The University Of Sheffield.

How can visual working memory be enhanced?

By:

Shuangke Jiang

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

The University of Sheffield

Faculty of Science

Department of Psychology

September 2023

Thesis Abstract

Working memory (WM), the ability to temporarily maintain information for further processing, is known to have limited capacity around four chunks of information (Cowan, 2001). The limited nature of WM and its critical involvement in our daily life, aging and several neuropsychological disorders, have motivated great interest in investigating effective WM enhancement interventions. However, the existing evidence in these interventions yielded discrepancies, raising concerns about the lack of theory-informed WM enhancement research, grounded in solid methodology, especially a formal and explicit assessment. Furthermore, the elusive results have questioned the replicability and robustness of previously reported effects, reflecting the replication and theory crises in the field. This PhD aims to address these concerns and crises. We first aim to empirically assess the robustness of the effects that are induced by two intervention methods, that is, transcranial direct current stimulation and WM training. We also aim to investigate the mechanisms underlying the cognitive changes from a recent theoretical account of cognitive training and transfer effects, that is, the capacity-efficiency mechanism (von Bastian et al., 2022). Finally, we aim to explore which contemporary computational visual WM models could better describe the limits of WM, especially when substantial cognitive changes are induced by interventions like WM training. The evaluation of these theory-driven, mathematically specified WM models could further facilitate a better understanding of the generalisability and replicability of implemented intervention programmes.

Acknowledgement

I started my PhD in October 2019, and I am submitting this thesis in September 2023. During this time, I had a special and challenging research time due to the global pandemic, and I could not achieve this milestone in my life without the support of amazing people around me.

First, I want to thank my first supervisor Dr. Claudia von Bastian. I have been feeling incredibly lucky to have you as my supervisor throughout my PhD. You have not only supervised my PhD but also exerted a huge positive influence on my life (as my academic mother). I have been constantly learning and receiving support from you. I have gained countless knowledge from your feedback, advice, and regular meetings and discussions with you. You have provided me with extra career opportunities and supported my applications for extra funding. Your rigour and passion for psychology research has inspired me and will keep guiding me in my future career. I cannot thank you enough! Second, I want to thank my second supervisor Dr. Myles Jones for your help, guidance, and supervision throughout this period. Third, I want to acknowledge my lab peers Kubra, Hakan, Billy, Eleanor, Alice, Fred and Jiawei for your support. Fourth, I want to acknowledge my cohort friends Anton, Melinda, Lily, Wenyi, Zhuo, and Bixuan for your support. Last but not least, I want to acknowledge the love and support from my family, especially my mother Shuangyue Yan, my best friend Wanyue Zhong, and my cute cat Cheeto!

Declaration and Note on Inclusion of Published Work

I, Shuangke Jiang, confirm that this thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means).

This thesis is in a publication format, and contains the following published work:

Chapter 2: Research presented in this chapter is currently under review and the preprint is available online at Research Square:

Jiang, S., Jones, M., & von Bastian, C. C. (2023). *Does transcranial direct current stimulation enhance visual working memory? A replication study*. Research Square. https://doi.org/10.21203/rs.3.rs-3192523/v1.

Chapter 3: Research presented in this chapter has been published and is available online at *Journal of Cognition*:

Jiang, S., Jones, M., & von Bastian, C. C. (2023). Mechanisms of cognitive change: Training improves the quality but not the quantity of visual working memory representations. *Journal of Cognition*, *6*(1), 1–20. https://doi.org/10.5334/joc.306

Conference Dissemination:

Chapter 2: Findings from this chapter have been presented in the following conferences:

Jiang, S., Jones, M., & von Bastian, C. C. (2020). *Proposal presentation - the tDCS effect on visual working memory over DLPFC and PPC* [Talk]. The 10th European Working Memory Symposium (EWoMS X), Online.

Jiang, S., Jones, M., & von Bastian, C. C. (2022). *TDCS effects on visual workingmemory over dorsolateral prefrontal cortex and posterior parietal cortex: A conceptual replication study of Wang et al. (2019)* [Talk]. Learning and Plasticity (LaP) Conference, Online.

Jiang, S., Jones, M., & von Bastian, C. C. (2022). *TDCS effects on visual working memory over dorsolateral prefrontal cortex and posterior parietal cortex: A*

conceptual replication study of Wang et al. (2019) [Talk]. Great Yorkshire Memory Meeting, York, UK.

Jiang, S., Jones, M., & von Bastian, C. C. (2022). *Effects of tDCS over DLPFC and PPC on visual working memory: A conceptual replication study of Wang et al. (2019)* [Poster]. The 22nd European Society for Cognitive Psychology (ESCoP) Conference, Lille, France.

Chapter 3: Findings from this chapter have been presented in the following conferences:

Jiang, S., Jones, M., & von Bastian, C. C. (2021). *Visual working memory training effects on the quantity and quality of representations* [Poster]. Psychonomic Society 62nd Annual Meeting, Online.

Jiang, S., Jones, M., & von Bastian, C. C. (2021). *Mechanisms of visual working memory training effects: Capacity and efficiency* [Talk]. Virtual Working Memory Symposium, Online.

Jiang, S., Jones, M., & von Bastian, C. C. (2023). *Mechanisms of visual working memory training effects: Capacity vs efficiency* [Talk]. Experimental Psychology Society (EPS) Meeting: University College London, London, UK.

Chapter 4: Findings from this chapter have been presented in the following conferences:

Jiang, S., Jones, M., & von Bastian, C. C. (2022). *Model selection to explain training-induced changes in visual working memory* [Poster]. Psychonomic Society 63rd Annual Meeting, Boston, USA.

Jiang, S., Jones, M., & von Bastian, C. C. (2023). *No single measurement model of visual working memory can explain training-induced change* [Poster]. The 23rd European Society for Cognitive Psychology (ESCoP) Conference, Porto, Portugal.

Table of Contents

List of Tables

List of Figures

Chapter 1 – General introduction

Contributions:

Shuangke Jiang (conceptualization, writing – original draft and editing)

Myles Jones (supervision and review)

Claudia von Bastian (supervision, review, and editing)

1.1 Working memory

Working memory (WM) is a cognitive system that provides access to temporarily maintained information for further processing (Miyake & Shah, 1999). WM is well known for its limited capacity. Specifically, around four chunks of information that can be simultaneously maintained at a time (Cowan, 2001). WM underpins a wide range of cognitive processes (for a review, see Barrett et al., 2004), such as perception (Agam & Sekuler, 2007; Teng & Kravitz, 2019), reasoning (Conway et al., 2003; Kyllonen et al., 1990; Oberauer et al., 2008), executive functions (Miyake et al., 2000), and intelligence (Engle et al., 1999; Fukuda et al., 2010). Furthermore, WM capacity is crucial for our general learning abilities and scholastic performance (Alloway & Alloway, 2010).

Like other fluid cognitive abilities, WM develops through the life span, and especially declines with age (Craik & Bialystok, 2006; Froudist-Walsh et al., 2018; Park et al., 2002). The early development of frontal-parietal neural networks is correlated with WM function in children (Alcauter et al., 2014; Fitch et al., 2016). Moreover, aging-related regional degeneration in neuronal structures (e.g., loss of dendrites and synapses) and dysregulation of neurotransmitters (e.g., decline in the efficiency of dopaminergic and cholinergic systems) have been shown to hinder long-term potentiation and neuroplasticity, and to contribute to WM impairment (Croxson et al., 2011; Froudist-Walsh et al., 2018; Sawaguchi & Goldman-Rakic, 1991; Störmer et al., 2012; Tsukada et al., 2005). Furthermore, deficits in WM are often a frequent concurrent symptom in several neurological diseases or psychological disorders such as Alzheimer's disease (AD) and mild cognitive impairment (MCI; Belleville et al., 2006), attention-deficit hyperactivity disorder (ADHD, Martinussen et al., 2005) and major depressive disorder (MDD, Rose & Ebmeier, 2006). Overall, the limited nature of WM, as well as its critical involvement in daily life, aging and several neuropsychological disorders have motivated great interest in investigating effective WM enhancement.

1.2 Theories of WM enhancement

Klingberg (2010) has proposed an influential theory of neural plasticity mechanisms underlying WM capacity enhancement. Cognitive plasticity has been regarded as the core foundation for transferable intervention benefits. At the neural level, enhanced WM capacity is correlated to changes in dopaminergic receptors that modulate synaptic plasticity (McNab et al., 2009). Furthermore, at the cortical level, enhanced WM capacity is correlated with stronger brain activity and functional connection between cortical regions, in particular frontal and parietal regions (Olesen et al., 2004). When two tasks share similar dopaminergic systems they recruit or require overlapping brain networks, the improvement in one task by forming new plasticity is more likely to be transferred to the other task (Constantinidis & Klingberg, 2016).

Following these previous accounts, WM enhancement interventions have long been aimed to enhance WM capacity and, thereby, potentially transfer to a broad range of related cognitive abilities (Jaeggi et al., 2008; Klingberg, 2010; Klingberg et al., 2002). However, the lack of consistent transfer benefits in existing evidence cannot be explained by such accounts that WM enhancement is a result of increased capacity (De Simoni & von Bastian, 2018; Melby-Lervåg & Hulme, 2013; Robison et al., 2017; Shipstead et al., 2012; von Bastian et al., 2019; S. Wang et al., 2019). To provide potential explanations for these discrepancies in past findings, a capacity-efficiency model has been introduced (von Bastian et al., 2022; von Bastian & Oberauer, 2014).

According to the capacity-efficiency model, two pathways, not mutually exclusive, have been proposed to describe how WM can be enhanced. One pathway is through expanding capacity as a general resource (e.g., increasing the quantity of representations that can be held activated in WM). Expanded capacity should generalise to other cognitive domains that rely on the same capacity limit. The other pathway is through enhancing

efficiency in using the existing capacity, for example, by applying the initially acquired expertise, such as strategies to reduce cognitive load or optimise attention allocation, to a new task. Enhanced efficiency will only lead to transfer if the acquired expertise can be applied in the context or content that share some similarity (Barnett & Ceci, 2002). In specific, expertise can be developed paradigm-specifically and/or stimulus-specifically to provide faster access to new cognitive routines when facing new stimulus types or paradigms. Paradigm-specific expertise will lead to better performance in tasks with the same surface structure but different types of stimuli (e.g., recall the orientation of triangles or the shape of rings). Stimulus-specific expertise will lead to better performance in tasks using the same type of stimuli but different paradigms (e.g., the orientation of triangles in a recall or recognition task). In the light of these theoretical frameworks, the potentials of interventions, such as non-invasive brain stimulation and computerised cognitive training, have been widely investigated regarding enhancing WM or preventing WM impairment (Goldthorpe et al., 2020; Karbach & Verhaeghen, 2015; Morrison & Chein, 2011; Siegert et al., 2021).

1.3 Transcranial direct current stimulation

Transcranial direct current stimulation (tDCS), as one of the most popular non-invasive brain stimulation techniques, passes weak (i.e., low intensity), directs current to cortical areas to facilitate or inhibit spontaneous neuronal network activity as a promising neuromodulation tool (Brunoni et al., 2012). When tDCS operates, the current flows inwards the brain via anodal electrodes and outwards via cathodal electrodes, inducing changes in the electric field that affect neuronal behaviours (Nord & Jonathan, 2015; Pelletier et al., 2015). Anodal stimulation is typically assumed to increase cortical excitability to enhance cognitive functions, whereas cathodal stimulation decreases excitability and, thus, inhibits brain activities (Bikson et al., 2012; Nitsche & Paulus, 2000). At the neural level, tDCS-induced membrane polarisation is related to a range of changes in action potential threshold and

timing, synaptic plasticity, neurotransmitter modulation, and even neural network coherence, which therefore influences the function of the nervous system, including WM (Jackson et al., 2016).

TDCS potentiates WM enhancement by directly targeting the neural substrates of the WM processes involved. One well-documented example is prefrontal-parietal stimulation. The prefrontal cortex plays a top-down control role over WM storage (Curtis & D'Esposito, 2003; Lara & Wallis, 2015). The prefrontal cortex does not maintain stimulus information in WM, yet it has access to that information and can reliably encode whether subsequent stimuli are targets or distractors (Pasternak et al., 2015; Stokes et al., 2013). At the same time, the posterior parietal cortex is the key neural locus of maintaining visual WM representations which is regulated by the prefrontal cortex (Lara & Wallis, 2015; Todd & Marois, 2004). Given the crucial roles of the dorsolateral prefrontal cortex (DLPFC) and the posterior parietal cortex (PPC) in regulating and maintaining WM representations (Curtis & D'Esposito, 2003; Ikkai & Curtis, 2011), stimulating these two brain regions may enhance WM performance.

Indeed, tDCS over DLPFC and PPC showed some beneficial effects (Arciniega et al., 2018; Baumert et al., 2020; Li et al., 2017). However, both DLPFC and PPC stimulation benefits failed to render replicable and robust empirical evidence (Dumont et al., 2021; Nikolin et al., 2018; Robison et al., 2017). Furthermore, at the meta-analytic level, only negligible to minimal effects of single-session tDCS have been observed. Specifically, previous reviews reported significant but relatively small tDCS effects to improve WM performance only in terms of reaction time (RT) but not accuracy (Brunoni & Vanderhasselt, 2014; Dedoncker et al., 2016; Hill et al., 2016), whereas other reviews showed no substantial effects of tDCS on WM enhancement at all (Horvath et al., 2015; Mancuso et al., 2016; Medina & Cason, 2017b).

