the PLV, the phases of the individual electrodes within the narrow band of interest (in this case, the theta band from 3 to 7 Hz) were first extracted via the Hilbert transform, and the difference between these instantaneous phases was then computed. The PLV takes values ranging from 0 to 1, where 0 indicates unsynchronized phases, and 1 indicates perfect phase synchronization. Similarly, an increase in the PLV value indicates an increase in synchronization between the two electrodes. The interest of this analysis was to determine whether connectivity between these two electrodes increased after stimulation. To compare theta and sham stimulation groups independently of individual baseline synchronization, the percentage change of both groups was calculated in the same manner as for PSD.

Lastly, to obtain an estimate of the mean theta frequency for each participant and condition, the "gravity frequency" (gF), or weighted mean of the band (Hooper, 2005; Klimesch, 1997, 1999) was calculated. This is obtained as the weighted sum of the spectral frequencies of interest, divided by the total theta power:

$$gF\ = \frac{\Sigma(a(f) \times f)}{\Sigma\, a(f)}, \tag{2}$$

where $a(f)$ are the power spectral at frequency f. In this case, the frequencies of interest were fixed in 3 to 7 Hz for all participants (i.e., theta band) and they were obtained in MATLAB in linear steps with the function *linspace* from 0 to Nyquist (i.e., 125) in N/2+1 steps, where N is the number of time points in the data delay. This returned fourteen continuous frequencies for which power was obtained and z-normalized before calculating gF. Again, the percentage of change was calculated to allow the comparison of the two stimulation groups.

Statistical analyses

Behavioral data were preprocessed in Rstudio (RStudio Team, 2022) using the library Tidyverse (Wickham et al., 2019). Data were analyzed with JASP 0.19.0.0 (JASP Team, 2022). A significant level of α = 0.05 was adopted. Post-hoc tests were performed to explore the significant interactions. Holm-Bonferroni correction was applied to correct for multiple comparisons. Besides, for all analyses, if Mauchly's sphericity test reached statistical significance, Greenhouse-Geisser correction was applied. First, it was tested whether stimulation affected performance on the delayed letter discrimination task, depending on baseline WMC and task difficulty. As preregistered, two Analyses of Variance (ANOVA) were performed on Response Times (RT) and accuracy (ACC) as dependent variables. Besides, two additional exploratory ANOVA were performed on discriminability (d') and Cowan's K value (Cowan, 2001; Rouder et al., 2008) to evaluate task performance. The mean theta power, the mean theta frequency, and the Phase Locking Value were used as dependent variables in the EEG analyses. The analyses were performed with *WMC groups* (high and low) and *Stimulation* (4hz and sham) as between-participant factors and *task difficulty* (low and high load) as within-participant factors. Besides, the *moment of recording* (pre-post) was also introduced in some analyses (specified in the *Results* section) as a within-participant factor.

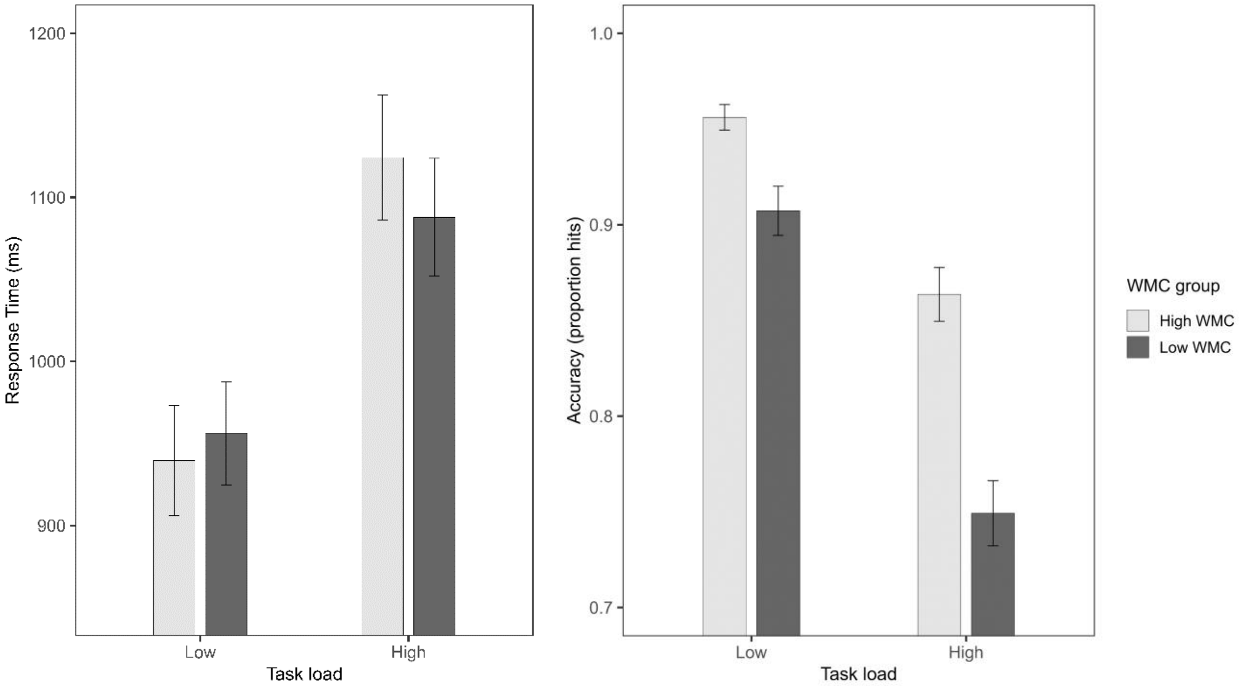
Results

tACS effects on WM performance

To evaluate the effects of stimulation on task performance, analyses were performed separately during the task and pre-post stimulation. On RTs during stimulation, results from the main effect of task difficulty were statistically significant, $F(1, 93) = 178.957$, $p < .001$, $\eta_p^2 = 0.658$, with faster RTs in the low load condition ($M_{low} = 948$ ms, $M_{high} = 1106$ ms). The interaction WM group × load was also significant, $F(1, 93) = 5.092$, $p = 0.026$, $\eta_p^2 = 0.052$ (see Figure 3), with slower RTs in the high WMC group compared to the low WMC group when the load was high. This must be interpreted considering the results in ACC, where the low WMC group performed worse than the high WMC group, especially in the most difficult trials. This might indicate that high WMC participants were slower because they were able to maintain all the information in WM and, to respond, they needed to perform a slower and more systematic search within their WM. In contrast, the fast response and low ACC of the low WMC group suggest that their responses were based on a faster and less accurate procedure, probably relying on familiarity due to the difficulty in maintaining all the information in their WM. However, no main effect or interaction involving stimulation reached significance (all $p \geq 0.153$). When the analysis was performed to compare the blocks before and after stimulation, no effect of stimulation was observed in RTs. Only the main effects of load (i.e., better performance in the low load condition), $F(1, 93) = 187.253$, $p < 0.001$, $\eta_p^2 = 0.668$, moment (i.e., better overall performance in the last task block), $F(1, 93) = 174.079$, $p < 0.001$, $\eta_p^2 = 0.652$, and the interaction WM group × load $F(1, 93) = 11.476$, $p = 0.001$, $\eta_p^2 = 0.110$, reached statistical significance.

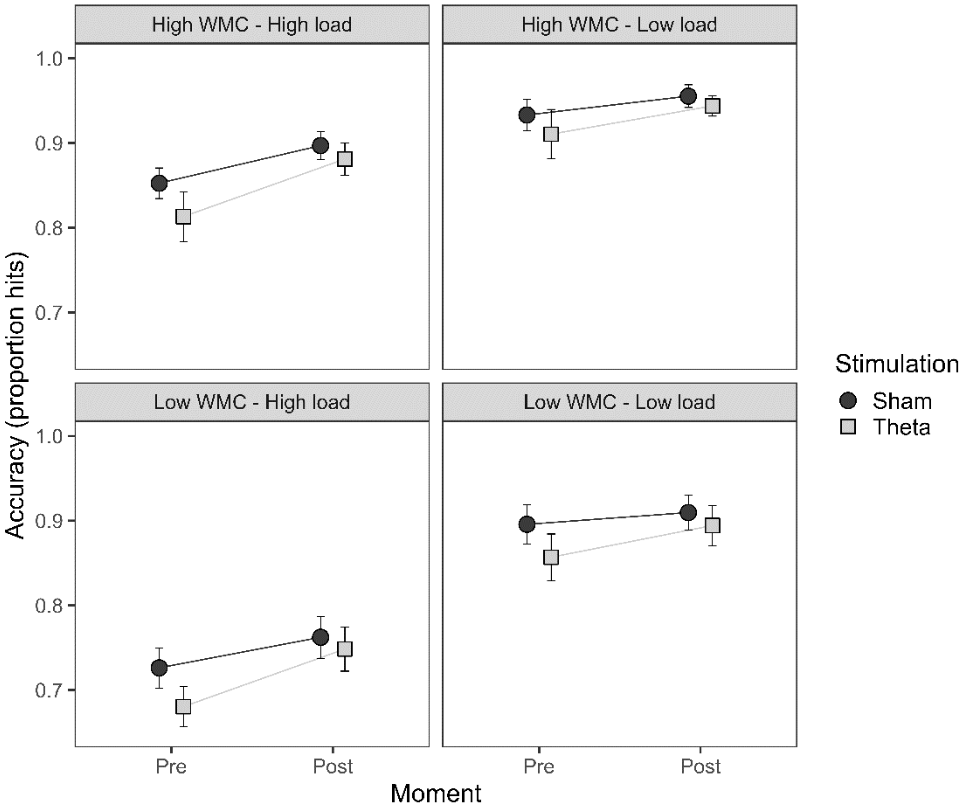
ACC analyses showed similar results. During stimulation, a significant WMC × load interaction reached statistical significance, $F(1, 93) = 19.319$, $p < .001$, $\eta_p^2 = 0.172$, with more pronounced differences between both WMC groups in the high load task condition (see Figure 3). These results show that the WMC groups and the task difficulty manipulations worked as intended because, overall, (1) the high WMC group performed better than the low WMC group at both high and low load conditions, $F(1, 93) = 22.439$, $p < .001$, $\eta_p^2 = 0.194$, and (2) accuracy was higher in the low load than the high load condition, $F(1, 93) = 282.201$, $p < .001$, $\eta_p^2 = 0.752$.

Figure 3. Working Memory Capacity group × Task difficulty interaction in response time (left panel) and accuracy (right panel). Error bars represent the standard error of the mean.



The WMC × task difficulty interaction was also observed when the ACC analyses were performed in the pre- and post-stimulation blocks, $F(1, 93) = 37.240$, $p < .001$, $\eta_p^2 = 0.286$. Additionally, the main effect of the moment was significant, $F(1, 93) = 19.828$, $p < .001$, $\eta_p^2 = 0.176$, showing an overall improvement in the task with practice ($M_{pre} = 0.83$, $M_{post} = 0.87$). Importantly, no main effects or interactions with stimulation were significant (all $p \geq 0.161$; see Figure 4). Not finding significant results of stimulation due to a ceiling effect was discarded considering that the average accuracy in the high load condition by the low WMC group (i.e., where effects were expected) was 0.75.

Figure 4. Working Memory Capacity group × task load interaction in accuracy split by theta (upper row) or sham (lower row) stimulation and before (left column) or after (right column) stimulation. Error bars represent the standard error of the mean.



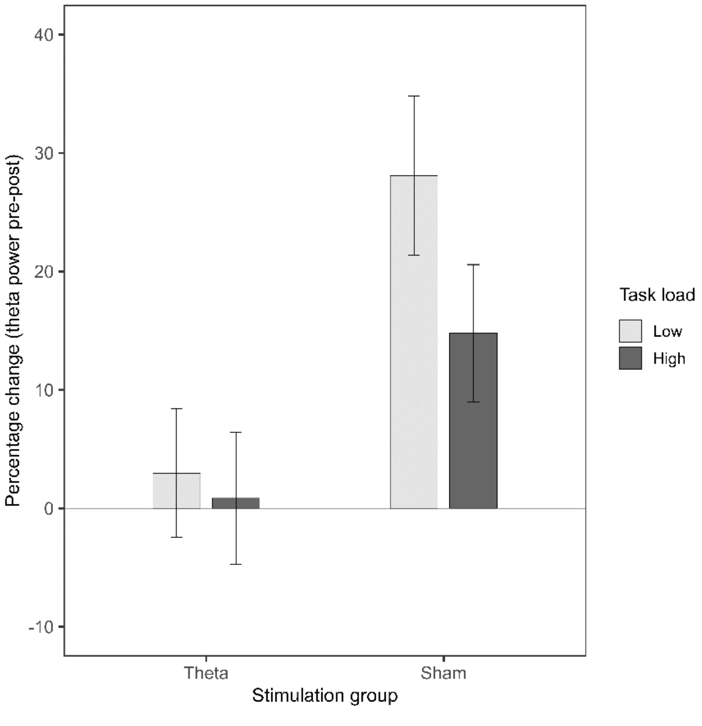
Although it was not preregistered, the same analyses were performed on d’ and K as dependent variables. Both analyses showed no significant main effect or interaction of stimulation (all $p \geq 0.153$). And, again, significant WMC × Load interactions were observed on d’ and K, with better performance in the case of the high WMC group in the more demanding condition, compared to the low WMC group. On d’, this interaction was only observed on the blocks pre- and post-stimulation, $F(1, 93) = 18.615$, $p < .001$, $\eta_p^2 = 0.167$, while it was significant both pre-post, $F(1, 93) = 26.050$, $p < .001$, $\eta_p^2 = 0.379$, and during stimulation, $F(1, 93) = 25.898$, $p < .001$, $\eta_p^2 = 0.218$, in K.

tACS Effects on Theta Power

Despite the absence of any behavioral effects, the impact of stimulation on brain oscillations was evaluated. Two participants were removed from the analyses because they lacked pre or post-stimulation EEG recording. First, the percentage change in resting-state theta power pre- and post-stimulation was calculated as the dependent variable. This analysis showed no significant effect of stimulation, $F(1, 91) = 0.118$, $p = 0.731$, $\eta_p^2 = 0.001$, WMC group, $F(1, 91) = 0.011$, $p = 0.916$, $\eta_p^2 = 1.236 \times 10^{-4}$, nor their interaction, $F(1, 91) = 1.497$, $p = 0.224$, $\eta_p^2 = 0.016$. However, the results differed when the analyses were performed during the task. For this, the same percentage change in theta power pre- and post-stimulation was calculated, but this time during the retention delay while participants performed the task. Then, it was introduced in a repeated measures ANOVA with WMC group and stimulation as between-subject factors and task difficulty as a within-subject factor. This analysis revealed a significant stimulation × task difficulty interaction, $F(1, 91) =$

18.615, $p = 0.046$, $\eta_p^2 = 0.043$, significant main effect of stimulation, $F(1, 91) = 6.043$, $p = 0.016$, $\eta_p^2 = 0.062$, and significant main effect of task difficulty, $F(1, 91) = 7.721$, $p = 0.007$, $\eta_p^2 = 0.078$. This analysis showed that the sham group experienced an increase in theta power after stimulation compared to the theta tACS group (M_{theta} = 1.92 %, M_{sham} = 21.44 %; see Figure 5). This effect was present in both task difficulty conditions but was particularly pronounced in the low load trials ($M_{low\text{-}sham}$ = 28.103 %, $M_{low\text{-}theta}$ = 2.969 %; $M_{high\text{-}sham}$ = 14.792 %, $M_{high\text{-}theta}$ = 0.862 %). Importantly, the theta stimulation group showed almost no increase in theta power, with no differences between load conditions, $t(94) = 0.538$, $p = 0.592$, $d = 0.052$.

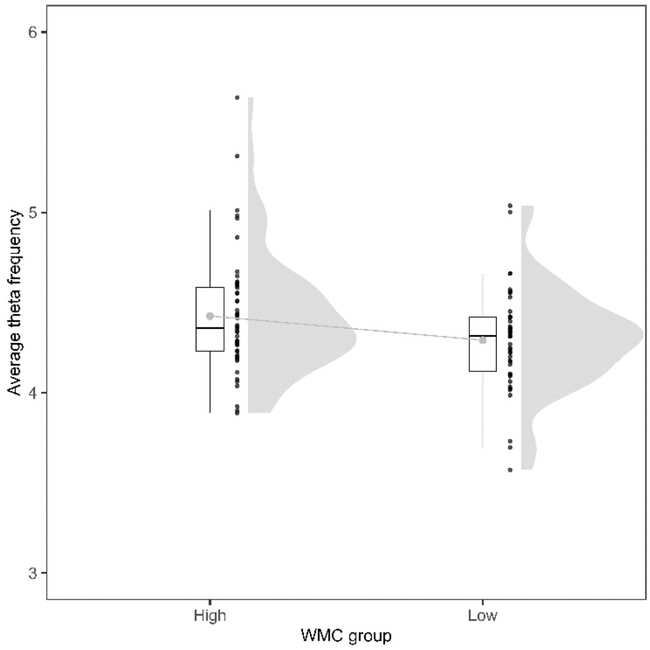
Figure 5. Stimulation group × task difficulty interaction in percentage theta change. Error bars represent the standard error of the mean.



Mean theta variations

Firstly, the prediction that low WMC participants must exhibit a higher mean theta frequency (allowing less gamma to couple and, thus, fewer items maintained) was tested. Then, the gF of the memory delay in the first EEG block (pre-stimulation) of both WMC groups was compared[3]. This analysis showed significant differences between the groups, $F(1, 93) = 4.335$, $p = 0.040$, $\eta_p^2 = 0.045$. However, the direction of the results was contrary to the theta-gamma coupling theory prediction: the high WMC participants exhibited a higher mean theta, compared to the low WMC participants ($M_{lowWMC} = 4.291$ Hz, $M_{highWMC} = 4.425$ Hz; see Figure 6). To test whether this result could depend on the high or low load trials, we again calculated gF but separately for 3-load and 5-load trials. Then, an ANOVA with task difficulty and WMC groups was performed again on the EEG block before stimulation, revealing no effect of task difficulty, $F(1, 93) = 1.894$, $p = 0.172$, $\eta_p^2 = 0.020$, or interaction with WMC group, $F(1, 93) = 0.801$, $p = 0.373$, $\eta_p^2 = 0.009$.

Figure 6. Raincloud plots of mean theta frequency during the retention delay in high and low WMC groups. The small dark points represent the mean of each participant in each condition. The light grey point joined by a grey line represents the means of both groups. The individual distributions for both WMC groups are represented in the split violin plots.



---

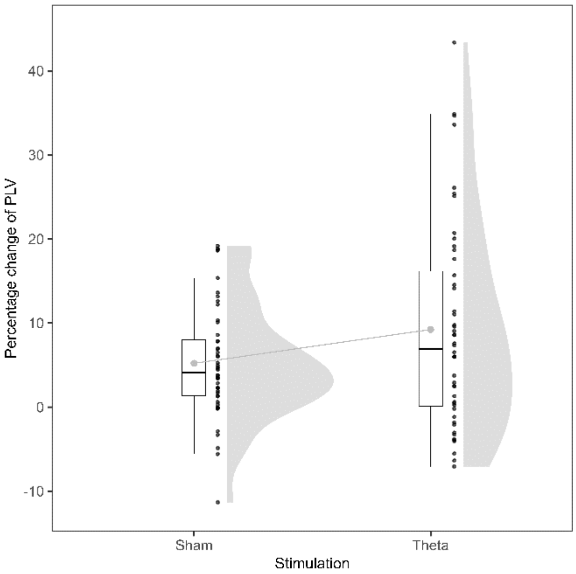
[3] The reported analyses were performed on all the trials but the same results were observed when the analyses were performed only on the hit trials (i.e. trials with successful item maintenance).

Secondly, the efficacy of tACS in modulating mean theta frequency was tested. For this, the percentage of change was calculated in both groups. Then, an ANOVA with WMC group and stimulation as between participants factors and percentage change as dependent variable was calculated. Results show no significant effect of stimulation, $F(1, 91) = 2.402$, $p = 0.125$, $\eta_p^2 = 0.026$, or WMC group x stimulation interaction, $F(1, 91) = 0.687$, $p = 0.409$, $\eta_p^2 = 0.007$. Raw EEG showed an overall decrease of the theta frequency in the last block of the task compared to the first block ($M_{first} = 4.361$ Hz, $M_{last} = 4.226$ Hz), independently of the stimulation. When the analysis was performed separately for the two types of load trials, no main effect of task load or interaction reached significance (all $p \geq 0.134$).

Phase Locking Value between Stimulation Electrodes

Changes in electrode synchronization during the WM delay were evaluated by calculating the percentage change in PLV before and after stimulation, both in the sham and theta stimulation groups. Initially, this calculation was performed using all trials. This analysis showed an overall increase in PLV in both the theta and sham groups ($M_{theta} = 9.36$ %, $M_{sham} = 5.35$ %). The difference between the two groups was only marginally significant, $F(1, 94) = 3.791$, $p = 0.055$, $\eta_p^2 = 0.039$. Since trials in which participants failed to respond correctly might exhibit lower synchronization, the same analysis was conducted using only the hit trials. Again, both groups showed an increase in PLV after the stimulation block ($M_{theta} = 9.23$ %, $M_{sham} = 5.21$ %); however, in this case, the difference between the two groups was statistically significant, $F(1, 94) = 4.037$, $p = 0.047$, $\eta_p^2 = 0.041$, with a greater increase in synchronization in the theta stimulation group (see Figure 7). Nevertheless, given the exploratory nature of the two analyses, these results should be interpreted with caution.

Figure 7. Raincloud plots of percentage change of Phase Locking Value (PLV) in sham and theta stimulation groups. The small dark points represent the mean of each participant in each condition. The light grey point joined by a grey line represents the means of both groups. The individual distributions for both stimulation groups are represented in the split violin plots.



Discussion

The role of individual factors, such as baseline working memory capacity, can play a crucial role in maximizing the benefits of electrical stimulation for modulating WM. In the present study, we consider this factor—WMC—to test the theory of theta-gamma coupling, which posits that lower theta frequency rhythms would facilitate greater gamma coupling, thereby allowing more information to be maintained in WM. To this end, we applied 4 Hz transcranial alternating current stimulation (tACS) to slow down theta rhythms in the fronto-parietal network. We also manipulated task difficulty, given previous studies that show tACS modulation effects primarily under challenging task conditions.

The behavioral results revealed the successful manipulation of both WMC groups and task difficulty. However, contrary to our hypotheses, participants receiving 4 Hz theta tACS did not exhibit any benefits over the sham group. These null effects were observed across measures of response time, accuracy, d’, and K. We replicated the results from Experiments 1 and 2A of Biel et al. (2022), given that they neither found behavioral modulation after tACS in a delayed letter discrimination task similar as the one used by Polania et al. (2012), i.e., the easy trials of the task used in the present experiment. However, contrary to the WM improvement found in their Experiment 2B, we did not observe any modulation in more demanding trials, despite using the same montage.

One aspect in which our experiment differs from that of Biel et al. (2022) and which could explain this divergence of results is that the task used was not the same. While they found a modulation in discriminability (d’) in a difficult 3-back task, we did not observe any effect on the d’ metric in the difficult version of the delayed letter discrimination task. In addition to item maintenance, the 3-back might require updating and controlling of WM representations (Gajewski et al., 2018; Rac-Lubashevsky & Kessler, 2016), as well as motor preparation (Pavlov & Kotchoubey, 2022). These central executive-dependent functions seem to be closely linked to fronto-parietal theta synchronization (Hanslmayr et al., 2008; Sauseng et al., 2005, 2010). Therefore, it is possible that tasks such as the 3-back—where executive control is involved—are more likely to be modulated by theta tACS. Against this argument, Vosskuhl et al., (2015) applied tACS and compared the effects of tACS on the 3-back task measured before and after stimulation and also found no modulation. However, the stimulation frequency was adapted to each participant and on average was close to 4 Hz (i.e., the same as we used), in contrast to the 6 Hz applied by Biel et al. (2022). Therefore, another possibility is that the theta modulation only shows benefits in n-back tasks with faster theta frequencies (Violante et al., 2017; but see Chander et al., 2016).

In contrast, EEG recordings revealed some differences between stimulation groups. First, contrary to our hypothesis, we observed a close to zero increase in theta power in the theta tACS group compared to the sham condition. This was an unexpected finding, as an increase in power at the stimulated frequency is typically seen after tACS in theta and other frequency bands (see Pavlov & Kotchoubey, 2022 and Vogeti et al., 2022, for recent reviews). While the effects of tACS are typically explained in terms of entrainment or long-range synchronization between stimulated regions, Spike-Timing Dependent Plasticity (STDP) has also been proposed as a mechanism by which tACS influences neural activity (Bi & Poo, 1998; Zaehle et al., 2010). STDP suggests that the timing between endogenous neural firing and tACS-induced oscillations determines synaptic changes. If a pre-synaptic event occurs just before a post-synaptic event, there is an increase in synaptic strength—i.e., long-term potentiation (LTP). Conversely, if a post-synaptic event precedes a pre-synaptic event, a synaptic weakening takes place—i.e., long-term depression (LTD) (Vogeti et al., 2022; Vossen et al., 2015; Zaehle et al., 2010). Applied to tACS, LTD could occur when the stimulation frequency is faster than the regular firing rate of the targeted network. Therefore, it might be that the observed attenuation in theta power was due to LTD resulting from a timing mismatch between the individual's endogenous theta rhythm and the tACS (Wischnewski & Schutter, 2017). Chander et al. (2016) also observed a decrease in theta power (and performance) after applying frontal 6 Hz tACS. In their case, the reduction was attributed to interference between the tACS stimulation and the participant's endogenous oscillations. Whether due to LTD or simply to a mismatch between oscillation rhythms, the lack of synchronization between endogenous and exogenous oscillations can lead to a reduction in power instead of the intended increase. We believe these factors should be considered in future studies.

Second, contrary to the theta-gamma coupling theory and our pre-registered hypothesis, we did not find that the high WMC group, who performed better on the task, exhibited a lower mean theta frequency before any stimulation. Instead, they showed a higher mean theta frequency compared to the low WMC group. This pattern of results aligns more closely with an alternative model regarding the role of theta in the maintenance of information in working memory (Herman et al., 2013; Van Vugt et al., 2014). According to this recent framework, a complete gamma burst represents a single item, and each theta cycle nests one gamma burst. When multiple items need to be maintained in WM, several theta cycles must be reactivated to refresh all the items, which might potentially explain the limited capacity of WM (Van Vugt et al., 2014). Faster theta cycles would therefore enable quicker refreshing of WM items, preventing their loss. A person capable of activating more items in a shorter time (i.e., with faster theta rhythms) should be able to maintain more information in their WM. Given the differences in mean theta frequency for the two WMC groups, the present results are more consistent with this explanation.

Evidence for the theta gamma coupling model comes from different sources including rodents and nonhuman primate invasive measures (Lisman & Jensen, 2013), and human intracranial EEG at the hippocampus (Axmacher et al., 2010; Chaieb et al., 2015). However, using human EEG the results are somewhat more mixed. While Sauseng et al. (2009) found cross-frequency synchronization at posterior electrodes, Kamiński et al. (2011) observed it but only at one frontal electrode and without correcting for multiple comparisons (see Chuderski, 2016). Furthermore, Maleninská et al. (2021) attempted to replicate Kamiński et al.'s (2011) results and, after applying corrections, found no evidence for the theory. Similar to Maleninská et al. (2021), the simple prediction that theta frequency should be lower in low WMC individuals was not observed in the present data. Also, contrary to the model, no theta frequency modulations were observed concerning the task difficulty manipulation in either WMC group. These results open a discussion on whether the hypotheses derived from this theory can be observed in human EEG data.

Besides, the mean theta frequency was not more reduced in the stimulation group. According to the theta-gamma coupling model, this could explain the lack of behavioral effects but the fact that no lower theta was observed in the high WMC group raises doubts about this reasoning. To be noted, in the context of the theta-gamma coupling model, tACS has been shown to reduce or increase theta power (Chander et al., 2016; Pahor & Jausovec, 2018; Vosskuhl et al., 2015) or phase synchronization (Alekseichuk et al., 2017; Chandler et al., 2017). Only two studies have examined the frequency peak modulations after theta tACS. Kleinert et al. (2017) studied the theta modulations in the resting state before and after 6 Hz stimulation and only found alpha modulations but not for the theta band. Vosskuhl et al. (2015) found an amplitude increase at 6.7 Hz after theta stimulation even though the average frequency of applied stimulation was 4.23 Hz. Then, the present results advocated caution in the use of tACS as a way of modulating specific frequencies, given that it might have no effect or an unexpected effect on frequencies different than the targeted ones.

Third, our exploratory analyses regarding the connectivity between the two stimulation electrodes showed a greater increase in phase synchrony following theta stimulation, consistent with previous findings (Alekseichuk et al., 2017; Chandler et al., 2017). However, this effect was only observed when error trials were excluded from the analyses. Although this might seem logical, as synchronization would be expected to be successful primarily in hit trials, the exploratory nature of these results warrants caution. A pre-registered replication study is necessary to confirm these findings.

Lastly, the present work has several limitations. First, only a few fronto-parietal electrodes were placed and analyzed but other studies have found relevant theta-gamma coupling EEG information in other electrode locations (e.g. Chander et al., 2016; Sauseng et al., 2009). As mentioned before, the comparison of the present results with those found by Biel et al. (2022) is limited given the difference in the tasks that were used. Other research could do a direct comparison between the results obtained using an n-back and a delayed letter response task, similar to Pahor & Jausovec (2018). Finally, the exploratory nature of some of our analyses requires replication before building stronger conclusions.