One of the reasons that could contribute to the observed elusive evidence for tDCS effects could be methodological variations. For example, researchers could choose either online or offline protocols where tDCS is administered during or before cognitive tasks, respectively. The variation of tDCS stimulation that was employed in single empirical studies, making it hard to identify optimal tDCS parameters for consistent benefits in WM enhancement. For instance, Fregni et al. (2005) have first reported small to medium effects of 10-minute, anodal-DLPFC stimulation to improve accuracy but not reaction time. In contrast, Hoy et al. (2013) found that 20-minute anodal-DLPFC could only reduce reaction time. Both studies used the common dosage (1–2 mA up to 20 minutes) with differences in duration time and online/offline protocols, which has led to opposite patterns of WM performance changes in accuracy, warranting further investigations or replications of previously reported tDCS-induced benefits with identical tDCS parameters.

Another reason could be due to the used WM tasks that tap unclear underlying WM architecture. Previous studies predominately used n-back tasks to investigate tDCS-induced benefits in WM by stimulating DLPFC (for reviews, see Brunoni & Vanderhasselt, 2014; Dedoncker et al., 2016; Hill et al., 2016; Horvath et al., 2015; Mancuso et al., 2016; Medina & Cason, 2017; for empirical examples, see Fregni et al., 2005; Ohn et al., 2008; Zaehle et al., 2011). In addition to the inconsistent benefits, n-back tasks have been criticised for lacking robust associations with other WM tasks (Jaeggi et al., 2010; Kane et al., 2007; Redick & Lindsey, 2013; Wilhelm et al., 2013). Therefore, it remains unclear to what extent n-back task performance reflects the limited capacity of WM, let alone tDCS-induced substantial changes in WM (Meule, 2017). Other empirical PPC-tDCS studies have used change-detection tasks to achieve more reliable measures of WM capacity estimates (Dai et al., 2019; Frost et al., 2021). Similarly, administering a continuous-reproduction paradigm combined with computational models provides measures of latent components of WM

representations, such as capacity and precision. This advanced approach further enables us to differentiate the latent WM representations from construct-irrelevant noise, especially motor noise which explains the main source of previously claimed tDCS benefits in WM. Testing the effects of interventions on these latent components could facilitate the understanding of tDCS-induced enhancement as well as the nature of WM (S. Wang et al., 2019; Zhang & Luck, 2008).

Overall, the inconsistency in prior evidence for tDCS effects could be due to the complex, variable tDCS parameters and unclear tapped WM components in the used WM tasks, making it hard to identify the optimal paradigm and true underlying mechanisms of tDCS effects. Therefore, investigating the replicable and robust effects of tDCS in enhancing WM is of great importance.

1.4 Cognitive training

An alternative intervention that has been designed to improve cognitive abilities is computerised cognitive training. Cognitive training typically involves the practice of one or more cognitive tasks over a short period of time, in repetitive or/and adaptive manner (Jaeggi et al., 2008; Moriya, 2019; von Bastian & Oberauer, 2014). The aim of cognitive training is to improve performance in both trained and untrained cognitive tasks. Improvement in the trained tasks refers to a *training effect*, whereas the generalisation of training effects to untrained tasks is called a *transfer effect*. Prior training research reported large and replicable training gains in the trained tasks whilst the transfer effect to untrained tasks and cognitive abilities has failed to yield consistent and robust evidence (Jaeggi et al., 2012; Karbach & Verhaeghen, 2015; Melby-Lervåg et al., 2016; Morrison & Chein, 2011; Shipstead et al., 2012; von Bastian et al., 2022).

Such inconsistencies could possibly stem from the methodological flaws from the early training studies. First, positive effects of WM training that were claimed in early WM

training research, spurring intensive research efforts on investigating WM training effects and the commercialization of numerous WM training programmes. However, these early claims often accompany a small sample size (for a review, see Redick et al., 2015). The small sample size translates into low statistical power (Button et al., 2013), which can lead to false-positive interpretations on the strength of WM training programmes. Second, early training interventions have been criticised for an overestimated effect due to ineffective control of placebo effects and expectancy effects with only a passive control group (Boot et al., 2013; Simons et al., 2016; von Bastian & Oberauer, 2014). Placebo effects may arise in interventions like WM training when the desired outcome is known or suggested to participants who are eager to improve their cognitive abilities. Foroughi et al., (2016) demonstrated that placebo effects, but not cognitive training, led to an improvement in intelligence. An active control group design that can match the levels of boredom, perceived effort, and training expectation with the treatment group could exert good control of placebo effects and expectation control, facilitating the investigation on reliable, true training effects and their underlying mechanisms (Foroughi et al., 2016; Redick, 2019; Simons et al., 2016; von Bastian & Oberauer, 2014).

In addition to addressing these concerns on methodology, a lack of theory-driven research on WM training can be reflected by these inconsistent findings on transfer effects. The metaphor that the brain is a muscle has often been used in many early training studies (Morra & Borella, 2015). Doing WM exercise makes capacity grow like a well-trained muscle. Therefore, it is assumed that WM training could work because of the increased number of discrete representations one can hold at a time, that is, increased capacity. Moreover, training on WM could lead to a broad transfer effect as increased muscle volume can afford heavier workloads in other complex cognitive tasks. However, the limited empirical evidence in transfer effects seems to speak against the muscular metaphor and the

notion of increased capacity only. Furthermore, emerging evidence in strategies use depending on the current task demands across different WM training sessions has brought attention to a different mechanism, such as enhanced efficiency by making a better use of the existing capacity (Gathercole et al., 2019; Redick et al., 2013; von Bastian & Oberauer, 2014). Therefore, a more comprehensive, falsifiable theoretical framework is needed to investigate when and why training and transfer effects may occur (Redick, 2019; Smid et al., 2020; von Bastian & Oberauer, 2014). The capacity-efficiency model proposed by von Bastian (2022) can potentially meet such a need by enabling us to test both views of increased capacity and enhanced efficiency through acquisition of strategies (also see section **1.2**).

To recap, the elusive evidence for transfer effects that were reported in previous WM training research has casted concerns on methodological limits, and more importantly the lack of comprehensive, falsifiable theoretical frameworks when assessing the training and transfer effects and their underlying mechanisms. Thus, theory-informed research grounded in solid methodology is pivotal for providing robust evidence for the true efficacy of WM training, and thus facilitating a better understanding of the generalisability and replicability of implementation training interventions (Redick, 2019; The Improved Clinical Effectiveness through Behavioural Research Group, 2006; von Bastian et al., 2022).

1.5 From explanatory theories to formal assessment of WM enhancement

Like other psychology research, intervention research on non-invasive brain stimulation and cognitive training aiming to improve WM, has been suffering from the 'replication crisis', that is, problematic failures or difficulties to replicate previously reported findings and/or form coherent theories developed from them (Open Science Collaboration, 2015; The Improved Clinical Effectiveness through Behavioural Research Group, 2006). The replication crisis has raised the concerns on not only the issues of methodology, statistics,

publication bias, but also the lack of robust measurement practice and theory development. The latter concern also refers to a 'theory crisis' which contributes to poor testability of theories and hypotheses (Borsboom et al., 2021; Eronen & Bringmann, 2021; Fried & Flake, 2018; Frischkorn & Popov, 2023; Maatman, 2021; Ngiam, 2023; Oberauer & Lewandowsky, 2019).

In addition to rigorous methodology, one way to address these crises is to have *open theory* – to guide researchers to conceptually analyse, specify, and formalise the explanatory theoretical frameworks that were previously proposed and debated using computational measurement models (Guest & Martin, 2021). *Explanatory models* provide theoretical explanations of the observed phenomena as well as the effects of experimental manipulations like interventions on WM. However, verbal descriptions of a theory do not necessarily lead to the same understanding between researchers. Taking one of the prominent debates in WM as an example, a discrete-slot view assumes an all-or-none information state, that WM is limited by the number of items that can be remembered with certain resolution of the remembered item information or not (Zhang & Luck, 2008); whilst a variable-precision view denies this all-or-none notion, suggesting an alternative explanation with the variation in memory strength due to noise in representations (Bays & Husain, 2008). If hypothesised WM capacity were tested without explicitly refined theory, claims that were drawn from the observed patterns of data could be mis-interpreted and/or mis-communicated by different research teams, which can be a precursor to the replication crisis (Guest & Martin, 2021; Ngiam, 2023). To help to make the interpretations or predictions of an effect more reproducible, *measurement models* of WM can explicitly specify and quantitatively measure theoretically meaningful latent variables as their free parameters, such as capacity vs precision of WM, and thus enable other researchers to replicate the reported relationship (Farrell & Lewandowsky, 2018; Oberauer et al., 2018).

Despite the critical role of measurement models in open theory to address the replication and theory crises, the use of computational measurement models in investigating the underlying mechanisms of WM enhancement is still in its infancy with three aspects. First, one measurement model is often arbitrarily chosen without giving strong theoretical reasons and/or quantitatively comparing which possible measurement models most adequately capture cognitive changes in WM. Such practice can potentially prolong the current debates in theories if the inference based on an arbitrarily chosen measurement model was in turn used to support the original theoretical assumptions of the chosen model and against the unchosen competing models. Second, fewer existing studies with model comparison compared more than two models (often the Standard Mixture Model and the Swap Model were compared), which has limited practical implications for other widely used, adequate models. Third and more importantly, the majority of modelling studies are based on single-sessional studies (Bays, 2016; Bays & Husain, 2008; Oberauer, 2021; Schurgin et al., 2020; Tomić & Bays, 2022; van den Berg et al., 2014; Williams et al., 2022; Zhang & Luck, 2008). However, intervention research often requires an accurate quantification of performance changes across multiple testing sessions to make inferences on the training effects and their underlying mechanisms. Therefore, evaluating the application of a set of existing measurement models accounting for possible changes in different test sessions is needed.

1.6 Aims of this PhD thesis

Overall, the existing evidence in WM enhancement interventions such as non-invasive brain stimulation and cognitive training yielded discrepancies, reflecting the lack of theory-informed research grounded in solid methodology, especially a formal, explicit assessment of these theories. To fill this gap, my PhD thesis has three core aims. In the context of the replication crisis, the first aim is to empirically assess the robustness of the

effects that are induced by intervention methods, such as tDCS and WM training. In light of the theory crisis, the second aim is to investigate the mechanisms underlying the cognitive changes from a recent theoretical account of cognitive training and transfer effects, that is, the capacity-efficiency mechanism (von Bastian et al., 2022; von Bastian & Oberauer, 2014). Last but not least, we aim to explore which contemporary computational visual WM models could better describe the limits of WM, especially when substantial WM changes are induced by interventions like WM training. The evaluation of these theory-driven, mathematically specified WM models could further facilitate a better understanding of the generalisability and replicability of implemented intervention programmes.

In chapter 2, we aim to replicate the positive effects of tDCS that have been previously reported by Wang et al. (2019) after accounting for the possible problematic design of the original study. In chapter 3, we will investigate the effects of WM training and the underlying mechanisms of cognitive changes that are induced by training. In chapter 4, we will compare the contemporary VWM measurement models to select which model could capture VWM performance when intervention-induced changes are present or absent. Chapter 5 will highlight the main findings of this PhD thesis and the theoretical and practical implications of this specific research topic. Moreover, this chapter will reflect on directions for future research by discussing the limitations of this thesis.

Chapter 2 – Does transcranial direct current stimulation enhance visual working memory? A replication study

Contributions:

Shuangke Jiang (conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, validation, visualization, writing—original draft, writing—review & editing)

Myles Jones (conceptualization, funding acquisition, methodology, resources, supervision, writing—review & editing)

Claudia von Bastian (conceptualization, funding acquisition, methodology, resources, software, supervision, writing—review & editing)

Research presented in this chapter is currently under review and the preprint is available online at Research Square:

Jiang, S., Jones, M., & von Bastian, C. C. (2023). *Does transcranial direct current stimulation enhance visual working memory? A replication study*. Research Square. https://doi.org/10.21203/rs.3.rs-3192523/v1.

Findings from this chapter have been presented in the following conferences:

Jiang, S., Jones, M., & von Bastian, C. C. (2020). *Proposal presentation - the tDCS effect on visual working memory over DLPFC and PPC* [Talk]. The 10th European Working Memory Symposium (EWoMS X), Online.

Jiang, S., Jones, M., & von Bastian, C. C. (2022). *TDCS effects on visual workingmemory over dorsolateral prefrontal cortex and posterior parietal cortex: A conceptual replication study of Wang et al. (2019)* [Talk]. Learning and Plasticity (LaP) Conference, Online.

Jiang, S., Jones, M., & von Bastian, C. C. (2022). *TDCS effects on visual working memory over dorsolateral prefrontal cortex and posterior parietal cortex: A conceptual replication study of Wang et al. (2019)* [Talk]. Great Yorkshire Memory Meeting, York, UK.

Jiang, S., Jones, M., & von Bastian, C. C. (2022). *Effects of tDCS over DLPFC and PPC on visual working memory: A conceptual replication study of Wang et al. (2019)* [Poster]. The 22nd European Society for Cognitive Psychology (ESCoP) Conference, Lille, France.

Case study based on research in this chapter has received the University of Sheffield Open Research Prize 2023 (individual PGR category):

Jiang, Shuangke (2023). Open Research Case Study: Shuangke Jiang - Preregistration, sharing materials, and conducting replication studies in Psychology. The University of Sheffield. Report. https://doi.org/10.15131/shef.data.23702061.v1

Abstract

In recent years, non-invasive brain stimulation has been highlighted as a possible intervention to induce cognitive benefits, including on visual working memory (VWM). However, findings are inconsistent, possibly due to methodological issues. A recent high-profile study by Wang et al. (2019) reported that anodal transcranial direct current stimulation (tDCS) over posterior parietal cortex (PPC), but not over dorsolateral prefrontal cortex (DLPFC), selectively improved VWM capacity but not precision, especially at a high VWM load. Given the broad implications of this finding, it is imperative to test its replicability. Thus, in the current pre-registered conceptual replication study, we accounted for the key potential methodological issues in the original study and tested an adequate number of subjects required to demonstrate the previously reported effects (n=48 compared to n=20). Participants underwent counterbalanced PPC, DLPFC and sham stimulation before completing 360 trials of a continuous orientation-reproduction task. We failed to replicate the selective effect of PPC stimulation. Instead, our results showed little credible evidence for effects of tDCS regardless of stimulation region and VWM load. The absence of tDCS effects was largely supported by substantial to strong Bayesian evidence. Therefore, our results challenge previously reported benefits of anodal PPC-tDCS on VWM.

Keywords: visual work memory, tDCS, DLPFC, PPC

2.1 Introduction

Visual working memory (VWM) refers to the active maintenance of visual information needed for higher cognitive processing in the present moment (Luck & Vogel, 2013). Typically, WM is limited to maintaining three to four chunks of information (Cowan, 2001). Like other fluid cognitive abilities, WM declines with age (Craik & Bialystok, 2006; Park et al., 2002). Furthermore, deficits in WM often occur with neurological diseases and psychological disorders. The limited capacity of WM, and its critical involvement in many disorders, has stimulated intensive research efforts into the effectiveness of WM enhancement interventions.

In particular, there is growing interest in affordable and non-invasive brain stimulation techniques, such as transcranial direct current stimulation (tDCS). Benefits of tDCS have been demonstrated for healthy young adults (Fregni et al., 2005; Ohn et al., 2008; Hsu et al., 2014; Johnson et al., 2022; Jones & Berryhill, 2012; Karthikeyan et al., 2021; Tseng et al., 2012; Zaehle et al., 2011), healthy older adults (for a review, see Siegert et al., 2021) as well as clinical cohorts with mild cognitive impairment and early Alzheimer's disease (for reviews, see Chen et al., 2022; Hsu et al., 2015), attention-deficit hyperactivity disorder (for reviews, see Cosmo et al., 2020; Salehinejad et al., 2020) and major depressive disorder (for reviews, see Brunoni et al., 2012; Woodham et al., 2021).

Typically, tDCS delivers weak currents from anode to cathode through the skull, generating electric fields to modulate cortical activities and facilitate neuroplasticity (Roche et al., 2015; Ruffini et al., 2013). Anodal stimulation is assumed to increase cortical excitability to enhance cognitive functions, whereas cathodal stimulation decreases excitability and, thus, inhibits brain activities (Bikson et al., 2012; Nitsche & Paulus, 2000). However, such polarity-specific effects of tDCS are likely an oversimplification when considering complex cognitive functions like VWM. For example, whereas excitatory effects

of anodal stimulation are largely robust, inhibitory effects of cathodal stimulations are less consistent when it comes to studies investigating complex cognition rather than motor effects (for a review, see Jacobson et al., 2012). Taken together, regardless of inconsistent cathodal effects, anodal stimulation has been shown to consistently modulate the neural activities in the target brain regions and, thus, is a promising avenue to enhance the corresponding cognitive functions.

Given that neural activation of frontal-parietal brain regions is known to be involved in the maintenance of VWM representations (Curtis & D'Esposito, 2003; Ikkai & Curtis, 2011), a growing body of research has investigated the possible VWM benefits of anodal stimulation of the dorsolateral prefrontal cortex (DLPFC) and posterior parietal cortex (PPC) (Arciniega et al., 2018; Baumert et al., 2020; Li et al., 2017; Wang et al., 2019). However, some previous studies have showed null effects of tDCS for both DLPFC and PPC stimulation (Dumont et al., 2021; Nikolin et al., 2018; Robison et al., 2017). Some meta-analyses quantifying the effectiveness of tDCS across multiple studies report medium effects of single-session tDCS on VWM (Brunoni & Vanderhasselt, 2014; Dedoncker et al., 2016; Wischnewski et al., 2021), while others report only negligible effects of single-session tDCS (Hill et al., 2016; Horvath et al., 2015; Mancuso et al., 2016; Medina & Cason, 2017a). These inconsistencies on the meta-analytic level point to several critical caveats of meta-analyses. Specifically, any conclusions drawn from meta-analysis depend on the included primary studies. First, if the included primary studies largely reported only positive effects, together with overestimated study-level effect sizes (Halsey et al., 2015), it can lead to high false-positive rate of meta-analyses (Kvarven et al., 2020; von Bastian et al., 2019). Second, tDCS studies vary widely in their design, such as administering online or offline protocols (Živanović et al., 2021), stimulating different regions (Wischnewski et al., 2021), or using different VWM paradigms which may require different cognitive processes to one another

(Saucedo-Marquez et al., 2013). These methodological variations could have contributed to the inconsistencies observed across both single studies and meta-analyses. Therefore, replications of those studies that reported positive results, using the same parameters and cognitive paradigms, may yield more conclusive evidence as to whether tDCS is effective or not.

The present, pre-registered replication study, therefore, focuses on a particularly high-profile study by Wang et al. (2019) who recently reported selective benefits of anodal tDCS over the PPC, but not DLPFC, on VWM. Wang et al. used a continuous-reproduction VWM paradigm and fitted the mathematical standard mixture-model (Zhang & Luck, 2008) to estimate VWM capacity (quantity of representations maintained in VWM) and precision (quality of those representations). In this task, participants memorised the orientations of 2, 4, or 6 bars on a screen. After either a short (100 ms) or long (1000 ms) interval, participants were asked to reproduce the orientation of one of the bars by mouse-click. The deviation of the reproduced orientation from the original orientation was then used to estimate VWM capacity and precision for each participant, interval duration, set size, and stimulation condition. Wang et al. tested the effects of 15-min anodal tDCS over the left DLPFC and the right PPC relative to a sham condition with a within-subjects design in 20 participants. After excluding two participants due to their poor performance at set size 6, Wang et al. observed a selective increase in VWM capacity for the long retention interval at this set size after PPC stimulation relative to sham, but not after DLPFC stimulation, at any other set size, short retention interval, or on VWM precision.

Wang et al. (2019) interpreted these findings as “causal evidence” (p. 535) of the role of the PPC for VWM functioning. They further argued that “tDCS could be used as promising non-invasive method to enhance [VWM]” (Wang et al., 2019, p. 535). Indeed, Wang et al.’s findings have several significant theoretical and practical implications to the

fields of VWM and tDCS. First, the findings from this study falsified the role of the anodal stimulation on DLPFC in improving VWM, thereby contradicting previous studies (Andrews et al., 2011; Fregni et al., 2005; Ohn et al., 2008; Zaehle et al., 2011). Second, by showing that tDCS selectively increases the capacity, but not the precision, of representations held in VWM, Wang et al.'s (2019) findings strongly favour theories conceptualizing the capacity limit of VWM as discrete memory slots (Zhang & Luck, 2008) over those assuming a flexible, continuous resource (Schneegans & Bays, 2016; Van Den Berg et al., 2012). Third, the promising benefits of PPC-tDCS suggests that VWM capacity can be expanded with a non-invasive, cost-effective method, with strong practical implications for clinical tDCS applications. Importantly, by employing a sham-control and through the null effects of DLPFC stimulation, Wang et al. (2019) excluded the possibility that these changes were driven by placebo effects or global excitability with tDCS (Dawood et al., 2019). Furthermore, Wang et al.'s (2019) additional control for sensory memory also ruled out the possibility that these changes were due to mere sensory processes or attentional regulation.

Given these far-reaching implications, it is imperative to ensure that Wang et al.'s (2019) findings are robust and replicable. Replication studies can verify the reliability of the originally reported effects (Simons, 2014), and test the generalizability across conditions that inevitably differ from the original study (Nosek & Errington, 2020). This is particularly critical in the present replication study because, despite its important findings and implications, several aspects of Wang et al.'s (2019) study are potentially problematic and warrant further investigation. First, Wang et al. retained only a small sample of 18 participants for analysis. The small sample size translates into low statistical power even for moderate effect sizes, and low statistical power can lead to false-positive findings (Button et al., 2013). The reported effect size is very large (d = 1.03), but this may reflect an overestimation due to low statistical power (Halsey et al., 2015). Second, Wang et al. (2019)

administered only 60 trials per design cell. Such relatively small numbers of trials increase bias and noise variance and thus reduce the precision of the parameter estimation (Lerche et al., 2017; Wiecki & Frank, 2013). Third, the counterbalancing of conditions was likely incomplete in Wang et al. (2019) and, thus, their design did not adequately control for possible carryover (e.g., practice) effects across sessions. Specifically, the study entailed three sessions (DLPFC, PPC, and sham stimulation), resulting in at least six possible sequences requiring counterbalancing. However, with 20 participants completing the experiment (and 18 included in the analysis), it is impossible to assign an equal number of participants to all sequences. Consequently, it cannot be excluded that carryover effects contributed to the previously reported effects. Finally, like many other studies, Wang et al. (2019) used rotated bars as stimuli. Thus, the unique angles of their stimuli effectively ranged only from 0° to 180°, leaving room for developing task-specific strategies. For example, participants may have realised they could simply memorise the location of just one end of the bar rather than the actual orientation of the bar, thereby making the task considerably easier.

To address these potential issues of their study, in this pre-registered experiment, we aimed to replicate Wang et al.'s (2019) study, using a bigger sample size, larger number of trials, complete counterbalancing, and stimuli that use the full space of possible angles (360°). Our pre-registered hypotheses (https://osf.io/n9fkp) based on Wang et al.'s findings were as follows:

Hypothesis 1: PPC stimulation will increase VWM capacity more than DLPFC stimulation. This effect is particularly pronounced at a high difficulty level of the task (i.e., set size 6).

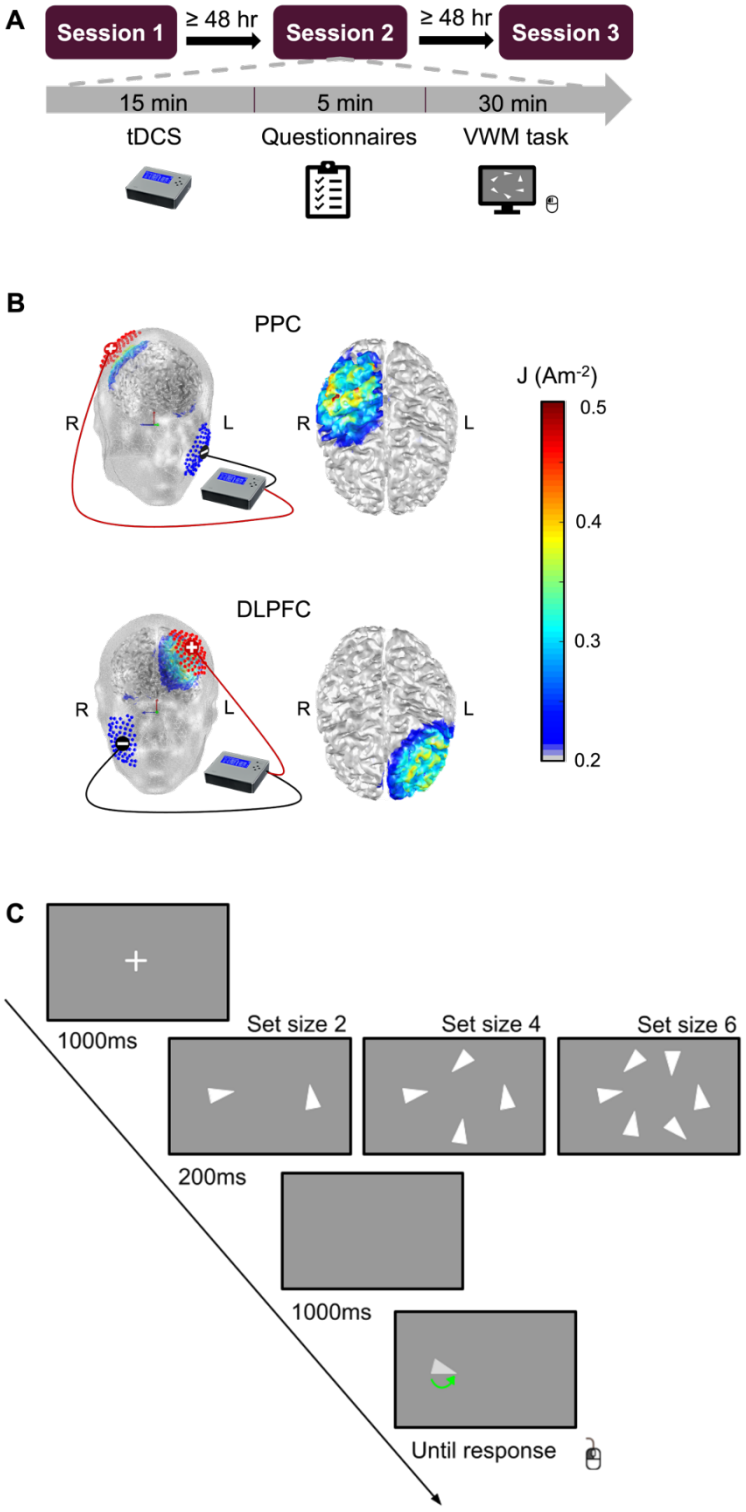
Hypothesis 2: Neither PPC nor DLPFC stimulation will improve VWM precision.