In summary, the present study shows that focal in-phase 4 Hz tACS does not modulate performance during a WM task that primarily relies on the maintenance of information, even when task difficulty and baseline WM capacity are controlled. Electrophysiological results indicated no power increase following stimulation, in contrast to the sham group, which displayed a strong increase in mean theta power over time on task. This lack of effect might be due to insufficient synchronization between the endogenous and exogenous frequency rhythms. EEG data also revealed stronger phase synchronization between the stimulation electrodes when participants responded correctly, suggesting an induced phase synchronization by tACS. However, the results are not entirely consistent with the theta-gamma coupling model, particularly concerning (1) the higher mean theta frequency observed in high WMC participants, who performed better in the WM task, and (2) the absence of mean theta frequency modulation associated with WM load during the task. Therefore, the theta-gamma coupling model, which is the most established framework for explaining WM maintenance, may need to be revisited.

References

Abellaneda-Pérez, K., Vaqué-Alcázar, L., Perellón-Alfonso, R., Bargalló, N., Kuo, M.-F., Pascual-Leone, A., Nitsche, M. A., & Bartrés-Faz, D. (2020). Differential tDCS and tACS Effects on Working Memory-Related Neural Activity and Resting-State Connectivity. *Frontiers in Neuroscience*, *13*. https://doi.org/10.3389/fnins.2019.01440

Alekseichuk, I., Falchier, A. Y., Linn, G., Xu, T., Milham, M. P., Schroeder, C. E., & Opitz, A. (2019). Electric field dynamics in the brain during multi-electrode transcranial electric stimulation. *Nature Communications*, *10*(1), 2573. https://doi.org/10.1038/s41467-019-10581-7

Alekseichuk, I., Pabel, S., Antal, A., & Paulus, W. (2017). Intrahemispheric theta rhythm desynchronization impairs working memory. *RESTORATIVE NEUROLOGY AND NEUROSCIENCE*, *35*(2), 147-158. https://doi.org/10.3233/RNN-160714

Alekseichuk, I., Turi, Z., de Lara, G., Antal, A., & Paulus, W. (2016). Spatial Working Memory in Humans Depends on Theta and High Gamma Synchronization in the Prefrontal Cortex. *CURRENT BIOLOGY*, *26*(12), 1513-1521. https://doi.org/10.1016/j.cub.2016.04.035

Antal, A., Boros, K., Poreisz, C., Chaieb, L., Terney, D., & Paulus, W. (2008). Comparatively weak after-effects of transcranial alternating current stimulation (tACS) on cortical excitability in humans. *Brain Stimulation*, *1*(2), 97-105. https://doi.org/10.1016/j.brs.2007.10.001

Antal, A., Luber, B., Brem, A., Bikson, M., Brunoni, A., Kadosh, R., Dubljevic, V., Fecteau, S., Ferreri, F., Floel, A., Hallett, M., Hamilton, R., Herrmann, C., Lavidor, M., Loo, C., Lustenberger, C., Machado, S., Miniussi, C., Moliadze, V., … Paulus, W. (2022). Non-invasive brain stimulation and neuroenhancement. *CLINICAL NEUROPHYSIOLOGY PRACTICE*, *7*, 146-165. https://doi.org/10.1016/j.cnp.2022.05.002

Arciniega, H., Gözenman, F., Jones, K. T., Stephens, J. A., & Berryhill, M. E. (2018). Frontoparietal tDCS Benefits Visual Working Memory in Older Adults With Low Working Memory Capacity. *Frontiers in Aging Neuroscience*, *10*. https://doi.org/10.3389/fnagi.2018.00057

Axmacher, N., Henseler, M. M., Jensen, O., Weinreich, I., Elger, C. E., & Fell, J. (2010). Cross-frequency coupling supports multi-item working memory in the human hippocampus. *Proceedings of the National Academy of Sciences*, *107*(7), 3228-3233. https://doi.org/10.1073/pnas.0911531107

Bender, M., Romei, V., & Sauseng, P. (2019). Slow Theta tACS of the Right Parietal Cortex Enhances Contralateral Visual Working Memory Capacity. *BRAIN TOPOGRAPHY*, *32*(3), 477-481. https://doi.org/10.1007/s10548-019-00702-2

Bi, G., & Poo, M. (1998). Synaptic Modifications in Cultured Hippocampal Neurons: Dependence on Spike Timing, Synaptic Strength, and Postsynaptic Cell Type. *Journal of Neuroscience*, *18*(24), 10464-10472. https://doi.org/10.1523/JNEUROSCI.18-24-10464.1998

Biel, A. L., Sterner, E., Röll, L., & Sauseng, P. (2022). Modulating verbal working memory with fronto-parietal transcranial electric stimulation at theta frequency: Does it work? *European Journal of Neuroscience*, *55*(2), 405-425. https://doi.org/10.1111/ejn.15563

Booth, S., Taylor, J., Brown, L., & Pobric, G. (2022). The effects of transcranial alternating current stimulation on memory performance in healthy adults: A systematic review. *CORTEX*, *147*, 112-139. https://doi.org/10.1016/j.cortex.2021.12.001

Bortoletto, M., Rodella, C., Salvador, R., Miranda, P. C., & Miniussi, C. (2016). Reduced Current Spread by Concentric Electrodes in Transcranial Electrical Stimulation (tES). *Brain Stimulation*, *9*(4), 525-528. https://doi.org/10.1016/j.brs.2016.03.001

Brooks, H., Goodman, M. S., Bowie, C. R., Zomorrodi, R., Blumberger, D. M., Butters, M. A., Daskalakis, Z. J., Fischer, C. E., Flint, A., Herrmann, N., Kumar, S., Mah, L., Mulsant, B. H., Pollock, B. G., Voineskos, A. N., & Rajji, T. K. (2020). Theta–gamma coupling and ordering information: A stable brain–behavior relationship across cognitive tasks and clinical conditions. *Neuropsychopharmacology*, *45*(12), 2038-2047. https://doi.org/10.1038/s41386-020-0759-z

Buzsáki, G., & Draguhn, A. (2004). Neuronal Oscillations in Cortical Networks. *Science*, *304*(5679), 1926-1929. https://doi.org/10.1126/science.1099745

Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2006). High Gamma Power Is Phase-Locked to Theta Oscillations in Human Neocortex. *Science*, *313*(5793), 1626-1628. https://doi.org/10.1126/science.1128115

Chaieb, L., Leszczynski, M., Axmacher, N., Höhne, M., Elger, C. E., & Fell, J. (2015). Theta-gamma phase-phase coupling during working memory maintenance in the human hippocampus. *Cognitive Neuroscience*, *6*(4), 149-157. https://doi.org/10.1080/17588928.2015.1058254

Chander, B. S., Witkowski, M., Braun, C., Robinson, S. E., Born, J., Cohen, L. G., Birbaumer, N., & Soekadar, S. R. (2016). tACS Phase Locking of Frontal Midline Theta Oscillations Disrupts Working Memory Performance. *Frontiers in Cellular Neuroscience*, *10*. https://doi.org/10.3389/fncel.2016.00120

Chandler, C. M., Follett, M. E., Porter, N. J., Liang, K. Y., Vallender, E. J., Miller, G. M., Rowlett, J. K., & Platt, D. M. (2017). Persistent negative effects of alcohol drinking on aspects of novelty-directed behavior in male rhesus macaques. *Alcohol*, *63*, 19-26. https://doi.org/10.1016/j.alcohol.2017.03.002

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J.-D. (2017). The Distributed Nature of Working Memory. *Trends in Cognitive Sciences*, *21*(2), 111-124. https://doi.org/10.1016/j.tics.2016.12.007

Chuderski, A. (2016). Fluid Intelligence and the Cross-Frequency Coupling of Neuronal Oscillations. *SPANISH JOURNAL OF PSYCHOLOGY*, *19*. https://doi.org/10.1017/sjp.2016.86

Clayton, M. S., Yeung, N., & Cohen Kadosh, R. (2019). Electrical stimulation of alpha oscillations stabilizes performance on visual attention tasks. *Journal of Experimental Psychology: General*, *148*(2), 203. https://doi.org/10.1037/xge0000502

Cohen, M. X. (2014). *Analyzing Neural Time Series Data: Theory and Practice*. MIT Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87-114. https://doi.org/10.1017/S0140525X01003922

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. https://doi.org/10.1016/j.jneumeth.2003.10.009

D’Esposito, M., & Postle, B. R. (2015). The Cognitive Neuroscience of Working Memory. *Annual Review of Psychology*, *66*(Volume 66, 2015), 115-142. https://doi.org/10.1146/annurev-psych-010814-015031

Elyamany, O., Leicht, G., Herrmann, C. S., & Mulert, C. (2021). Transcranial alternating current stimulation (tACS): From basic mechanisms towards first applications in psychiatry. *European Archives of Psychiatry and Clinical Neuroscience*, *271*(1), 135-156. https://doi.org/10.1007/s00406-020-01209-9

Fernández, A., Pinal, D., Díaz, F., & Zurrón, M. (2021). Working memory load modulates oscillatory activity and the distribution of fast frequencies across frontal theta phase during working memory maintenance. *Neurobiology of Learning and Memory*, *183*, 107476. https://doi.org/10.1016/j.nlm.2021.107476

Fertonani, A., Ferrari, C., & Miniussi, C. (2015). What do you feel if I apply transcranial electric stimulation? Safety, sensations and secondary induced effects. *Clinical Neurophysiology*, *126*(11), 2181-2188. https://doi.org/10.1016/j.clinph.2015.03.015

Feurra, M., Galli, G., Pavone, E., Rossi, A., & Rossi, S. (2016). Frequency-specific insight into short-term memory capacity. *JOURNAL OF NEUROPHYSIOLOGY*, *116*(1), 153-158. https://doi.org/10.1152/jn.01080.2015

Fröhlich, F., Sellers, K. K., & Cordle, A. L. (2015). Targeting the neurophysiology of cognitive systems with transcranial alternating current stimulation. *Expert Review of Neurotherapeutics*, *15*(2), 145-167. https://doi.org/10.1586/14737175.2015.992782

Gajewski, P. D., Hanisch, E., Falkenstein, M., Thönes, S., & Wascher, E. (2018). What Does the n-Back Task Measure as We Get Older? Relations Between Working-Memory Measures and Other Cognitive Functions Across the

Lifespan. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02208

Gazzaley, A., Rissman, J., & D’esposito, M. (2004). Functional connectivity during working memory maintenance. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(4), 580-599. https://doi.org/10.3758/CABN.4.4.580

Hanslmayr, S., Pastötter, B., Bäuml, K.-H., Gruber, S., Wimber, M., & Klimesch, W. (2008). The Electrophysiological Dynamics of Interference during the Stroop Task. *Journal of Cognitive Neuroscience*, *20*(2), 215-225. Journal of Cognitive Neuroscience. https://doi.org/10.1162/jocn.2008.20020

Helfrich, R. F., Schneider, T. R., Rach, S., Trautmann-Lengsfeld, S. A., Engel, A. K., & Herrmann, C. S. (2014). Entrainment of Brain Oscillations by Transcranial Alternating Current Stimulation. *Current Biology*, *24*(3), 333-339. https://doi.org/10.1016/j.cub.2013.12.041

Hemmerich, K., Lupiáñez, J., Luna, F. G., & Martín-Arévalo, E. (2023). The mitigation of the executive vigilance decrement via HD-tDCS over the right posterior parietal cortex and its association with neural oscillations. *Cerebral Cortex*, bhac540. https://doi.org/10.1093/cercor/bhac540

Herman, P. A., Lundqvist, M., & Lansner, A. (2013). Nested theta to gamma oscillations and precise spatiotemporal firing during memory retrieval in a simulated attractor network. *Brain Research*, *1536*, 68-87. https://doi.org/10.1016/j.brainres.2013.08.002

Holz, E. M., Glennon, M., Prendergast, K., & Sauseng, P. (2010). Theta–gamma phase synchronization during memory matching in visual working memory.

NeuroImage, *52*(1), 326-335. https://doi.org/10.1016/j.neuroimage.2010.04.003

Hooper, G. S. (2005). Comparison of the distributions of classical and adaptively aligned EEG power spectra. *International Journal of Psychophysiology*, *55*(2), 179-189. https://doi.org/10.1016/j.ijpsycho.2004.07.008

Hoy, K., Bailey, N., Arnold, S., Windsor, K., John, J., Daskalakis, Z., & Fitzgerald, P. (2015). The effect of gamma-tACS on working memory performance in healthy controls. *BRAIN AND COGNITION*, *101*, 51-56. https://doi.org/10.1016/j.bandc.2015.11.002

Hsu, T., Juan, C., & Tseng, P. (2016). Individual Differences and State-Dependent Responses in Transcranial Direct Current Stimulation. *FRONTIERS IN HUMAN NEUROSCIENCE*, *10*. https://doi.org/10.3389/fnhum.2016.00643

JASP Team. (2022). *JASP (Version 0.19.0.0) [Computer software]* [Software]. https://jasp-stats.org/

Jausovec, N., & Jausovec, K. (2014). Increasing working memory capacity with theta transcranial alternating current stimulation (tACS). *BIOLOGICAL PSYCHOLOGY*, *96*, 42-47. https://doi.org/10.1016/j.biopsycho.2013.11.006

Jaušovec, N., Jaušovec, K., & Pahor, A. (2014). The influence of theta transcranial alternating current stimulation (tACS) on working memory storage and processing functions. *Acta Psychologica*, *146*, 1-6. https://doi.org/10.1016/j.actpsy.2013.11.011

Jensen, O., & Lisman, J. E. (1996). Novel lists of 7 +/- 2 known items can be reliably stored in an oscillatory short-term memory network: Interaction with long-term memory. *Learning & Memory*, *3*(2-3), 257-263.

Jones, K., Arciniega, H., & Berryhill, M. (2019). Replacing tDCS with theta tACS provides selective, but not general WM benefits. *BRAIN RESEARCH*, *1720*. https://doi.org/10.1016/j.brainres.2019.146324

Kamiński, J., Brzezicka, A., & Wróbel, A. (2011). Short-term memory capacity (7 ± 2) predicted by theta to gamma cycle length ratio. *Neurobiology of Learning and Memory*, *95*(1), 19-23. https://doi.org/10.1016/j.nlm.2010.10.001

Kleinert, M., Szymanski, C., & Muller, V. (2017). Frequency-Unspecific Effects of theta-tACS Related to a Visuospatial Working Memory Task. *FRONTIERS IN HUMAN NEUROSCIENCE*, *11*. https://doi.org/10.3389/fnhum.2017.00367

Klimesch, W. (1997). EEG-alpha rhythms and memory processes. *International Journal of Psychophysiology*, *26*(1), 319-340. https://doi.org/10.1016/S0167-8760(97)00773-3

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, *29*(2), 169-195. https://doi.org/10.1016/S0165-0173(98)00056-3

Köster, M., Friese, U., Schöne, B., Trujillo-Barreto, N., & Gruber, T. (2014). Theta–gamma coupling during episodic retrieval in the human EEG. *Brain Research*, *1577*, 57-68. https://doi.org/10.1016/j.brainres.2014.06.028

Krause, B., & Cohen Kadosh, R. (2014). Not all brains are created equal: The relevance of individual differences in responsiveness to transcranial electrical stimulation. *Frontiers in Systems Neuroscience*, *8*, 25. https://doi.org/10.3389/fnsys.2014.00025

Lachaux, J.-P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, *8*(4), 194-208.

https://doi.org/10.1002/(SICI)1097-0193(1999)8:4<194::AID-HBM4>3.0.CO;2-C

Lisman, J. E., & Jensen, O. (2013). The Theta-Gamma Neural Code. *Neuron*, *77*(6), 1002-1016. https://doi.org/10.1016/j.neuron.2013.03.007

Maleninská, K., Rudolfová, V., Šulcová, K., Koudelka, V., Brunovský, M., Horáček, J., & Nekovářová, T. (2021). Is short-term memory capacity (7±2) really predicted by theta to gamma cycle length ratio? *Behavioural Brain Research*, *414*, 113465. https://doi.org/10.1016/j.bbr.2021.113465

Miller, E. K., Lundqvist, M., & Bastos, A. M. (2018). Working Memory 2.0. *Neuron*, *100*(2), 463-475. https://doi.org/10.1016/j.neuron.2018.09.023

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*(1), 46-59. https://doi.org/10.1002/hbm.20131

Pahor, A., & Jausovec, N. (2018). The Effects of Theta and Gamma tACS on Working Memory and Electrophysiology. *FRONTIERS IN HUMAN NEUROSCIENCE*, *11*. https://doi.org/10.3389/fnhum.2017.00651

Pavlov, Y. G., & Kotchoubey, B. (2022). Oscillatory brain activity and maintenance of verbal and visual working memory: A systematic review. *Psychophysiology*, *59*(5), e13735. https://doi.org/10.1111/psyp.13735

Polania, R., Nitsche, M., Korman, C., Batsikadze, G., & Paulus, W. (2012). The Importance of Timing in Segregated Theta Phase-Coupling for Cognitive Performance. *CURRENT BIOLOGY*, *22*(14), 1314-1318. https://doi.org/10.1016/j.cub.2012.05.021

Psychology Software Tools, Inc. (2016). *E-Prime 3.0* (Retrieved from https://support.pstnet.com/) [Software].

Rac-Lubashevsky, R., & Kessler, Y. (2016). Decomposing the n-back task: An individual differences study using the reference-back paradigm. *Neuropsychologia*, *90*, 190-199. https://doi.org/10.1016/j.neuropsychologia.2016.07.013

Reinhart, R. M. G., & Nguyen, J. A. (2019). Working memory revived in older adults by synchronizing rhythmic brain circuits. *Nature Neuroscience*, *22*(5), 820-827. https://doi.org/10.1038/s41593-019-0371-x

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences*, *105*(16), 5975-5979. https://doi.org/10.1073/pnas.0711295105

Roux, F., & Uhlhaas, P. J. (2014). Working memory and neural oscillations: Alpha–gamma versus theta–gamma codes for distinct WM information? *Trends in Cognitive Sciences*, *18*(1), 16-25. https://doi.org/10.1016/j.tics.2013.10.010

RStudio Team. (2022). *RStudio: Integrated Development Environment for R* (Versión 2022.02.3) [Software]. RStudio, Inc. http://www.rstudio.com/

Rudroff, T., Workman, C. D., Fietsam, A. C., & Kamholz, J. (2020). Response Variability in Transcranial Direct Current Stimulation: Why Sex Matters. *Frontiers in Psychiatry*, *11*. https://www.frontiersin.org/articles/10.3389/fpsyt.2020.00585

Sahu, P., & Tseng, P. (2021). Frontoparietal theta tACS nonselectively enhances encoding, maintenance, and retrieval stages in visuospatial working memory. *NEUROSCIENCE RESEARCH*, *172*, 41-50. https://doi.org/10.1016/j.neures.2021.05.005

Salehinejad, M. A., Wischnewski, M., Ghanavati, E., Mosayebi-Samani, M., Kuo, M.-F., & Nitsche, M. A. (2021). Cognitive functions and underlying parameters of human brain physiology are associated with chronotype. *Nature Communications*, *12*(1), 4672. https://doi.org/10.1038/s41467-021-24885-0

Saturnino, G. B., Madsen, K. H., Siebner, H. R., & Thielscher, A. (2017). How to target inter-regional phase synchronization with dual-site Transcranial Alternating Current Stimulation. *NeuroImage*, *163*, 68-80. https://doi.org/10.1016/j.neuroimage.2017.09.024

Sauseng, P., Griesmayr, B., Freunberger, R., & Klimesch, W. (2010). Control mechanisms in working memory: A possible function of EEG theta oscillations. *Neuroscience & Biobehavioral Reviews*, *34*(7), 1015-1022. https://doi.org/10.1016/j.neubiorev.2009.12.006

Sauseng, P., Klimesch, W., Heise, K. F., Gruber, W. R., Holz, E., Karim, A. A., Glennon, M., Gerloff, C., Birbaumer, N., & Hummel, F. C. (2009). Brain oscillatory substrates of visual short-term memory capacity. *Current Biology: CB*, *19*(21), 1846-1852. https://doi.org/10.1016/j.cub.2009.08.062

Sauseng, P., Klimesch, W., Schabus, M., & Doppelmayr, M. (2005). Fronto-parietal EEG coherence in theta and upper alpha reflect central executive functions of working memory. *International Journal of Psychophysiology*, *57*(2), 97-103. https://doi.org/10.1016/j.ijpsycho.2005.03.018

Sauseng, P., Peylo, C., Biel, A. L., Friedrich, E. V. C., & Romberg-Taylor, C. (2019). Does cross-frequency phase coupling of oscillatory brain activity contribute to a better understanding of visual working memory? *British Journal of Psychology*, *110*(2), 245-255. https://doi.org/10.1111/bjop.12340

Schwab, B. C., Misselhorn, J., & Engel, A. K. (2019). Modulation of large-scale cortical coupling by transcranial alternating current stimulation. *Brain Stimulation*, *12*(5), 1187-1196. https://doi.org/10.1016/j.brs.2019.04.013

Tseng, P., Chang, Y.-T., Chang, C.-F., Liang, W.-K., & Juan, C.-H. (2016). The critical role of phase difference in gamma oscillation within the temporoparietal network for binding visual working memory. *Scientific Reports*, *6*(1), 32138. https://doi.org/10.1038/srep32138

Tseng, P., Iu, K.-C., & Juan, C.-H. (2018). The critical role of phase difference in theta oscillation between bilateral parietal cortices for visuospatial working memory. *Scientific Reports*, *8*(1), 349. https://doi.org/10.1038/s41598-017-18449-w

Van Vugt, M. K., Chakravarthi, R., & Lachaux, J.-P. (2014). For whom the bell tolls: Periodic reactivation of sensory cortex in the gamma band as a substrate of visual working memory maintenance. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00696

Violante, I. R., Li, L. M., Carmichael, D. W., Lorenz, R., Leech, R., Hampshire, A., Rothwell, J. C., & Sharp, D. J. (2017). Externally induced frontoparietal synchronization modulates network dynamics and enhances working memory performance. *eLife*, *6*, e22001. https://doi.org/10.7554/eLife.22001

Vogeti, S., Boetzel, C., & Herrmann, C. S. (2022). Entrainment and Spike-Timing Dependent Plasticity – A Review of Proposed Mechanisms of Transcranial

Alternating Current Stimulation. *Frontiers in Systems Neuroscience*, *16*. https://doi.org/10.3389/fnsys.2022.827353

Vossen, A., Gross, J., & Thut, G. (2015). Alpha Power Increase After Transcranial Alternating Current Stimulation at Alpha Frequency (α-tACS) Reflects Plastic Changes Rather Than Entrainment. *Brain Stimulation*, *8*(3), 499-508. https://doi.org/10.1016/j.brs.2014.12.004

Vosskuhl, J., Huster, R., & Herrmann, C. (2015). Increase in short-term memory capacity induced by down-regulating individual theta frequency via transcranial alternating current stimulation. *FRONTIERS IN HUMAN NEUROSCIENCE*, *9*. https://doi.org/10.3389/fnhum.2015.00257

Weinrich, C. A., Brittain, J.-S., Nowak, M., Salimi-Khorshidi, R., Brown, P., & Stagg, C. J. (2017). Modulation of Long-Range Connectivity Patterns via Frequency-Specific Stimulation of Human Cortex. *Current Biology*, *27*(19), 3061-3068.e3. https://doi.org/10.1016/j.cub.2017.08.075

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wischnewski, M., & Schutter, D. J. L. G. (2017). After-effects of transcranial alternating current stimulation on evoked delta and theta power. *Clinical Neurophysiology*, *128*(11), 2227-2232. https://doi.org/10.1016/j.clinph.2017.08.029

Wolinski, N., Cooper, N., Sauseng, P., & Romei, V. (2018). The speed of parietal theta frequency drives visuospatial working memory capacity. *PLOS BIOLOGY*, *16*(3). https://doi.org/10.1371/journal.pbio.2005348

Wu, Y.-J., Tseng, P., Chang, C.-F., Pai, M.-C., Hsu, K.-S., Lin, C.-C., & Juan, C.-H. (2014). Modulating the interference effect on spatial working memory by applying transcranial direct current stimulation over the right dorsolateral prefrontal cortex. *Brain and Cognition*, *91*, 87-94. https://doi.org/10.1016/j.bandc.2014.09.002

Zaehle, T., Rach, S., & Herrmann, C. S. (2010). Transcranial Alternating Current Stimulation Enhances Individual Alpha Activity in Human EEG. *PLOS ONE*, *5*(11), e13766. https://doi.org/10.1371/journal.pone.0013766

Chapter 4: Distinguishing Between Temporary and Permanent Removal in Verbal Working Memory

Abstract

We employed the retro-cue paradigm to examine performance improvements resulting from permanent versus temporary removal in verbal working memory (WM). Permanent removal entails discarding a subset of WM representations marked as definitively irrelevant, while temporary removal involves momentarily setting aside the uncued subset of representations from the attentional focus, preserving accessibility for later refocusing. We observed that both permanent and temporary removal led to marked progressive reductions in reaction time and errors across cue-target intervals (200, 400, 800, and 1600 ms), reflecting the gradual simplification of the search set following informative cues. Although removal conditions did not differ in accuracy, responses were slower in the temporary removal condition, especially at the longest interval. A key finding was that performance in the temporary removal condition, but not in the permanent removal condition, was modulated by the presentation order of the target's memory set. This order effect was also observed in a non-removal control condition where double retro-cues marked all presented information as relevant. We suggest that order effects depend on maintaining the integrity of the retrieval structure (all the contextual cues) needed to guide attentional access to specific representations, including those provisionally set aside in the temporary removal condition. We conclude that the primary distinction between permanent and temporary removal processes is that only permanent removal simplifies the retrieval structure by eliminating unnecessary contextual cues, resulting in a greater reduction in the complexity of the search set compared to temporary removal.

Keywords: Verbal working memory, retro-cue paradigm, permanent removal, temporary removal, time-course, binding, working memory updating.

Introduction

Working memory (WM) is the cognitive system that enables a limited set of mental representations to be readily available for ongoing cognitive tasks. Traditional models posited the existence of specialized stores (e.g., the phonological store) dedicated to holding WM representations (Atkinson & Shiffrin, 1968; Baddeley & Hitch, 1974). The prevailing view today, however, is that WM and declarative long-term memory (LTM) share a common representational substrate, with WM representations being in a transitory state that allows them to be more immediately available than LTM representations (Cowan, 1995; Oberauer, 2002, 2009; Ruchkin et al., 2003). This second view usually comes with the additional assumption that, within WM, representations may differ in the degree to which they can be accessed and manipulated. The most crucial distinction is between representations currently attended (or currently included in an attentional refreshing loop) and those that have been temporarily left aside but remain in a state of privileged accessibility (Mallett & Lewis-Peacock, 2018; Oberauer, 2002; Oberauer & Awh, 2022; Oberauer & Hein, 2012; Olivers et al., 2011). These two levels of accessibility of WM representations may correspond to two different underlying neural mechanisms, persistent neural firing and some form of short-term synaptic plasticity (Masse et al., 2020; Miller et al., 2018; but see Stokes et al., 2020, for alternatives). Congruently with this view, some studies trying to decode WM representations from fMRI and electrophysiological neural signals have found that only attended representations can be effectively decoded (at least, with the current technical means), while non-attended WM contents can transition from an undecodable state to a decodable one as it shifts from irrelevant to relevant for the ongoing mental operation (LaRocque et al., 2013, 2017; Lewis-Peacock et al., 2012; Rose et al., 2016; Sprague et al., 2016; Wolff et al., 2017; but see Christophel et al., 2017, 2018). These findings reinforce the pertinency of distinguishing between attended (neurally active) and non-attended (dormant but still privileged) WM representations.

In real life, there are many situations in which a subset of WM representations must be temporarily set aside to minimize interference with the current mental operation, but in a way that allows these contents to be maintained in WM for later attentional refocusing. Consider, for example, situations such as mentally solving a reasoning problem, comprehending an intricate sentence we have just heard, or mentally performing an arithmetic operation. Typically, these tasks are segmented into steps that require focusing on a different portion of the information held in WM. Therefore, as one progresses through these steps, different subsets of representations are dynamically transferred in and out of the attentional focus. We will use the term "temporary removal" to refer to the operation of eliminating from the attentional focus WM representations that are currently unnecessary, but that will need to be refocused in a forthcoming step of the current task (Günseli et al., 2015; Lewis-Peacock et al., 2018). We assumed that these temporarily unattended representations remain within WM in a dormant but privileged state of easy accessibility.

In contrast to temporary removal, there are also many instances during the execution of WM tasks where a portion of the attended information ceases to be relevant for good. In a mental calculation task, for example, once a partial operation is completed and its result is maintained in WM, information about that partial operation becomes unnecessary and can be permanently removed from WM. Due to the limited capacity of WM (Cowan, 2010), having a mechanism dedicated to eliminating contents that have become definitively irrelevant seems crucial to maintaining the system's proper functioning (Lewis-Peacock et al., 2018). Some evidence for such a beneficial effect of permanent removal comes from studies employing the retro-cue paradigm (Griffin & Nobre, 2003) with 100% valid cues. The procedure in these studies involved the supply, after the initial encoding, of a cue pointing out a set of the presented items as the only relevant for the imminent memory test. The assumption is that, following cue presentation, attention focuses on the relevant information while a removal mechanism operates on the uncued WM representations. With this procedure, it has been observed that the performance cost associated with the irrelevant set progressively diminished during the first seconds immediately following the cue, which has been interpreted as reflecting the time course of the removal of the uncued contents from WM (Souza, Rerko, & Oberauer, 2014; Oberauer, 2001, 2002, 2018). Since irrelevant sets in these situations are marked as definitively unnecessary, we term the type of removal involved as “permanent removal” (Lewis-Peacock et al., 2018).