2.2 Results

Figure 2.1 provides an overview of the study procedure, tDCS setup, and VWM task used. During each experimental session, participants first received either active PPC stimulation, active DLPFC stimulation or sham. Following short questionnaires, post-tDCS VWM performance was measured using a continuous orientation-reproduction task. During each trial of the VWM task, participants memorised the orientations of 2, 4, or 6 triangles on a screen. In this replication, we included only the relevant maintenance condition with a long (1000 ms) retention interval, after which participants were asked to reproduce the orientation of one of the triangles. The deviation of the reproduced orientation from the original orientation was then used to estimate VWM capacity and precision for each participant, set size, and stimulation condition.

Figure 2.2

Study Overview



Note. Panel A: Study procedure. Panel B: tDCS montages on head models (left) and current density distributions from the superior view (right). Red electrodes with a cross: anode; black

electrodes with a line: cathode. Panel C: Continuous orientation-reproduction task. PPC = right posterior parietal cortex; DLPFC = left dorsolateral prefrontal cortex; L = left; R = right.

2.2.1 No Evidence for Enhanced VWM Capacity and Precision by tDCS

Table 2.1 lists the descriptive statistics for tDCS effects on capacity and precision relative to sham. Wang et al. (2019) reported selective effects of tDCS relative to sham stimulation over the PPC, but not the DLPFC, on VWM capacity, but not precision. To test whether these effects can be replicated in our study, like Wang et al. (2019), we computed the differences in performance between the stimulation and the sham condition for each participant and set size. Using these difference scores as dependent variable, we then ran analyses of variance (ANOVAs) with the two within-subjects factors set size (2, 4, 6) and stimulation region (PPC and DLPFC) for each capacity and precision. Bayes factors (BFs) using the default priors (Cauchy distribution with $r = 0.5$) and Monte Carlo setting (iterations = 10,000) were calculated to evaluate the strength of evidence for the absence or presence of effects (Ly et al., 2016; Rouder et al., 2012). BF_{10} refers to the evidence in favour of the alternative hypothesis that capacity changes relative to sham not equal to zero, against the null effect that capacity changes equal to zero.

Table 2.1

*Descriptive Statistics of Performance Changes Relative to Sham (*N = *48)*

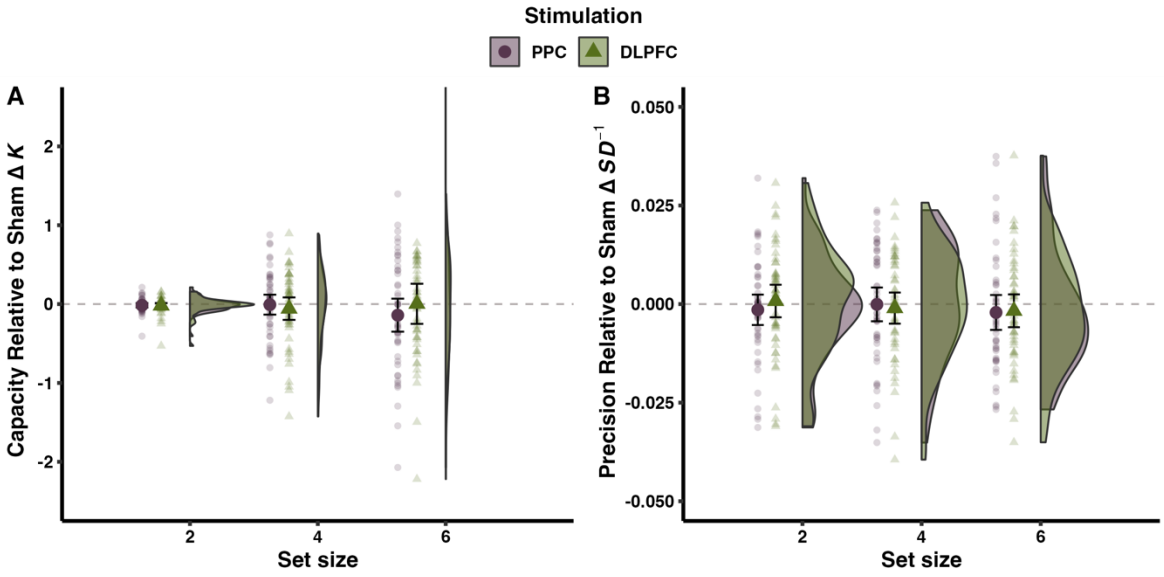
Variable	PPC			DLPFC		
	M	*SD*	*CI*	*M*	*SD*	*CI*
Capacity (Δ*K*)						
Set size 2	-0.01	0.09	0.03	-0.02	0.11	0.03
Set size 4	-0.01	0.43	0.13	-0.06	0.49	0.14
Set size 6	-0.14	0.72	0.21	< 0.01	0.88	0.25
Precision (ΔSD^{-1})						
Set size 2	< -0.01	0.01	< 0.01	< 0.01	0.01	< 0.01
Set size 4	< -0.01	0.01	< 0.01	< -0.01	0.01	< 0.01
Set size 6	< -0.01	0.01	< 0.01	< -0.01	0.01	< 0.01

Note. Capacity ranges from 0 to the set size; precision ranges from 0 to ∞. M = mean; SD = standard deviation; CI = 95% confidence interval of the mean value. PPC = right posterior parietal cortex; DLPFC = left dorsolateral prefrontal cortex.

As shown in Figure 2.2, we found no evidence for tDCS-induced changes in capacity or precision. For capacity, in contrast to Wang et al. (2019), we found no significant main effects of stimulation region, $F(1, 47) = 0.30$, $p = .584$, $\eta_G^2 < .01$, $\eta_p^2 = .01$, $BF_{10} = 1/21.30 \pm 1.05\%$, and set size, $F(2, 94) = 0.17$, $p = .763$, $\eta_G^2 < .01$, $\eta_p^2 < .01$, $BF_{10} = 1/6.86 \pm 1.38\%$, and no interaction between stimulation region and set size, $F(2, 94) = 1.48$, $p = .233$, $\eta_G^2 = .01$, $\eta_p^2 = .03$, $BF_{10} = 1/6.46 \pm 1.85\%$. Notably, the absence of these effects was supported by substantial to strong Bayesian evidence. If anything, although non-significant, PPC stimulation even induced marginal decreases in capacity relative to sham, opposite to the observed improvements in the original study.

Figure 2.2

Transcranial Direct Current Stimulation Effects on Visual Working Memory Capacity and Precision Relative to Sham Stimulation



Note. Panel A: Changes in capacity relative to sham. Panel B: Changes in precision relative to sham. Left: Opaque symbols indicate group mean values, with the error bars representing 95% confidence intervals. Transparent symbols indicate individual data points. Right: Density distributions of the data for both groups. PPC = right posterior parietal cortex. DLPFC = left dorsolateral prefrontal cortex.

To directly replicate Wang et al.'s (2019) analysis on their main findings regarding capacity, we further ran one-sample t-tests against zero for each region of stimulation and set size (Table 2.2). Based on the pattern of results from Wang et al. (2019) that "enhanced memory capacity via tDCS was specific to PPC (not DLPFC) stimulation" (p. 533), we ran one-sided t-tests for PPC stimulation condition while two-tailed t-tests for DLPFC stimulation condition. Data at set size 2 violated the assumptions, therefore, the equivalent non-parametric one-sample Wilcoxon signed rank test were run for the condition of set size 2. BFs were calculated using default Monte Carlo setting (iterations = 10,000) and an informative prior based the reported significant effect sizes in Wang et al. (2019). Again,

based on the pattern of results from the original study, we used Bayes factors (BF_{+0} and BF_{10}) to quantify the strength of evidence for PPC and DLPFC stimulation, respectively. BF_{+0} refers to the evidence in favour of the alternative hypothesis that capacity changes relative to sham are greater than zero, against the null effect that capacity changes equal to zero. For the PPC condition, we used the reported effect size (d = 1.03) as informative prior. For the DLPFC condition, only the range of effect sizes (ds = 0.078–0.409) was reported in the original study. Thus, we used the biggest effect size value as the informative prior (Cauchy distribution with r = 0.409).

We observed neither PPC stimulation nor DLPFC stimulation effects compared to sham at any set size. Critically, in contrast to Wang et al.'s main finding that PPC stimulation increased relative capacity changes compared to zero at set size 6 with a large effect size (Cohen's d = 1.03), this effect was absent in our data, $t(47) = -1.35$, $p = .909$, Cohen's d = -0.20, 95% confidence interval (CI) = [-0.32, ∞), which was supported by strong Bayesian evidence, $BF_{+0} = 1/20.20 \pm < 0.01\%$. However, again, if anything, PPC stimulation tended toward *decreasing* capacity at set size 6. Furthermore, Wang et al. (2019) reported that the increase in capacity induced by PPC stimulation was significantly higher than that by DLPFC stimulation at set size 6, with a medium to large effect size (Cohen's d = 0.71)[1]. Different to the original study, we found no credible evidence of a significant difference in relative capacity changes between the effects of the two stimulation sites at set size 6, $t(47) = -1.03$, $p = .846$, Cohen's d = -0.15, 95% CI = [-0.38, ∞). The absence of this difference was supported by strong Bayesian evidence, $BF_{+0} = 1/12.14 \pm < 0.01\%$.

[1] Although this was not mentioned in analysis plan of the pre-registration, it is added in the actual analysis to fully test Hypothesis 1.

Table 2.2

One-Sample T-Tests for Capacity Changes Relative to Sham Against Zero

Stimulation	Set size	Statistical value	*P*	Effect sizes	BF	Error (%)
PPC	2	476.00	.874	0.17	1/5.65[a]	< 0.01
	4	-0.11	.546	-0.02	1/2.65[a]	< 0.01
	6	-1.35	.909	-0.20	1/20.20[a]	< 0.01
DLPFC	2	488.00	.424	0.12	1/2.13[b]	0.01
	4	-0.82	.419	-0.12	1/2.95[b]	0.01
	6	0.03	.976	< 0.01	1/3.92[b]	0.01

Note. df = 47. Bonferroni-corrected threshold of 0.0083. PPC: Right posterior parietal cortex; DLPFC: Left dorsolateral prefrontal cortex. [a] BF_{+0}: in favour of the alternative hypothesis that capacity changes relative to sham are greater than zero, against the null effect that capacity changes equal to zero; [b] BF_{10}: in favour of the alternative hypothesis that capacity changes relative to sham not equal to zero, against the null effect that capacity changes equal to zero.

Regarding tDCS effects on precision, we did not observe significant effects for any stimulation region or set size, consistent with the original results and our Hypothesis 2. A two-way repeated measures ANOVA showed no significant main effects of stimulation region, $F(1,47) = 0.15$, $p = .700$, $\eta_G^2 < .01$, $\eta_p^2 < .01$, $BF_{10} = 1/17.46 \pm 0.75\%$, or set size, $F(2,94) = 0.33$, $p = .718$, $\eta_G^2 < .01$, $\eta_p^2 = .01$, $BF_{10} = 1/7.17 \pm 1.18\%$. There was also no significant interaction effect, $F(2,94) = 0.85$, $p = .432$, $\eta_G^2 < .01$, $\eta_p^2 = .02$, $BF_{10} = 1/10.57 \pm 2.72\%$. The absence of these effects was supported by substantial to strong Bayesian evidence.

2.2.2 Summary

Like Wang et al. (2019), we observed no effects of tDCS on VWM capacity and precision induced by DLPFC stimulation. However, in contrast to the original study, PPC stimulation did not significantly enhance capacity selectively at set size 6, and also not at any other set size. The absence of tDCS effects was largely supported by substantial to strong Bayesian evidence.

2.3 Discussion

This preregistered study aimed to replicate the benefits of non-invasive stimulation on VWM that were recently reported by Wang et al. (2019). Wang et al. found that tDCS over the PPC, but not the DLPF, selectively improved capacity, but not precision, when VWM load was high (set size 6). While accounting for the methodological issues from the original study, we found no credible evidence of such selective effects. Stimulation over the right PPC improved neither capacity nor precision of representations in VWM performance. In contrast, if anything, our results indicated that when VWM load is high, capacity slightly, although not significantly, decreased after right PPC stimulation compared to the sham condition. Therefore, our Hypothesis 1 (improvements of capacity) was rejected, while Hypothesis 2 (no improvements in precision) was confirmed.

Our findings are consistent with recent studies that focused on other stimulation sites and used different paradigms, suggesting that the lack of an effect in our study is not specific to the present montage or paradigm. For instance, the absence of benefits of anodal PPC-tDCS in the present study is consistent with other recent research (Dumont et al., 2021; Robison et al., 2017) that used other types of montage (left PPC) and/or VWM paradigms (change-detection). Like Dumont et al. (2021), our findings were largely supported by substantial to strong Bayesian evidence, challenging the existing positive effects (Hsu et al., 2014; Tseng et al., 2012; Wang et al., 2019). Our results are also consistent with the results from Nikolin et al. (2018) that anodal DLPFC stimulation does not alter VWM performance, even using a different montage (i.e., cathode on the contralateral DLPFC) and VWM paradigm (n-back). Nikolin et al. further systematically tested different tDCS dosages, resulting in consistent null effects. Our findings add critically to the existing literature by demonstrating that it is unlikely that anodal DLPFC-tDCS produces improvements in VWM performance in healthy participants.