An intriguing question regarding temporary and permanent removal revolves around the similarities and differences between these two mechanisms. Both mechanisms appear to involve removing the information deemed unnecessary for the ongoing mental operation from the attentional focus while attention focuses on the relevant content. Differences, therefore, might concentrate on operations related to the accessibility level of the unattended information and how easily representations left aside can be refocused by attention when necessary. In this regard, it has been proposed that permanent removal operates by gradually disrupting the binding of items and their context through Hebbian antilearning (Lewis-Peacock et al., 2018; Oberauer, 2018). Unattended representations, therefore, are not obliterated but merely disconnected from their associated context. This would explain why irrelevant items, including those belonging to previous trials, generate strong interference when presented as negative probes in Sternberg-like tasks, even after long intervals since that information was flagged as irrelevant (Berman, Jonides, & Lewis, 2009; Campoy, 2011, 2012; Oberauer, 2001, 2018). Although the irrelevant representations were not obliterated during permanent removal, the fact that they were unbinding from the current context simplifies search processes and facilitates the encoding of new items by reducing cue overload (Dames & Oberauer, 2022; Festini & Reuter-Lorenz, 2014; Souza et al., 2014). From this perspective, the main difference between permanent and temporary removal would be that context-content bindings should remain intact after temporary removal to enable the subsequent retrieval of this information for the attentional focus (Koch et al., 2013; Lewis-Peacock et al., 2018). Considering that the enhancement in performance associated with permanent removal partially arises from the unbinding of items from their context, a process absent in temporary removal, it logically follows that temporary removal would likely result in a diminished performance improvement compared to permanent removal. This represents a simple prediction that aligns seamlessly with intuition. However, as far as we know, it hasn't been experimentally tested. The main purpose of this study is to address this gap.

We rely on the retro-cue paradigm with 100% valid cues outlined above. The condition designed to evaluate permanent removal mimics previous studies. Two sets of items were sequentially presented, followed by a cue marking one of the sets as the one relevant for the upcoming memory test, which was applied after a variable brief interval (the cue-target interval, CTI). Based on previous findings, performance was anticipated to improve as the CTI increased, reflecting the time course of uncued-set removal (Oberauer, 2018). The temporary removal condition was created by adding, immediately following the response, a second memory test on the other set, the one initially marked as irrelevant. The existence of this second test forced participants to maintain the irrelevant set in an accessible state during the first interval. Introducing a second response regarding previously irrelevant information is not new (LaRocque et al., 2013; Oberauer, 2005). What is novel is incorporating both single-response and double-response trials within the same experiment in a fully predictable manner (allowing participants to apply the most suitable removal mechanism in each case) and keeping all other experimental parameters equal to enable direct comparison. Since there were different CTIs, we could compare not only the global effect of permanent and temporary removal but also their temporal course.

In addition to these key conditions, we introduced two single-response control conditions to provide baseline comparisons. One had double retro-cues, so the two memory sets were marked as relevant. This control condition was needed to verify that removing the irrelevant set, either permanently or temporarily, enhances performance. In the other control condition, only one memory set was presented at encoding, followed by a cue redundantly marking that set as relevant (to keep the trial structure consistent). We anticipated observing a progressive convergence of performance in conditions with removal towards the performance in this second control condition as the amount of WM representations in the attentional focus became equated across CTIs.

Besides comparing the beneficial effect of permanent and temporary removal and their temporal courses, we aimed to investigate two additional issues. The first of these was not initially considered but gained prominence through the results of two pilot studies conducted before the one described here. In these pilots, we found that performance in the temporary removal condition, but not in the permanent removal condition, significantly depended on whether the cued set was the one presented first or second during initial encoding. Hence, we decided to design the present experiment considering this factor, which ultimately proved crucial for interpreting the results (as we elaborate in the Discussion section).

The second additional issue we aimed to address pertains to the second response in the temporary removal condition (the only one with two tests per trial). Specifically, we were interested in whether the length of the CTI preceding the first response would influence performance on the second test. If performance on the first test improves with longer CTIs due to participants having more time to complete the removal of the uncued set, then a prediction arises that the positive effect of CTI length on the first response may be accompanied by a negative mirror effect on the second response, when the removed set must be attentionally retrieved. We also tested this prediction.

Method

Participants

Fifty-six undergraduate students (43 females and 13 males; M age = 20.57 years; SD age = 2.31) from the University of Murcia participated for course credit. The number of participants was determined after conducting a simulated power analysis using the Superpower R package (Lakens & Caldwell, 2021), which yielded an estimated sample size of 52 to achieve a power of 0.80 for the three-way interaction of our 4 × 4 × 2 within-participants design. The final number was 56 because we needed a multiple of eight (see Procedure). Three participants initially included in the sample were replaced due to poor performance (see Results).

All participants reported normal-to-corrected vision. This study was approved by the University of Murcia Ethics Committee and was conducted according to the ethical standards of the 1964 Declaration of Helsinki.

Materials

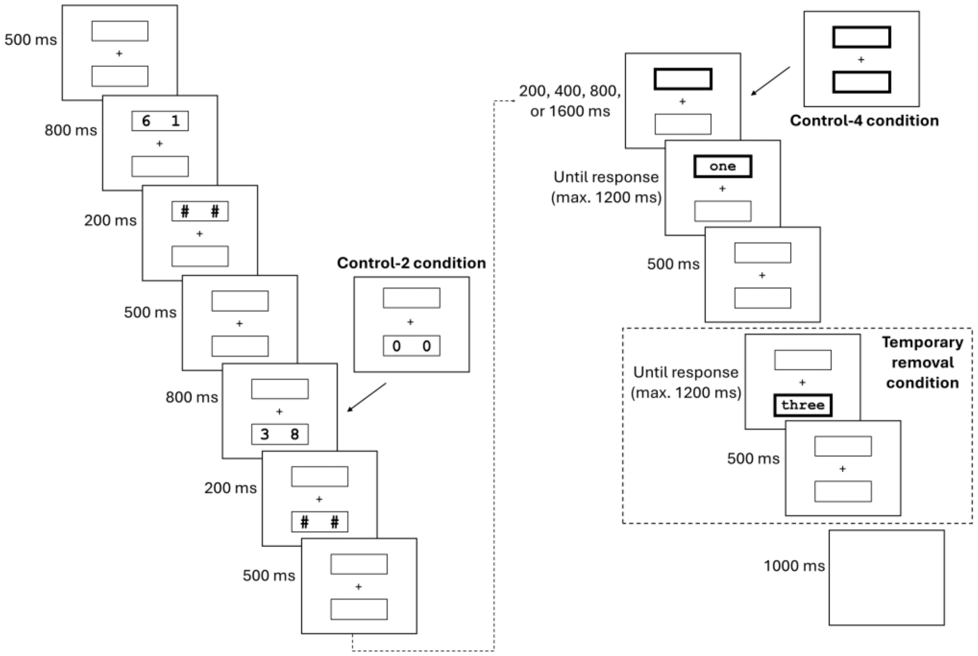
The experiment was controlled by a computer program written with E-Prime 3.0 (Schneider, Eschman, & Zuccolotto, 2002). Stimuli were presented on a 22” TFT monitor while responses were collected via Chronos devices. Participants were tested individually in sound-attenuated booths.

Procedure

The participants completed 192 experimental trials divided into four blocks of 48 trials. These blocks corresponded to the four experimental conditions:

permanent removal, temporary removal, control-4, and control-2. Preceding each block, participants received instructions and completed 16 practice trials.

Figure 1. Schematic representation of the experimental procedure.



Trials in the permanent-removal condition began with two empty rectangular frames with a narrow black border displayed against a white background (Figure 1). These frames appeared symmetrically above and below a centrally located fixation point. After 500 ms, two digits were simultaneously presented within one of the frames (the presentation frame 1) for 800 ms, followed by a 200-ms mask (two # signs). After an interval of 500 ms, two different digits were presented in the other frame (the presentation frame 2) following the same procedure. The four digits in each trial were randomly selected from 1 to 9 with the following restrictions: (a) All four digits had to be different; (b) no two consecutive numbers were allowed within a frame; (c) no digit could appear in the same position (left or right) as in the previous trial, regardless of whether it appeared in the same frame or not; and (d) the two digits in a frame were not allowed to be the same as the two numbers presented in a frame from the previous trial, regardless of whether it was the same frame or not. After 500 ms, the border of one of the presentation frames thickened (by a factor of 10), cueing participants that only the items that appeared in that frame were relevant for the upcoming memory test. Finally, after a CTI of 200, 400, 800, or 1600 ms, one of the two numbers that had appeared in the cued frame was displayed within the same frame in its alphabetical form. Participants were instructed to indicate, as fast and accurately as possible, the location (left or right) where that target number had previously appeared by pressing the left or right bottom of the Chronos device, respectively. Targets remained visible until response or until a maximum of 1200 ms elapsed. As feedback, the screen briefly flashed red (50 ms) following incorrect responses or failures to respond within the 1200-ms time window. The next trial commenced after a 1500-ms interval (subsuming the 50-ms feedback if provided).

Trials in the other three conditions differed only in the following aspects. In the temporary-removal condition, 500 ms after the first response, the border of the uncued frame thickened, and one of the numbers previously presented within that frame appeared in alphabetic form. Participants then provided a second response following the same procedure as in the initial one. In the control-4 condition, the two presentation frames were cued, marking the four digits as relevant for the upcoming test. In the control-2 condition, digits from the frame assigned to be the uncued one were replaced by zeros at the initial presentation, so only two digits had to be encoded and maintained.

The 48 trials of each block were designed to ensure equal representation of all the combinations of target presentation frame (1 or 2), target location (left or right), and CTI (200, 400, 800, or 1600 ms). At least one participant was randomly assigned to each of the 4! × 2 = 48 arrangements resulting from combining the administration order of the four blocks and the location of the presentation frame 1 (above or below the fixation point; note that the set appearing in the presentation frame 1 is the one presented first at initial encoding). Across participants and frame-1 locations, each of the four block types was administered the same number of times in each position.

Results

Data from three participants were excluded from the reported results because their proportion of trials with correct responses (0.58, 0.62, and 0.64) deviated from the mean by more than three standard deviations (the next lowest proportion was 0.83). In applying this same criterion to the second response of the temporary removal condition, no additional participant’s data were excluded.

The primary statistical tests were repeated measures analyses of variance (ANOVA). When Mauchly’s sphericity test reached statistical significance, the Greenhouse-Geisser correction was applied. Post-hoc analyses were corrected for multiple comparisons using the Holm-Bonferroni method. A significance level of 0.05 was adopted for all analyses. We began analyzing reaction times (RTs) in the four conditions (in the temporary removal condition, only RTs from the first response). Trials with no response (1.03 %), with RTs shorter than 250 ms (0.11 %), or with incorrect responses (5.11 %) were

Table 1. Post hoc comparisons between conditions across CTIs.

Comparison	Interval	M difference	*Cohen's d*	*t*	*p*
PR - TR	200 ms	-23 ms	-0.25	-2.430	.064
	400 ms	-12 ms	-0.14	-1.286	.597
	800 ms	-27 ms	-0.29	-2.771	.030
	1600 ms	-43 ms	-0.46	-4.415	<.001
	Total	-26 ms	-0.29	-3.624	<.001
PR – C2	200 ms	76 ms	0.82	7.872	<.001
	400 ms	34 ms	0.37	3.574	<.001
	800 ms	1 ms	0.02	0.143	.886
	1600 ms	2 ms	0.03	0.249	1.000
	Total	29 ms	0.31	3.935	<.001
PR – C4	200 ms	-51 ms	-0.56	-5.310	<.001
	400 ms	-99 ms	-1.08	-10.273	<.001
	800 ms	-136 ms	-1.48	-14.139	<.001
	1600 ms	-163 ms	-1.76	-16.857	<.001
	Total	-112 ms	-1.22	-15.483	<.001
TR – C2	200 ms	99 ms	1.08	10.302	<.001
	400 ms	47 ms	0.51	4.861	<.001
	800 ms	28 ms	0.31	2.913	.028
	1600 ms	45 ms	0.49	4.663	<.001
	Total	55 ms	0.60	7.559	<.001
TR – C4	200 ms	-28 ms	-0.30	-2.880	.024
	400 ms	-87 ms	-0.94	-8.987	<.001
	800 ms	-110 ms	-1.19	-11.369	<.001
	1600 ms	-120 ms	-1.30	-12.442	<.001
	Total	-86 ms	-0,93	-11.859	<.001
C2 – C4	200 ms	-127 ms	-1.38	-13.182	<.001
	400 ms	-134 ms	-1.45	-13.847	<.001
	800 ms	-138 ms	-1.50	-14.282	<.001
	1600 ms	-165 ms	-1.79	-17.106	<.001
	Total	-141 ms	-1.53	-19.418	<.001

Note: *p* values were corrected for multiple comparisons (30) using the Holm-Bonferroni method.
PR: permanent removal; TR: temporary removal; C2: control 2; C4: control 4.

excluded from the analysis. A 4 (condition) × 4 (CTI) × 2 (target frame) repeated measures ANOVA revealed an effect of condition, $F(3, 165) = 140.523$, $p < 0.001$, $\eta_p^2 = 0.719$, with post hoc comparisons (Table 1) indicating differences among the four conditions (control-4 > temporary removal > permanent removal > control-2).

There was also a condition × CTI interaction, $F(9, 495) = 20.604$, $p < 0.001$, $\eta_p^2 = 0.273$. While independent ANOVAs for each condition showed statistically significant effects of CTI in all the conditions (Table 2), inspection of Figure 2 and effect sizes in Table 2 suggests that the interaction resulted from a stronger effect of CTI on the removal conditions compared to the control conditions. Congruently, the four possible independent ANOVAs involving one removal condition and one control condition showed a condition × CTI interaction, all $F\text{s}(3, 165) \geq 18.994$, $p\text{s} < 0.001$. Importantly, the condition × CTI interaction reached statistical significance in an independent ANOVA that included only the two removal conditions, $F(3, 165) = 2.938$, $p = 0.035$, $\eta_p^2 = 0.051$. This interaction emerged because the difference between the two conditions became more pronounced at the longest CTI (Table 1). Finally, the global ANOVA revealed a condition × target frame interaction, $F(2.6, 145.0) = 11.546$, $p < 0.001$, $\eta_p^2 = 0.164$. As illustrated in Figure 3 and Table 3, and confirmed by individual ANOVAs for each condition (Table 2), this interaction resulted from the frame modulating RTs only on the temporary removal and control-4 conditions, with faster responses when the target appeared in frame 1 (i.e., when the target belonged to the set presented first) than when presented in frame 2 (i.e., when the target belonged to the set presented second). The advantage of frame 1 on the temporary removal and control 4 conditions faded in the longer CTI (Figure 3).

Consequently, the triple interaction in the global ANOVA was close to reaching the statistical significance level, $F(9, 495) = 1.714$, $p = 0.083$, $\eta_p^2 = 0.030$.

Table 2. Independent ANOVA results for RT in each condition.

Condition	Effect	F	df	p	η_p^2
PR	CTI	148.946	3, 165	< 0.001	0.730
	Frame	0.020	1, 55	0.888	0.000
	CTI × Frame	0.214	3, 165	0.886	0.004
TR	CTI	110.796	3, 165	< 0.001	0.668
	Frame	47.757	1, 55	< 0.001	0.465
	CTI × Frame	6.569	2.5, 138.1	< 0.001	0.107
C2	CTI	42.173	3, 165	< 0.001	0.434
	Frame	1.483	1, 55	0.228	0.026
	CTI × Frame	0.340	2.7, 147.7	0.774	0.006
C4	CTI	8.371	2.6, 147.3	< 0.001	0.132
	Frame	14.889	1, 55	< 0.001	0.213
	CTI × Frame	2.311	3, 165	0.078	0.040

Note: PR: permanent removal; TR: temporary removal; C2: control 2; C4: control 4.

Figure 2. Mean RT across conditions and CTIs. Error bars represent the standard error of the mean.

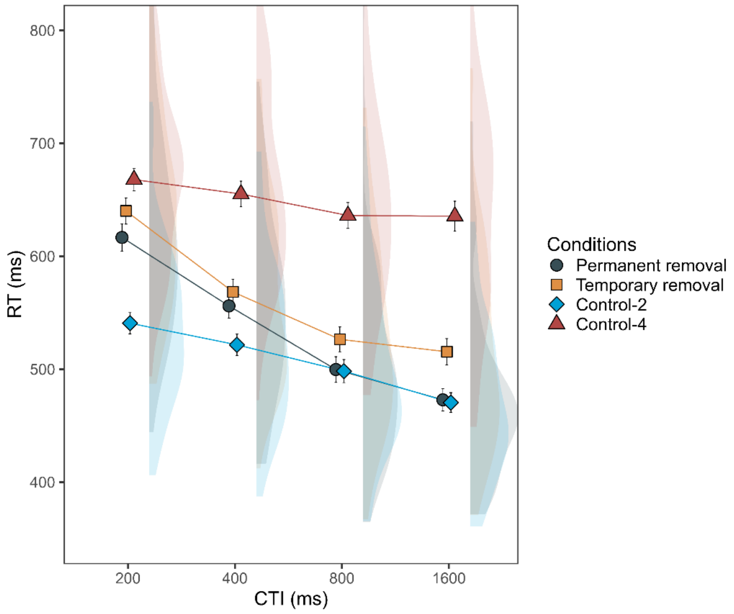


Figure 3. Mean RTs across presentation frames and conditions. Error bars represent the standard error of the mean.

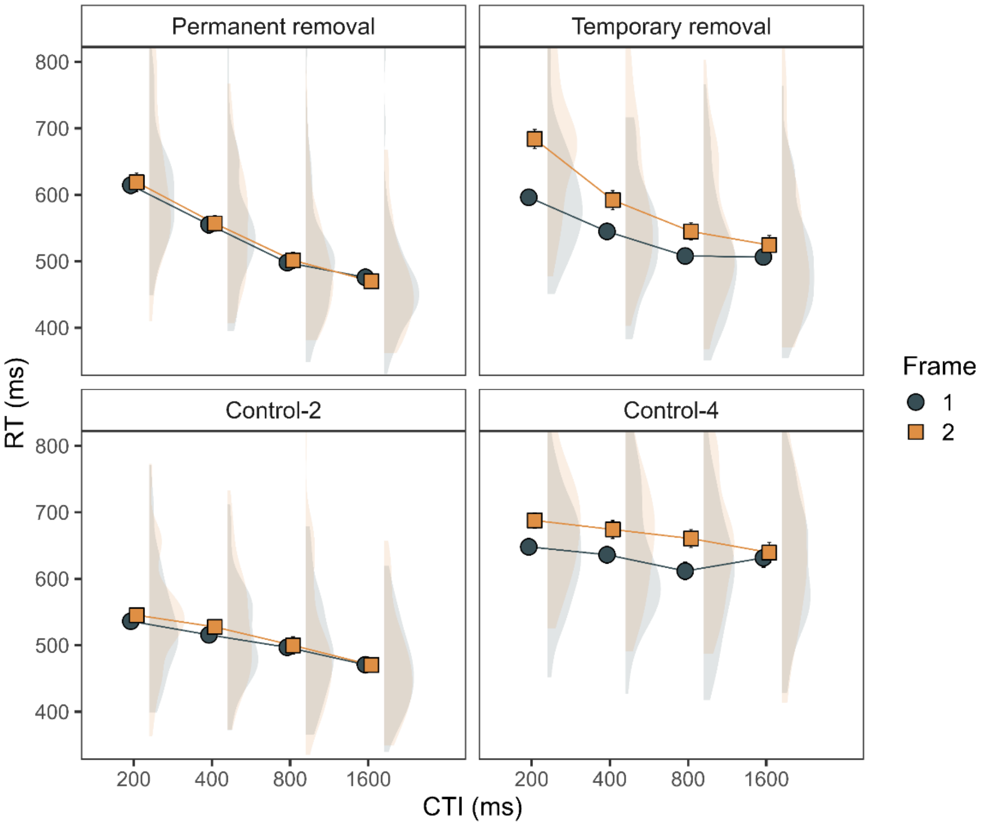


Table 3. Mean RT (in ms) and proportion of correct responses (accuracy) as a function of condition and presentation frame.

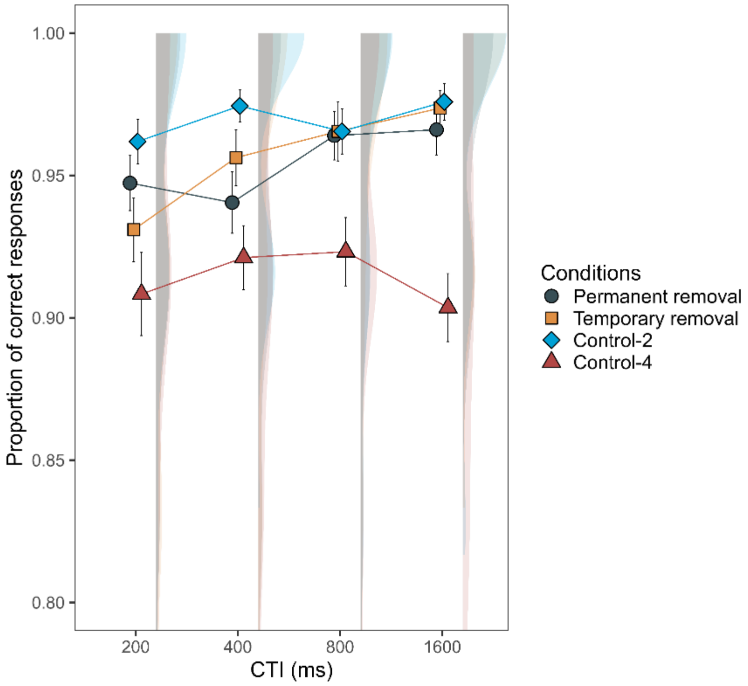
		Frame 1		Frame 2		Difference	
Measure	Condition	M	SD	M	SD	M	SD
RT	PR	534	79	536	76	3	47
	TR	534	73	582	88	48	52
	C2	498	61	504	71	6	35
	C4	626	84	664	87	38	66
Accuracy	PR	0.95	0.06	0.96	0.04	-0.01	0.06
	TR	0.97	0.05	0.95	0.07	0.02	0.06
	C2	0.97	0.04	0.97	0.04	0.00	0.06
	C4	0.92	0.07	0.91	0.08	0.02	0.08

Note: PR: permanent removal; TR: temporary removal; C2: control 2; C4: control 4.

An equivalent ANOVA on the proportion of correct responses (Figure 4) revealed statistically significant main effects of CTI, $F(3, 165) = 3.850$, $p = 0.011$, $\eta_p^2 = 0.065$, and condition, $F(2.6, 142.0) = 20.350$, $p < 0.001$, $\eta_p^2 = 0.270$. Post hoc tests revealed poorer performance in the control-4 condition than in the other three conditions, all ts ≥ 5.375, ps < 0.001. There was a tendency for better performance in the control-2 condition than in the two removal conditions, but differences did not reach the statistical significance level after correcting for multiple comparisons, both ts ≥ 1.708, $p \geq 0.145$. Importantly, the proportion of correct responses in the two conditions with removal was equivalent, $t < 1$, and not even a numerical advantage was found in favor of the condition with permanent removal over the temporary removal condition (Table 3). There was also a condition × frame interaction, $F(3, 165) = 3.134$, $p = 0.027$, $\eta_p^2 = 0.054$. Congruently with RT results, there was a tendency for better performance when the target appeared in frame 1 in the temporary removal and the control 4 conditions (Table 3), although, in this case, the post hoc test did not reach the statistical significance level after correcting for multiple comparisons, both ts ≥ 2.011, ps ≥ 0.136. No other main effect or interaction reached statistical significance.

The last set of analyses focused on the second response in the temporary removal condition (Figure 5). For the RT analysis, trials with no response (1.23 %), with RTs shorter than 250 ms (0.04 %), or with incorrect responses (6.81 %) were excluded. A 4 (CTI) × 2 (target frame) repeated measures ANOVA showed a main effect of CTI, $F(3, 165) = 12.466$, $p < 0.001$, $\eta_p^2 = 0.185$. Post hoc tests revealed that the effect of CTI was the consequence of slower responses when the CTI preceding the first response was the longest, all ts ≥ 3.708, ps < 0.001. There was also a main effect of frame, $F(1, 55) = 19.497$, $p < 0.001$, $\eta_p^2 = 0.262$, indicating that RTs were faster when targets appeared in frame 2 (i.e., when the cued set for the first response was the one presented first at initial encoding and, thus, the current target belonged to the set presented second). The CTI × frame interaction did not reach the statistical significance level, $F < 1$.

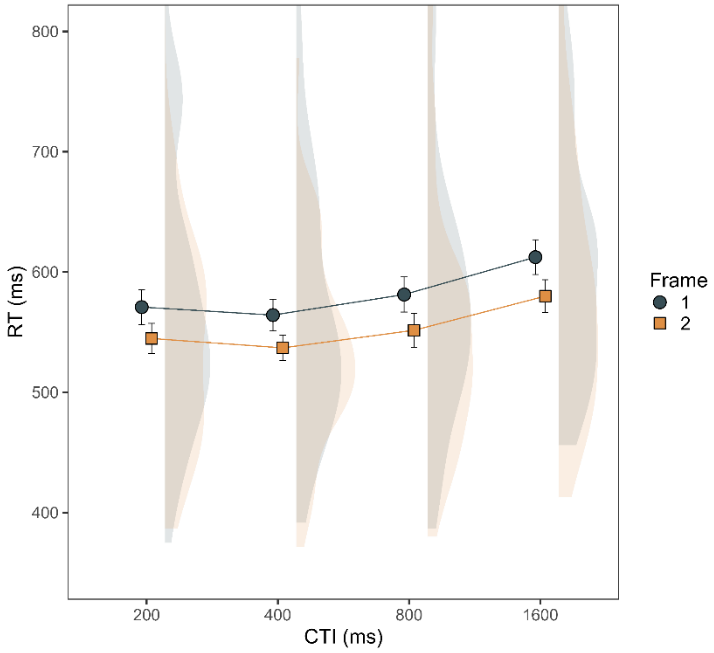
Figure 4. Proportion of correct responses across conditions and CTIs. Error bars represent the standard error of the mean.



The last set of analyses focused on the second response in the temporary removal condition (Figure 5). For the RT analysis, trials with no response (1.23 %), with RTs shorter than 250 ms (0.04 %), or with incorrect responses (6.81 %) were excluded. A 4 (CTI) × 2 (target frame) repeated measures ANOVA showed a main effect of CTI, $F(3, 165) = 12.466$, $p < 0.001$, $\eta_p^2 = 0.185$. Post hoc tests revealed that the effect of CTI was the consequence of slower responses when the CTI preceding the first response was the longest, all $ts \geq 3.708$, $ps < 0.001$. There was also a main effect of frame, *F*(1, 55) = 19.497, $p < 0.001$, $\eta_p^2 = 0.262$, indicating that RTs were faster when targets appeared in frame 2 (i.e., when the cued set for the first response was the one presented first at initial encoding and, thus, the current target belonged to the set presented second). The CTI × frame interaction did not reach the statistical significance level, $F < 1$.

An equivalent ANOVA on the proportion of incorrect responses revealed statistically significant effects of CTI, *F*(1, 165) = 3.825, $p = 0.011$, $\eta_p^2 = 0.065$, and frame, *F*(1, 55) = 7.135, $p = 0.010$, $\eta_p^2 = 0.115$. Congruently with the RT analysis just described, this main effect resulted from lower error rates for targets presented in frame 2 and poorer accuracy when the CTI preceding the previous response was the longest.

Figure 5. Mean RTs on the second response of the temporary removal condition. Error bars represent the standard error of the mean.



Discussion

We employed the retro-cue paradigm to investigate the following main question: Does permanently discarding a subset of WM contents marked as irrelevant offer additional benefit for WM performance compared to just temporarily putting them aside? Answering this apparently straightforward question requires careful examination of nuanced patterns. The first relevant observation is that performance in the permanent and temporary removal conditions did not differ in terms of accuracy. This fact suggests an equivalent quality of WM representations for the relevant set, including items, contextual cues, and the binding between them, as well as an equal precision in the process of searching for the required information. In contrast to accuracy, differences between the two removal conditions did appear in RT terms, with slower responses in the temporary removal condition. To interpret this, we must consider two additional observations: the evolution of RTs across cue-target intervals (CTIs) and the modulatory effect of the order of presentation of the cued set. We will explore these issues in turn.

Consistently with previous studies (Oberauer, 2018; Tortajada et al., in press), RTs in the permanent removal condition decreased across CTIs, progressively diverging from the control-4 condition, in which the irrelevant set could not be discarded in advance, and ultimately converging with the control-2 condition, in which only the relevant set was presented at initial encoding. This pattern suggests that participants used the CTI to focus attention on the relevant set and remove the irrelevant contents, progressively reducing the complexity of the search set and, consequently, the time required to locate the required information. Importantly, although responses were generally slower in the temporary removal condition compared to the permanent removal condition, the progression of both conditions was similar during the first three intervals, with a slight increase in the advantage of permanent removal at the longer interval. This observation seems to rule out some potential explanations for the longer RTs in the temporary removal condition. For example, it could be argued that participants in the temporary removal condition delayed the onset of the removal operation to strengthen the representation of the irrelevant set before removing it, thereby facilitating its later retrieval. However, this interpretation would predict a reduction of the difference between the two removal conditions at the longer intervals once RT in the permanent removal condition reached the minimum level informed by the control-2 condition, which is the opposite pattern to what was found.

Regarding the effect of the order of presentation of the cued set, we found that, in the conditions in which the two memory sets should be maintained in an accessible state (i.e., in temporary removal and control-4 conditions), performance was better when the target belonged to the set presented first; in contrast, when only one set should be maintained available (i.e., in the permanent removal and control-2 conditions), performance was not modulated by the order of presentation of the relevant information. A related observation is that the permanent removal condition showed better performance than the temporary removal condition only when the tested set was the one presented second; when the target belonged to the set presented first, the two removal conditions yielded equivalent RTs. In our opinion, this modulatory effect of the order of presentation is key to understanding the underlying differences between the temporary and permanent removal conditions.