Our findings highlight the importance of replication studies investigating robust tDCS effects. Wang et al.'s findings would have far-reaching theoretical and practical implications for the scientific understanding of both tDCS effects and VWM processes. Our findings mirror the replication crisis that replication effect sizes are typically only a quarter or halve of the magnitude of original effects (Klein et al., 2018; Open Science Collaboration, 2015). Our replication attempt—which disconfirmed tDCS-induced increases in VWM capacity by stimulating right PPC—can serve as a starting point for more replications to further test the veracity of such tDCS effects (Brandt et al., 2014).

In addition to inevitable differences like the characteristics of recruited participants and lab environment, our study used slightly different stimuli than the original study. Hence, our conceptual disconfirmation of the tDCS-PPC effects also questions the generalization of tDCS benefits. Similarly, Robison et al. (2017) recently failed to conceptually replicate the positive tDCS effects over PPC and DLPFC that were reported by Li et al. (2017), using a design similar but not identical to the original one. Altogether, these two examples of conceptual replication attempts are likely only the tip of the iceberg of a lack of replicability and generalization in tDCS research. Importantly though, any single replication study does not rule out that tDCS may benefit cognitive performance in general (Hedges & Schauer, 2019; Maxwell et al., 2015).Therefore, more replications using the same tDCS setups are needed to advance this promising area of research.

2.3.1 Conclusion

We did not observe any benefits of single-session, anodal parietal or prefrontal tDCS on VWM capacity and precision. In particular, we failed to replicate the selective, large effect of parietal tDCS in increasing VWM capacity at a big set size that was reported by Wang et al. (2019). Indeed, the empirical evidence from our study consistently favoured the absence of any cognitive benefits after tDCS regardless of stimulation site and task difficulty.

Considering the complexity of tDCS parameters and setups, our null findings highlight the critical importance of conducting replications for building a robust and informative body of evidence on the effectiveness of non-invasive brain stimulation on cognitive performance.

2.4 Method

This experiment and our hypotheses were pre-registered on the Open Science Framework (https://osf.io/n9fkp). A pilot study served to test the feasibility of the study, the safety of current tDCS setup and the feasible workflow of the analysis. These pilot data were not included in the analyses of the present study. The study was approved by the University of Sheffield Research Ethics Committee.

2.4.1 Participants

A total of 48 healthy young adults were recruited (31 females, 17 males, all right-handed, 22.65 $\pm$ 4.34 years old, range 18 – 33 years). All participants were retained for analysis. We chose this sample size for two reasons. First, although Wang et al. (2019) reported a large effect size of Cohen's d = 1.03, yielding a (post-hoc) power of 1 - β = 0.98 for their included sample of 18 participants for analysis, simulations have shown that effect sizes are often overestimated for such small samples (Halsey et al., 2015). Therefore, we ran an *a priori* power analysis based on a more conservative medium effect size of Cohen's d = 0.50, a power of 1 - β = 0.90 and an α-level of 0.05, resulting in a minimum sample size of 44 (G*power 3.1; Faul et al., 2007). Second, fully counterbalancing the stimulation conditions (i.e., DLPFC, PPC and PPC/DLPFC sham) across three sessions results in 12 possible combinations; therefore, we recruited 48 healthy participants, which is a multiple of 12 and more than twice larger than the sample size in the original study.

The inclusion criteria were similar to those in Wang et al. (2019): all participants had normal or corrected-to-normal vision, no metallic implant, and no history of any neurological or psychiatric illness. In addition, in the present study, only participants who were proficient in

English and educated to A-level or higher were included. Furthermore, we excluded participants who self-reported that they underwent neurostimulation within the past week, were on medication with known cognitive side-effects, in particular on memory and attention, or were currently using recreational drugs (e.g., cannabis, cocaine, or methamphetamines), or were pregnant. Participants were recruited through university volunteer systems, social media (e.g., Facebook), display of flyers, and word-of-mouth. Participants were compensated with £15 or £5 and course credits.

2.4.2 Procedure

This lab-based study used a within-subjects, randomized and single-blinded design (Figure 2.1 A). All participants came to the lab for three sessions. Each session lasted about 1 hour, with an intersession-interval of at least 48 hours to allow for any possible after-effects of tDCS to return to baseline ('wash out'). Upon arrival at their first session, participants gave their written informed consent for their participation and completed a questionnaire on demographic information (age, sex, main language, handedness, and education level). At each session, participants first received the tDCS. Next, they completed short post-stimulation ratings (Appendix A) on their current pain, attention, and fatigue levels, followed by a tDCS adverse-effects questionnaire (Appendix B). Next, participants completed a computerised VWM task. In addition, to measure expectation effects, at the end of the third and final session, participants were asked to guess whether they had received active or sham stimulation at each session (Appendix C)[2].

[2] Note that these data are not reported in this article, but results from analyses of these data can be found in the Supplementary Materials on OSF (also see Appendix F). Overall, the current tDCS setup did not lead to any severe adverse effect, which indicates the safety of the montage and paradigm. The pattern of results confirms that sham stimulation provided a good level of blindness, thereby preventing placebo effects.

2.4.3 TDCS Setup

A battery-driven TCT Research tDCS 1ch device was used to deliver direct current via two saline-solution-soaked sponge electrodes (electrodes size: 5 x 7 cm^2; https://trans-cranial.com). Figure 2.1 B illustrates the current density model for the two active stimulation conditions simulated by MATLAB-based COMETS toolbox (Jung et al., 2013). Identical to the original study, the anodal electrode was placed at the target stimulation brain regions, that is, either the left DLPFC or the right PPC, while the cathodal electrode was placed on the contralateral cheek. In each session, participants received one of the three types of stimulation (active DLPFC, active PPC, or sham) for 15 minutes. The order of the three stimulation conditions was counterbalanced across participants.

For half of the participants, sham stimulation was on the left DLPFC, and for the other half sham stimulation was on the right PPC. In the active stimulations, the tDCS current linearly reached 2mA within the first 30 s (20 s in the original study) and then remained stable until the last 30 s when the current gradually decreased until tDCS was turned off. The sham stimulation followed the same procedure, except that the tDCS was pre-set to turn off after 30 s. This procedure produces the expected typical 'tingling' sensation on the scalp and, thus, provides an effective control condition to minimise placebo effects. Regardless of the stimulation type, identical beeping sounds were generated at the beginning and the end of stimulation.

To locate the stimulation regions, individuals' head sizes (see Appendix D) were measured using a soft measure tape and wax pencil. EZ-EEG (http://clinicalresearcher.org/eeg/; Beam et al., 2009) was used to locate the left DLPFC (F3) accurately and efficiently from the nasion-inion, tragus-tragus and circumference lengths. The right PPC (P4) was located at the symmetrical point of left DLPFC (F3), centring at Cz, according to the international 10-20 system (Klem et al., 1999).

2.4.4 VWM Task

Post-tDCS VWM performance was measured using a continuous orientation-reproduction task (Figure 2.1 C). Each trial began with a fixation cross displayed centrally for 1000 ms. Next, an array of two, four or six randomly orientated (0-360°) isosceles triangles that were arranged in a circular manner appeared on the screen for 200 ms. After a 1000 ms blank screen, one of the displayed triangles was randomly selected as the target stimulus and presented in a random orientation. Participants were instructed to reproduce the original orientation using the mouse. Reaction time and recall errors (i.e., angular distance between the targeted orientation and reported orientation) were recorded. Note that we did not manipulate the blank interval duration as Wang et al. (2019) did, as they did not observe any effect of tDCS in their short-interval (100 ms, labelled sensory memory) condition.
During each session, participants first completed 30 practice trials (10 practice trials per set size, intermixed) with feedback. For this feedback, the original stimulus array was shown, overlaid by the reproduced angle in green for recall errors smaller than 15 degrees, in orange for errors between 15 and 45 degrees, and in red for errors larger than 45 degrees. Next, participants completed 360 trials without feedback (120 trials per set size, intermixed). The VWM task was executed with Tatool Web (www.tatool-web.com, von Bastian et al., 2013).

2.4.5 Model Fitting

First, we calculated recall errors for each set size and stimulation condition and fitted computational models to these recall errors. Specifically, we compared fits of the Standard Mixture Model (SMM; Zhang & Luck, 2008) and Swap Model (SM; Bays & Husain, 2008) using the MATLAB MemToolbox (Suchow et al., 2013). Following Wang et al.'s (2019) procedure, we computed the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) to indicate relative fits of the models to the data separately for each participant, set size, and stimulation condition. As shown in Table 2.3, overall, both the

AIC and BIC favoured the SMM over the SM (60.65% and 87.04%, respectively). The AIC and BIC values of each participant in all conditions are listed in Appendix E, Table E1 and Table E2, respectively. Then, we used winning model (i.e., SMM) to estimate the capacity and precision parameters using maximum likelihood estimation. The SMM assumes a mixture of two components: a uniform distribution and a circular von Mises distribution. The height of the uniform distribution (g) represents random guess responses, which is used to calculate the probability of retrieving the target stimulus ($Pm = 1\text{-}g$). Capacity (K) is the product of Pm and the set size ($K = Pm*N$). The standard deviation (SD) of the von Mises distribution represents the precision of the retrieved representation of the target stimulus. A smaller SD is interpreted as higher precision. The precision is denoted by the inverse of the SD (SD^{-1}). Following Wang et al.'s (2019) procedure, normalised values (ΔK and ΔSD^{-1}) were used for testing the hypotheses. Normalised values were computed for statistical analyses by subtracting capacity K and precision SD^{-1} in the sham condition from those in the active PPC and DLPFC conditions for each set size and participant. All statistical analyses were performed with R Statistical software (v4.1.3; R Core Team, 2022) and R packages rstatix (Kassambara, 2021), ez (Lawrence, 2016), lsr (Navarro, 2015), and BayesFactor (Morey & Rouder, 2021).

Table 2.3

Summary of Model Fits Favoured the Standard Mixture Model Over the Swap Model

Stimulation	Set size	AIC (%)	BIC (%)
Sham	2	83.33	95.83
	4	50.00	85.42
	6	64.58	87.50
PPC	2	85.42	97.92
	4	56.25	81.25
	6	33.33	72.92
DLPFC	2	79.17	95.83
	4	45.83	85.42
	6	47.92	81.25
Overall	–	60.65	87.04

Note. AIC: Akaike information criterion; BIC: Bayesian information criterion.

Chapter 3 – Mechanisms of cognitive change: Training improves the quality but not the quantity of visual working memory representations

Contributions:

Shuangke Jiang (conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, validation, visualization, writing—original draft, writing—review & editing)

Myles Jones (supervision, writing—review & editing)

Claudia von Bastian (conceptualization, funding acquisition, methodology, resources, software, supervision, writing—review & editing)

Research presented in this chapter has been published and is available online at *Journal of Cognition*:

Jiang, S., Jones, M., & von Bastian, C. C. (2023). Mechanisms of cognitive change: Training improves the quality but not the quantity of visual working memory representations. *Journal of Cognition*, *6*(1), 1–20. https://doi.org/10.5334/joc.306

Findings from this chapter have been presented in the following conferences:

Jiang, S., Jones, M., & von Bastian, C. C. (2021). Visual working memory training effects on the quantity and quality of representations [Poster]. Psychonomic Society 62nd Annual Meeting, Online.

Jiang, S., Jones, M., & von Bastian, C. C. (2021). Mechanisms of visual working memory training effects: Capacity and efficiency [Talk]. Virtual Working Memory Symposium, Online.

Jiang, S., Jones, M., & von Bastian, C. C. (2023). Mechanisms of visual working memory training effects: Capacity vs efficiency [Talk]. Experimental Psychology Society (EPS)

Meeting: University College London, London, UK.

A research proposoal based on this chapter has received 2021 J. Frank Yates Student Conference Award:

Jiang, S., Jones, M., & von Bastian, C. C. (2021). *Visual working memory training effects on the quantity and quality of representations* [Poster]. Psychonomic Society 62nd Annual Meeting, Online.

Abstract

As of yet, visual working memory (WM) training has failed to yield consistent cognitive benefits to performance in untrained tasks, despite large improvements in trained tasks. Investigating the mechanisms underlying training effects can help explain these inconsistencies. In this pre-registered, pre-test/post-test online training study, we examined how training affects the quantity and quality of representations in visual WM using continuous-reproduction tasks. $N = 64$ young healthy adults were randomly assigned to an experimental group or an active control group to complete four training sessions of practice in an orientation-reproduction or visual search task, respectively. We observed that, in the trained task, only the quality, but not the quantity, of visual WM representations significantly increased in the experimental group relative to the control group. These improvements did not generalise to untrained stimuli or paradigms. Therefore, our findings suggest that training gains are not driven by enhanced capacity. Instead, gains in the quality of visual WM representations that are tied to specific stimuli and paradigms may reflect enhanced efficiency in using the existing visual WM capacity.

Keywords: visual work memory, quantity and quality, training mechanism, capacity and efficiency

3.1 Introduction

Working memory (WM) is a cognitive system providing temporary access to representations that are needed for complex cognition in the present moment. WM has a limited capacity of around four chunks of information that can be simultaneously maintained at a time (Cowan, 2001). The individual limit of WM capacity is strongly correlated with reasoning (Conway et al., 2003; Engle et al., 1999; Oberauer et al., 2008), executive functions (Miyake et al., 2000), and a range of other cognitive abilities (for a review, see Barrett et al., 2004). Furthermore, neurocognitive disorders such as ADHD (Martinussen et al., 2005) and age-related cognitive declines (Park et al., 2002) often go along with WM impairments.