We will articulate our interpretation as follows. (1) RTs in the present task were mainly determined by the complexity of the search set. We posit that the complexity of the search set is directly related to both the size of the retrieval structure (the number of contextual cues) that participants maintain to guide attentional access to a specific representation or representation set and the number of active representations bound to this retrieval structure. (2) In the two removal conditions, the complexity of the search set diminished progressively during the CTI, whereas it remained constant in the two control conditions. That is why RTs decreased more markedly across intervals in the removal conditions. (3) Simplifying the search set in the permanent removal condition included reducing the retrieval structure. We could conceive, for instance, that item representations are bound to two contextual cues representing the two memory sets (or the two presentation frames), which, in turn, are connected to a global trial-level context. The permanent removal process would entail unbinding the set-level contextual cue from the global context, eventually equalizing the complexity of the search set to that in the control-2 condition, in which only one set-level contextual cue was established at initial encoding. (4) Simplification of the search set in the temporary removal condition was achieved by withdrawing attention from the irrelevant representations and, consequently, reducing the number of active representations bound to the retrieval structure. However, no reduction in the size of the retrieval structure is possible because the initially unneeded contextual cues are required later to retrieve the uncued set for the second test. (5) Whenever the whole retrieval structure was maintained (i.e., in the temporary removal and control-4 conditions), an order effect appeared because of the well-known tendency in verbal WM to scan contents sequentially (Kessler & Oberauer, 2015). In summary, therefore, our proposal posits that, compared to temporary removal, permanent removal enables superior WM performance by simplifying the context utilized for accessing item representations via context-content associations.

The observation that, when contents are accessed through the whole retrieval structure, there is an advantage for the information presented first is congruent with the common assumption that verbal information in WM is encoded, maintained, and scanned in a serial-ordered way (Kessler & Oberauer, 2015; Majerus, 2019). Early evidence revealed that when participants must respond to whether a probe stimulus belonged to a list presented right before, RTs increase with the list length, suggesting that representations are scanned sequentially (Sternberg, 1966). From a neurophysiology point of view, it has been proposed that the serial organization of representations in WM responds to a neural substrate in which individual items are presented in gamma cycles sequentially coupled in theta cycles (Lisman & Jensen, 2013). In this regard, Ideriha and Ushiyama (2024) found that recall of sequential information from WM follows a theta rhythm. Bahramisharif et al. (2018) used intracranial recording to show that item-specific gamma activation was coupled to theta in a position-dependent manner. Moreover, the benefit of testing in the same order in which the information was encoded is also well-known. For instance, recall is more accurate when it is done in the same order as encoding (Oberauer, 2003), and recognition is faster when items are tested in the forward order, as compared to random or backward order (Lange et al., 2010, 2011). There is also a trend to free recall in the same order as information was encoded even when it is not explicitly requested (Klein et al., 2005). These findings have been accounted for by models that include a chaining effect (Ebbinghaus, 1885; Logan & Cox, 2021). These models propose that representations are encoded in WM tightly linked to a context, being the previously encoded items part of this context. This causes the previous item to serve as a cue for retrieving the present item, but not vice versa (Kahana & Caplan, 2002; Nairne et al., 2007). This latter suggestion is pivotal for interpreting the pattern of results observed in the second response of the temporary removal condition, where participants were required to respond to the set initially marked as irrelevant (see below). An observation for which we have no explanation is that in the two conditions with an effect associated with the presentation frame (temporary removal and control-4 conditions), this effect disappears at the longest interval. Despite the lack of an explanation, the fact that the same pattern emerged in both conditions is consistent with our proposal that the temporary removal and control-4 conditions share the necessity to maintain the retrieval structure intact, leading to the emergence of an equivalent pattern associated with the presentation frame in both conditions.

The last aspect to consider is our prediction regarding the second response in the temporary removal condition. This prediction was that longer intervals preceding the first response would result in slower responses in the second test. The logic behind this prediction is that longer intervals entail deeper elimination of the irrelevant set prior to response one, consequently leading to a worse starting point for the subsequent reactivation of that set. The results confirmed this prediction but also revealed an order of presentation effect characterized by shorter

RTs when the evaluated set was the one presented second during the initial encoding. This observation is consistent with the notion that the set presented first acted as a retrieval cue for the set presented second but not the other way around, as posited by chain models. Therefore, the complete pattern in the temporary removal condition is clear: optimal performance was achieved when the two memory sets were evaluated in their presentation order.

To summarize, our study showed that both permanent and temporary removal led to a marked progressive improvement in performance in terms of both RTs and accuracy. This improvement presumably results from the gradual reduction of the search set's complexity following the presentation of informative retro-cues. In temporary removal situations, irrelevant representations are deactivated, but the entire retrieval structure is preserved because contextual cues bound to the irrelevant set will be necessary later to guide attentional refocusing. In the permanent removal condition, however, the retrieval structure can be simplified by eliminating or unbinding unnecessary contextual cues. We suggest that this simplification in the retrieval structure with permanent removal accounts for the absence of order effects and leads to shorter RTs, representing the main difference between permanent and temporary removal processes.

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes. In K. W. Spence & J. T. Spence (Eds.), *Psychology of Learning and Motivation* (Vol. 2, pp. 89-195). Academic Press. https://doi.org/10.1016/S0079-7421(08)60422-3.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In *Psychology of Learning and Motivatio*n (Vol. 8, pp. 47-89). Elsevier. https://doi.org/10.1016/S0079-7421(08)60452-1

Bahramisharif, A., Jensen, O., Jacobs, J., & Lisman, J. (2018). Serial representation of items during working memory maintenance at letter-selective cortical sites. *PLOS Biology, 16*(8), e2003805. https://doi.org/10.1371/journal.pbio.2003805

Berman, M. G., Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 317–333. https://doi.org/10.1037/a0014873.

Campoy, G. (2011). Retroactive interference in short-term memory and the word-length effect. *Acta Psychologica, 138*, 135-142. https://doi.org/10.1016/j.actpsy.2011.05.016.

Campoy, G. (2012). Evidence for decay in verbal short-term memory: A commentary on Berman, Jonides, and Lewis (2009). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 1129-1136. https://doi.org/10.1037/a0026934.

Christophel, T. B., Iamshchinina, P., Yan, C., Allefeld, C., & Haynes, J.-D. (2018). Cortical specialization for attended versus unattended working memory. *Nature Neuroscience, 21*(4), Article 4. https://doi.org/10.1038/s41593-018-0094-4

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J.-D. (2017). The Distributed Nature of Working Memory. *Trends in Cognitive Sciences, 21*(2), 111-124. https://doi.org/10.1016/j.tics.2016.12.007

Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford University Press.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science, 51-57*. https://doi.org/10.1177/0963721409359277.

Dames, H., & Oberauer, K. (2022). Directed forgetting in working memory. *Journal of Experimental Psychology. General, 151*(12), 2990-3008. https://doi.org/10.1037/xge0001256

Ebbinghaus, H. (1885). Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie. Leipzig: Duncker & Humblot.

Festini, S. B., & Reuter-Lorenz, P. A. (2014). Cognitive control of familiarity: Directed forgetting reduces proactive interference in working memory. *Cognitive, Affective, & Behavioral Neuroscience, 14*(1), 78-89. https://doi.org/10.3758/s13415-013-0231-1

Griffin, I. C., & Nobre, A. C. (2003). Orienting Attention to Locations in Internal Representations. *Journal of Cognitive Neuroscience, 15*(8), 1176-1194. https://doi.org/10.1162/089892903322598139

Günseli, E., van Moorselaar, D., Meeter, M., & Olivers, C. N. L. (2015). The reliability of retro-cues determines the fate of noncued visual working memory representations. *Psychonomic Bulletin & Review, 22*(5), 1334-1341. https://doi.org/10.3758/s13423-014-0796-x

Ideriha, T., & Ushiyama, J. (2024). Behavioral fluctuation reflecting theta-rhythmic activation of sequential working memory. *Scientific Reports, 14*, 1. https://doi.org/10.1038/s41598-023-51128-7.

Kahana, M. J., & Caplan, J. B. (2002). Associative asymmetry in probed recall of serial lists. *Memory & Cognition, 30*(6), 841-849. https://doi.org/10.3758/bf03195770

Kessler, Y., & Oberauer, K. (2015). Forward scanning in verbal working memory updating. *Psychonomic Bulletin & Review, 22*, 1770-1776.

Klein, K. A., Addis, K. M., & Kahana, M. J. (2005). A comparative analysis of serial and free recall. *Memory & Cognition, 33*(5), 833-839. https://doi.org/10.3758/BF03193078.

Koch, G., Ponzo, V., Lorenzo, F. D., Caltagirone, C., & Veniero, D. (2013). Hebbian and Anti-Hebbian Spike-Timing-Dependent Plasticity of Human Cortico-Cortical Connections. *Journal of Neuroscience, 33*(23), 9725-9733. https://doi.org/10.1523/JNEUROSCI.4988-12.2013.

Lakens, D., & Caldwell, A. R. (2021). Simulation-Based Power Analysis for Factorial Analysis of Variance Designs. Advances in Methods and Practices in *Psychological Science, 4*(1), 2515245920951503. https://doi.org/10.1177/2515245920951503

Lange, E. B., Cerella, J., & Verhaeghen, P. (2011). Ease of access to list items in short-term memory depends on the order of the recognition probes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(3), 608-620. https://doi.org/10.1037/a0022220

Lange, E. B., Verhaeghen, P., & Cerella, J. (2010). Dual representation of item positions in verbal short-term memory: Evidence for two access modes. *European Journal of Cognitive Psychology, 22*(3), 463-479. https://doi.org/10.1080/09541440903155658

LaRocque, J. J., Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2013). Decoding Attended Information in Short-term Memory: An EEG Study. *Journal of Cognitive Neuroscience, 25*(1), 127-142. https://doi.org/10.1162/jocn_a_00305

LaRocque, J. J., Riggall, A. C., Emrich, S. M., & Postle, B. R. (2017). Within-Category Decoding of Information in Different Attentional States in Short-Term Memory. *Cerebral Cortex, 27*(10), 4881-4890. https://doi.org/10.1093/cercor/bhw283

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural Evidence for a Distinction between Short-term Memory and the Focus of Attention. *Journal of Cognitive Neuroscience, 24*(1), 61-79. https://doi.org/10.1162/jocn_a_00140

Lewis-Peacock, J. A., Kessler, Y., & Oberauer, K. (2018). The removal of information from working memory. *Annals of the New York Academy of Sciences, 1424*(1), 33-44. https://doi.org/10.1111/nyas.13714

Lisman, J. E., & Jensen, O. (2013). The Theta-Gamma Neural Code. *Neuron, 77*(6), 1002-1016. https://doi.org/10.1016/j.neuron.2013.03.007

Logan, G. D., & Cox, G. E. (2021). Serial memory: Putting chains and position codes in context. *Psychological Review, 128*, 1197–1205. https://doi.org/10.1037/rev0000327.

Majerus, S. (2019). Verbal working memory and the phonological buffer: The question of serial order. *Cortex, 112*, 122-133. https://doi.org/10.1016/j.cortex.2018.04.016

Mallett, R., & Lewis-Peacock, J. A. (2018). Behavioral decoding of working memory items inside and outside the focus of attention. *Annals of the New York Academy of Sciences, 1424*(1), 256-267. https://doi.org/10.1111/nyas.13647

Masse, N. Y., Rosen, M. C., & Freedman, D. J. (2020). Reevaluating the Role of Persistent Neural Activity in Short-Term Memory. *Trends in Cognitive Sciences, 24*(3), 242-258. https://doi.org/10.1016/j.tics.2019.12.014

Miller, E. K., Lundqvist, M., & Bastos, A. M. (2018). Working Memory 2.0. *Neuron, 100*(2), 463-475. https://doi.org/10.1016/j.neuron.2018.09.023

Nairne, J. S., Ceo, D. A., & Reysen, M. B. (2007). The mnemonic effects of recall on immediate retention. *Memory & Cognition, 35*(1), 191-199. https://doi.org/10.3758/bf03195954

Oberauer, K. (2001). Removing irrelevant information from working memory: A cognitive aging study with the modified Sternberg task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(4), 948.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 411-421.

Oberauer, K. (2003). Understanding serial position curves in short-term recognition and recall. *Journal of Memory and Language, 49*(4), 469-483. https://doi.org/10.1016/S0749-596X(03)00080-9

Oberauer, K. (2005). Control of the contents of working memory—A comparison of two paradigms and two age groups. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(4), 714–728. https://doi.org/10.1037/0278-7393.31.4.714.

Oberauer, K. (2009). Design for a Working Memory. In P*sychology of Learning and Motivation* (Vol. 51, pp. 45-100). Academic Press. https://doi.org/10.1016/S0079-7421(09)51002-X.

Oberauer, K. (2018). Removal of irrelevant information from working memory: Sometimes fast, sometimes slow, and sometimes not at all. *Annals of the New York Academy of Sciences, 1424*(1), 239-255.

Oberauer, K., & Awh, E. (2022). Is There an Activity-silent Working Memory? *Journal of Cognitive Neuroscience, 34*(12), 2360-2374. https://doi.org/10.1162/jocn_a_01917

Oberauer, K., & Hein, L. (2012). Attention to Information in Working Memory. *Current Directions in Psychological Science, 21*(3), 164-169. https://doi.org/10.1177/0963721412444727

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences, 15*(7), 327-334. https://doi.org/10.1016/j.tics.2011.05.004

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science, 354*(6316), 1136-1139. https://doi.org/10.1126/science.aah7011

Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain Sciences, 26*(6), 709-728. https://doi.org/10.1017/S0140525X03000165

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime reference guide*. Psychology Software Tools, Inc.

Souza, A. S., Rerko, L., & Oberauer, K. (2014). Unloading and reloading working memory: Attending to one item frees capacity. *Journal of Experimental Psychology: Human Perception and Performance, 40*(3), 1237-1256. https://doi.org/10.1037/a0036331

Sprague, T. C., Ester, E. F., & Serences, J. T. (2016). Restoring Latent Visual Working Memory Representations in Human Cortex. *Neuron, 91*(3), 694-707. https://doi.org/10.1016/j.neuron.2016.07.006

Sternberg, S. (1966). High-Speed Scanning in Human Memory. *Science, 153*(3736), 652-654. https://doi.org/10.1126/science.153.3736.652.

Stokes, M. G. (2015). 'Activity-silent' working memory in prefrontal cortex: A dynamic coding framework. *Trends in Cognitive Sciences, 19*(7), 394-405. https://doi.org/10.1016/j.tics.2015.05.004

Stokes, M. G., Muhle-Karbe, P. S., & Myers, N. E. (2020). Theoretical distinction between functional states in working memory and their corresponding neural states. *Visual Cognition, 28*(5-8), 420-432. https://doi.org/10.1080/13506285.2020.1825141

Tortajada, M., Fahrenfort, J. J., Sandoval-Lentisco, A., Martínez-Pérez, V., Palmero, L. B., Melero, L. F., Campoy, G., & Olivers, C. N. L. (in press). Decoding Load or Selection in Visuospatial Working Memory? *Psychophysiology.*

Wolff, M. J., Jochim, J., Akyürek, E. G., & Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience, 20*(6), 864-871. https://doi.org/10.1038/nn.4546.

Chapter 5: General Discussion and Conclusions

Despite the early interest in the set of processes currently known as “working memory”, the study of its functioning remains a major research focus in the fields of neuroscience and cognitive psychology. The development of new techniques and analytical methods in recent years has facilitated a better understanding of the subprocesses essential for its proper functioning, as well as the limitations inherent to WM. This dissertation has focused on a deeper understanding of three subprocesses that, although they may initially appear independent, are found to be highly interconnected, making it challenging to disentangle them: selection (Chapter 2), maintenance (Chapter 2 and Chapter 3), and removal (Chapter 2 and Chapter 4).

The key findings of this dissertation are outlined below:

- Permanent removal after a retro-cue takes around half a second to be completed in visuospatial WM (**Chapter 2**).
- Load can be decoded using Multivariate Pattern Analysis (MVPA) from EEG right after encoding. However, MVPA is highly sensitive to other factors such as attentional selection or eye movements, which precludes the decoding of load reductions during contents removal after a retro-cue in visuospatial WM (**Chapter 2**).
- MVPA decoding of EEG and EOG signals suggests an important role of attentional selection when WM contents need to be updated (**Chapter 2**).

- Despite the successful baseline WM capacity and task difficulty manipulations, no significant improvements in performance with 4 Hz tACS were obtained (**Chapter 3**).
- The sham group showed a theta power increase with time on task, but contrary to the expected results, theta power did not increase in the active tACS condition, which could be explained by a mismatch between the endogenous and exogenous stimulation frequencies (**Chapter 3**).
- The EEG results do not serve as evidence for the theta-gamma theory because (1) the high WMC group showed faster mean theta frequencies—which aligns more with a model where faster theta cycles enable better WM maintenance rather than the theta-gamma coupling theory—and (2) no modulation of theta frequency related to WM load was observed (**Chapter 3**).
- There was an increase in phase synchrony between stimulation electrodes following tACS, but this effect was only observed when error trials were excluded, suggesting a potential effect of tACS on phase synchronization (**Chapter 3**).
- Removal of information from WM, regardless of whether it is temporary or permanent, improves performance when the time to remove is long enough, observed in faster reaction times, fewer errors, and better discrimination (**Chapters 2 and 4**).
- How temporary and permanent removal operate differs. Permanent removal allows the reduction of the retrieval structure, which turns into more rapid access to the remaining contents since the context used to retrieve the items is simplified. However, temporary removal works by withdrawing attention

from the irrelevant items, which reduces the number of active representations but not the retrieval structure per se (**Chapter 4**).

- The influence of encoding order on subsequent retrieval aligns with the serial organization of information in verbal WM. This supports the idea that items are encoded and accessed in a sequential manner, where earlier encoded items facilitate the retrieval of later ones (**Chapter 4**).
- Information temporarily removed for a longer time allows a deeper removal, negatively affecting its posterior retrieval (**Chapter 4**).

Resumen en español

Introducción

La memoria de trabajo (MT) es un concepto que surge de la distinción entre memoria a corto plazo (MCP) y memoria a largo plazo (MLP). Originalmente descrita en los años 60 por Miller y colaboradores como un sistema temporal para la ejecución de planes, esta idea fue refinada por Atkinson y Shiffrin, y posteriormente por Baddeley y Hitch, quienes destacaron su papel no solo en el almacenamiento temporal, sino también en la manipulación de información, tal y como se considera en la actualidad.

La MT no es homogénea, sino que presenta diferentes estados de activación dependiendo de la atención dirigida a la información presente en esta. Aunque distintos modelos estudian el funcionamiento de estos estados, la presente tesis se basa en el modelo *three embedded-processes* de Oberauer, que distingue entre un componente amplio que activa la información de la MLP, la llamada región de acceso directo con información relevante para la tarea actual y un foco atencional que selecciona principalmente el elemento más importante para la tarea. Por lo tanto, partiendo de este modelo, se estudia el funcionamiento de la selección de contenidos de estos distintos estados de activación, el mantenimiento de esta información en un estado accesible y, por último, la eliminación de estas representaciones de los distintos estados de la MT.

El primer proceso estudiado es la selección de información dentro de la MT, que se podría incluir dentro del término más amplio denominado "atención selectiva interna", el cual abarca varios procesos clave en la MT, como la selección, consolidación o priorización de información. A diferencia de lo que se creía tradicionalmente, la atención selectiva interna no está restringida a procesos transitorios y de corta duración, sino que tiene un papel importante en la actualización flexible de la relevancia de los contenidos de la MT durante periodos más prolongados. Este aspecto es explorado a fondo en el **capítulo 3**, donde se aplica análisis multivariado de patrones (MVPA) a la señal de electroencefalografía (EEG) para explorar el impacto que los procesos atencionales de selección interna tienen sobre la información mantenida en la MT.

Otro proceso relevante para el funcionamiento de la MT es el mantenimiento de la información. Se ha propuesto que la información se puede mantener en dos estados: un estado activo, donde los elementos son atendidos mediante actividad neuronal persistente (aunque distintos estudios desafían esta idea), y un estado pasivo, donde la información no priorizada permanece inactiva pero puede ser reactivada cuando se vuelve relevante. Sin embargo, existen estudios que desafían estas ideas. Dada la falta de consenso sobre estos mecanismos, se necesitan técnicas sensibles, como el MVPA sobre datos de EEG o de resonancia magnética, para examinar estas posibles representaciones, tal como se analiza en el **capítulo 2** de esta tesis.

El rol que tienen las oscilaciones en el mantenimiento de la información en la MT también ha sido muy estudiado. En concreto, la interacción entre oscilaciones de ritmo lento (theta) y rápido (gamma)—*Theta-gamma coupling theory*—parecen ser un sustrato neural importante que sostiene el mantenimiento de las representaciones. En el **capítulo 3** de la presente tesis se utiliza estimulación transcraneal por corriente alterna con el objetivo de reducir la frecuencia theta de los individuos y tratar de mejorar el mantenimiento de ítems en la MT en participantes cuya capacidad de partida era baja.

Por último, dado que la MT tiene una capacidad limitada, la eliminación de la información irrelevante de forma óptima es crucial. El uso de distintos paradigmas de investigación ha demostrado que la eliminación permite la mejora del rendimiento en tareas de MT al reducir la cantidad de información irrelevante, permitiendo una codificación más eficiente de nueva información y acelerando las respuestas relacionadas con la información relevante. Sin embargo, hay situaciones en las que no se requiere de una eliminación definitiva, simplemente temporal. En este sentido, la eliminación permanente ocurre cuando la información es completamente eliminada de la MT, mientras que la eliminación temporal se refiere a situaciones en las que la información se deja de lado temporalmente pero puede necesitarse más adelante. El estudio de la eliminación de información de la MT se ha centrado en la eliminación permanente, mientras que la temporal permanece muy inexplorada. El **capítulo 4** de la presente tesis busca examinar cómo la eliminación temporal y permanente afectan diferencialmente a la recuperación de información en la MT y cómo estos procesos se desarrollan a lo largo del tiempo.

Por tanto, la presente tesis se centra en caracterizar mejor los mecanismos subyacentes en la selección, mantenimiento y eliminación de información en la MT, utilizando métodos tanto conductuales como electrofisiológicos a lo largo de tres estudios presentados en los tres capítulos centrales del trabajo.

Capítulo 2: decodificando carga o selección en la memoria de trabajo visoespacial?

El capítulo 2 se centró en investigar las dinámicas temporales de la eliminación de contenido de la memoria de trabajo, en este caso, de información visoespacial. Investigaciones previas habían demostrado que era posible decodificar la carga de la MT a partir de la señal electroencefalográfica. Sin embargo, hasta el momento no se había testeado si se podía decodificar también las variaciones en la carga mantenida a lo largo del tiempo. Por ello, el objetivo principal era determinar si los cambios en la carga de la MT (o en la cantidad de ítems mantenidos) podrían ser monitorizados durante la eliminación de información. Para ello, se utilizó MVPA sobre la señal de EEG y también sobre la señal de electrooculografía (EOG). Específicamente se hipotetizó: (1) un mejor rendimiento tras la eliminación de la información irrelevante a nivel conductual, igualándose a la condición en la que desde el principio se debían mantener baja carga; (2) una decodificación en la señal de EEG de las condiciones de alta y baja carga antes de cualquier señal que señalara la información relevante; y (3) en paralelo con el rendimiento conductual, que la clasificación de la señal de EEG de la condición de eliminación mostrara una transición de ser clasificada como alta carga a baja carga una vez que la eliminación se hubiera completado.

El experimento se compuso de dos sesiones. En ambas sesiones los participantes debían completar la misma tarea de retro-cue visoespacial. Sin embargo, en la primera sesión, se manipuló el intervalo entre la cue y el target (CTI) para estudiar el curso temporal de la eliminación, mientras que en la segunda sesión sólo se estudió este curso temporal desde un punto de vista electrofisiológico manteniendo este intervalo fijo en un segundo.

Los resultados fueron los siguientes. En primer lugar, los resultados conductuales mostraron que la eliminación se completó en medio segundo, al contrario que otros estudios con información verbal y un mayor número de estímulos en los que se necesitó alrededor de un segundo para ser completada. La carga pudo ser decodificada de la señal de EEG durante el periodo de mantenimiento en la MT mediante MVPA. Sin embargo, los resultados durante el periodo en el que se produjo la eliminación de información muestran que MVPA es una técnica de análisis muy sensible a factores como la selección atencional o los movimientos oculares, lo cual imposibilitó la decodificación de las reducciones de la carga en la MT tras la retro-cue. Además, los resultados muestran un rol importante de la selección atencional durante el periodo de eliminación que no se restringió al momento inicial, sino que se pudo decodificar a lo largo de todo el intervalo, en línea con las conceptualizaciones más recientes de la atención selectiva interna como detectable durante periodos de tiempo prolongados.

Capítulo 3: cuestionando las predicciones del acoplamiento theta-gamma: ineficacia de la estimulación transcraneal de corriente alterna (tACS) a 4 hz en el mantenimiento de la memoria de trabajo y la modulación de la frecuencia theta.

El objetivo general de este estudio fue estudiar el papel de las oscilaciones en el mantenimiento de las representaciones en la MT. Dada la atención recibida por la teoría *theta-gamma coupling* como sustrato electrofisiológico del mantenimiento, se aplica tACS a 4 Hz con el fin de enlentecer las frecuencias theta de los participantes ya que, según esta teoría, esto permitiría un mayor acoplamiento de gamma y, con ello, un mantenimiento de más ítems en la MT. Esta estimulación se aplica a dos grupos de participantes: con alta y baja capacidad base de MT. También se manipuló la dificultad de la tarea mediante ensayos de baja y alta carga dados los resultados de estudios previos donde solo se encontró efectos de la estimulación ante tareas difíciles. Además, se recoge señal de EEG en estado de reposo y durante la ejecución de la tarea tanto antes como después de la estimulación para estudiar los efectos que esta puede tener sobre las oscilaciones cerebrales, independientemente del efecto conductual observable.

Por lo tanto, este estudio tiene dos objetivos específicos: primero, probar la predicción derivada de la teoría *theta-gamma coupling* de que una frecuencia theta más baja facilitaría el mantenimiento de más información en la MT; y, segundo, examinar la influencia de las diferencias individuales y la dificultad de la tarea en la eficacia de la tACS. Se hipotetizó: (1) un mayor beneficio de la estimulación en los participantes con baja MT, en los ensayos más difíciles; (2) una menor frecuencia theta media en los participantes con mayor capacidad de MT de base; y (3) una modulación de los distintos correlatos electrofisiológicos tras la estimulación, como una reducción de la frecuencia theta o un aumento del *power*, entre otros.

Comportamentalmente, los resultados mostraron una correcta selección de los grupos de MT, así como una manipulación correcta de la dificultad de la tarea. Sin embargo, no se observó ninguna mejora ni empeoramiento en la ejecución tras la aplicación de la estimulación tACS a 4 Hz. En cuanto a los resultados de EEG, el grupo sham (control) mostró un incremento en el *power* theta en el último bloque de la tarea con respecto al primero, al contrario que el grupo que recibió estimulación theta, que mantuvo sus niveles de theta *power* estables a lo largo de la tarea, lo cual podría explicarse por una discrepancia entre las frecuencias de estimulación endógena y exógena. Los resultados de EEG de este experimento no aportan evidencia a favor de la teoría *theta-gamma coupling*, ya que el grupo de alta MT mostró frecuencias theta medias más rápidas, al contrario de lo que predice la teoría. Además, no hubo ninguna modulación relacionada con la manipulación de la carga de los distintos ensayos. Por último, se observó un incremento en la sincronización de las fases entre los electrodos de estimulación tras la tACS en los ensayos sin errores, lo que parece sugerir un efecto de la tACS en la sincronización de las fases de las regiones estimuladas.

Capítulo 4: distinción entre la eliminación temporal y permanente en la memoria de trabajo verbal

En este último capítulo, el objetivo principal fue comparar los efectos de la eliminación temporal y permanente de información de la MT. Concretamente, el interés era estudiar los efectos que estos tipos de eliminación podían producir sobre la posterior recuperación del resto de información mantenida en la MT, una comparación que no había sido estudiada hasta el momento. Como en el capítulo 2, se utilizó el paradigma de retro-cue donde se manipuló el intervalo entre la cue y el target para poder estudiar el curso temporal de esta eliminación. Por último, se exploró la influencia que eliminar temporalmente una información durante un tiempo prolongado puede tener en su posterior recuperación. Se hipotetizó que (1) ambos tipos de eliminación dieran lugar a mejoras en el rendimiento cuando el tiempo disponible para eliminar la información era largo, aunque se esperaba que esta mejora fuera superior en la condición de eliminación permanente; y (2) un deterioro del rendimiento para dar la segunda respuesta cuando el intervalo de eliminación temporal había sido más largo.

Por lo tanto, se llevó a cabo un experimento conductual donde los participantes debían recordar inicialmente dos pares de números y solo uno de los pares se marcaba como relevante, pudiendo eliminar el otro par de manera definitiva o temporal para responder seguidamente a una pregunta sobre la localización de uno de esos dos números. En el caso de eliminación temporal, se hacía una segunda pregunta acerca del par inicialmente irrelevante, de forma que debía ser recuperado. Además, dos condiciones en las que se debía codificar solo un par desde el principio o recordar los dos pares hasta el final fueron introducidas como método de control. Por último, se manipuló el intervalo entre la cue y el target (CTI) para estudiar la evolución del curso temporal en las distintas condiciones.

Los resultados mostraron que tanto la eliminación permanente como la temporal llevaron a reducciones progresivas significativas en el tiempo de reacción y en los errores, mostrando una simplificación de la cantidad de información mantenida en la MT. Los tiempos de reacción fueron más lentos en la condición de eliminación temporal, como se esperaba. Un aspecto clave para interpretar correctamente los resultados fue que el rendimiento se vio modulado por el orden de presentación del conjunto de ítems únicamente en la condición de eliminación temporal. Este efecto de orden parece reflejar el mantenimiento de la estructura de recuperación intacta (contextos para la recuperación) que es necesaria para guiar la atención sobre las representaciones de la MT, incluso aquellas que se habían apartado temporalmente. Por lo tanto, parece que la principal diferencia entre ambos tipos de eliminación es la reducción de la estructura de recuperación únicamente en el caso de la eliminación permanente.