The central role ascribed to WM in human cognition has motivated research into training interventions aiming to enhance WM capacity and, thereby, potentially also reasoning and other related cognitive abilities (Jaeggi et al., 2008; Klingberg, 2010; Klingberg et al., 2002). WM training typically involves repeated practice on one or more WM tasks over a short period of time, aiming to improve performance in trained and untrained cognitive tasks. The improvements in related yet untrained cognitive abilities are referred to as transfer effects. However, so far, WM training has failed to yield consistent and robust cognitive benefits (Jaeggi et al., 2012; Karbach & Verhaeghen, 2015; Melby-Lervåg et al., 2016; Morrison & Chein, 2011; Shipstead et al., 2012; von Bastian et al., 2022). Although previous research reported large and replicable gains in the trained WM tasks, transfer effects on untrained tasks remain inconsistent and elusive. A focus on the theoretical mechanisms underlying training gains can yield important insights for when and why transfer effects may occur (Redick, 2019; Smid et al., 2020; von Bastian & Oberauer, 2014).

The capacity-efficiency model of cognitive training and transfer effects (von Bastian et al., 2022; von Bastian & Oberauer, 2014) provides a framework for explaining these

inconsistencies in past findings by proposing two, not mutually exclusive, pathways of how training may induce change. One pathway is through expanding cognitive capacity itself. Expanded capacity should generalise to any untrained tasks that draw on the same capacity limit. WM training-induced enhancements of capacity would be reflected by an increased quantity of representations that are simultaneously maintained in WM. These improvements would be expected to yield broad benefits across a range of related cognitive abilities. However, given the lack of broad and robust transfer effects, it is unlikely that training expands working memory capacity (von Bastian et al., 2022).

The other pathway is through enhancing efficiency in using the available capacity. Mechanisms of enhanced efficiency can be broadly grouped into compression and optimisation. Compression is to learn the regularities of information and making use of observed redundancies to reduce the overall cognitive load (Bavelier et al., 2012; Brady et al., 2009). Compression-based efficiency can be *paradigm-specific* through learning the necessary routines and effective strategies for completing an ongoing task. For example, performance can be boosted by strategies such as chunking (e.g., remembering the three digits 8, 1, and 9 as one number 819). In addition, better metacognitive skills, such as improved introspection about self-performance in an ongoing task (Carpenter et al., 2019) could facilitate applying effective task strategies to a different context (Belleville et al., 2014). Compression can also be *stimuli-specific*, for example through gaining a level of perceptual expertise that allows for more efficient coding of the stimuli (Curby & Gauthier, 2007) by increasing the precision of their representations in WM (Scolari et al., 2008). Finally, efficiency can also be enhanced by optimizing attention allocation to different stimuli or task sets (De Simoni & von Bastian, 2018; Zerr et al., 2021). In contrast to the broad benefits that are expected to result from expanding capacity, enhanced efficiency is expected to be useful only in contexts where these efficiency mechanisms can be applied as well.

There is tentative evidence for training-induced enhancements in efficiency. For example, De Simoni and von Bastian (2018) found that the majority of participants reported the acquisition of paradigm-specific strategies during training, including cognitive load-reducing strategies such as remembering only one of two items of a pair in an associative memory task. De Simoni and von Bastian also found that participants improved selectively in remembering which items they have encountered (i.e., item recognition) but not their current context (i.e., item recollection; e.g., the item's location on the screen). De Simoni and von Bastian speculated that these improvements in recognition were possibly due to training-induced acquisition of stimuli-specific expertise by which the precision of the item representations in memory was enhanced (see also Olson et al., 2005), thereby increasing success of retrieval. In the present study, we focus on investigating to what extent the acquisition of paradigm-specific and stimuli-specific expertise transfer to other contexts. Paradigm-specific expertise may lead to better performance in tasks with the same surface structure but different stimuli (e.g., recall the orientation of triangles or the shape of rings). Stimuli-specific expertise may lead to better performance in tasks using the same stimuli but different paradigms (e.g., the orientation of triangles in a recall or recognition task).

To distinguish training effects through capacity from those through efficiency, WM models that differentiate between the quantity and the quality of representations maintained in WM are useful (Alvarez & Cavanagh, 2004; Awh et al., 2007; Fougnie et al., 2010; Olson & Jiang, 2002; Zhang & Luck, 2008). This distinction between the quantity (the number of remembered items) and quality (the precision of these items) has been supported by neural evidence demonstrating a dissociative role of different parietal-occipital subregions. Specifically, the inferior intraparietal sulcus (IPS) has been found to track the number of items at different locations, whereas the superior IPS and lateral occipital complex encoded the precision of the attended items (Todd & Marois, 2004; Xu & Chun, 2006). Furthermore,

WM quantity, but not quality, shows a strong connection with fluid intelligence (Fukuda et al., 2010).

To date, only few existing studies have investigated training-induced changes specifically in the quantity and quality of visual WM representations (Buschkuehl et al., 2017; Moriya, 2019; Ovalle Fresa & Rothen, 2019; K. Wang & Qian, 2021), and most of the existing studies offer only crude estimates of changes in quantity and quality of visual WM representations. For example, Moriya (2019) distinguished between the quantity and quality of visual WM representations using two versions of change-detection tasks, in which participants were asked to compare two memory arrays and detect whether they are identical or not. Moriya's tasks varied in the extent to which the deviating stimulus differed from the memoranda: 45° in the quantity version, vs. 5° in the quality versions of the task. Moriya found significant effects of training for both the quantity and the quality versions of the change-detection tasks, but with asymmetric patterns of transfer: whereas training of the quantity task led to strong transfer to the quality version, training of the quality task yielded only weak transfer to the quantity version. However, performance changes in quantity and quality of visual WM were estimated by the same parameter (i.e., Pashler's *k*, 1988) and, thus, conclusion about the two types of visual WM representations could only be drawn indirectly. Similarly, Wang and Qian (2021) reported training effects of the same change-detection paradigm on the quantity of visual WM representations as well as transfer effects on the quality of visual WM representations, measured by a trained orientation-changed detection task and an untrained orientation continuous-reproduction task, respectively. However, Wang and Qian measured the quality of visual WM representations using the overall recall error which mixes quantity and quality of visual WM representations.

Buschkuehl et al. (2017) trained participants in one of two variants of a colour-change detection task. Different to the Moriya (2019) and Wang and Qian (2021), Buschkuehl et al.

(2017) used transfer tasks that allowed for estimating the precision of WM representations. Despite substantial training improvements in change-detection performance, the authors found no transfer of these improvements to the precision of representations of colour and spatial features. However, like the other existing studies, Buschkuehl et al. did not use training tasks that allowed for distinguishing changes in the quantity from changes in the quality.

Continuous-reproduction tasks, in which participants were asked to memorise and later reproduce features of stimuli on continuous dimensions (e.g., orientation or shape), probe high-resolution contents of visual WM directly (Gorgoraptis et al., 2011; Ma et al., 2014; Wilken & Ma, 2004; Zhang & Luck, 2008). The dependent variable, that is, the difference between the original and the reproduced feature can then be used to estimate the quantity (or capacity) and quality (or precision) of visual WM representations using computational models such as the standard mixture model (SMM; Zhang & Luck, 2008). The SMM assumes a mixture of two components: a uniform distribution representing random guesses, and the standard deviation of a von Mises distribution (a circular normal distribution) around the target, representing that remembered information is remembered with a certain degree of precision. For example, Ovalle Fresa and Rothen (2019) used a continuous colour-reproduction task to train participants in visual long-term memory and applied the SMM. After six training sessions over the course of three days, participants' precision in both visual long-term memory and visual short-term memory improved significantly. However, Ovalle Fresa and Rothen focused on long-term memory training, and did not assess transfer to substantially different stimuli and paradigms. Therefore, taken together, it remains unclear whether WM training effects are due to changes in quantity or quality of visual WM representations, and to what extent these changes are specific to the trained paradigm or stimuli. The present study fills this gap.

3.1.1 Present Study

This pre-registered study investigated the mechanisms of training gains by distinguishing between quantity and quality of representations in visual WM. We administered a continuous orientation-reproduction training task for four training sessions. To examine the capacity-efficiency model and its proposed mechanisms of training and transfer effects, we used the SMM (Zhang & Luck, 2008) to estimate changes in the quantity (i.e., capacity) and the quality (i.e., precision) of visual WM representations from pre-test to post-test and during training. Furthermore, we assessed transfer to two untrained tasks (shape reproduction and orientation-change detection). All effects in the experimental training group were evaluated relative to an active control group practising visual search, which has been shown to demand only minimal visual WM (Wolfe & Horowitz, 1998; Woodman et al., 2001). Including an active control group controls for placebo effects and expectancy effects (Foroughi et al., 2016; Simons et al., 2016; von Bastian & Oberauer, 2014).

Table 3.1

Hypotheses

Mechanism	Trained task (ORT)		Untrained stimuli (SRT)		Untrained paradigm (ODT)
	Quantity	Quality	Quantity	Quality	Performance
Capacity	Increase	-	Increase	-	Increase
Efficiency					
Paradigm-specific expertise	-	Increase	-	Increase	No change
Stimulus-specific expertise	-	Increase	-	No change	Increase

Note. All performance changes are relative to changes observed in the active control group. Hyphens (-) refer to possible concurrent improvements. ORT: orientation-reproduction task; SRT: shape-reproduction task; ODT: orientation-change detection task.

Our pre-registered hypotheses[3] (https://osf.io/mk8fa) are summarised in Table 3.1 and stated as follows:

(1) If visual WM training-induced performance gains reflect increased visual WM capacity, the experimental group will show larger gains in the quantity of visual WM representations in the trained task (orientation reproduction) and in the untrained structurally similar task (shape reproduction) as well as improved performance in the untrained structurally different task (orientation-change detection) above and beyond any improvements observed in the active control group.

(2) If visual WM training-induced performance gains reflect acquisition of paradigm-specific expertise, the experimental group will show larger gains than the active control group in the quality of visual WM representations in the trained task (orientation reproduction) and

[3] Hypotheses 2 and 3 were slightly reworded (while keeping the identical meaning) to facilitate understanding. Furthermore, paradigm-specific expertise was labelled task-specific expertise in the pre-registration.

in the untrained, structurally similar task (shape reproduction), but no performance gains in the untrained, structurally different task (orientation-change detection).

If, in addition to these improvements in quality, we would observe training-specific gains in the quantity of visual WM representations in both reproduction tasks, it would suggest that paradigm-specific expertise (e.g., strategies) hindered transfer to the structurally different task. If those training-induced quantity gains were observed in just one of the reproduction tasks, it would suggest that training-induced performance gains were primarily driven by gains in paradigm-specific expertise.

(3) If visual WM training-induced performance gains reflect acquisition of stimuli-specific expertise, the experimental group will show larger gains than the active control group in the quality of visual WM representations in the trained task (orientation reproduction) only, without any improvements in the quality of visual WM representations in the untrained, structurally similar task (shape reproduction). If this increased quality of visual WM representations is observed in the trained task but not in the shape-reproduction task, alongside increased visual WM performance in the orientation-change detection task, it would suggest that stimuli-specific expertise transferred across paradigms.

Importantly, these hypotheses were not mutually exclusive as increases in visual WM capacity and acquisition of stimuli-specific and task-specific expertise may co-occur (von Bastian et al., 2023).

3.2 Method

This online training study used a pre-test-post-test, randomised-controlled design. Participants who had completed the pre-test were randomly assigned to the experimental group or the active control group where they practised an orientation-reproduction task or a visual search task, respectively, for four training sessions. Most participants (87% of the final sample included in the analysis) completed the four training sessions over four consecutive

days. Participants who missed a day were retained until they completed their sessions or withdrew. To ensure that participants could maximally complete one training session per day, they received a website link for the next day's session only after they had completed the previous session. After the training sessions, participants completed the post-test. The pre-test and post-test were designed to assess training effects on performance in the orientation-reproduction task and visual search task, as well as transfer effects to a shape-reproduction task and an orientation-change detection task.

This experiment and its hypotheses were pre-registered on the Open Science Framework (https://osf.io/mk8fa). Pilot data from six participants were collected before the pre-registration. The pilot study served to test the feasibility of the study and the compatibility between the recruitment platform Prolific (https://www.prolific.co) and the experiment software Tatool Web (www.tatool-web.com, von Bastian et al., 2013). As the pilot study was successful with no further changes to the study materials, the pilot data were included in the current study. The study was approved by the University of Sheffield Research Ethics Committee.

3.2.1 Participants

The target sample size was 100 participants at post-test. An *a priori* power analysis assuming a small to medium within-between interaction effect size (Cohen's $f = 0.15$) and power of $1\text{-}ß = 0.80$ suggested a sample size of $N = 90$, which we increased by 10 participants to account for possible dropouts. We recruited 108 healthy participants, aged from 18 to 35, to take part in a study on "Cognitive training" that was advertised on Prolific. We pre-screened participants by customising the allow list according to our pre-registered inclusion and exclusion criteria. After signing up for the study, participants gave online consent to taking part in the study by clicking a button. All participants who met the inclusion criteria and completed the study received £17.40. Before the start of recruitment, a list of

group assignments was randomly generated on GraphPad (https://www.graphpad.com/quickcalcs/randomize2/). Following this pre-generated list, participants who completed the pre-test were randomly assigned to either an experimental group or an active control group. Participants were blind to the group condition.

The flow chart in Figure 3.1 illustrates participant recruitment, attrition, and retention. Eight participants (four from each group) dropped out, without giving a specific reason, after completing the pre-test. We replaced these eight participants who dropped out, so that we reached the target sample size of $N = 100$ participants who completed the post-test. After concluding data collection, data from 36 participants were excluded from analysis. Data from two participants in the experimental group were partially missing due to technical issues and, therefore, these data were excluded. In addition, although we instructed them otherwise, we noticed that some participants completed some sessions (pre-test, post-test or training) multiple times. We excluded all participants (11 per training group) for whom the number of additional trials exceeded 10% for any task (12 trials per task). Furthermore, seven participants from the experimental group and five from the active control group were excluded according to pre-registered criteria using reaction times (RT) and omission errors designed to identify participants who did not follow instructions in an online experiment setting[4]. Of the remaining 64 participants included in the analysis, 30 were in the experimental group and 34 were in the control group. Sensitivity analyses which included all these 12 participants who were excluded due to pre-registered criteria showed similar patterns of results and, thus, led to the same conclusions. Table 3.2 lists the participants'

[4] Participants were excluded with any of the following three data patterns: RT of less than 1500 ms in 1/3 of the trials in the orientation-reproduction task and in the shape-reproduction task; RT of less than 300 ms in 1/3 of the trials in the orientation-change detection task and in the visual search task; omission errors in 1/3 of the trials in the visual search task.

demographics. Overall, the groups were comparable regarding their gender and age, but the evidence for the absence of group differences was ambiguous.

Figure 3.1

Participant Flow Chart

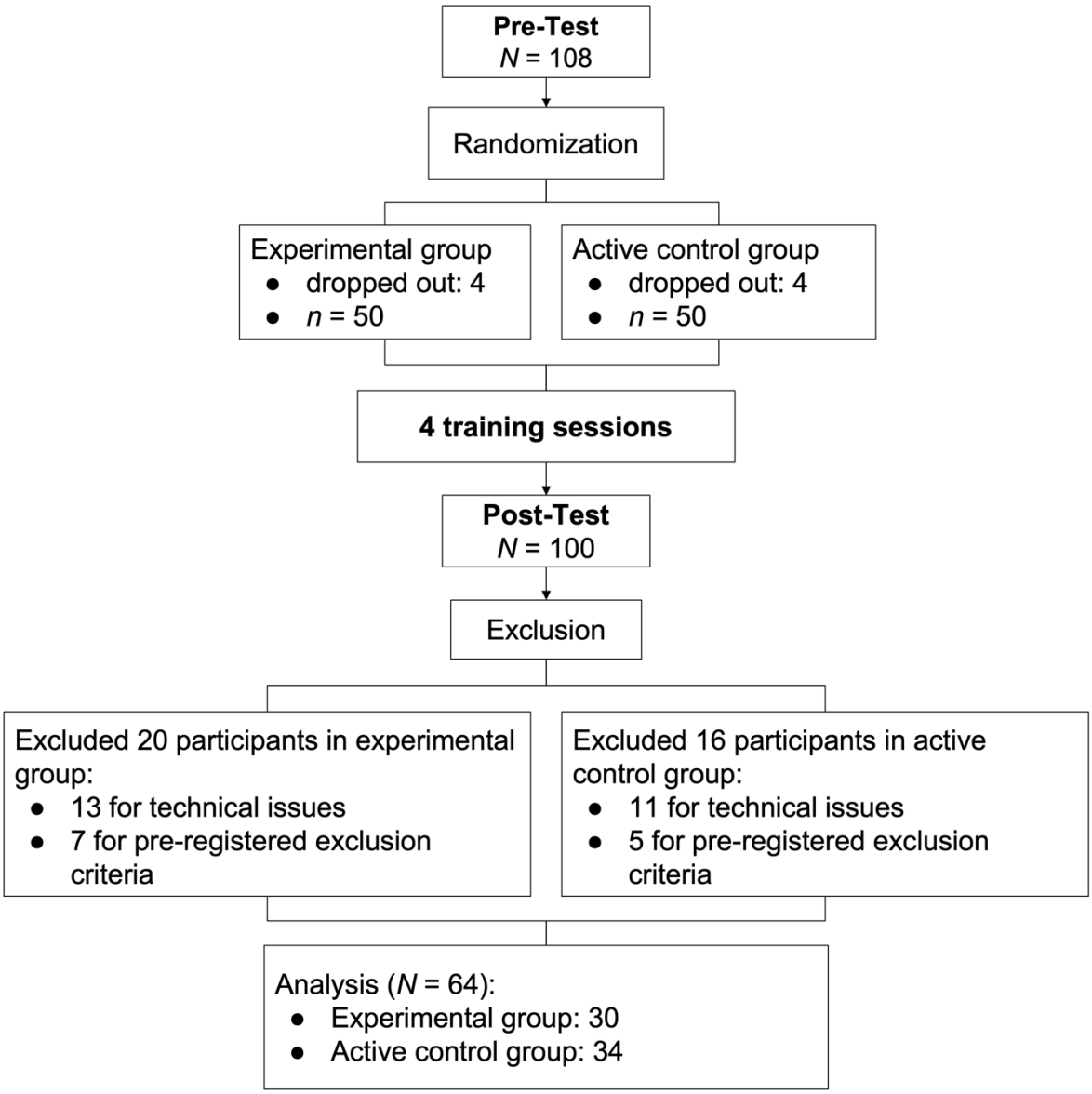


Table 3.2

Participant Demographics as a Function of Groups

Measure	Group		Comparison		
	Experimental	Active Control	Statistical Value	*p*	$BF_{10} \pm$ error %
Group size: *n*	30	34			
Gender: female/male/non-binary	8/22/0	17/17/0	2.73	.098	3.40 ± 0.00
Age: *M* (*SD*)	22.73 (3.92)	21.94 (2.52)	0.33	.745	1/2.62 ± 0.00

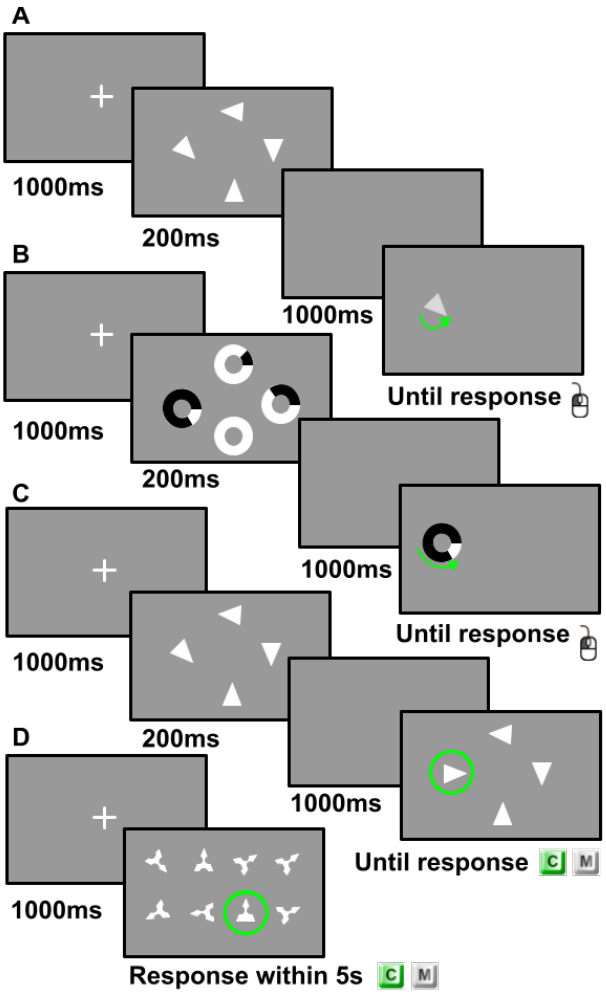
Note. Gender differences were tested with a chi-squared test and age differences with Yuen's t-test.

3.2.2 Materials

Figure 3.2 illustrates the training and transfer tasks. In pre-test and post-test, each experimental task comprised 20 practice trials and 120 testing trials with a set size of the stimulus array of 4 items in the visual WM tasks, and 16 items in the visual search task. The order of representing different experimental tasks was random. Pre-test and post-test took approximately 40 min each. Participants underwent four training sessions. Each training session consisted of 360 trials, with 120 trials per set size (2, 4, and 6 in the orientation-reproduction task, and 8, 16, and 24 in the visual search task). Set sizes were intermixed within each session. Each training session lasted approximately 30 min.

Figure 3.2

Training and Transfer Tasks



Note. Panel A: Orientation-reproduction task at set size four. Panel B: Shape-reproduction task at set size 4. Panel C: Orientation-change detection task at set size 4 in change condition. Panel D: Visual search task at set size eight in the change condition.

3.2.2.1 Orientation-Reproduction Task

Each trial began with a fixation cross displayed centrally for 1000 ms. Next, an array of randomly orientated (0-360°) isosceles triangles was arranged in a circular manner and appeared on the screen for 200 ms, followed by a 1000 ms blank screen. Then, one of the displayed triangles was randomly selected as the target stimulus and presented in a random

orientation. Participants were instructed to reproduce the original orientation by rotating the triangle with the computer mouse and click the left mouse-button to record their response.

We measured recall errors, that is, the difference in degrees between the reproduced orientation and the target orientation, ranging from -π to π, to estimate capacity and efficiency parameters by fitting the SMM (Zhang & Luck, 2008) using the MemToolbox (Suchow et al., 2013)[5]. The SMM consists of two components, a von Mise distribution approximating a circular normal distribution, and a uniform distribution:

$$P(x) = (1 - g)\frac{1}{2\pi I_0(\kappa)}e^{\kappa \cdot cos(x)} + g\frac{1}{2\pi}, \quad (1)$$

where x is the response, g is the proportion of random guess responses, κ is the concentration parameter of the von Mises distribution, and $I_0(\kappa)$ is the modified Bessel function of order 0. The SMM assumes that the target can either be recalled with a certain precision or not at all, leading to random guesses. Therefore, the probability of remembering the target (Pm) is calculated as

$$Pm = 1 - g. \quad (2)$$

The quantity of representations in visual WM, that is, capacity K is computed as the product of the probability of remembering the target and the set size N:

$$K = Pm \times N. \quad (3)$$

Finally, the quality of representations in WM, that is, precision, is computed as the inverse of the standard deviation (SD^{-1}) of the von Mises distribution, which was converted from the concentration parameter κ.

[5] As noted in the pre-registration, we also explored fitting other existing visual WM models, such as, swap model (Bays, 2016), signal discrimination model (Oberauer, 2021), and target confusability competition model (Schurgin et al., 2020), to the data and conducted a series of systematic model comparisons. Overall, the SMM turned out to be the best fitting model for pre-test to post-test changes and, therefore, is reported here. The model comparisons will be reported elsewhere as this would exceed the scope of the present study.

3.2.2.2 Shape-Reproduction Task

Following a central fixation cross for 1000 ms, an array of black ring-shaped objects with varying proportions filled in white were distributed on the screen in a circular manner for 200 ms. After a 1000 ms blank screen, one of the displayed objects was randomly selected as the target stimulus. The target stimulus was presented in black colour with a white bar. Participants were instructed to reproduce the original proportion of the white segment by rotating and left clicking the mouse. As for the orientation-reproduction task, capacity and precision were estimated based on the recall errors using the SMM.

3.2.2.3 Orientation-Change Detection Task

After a fixation cross presented centrally for 1000 ms, an array of randomly orientated (0-360°) isosceles triangles appeared on the screen for 200 ms, followed by a 1000 ms blank screen. Immediately afterwards, a second array was presented until response. In half of the trials, the two arrays were identical. In the other half of the trials, one of the triangles in the second array was randomly selected and presented in a randomly selected, different orientation. Participants were instructed to press the 'C' or 'M' key of the keyboard to respond to a detection of change or match respectively. To measure visual WM capacity, we computed Pashler's *k* (Pashler, 1988) for whole-display tasks using Equation 1 (Pashler, 1988; Rouder et al., 2011):

$$k = \frac{H-FA}{1-FA} \times N, \tag{4}$$

where *H* and *FA* are the hit and false alarm rates and *N* is the display set size.

3.2.2.4 Visual Search Task

On each trial, participants first saw a fixation cross for 1000 ms. Then, an array of isosceles triangles with two or three semi-circular gaps, pointing to random directions, was presented. In half of the trials, all triangles had three gaps. In the other half of the trials, one of the triangles had only two gaps. Participants were instructed to press the 'M' key of the

keyboard within 5 s if all triangles had three gaps, or to press the 'C' key if one of the triangles only had two gaps. The overall accuracy which is calculated by the proportion of correct responses excluding omission errors (no response given after 5000 ms), as well as the mean reaction time (RT) for correct responses were measured and used for analysis.

3.3 Results

In addition to frequentist significance tests (including t-tests and analyses of variance, ANOVAs), Bayes factors (BFs) using the default priors from the BayesFactor package (Cauchy distribution with $r = 0.5$ for ANOVAs, $r = 0.707$ for t-tests; Poisson distribution for chi-square tests with $a = 1$) were calculated to evaluate the strength of evidence for the absence or presence of effects (Ly et al., 2016; Rouder et al., 2012). Table 3.3 lists the categorical labels for describing the strength of evidence adapted from Wetzels and Wagenmakers (2012). As most of the data violated the assumption of normality, we ran robust Yuen t-tests (Yuen, 1974) and report Algina-Keselman-Penfield robust effect sizes, δ_t (Algina et al., 2005). We calculated and report both general effect sizes, η^2_G and partial effect sizes, η^2_p, for ANOVAs to facilitate further use in power analyses and meta-analyses (Lakens, 2013). All statistical analyses were performed with R Statistical software (v4.1.3; R Core Team, 2022). The R packages rstatix (Kassambara, 2021) and ez (Lawrence, 2016) were used for frequentist significance tests. BayesFactor (Morey & Rouder, 2021) and WRS2 (Mair & Wilcox, 2020) were used for Bayesian and robust statistical tests.

Table 3.3

Categorical Labels for Describing the Strength of Bayesian Evidence

Bayes factors		
H_{10}	H_{01}	Categorical labels
>100	<1/100	Decisive
30 to 100	1/100 to 1/30	Very strong
10 to 30	1/30 to 1/10	Strong
3 to 10	1/10 to 1/3	Substantial
1 to 3	1/3 to 1	Ambiguous
1	1	No evidence

Note. Adapted from Wetzels and Wagenmakers (2012). H_{10} = evidence in favour of the alternative hypothesis; H_{01} = evidence in favour of the null hypothesis.

3.3.1 Training Performance

Table 3.4 lists the descriptive statistics for the experimental group and the active control group in the orientation reproduction and visual search tasks during training. To analyse performance changes during training, we ran a repeated-measures ANOVA with the within-subjects factors Time (training session 1 to 4) and Set Size (2, 4, 6).

Table 3.4

Descriptive Statistics of Performance During Training

	Training Session							
	1		2		3		4	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Experimental Group (n = 30)								
Capacity (K)								
Set Size 2	1.88	0.15	1.88	0.19	1.89	0.12	1.89	0.16
Set Size 4	2.76	0.71	2.92	0.74	2.93	0.73	2.87	0.79
Set Size 6	2.97	1.29	3.15	1.30	3.28	1.29	3.25	1.33
Precision (SD^{-1})								
Set Size 2	0.08	0.02	0.09	0.02	0.09	0.02	0.09	0.02
Set Size 4	0.06	0.02	0.07	0.02	0.07	0.02	0.07	0.02
Set Size 6	0.06	0.01	0.06	0.02	0.07	0.02	0.07	0.02
Active Control Group (n = 34)								
Accuracy								
Set Size 8	0.91	0.07	0.92	0.09	0.94	0.05	0.93	0.05
Set Size 16	0.84	0.07	0.84	0.10	0.87	0.08	0.86	0.08
Set Size 24	0.73	0.09	0.73	0.09	0.77	0.09	0.77	0.09
RT (ms)								
Set Size 8	2056	326	1998	305	1890	282	1918	321
Set Size 16	2871	422	2806	401	2712	453	2707	428
Set Size 24	3257	468	3192	452	3101	499	3066	447

Note. Capacity ranges from 0 to the set size; precision ranges from 0 to ∞. RT = mean reaction time.

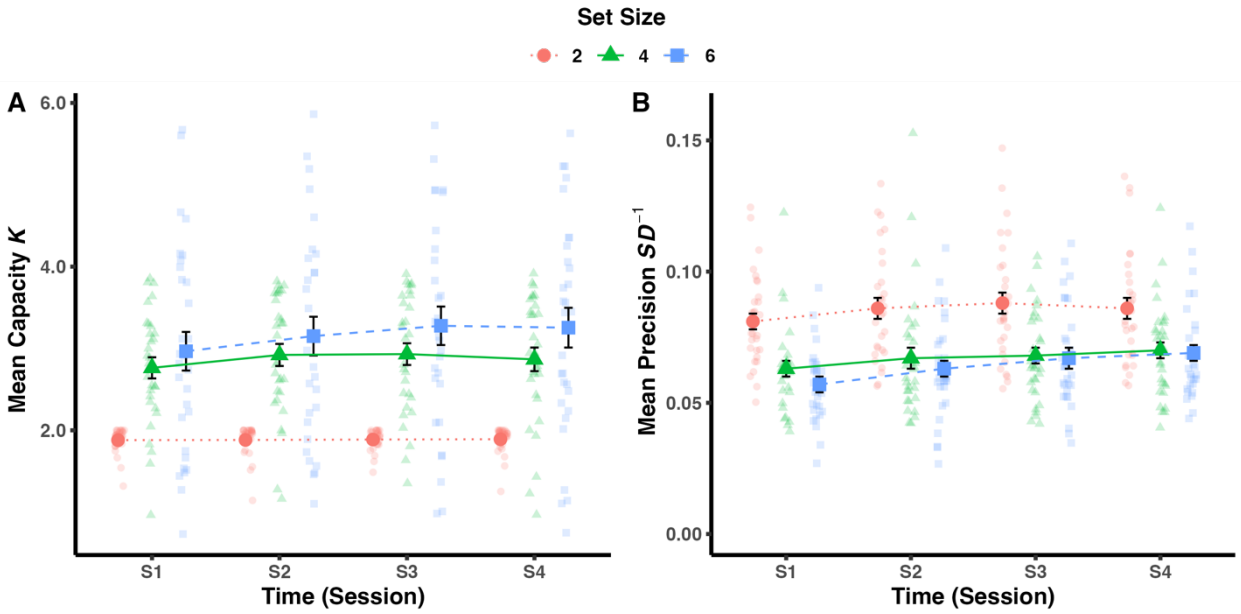
3.3.1.1 Orientation Reproduction

Figure 3.3 illustrates estimates of capacity and precision in the experimental group for each training session at set size levels 2, 4, and 6. There was a significant effect of Set Size on both capacity, $F(2,58) = 39.28$, $p < .001$, $\eta^2_G = .49$, $\eta^2_p = .58$, $BF_{10} > 100 \pm 0.66\%$, and precision, $F(2, 58) = 46.12$, $p = < .001$, $\eta^2_G = .37$, $\eta^2_p = .61$, $BF_{10} > 100 \pm 0.67\%$. We observed a significant effect of Time on precision, $F(3,87) = 6.56$, $p < .001$, $\eta^2_G = .06$, $\eta^2_p = .18$, $BF_{10} = 3.03 \pm 0.59\%$, but not on capacity, $F(3,87) = 2.17$, $p = .097$, $\eta^2_G = .01$, $\eta^2_p = .07$, $BF_{10} = 1/34.51 \pm 0.83\%$. Furthermore, there was no interaction between Time and Set Size for capacity, $F(6, 174) = 1.94$, $p = .078$, $\eta^2_G = .01$, $\eta^2_p = .06$, $BF_{10} = 1/60.01 \pm 2.56\%$, or precision, $F(6, 174) = 0.97$, $p = .444$, $\eta^2_G = .01$, $\eta^2_p = .03$, $BF_{10} = 1/41.37 \pm 1.73\%$. Taken together, we

observed an effect of Set Size on capacity and precision that replicates the set size effect typically observed in visual WM, that is, the bigger the set size, the lower the probability of retrieving an item and its precision. In addition, there was only substantial evidence for significant performance improvement in precision during training.

Figure 3.3

Estimates of Capacity and Precision in the Experimental Group Over Four Training Sessions



Note. Panel A: Estimates of capacity. Panel B: Estimates of precision. Data points with reduced opacity show individual estimates, solid data points represent group means. S1 to S4 = training session 1 to 4.

3.3.1.2 Visual Search

During visual search training, there was a significant effect of Set Size on both accuracy, $F(2,66) = 154.73$, $p < .001$, $\eta^2_G = .68$, $\eta^2_p = .82$, $BF_{10} > 100 \pm 0.81\%$ and mean RTs, $F(2,66) = 330.18$, $p < .001$, $\eta^2_G = .82$, $\eta^2_p = .91$, $BF_{10} > 100 \pm 6.87\%$. We also observed an effect of Time on accuracy, $F(3,99) = 8.50$, $p < .001$, $\eta^2_G = .08$, $\eta^2_p = .20$, with, however, ambiguous Bayesian evidence, $BF_{10} = 1/1.36 \pm 0.85\%$, and mean RTs, $F(3,99) = 7.54$, $p < .001$, $\eta^2_G = .08$, $\eta^2_p = .19$, $BF_{10} = 1/6.99 \pm 0.49\%$. Furthermore, there was no interaction between Time

and Set Size for accuracy, $F(6,198) = 1.32$, $p = .249$, $\eta^2_G = .01$, $\eta^2_p = .04$, $BF_{10} = 1/61.71 \pm 1.73\%$, or mean RTs, $F(6,198) = 0.48$, $p = .823$, $\eta^2_G < .01$, $\eta^2_p = .01$, $BF_{10} = 1/143.26 \pm 2.27\%$. Taken together, we observed the set size effect in visual search with ambiguous evidence for performance improvements during training.

3.3.2 Cognitive Performance Changes from Pre-Test to Post-Test

Table 3.5 lists the descriptive statistics for the training and transfer tasks administered at pre-test and post-test. First, we tested whether the experimental group and the active control group were comparable at baseline based on their pre-test performance using two-tailed *t*-tests (Table 3.6). Next, we assessed training and transfer effects by running two-way mixed ANOVAs separately for each dependent variable, with the within-subjects factor Time (pre-test, post-test), the between-subjects factor Group (experimental group, active control group), and their interaction. Table 3.7 provides an overview of the results of these analyses. For testing our hypotheses, we were primarily interested in the Time x Group interaction.

Table 3.5

Descriptive Statistics of Cognitive Performance at Pre-Test and Post-Test

Variable	Group							
	Experimental				Active Control			
	Pre-test		Post-test		Pre-test		Post-test	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
	Training tasks							
Orientation reproduction								
Capacity (*K*)	2.57	0.77	2.89	0.73	2.35	0.91	2.67	0.69
Precision (SD^{-1})	0.06	0.01	0.07	0.02	0.06	0.02	0.05	0.01
Visual search								
Accuracy	0.76	0.14	0.81	0.13	0.78	0.09	0.86	0.10
RT (ms)	2973	849	2985	633	3101	475	2636	509
	Transfer tasks							
Shape reproduction								
Capacity (*K*)	2.26	0.76	2.10	0.84	2.22	0.68	2.30	0.71
Precision (SD^{-1})	0.05	0.02	0.06	0.03	0.04	0.02	0.04	0.03
Orientation-Change detection								
Capacity (Pashler's *k*)	2.09	1.12	2.37	0.70	2.05	0.82	2.01	0.72

Note. Pashler's *k* can range from 0 to set size. RT = mean reaction time.

3.3.2.1 Baseline Comparisons

There were no significant group differences, though the evidence was ambiguous for capacity in the orientation-reproduction task and precision in the shape-reproduction task, with participants in the active control group showing numerically slightly lower capacity in the former task and lower precision in the latter task at pre-test than participants in the experimental group.

Table 3.6

Statistical Group Comparisons at Baseline

Variable	*df*	*t*	*p*	δ_t	BF_{10} ± error %
			Training tasks		
Orientation reproduction					
Capacity (*K*)	36.72	0.51	.610	-0.13	1/2.42 ± 0.01
Precision (SD^{-1})	37.72	0.71	.484	-0.18	1/3.91 ± 0.01
Visual search					
Accuracy	29.33	0.30	.766	0.08	1/2.93 ± 0.01
RT (ms)	25.34	0.01	.993	0.00	1/3.06 ± 0.01
			Transfer tasks		
Shape reproduction					
Capacity (*K*)	33.55	0.38	.707	-0.10	1/3.84 ± 0.01
Precision (SD^{-1})	38.00	1.11	.274	-0.28	1/2.21 ± 0.01
Orientation-Change detection					
Capacity (K)	37.49	0.54	.595	-0.14	1/3.87 ± 0.01

Table 3.7

Analysis of Variance Effects of Training on Cognitive Performance

Variable/Effect	*F*	*p*	η^2_G	η^2_p	BF_{10} ± error %
Orientation reproduction					
Capacity					
Time	18.12	**< .001**	.04	.23	> 100 ± 2.04
Group	1.53	.221	.02	.02	1/1.58 ± 1.89
Time x Group	0.00	.974	< .01	< .01	1/4.25 ± 3.26
Precision					
Time	3.05	.086	.01	.05	1/2.78 ± 2.21
Group	5.68	**.020**	.07	.08	2.87 ± 1.60
Time x Group	25.63	**< .001**	.07	.29	> 100 ± 4.11
Visual search					
Accuracy					
Time	24.79	**< .001**	.08	.29	> 100 ± 0.84
Group	2.55	.116	.03	.04	1/1.23 ± 2.06
Time x Group	1.55	.218	.01	.02	1/2.03 ± 4.33
Reaction time					
Time	8.22	**.006**	.03	.12	6.61 ± 0.99
Group	0.67	.417	.01	.01	1/2.80 ± 2.08
Time x Group	9.09	**.004**	.04	.13	10.96 ± 2.40
Shape reproduction					
Capacity					
Time	0.15	.704	< .01	< .01	1/5.18 ± 1.28
Group	0.28	.596	< .01	< .01	1/3.31 ± 0.98
Time x Group	1.36	.249	.01	.02	1/2.23 ± 3.69
Precision					
Time	1.12	.293	.01	.02	1/3.47 ± 1.05
Group	4.72	**.034**	.05	.07	1.63 ± 0.80
Time x Group	1.72	.195	.01	.03	1/1.90 ± 2.33
Orientation-Change detection					
Capacity					
Time	1.83	0.181	0.01	0.03	1/2.79 ± 1.00
Group	1.06	0.306	0.01	0.02	1/1.98 ± 0.55
Time x Group	3.12	0.082	0.01	0.05	1/1.05 ± 2.56

Note. BF_{10} = Bayes factor in favour of the alternative hypothesis. Degrees of freedom df_1 and df_2 were 1, 62 respectively.

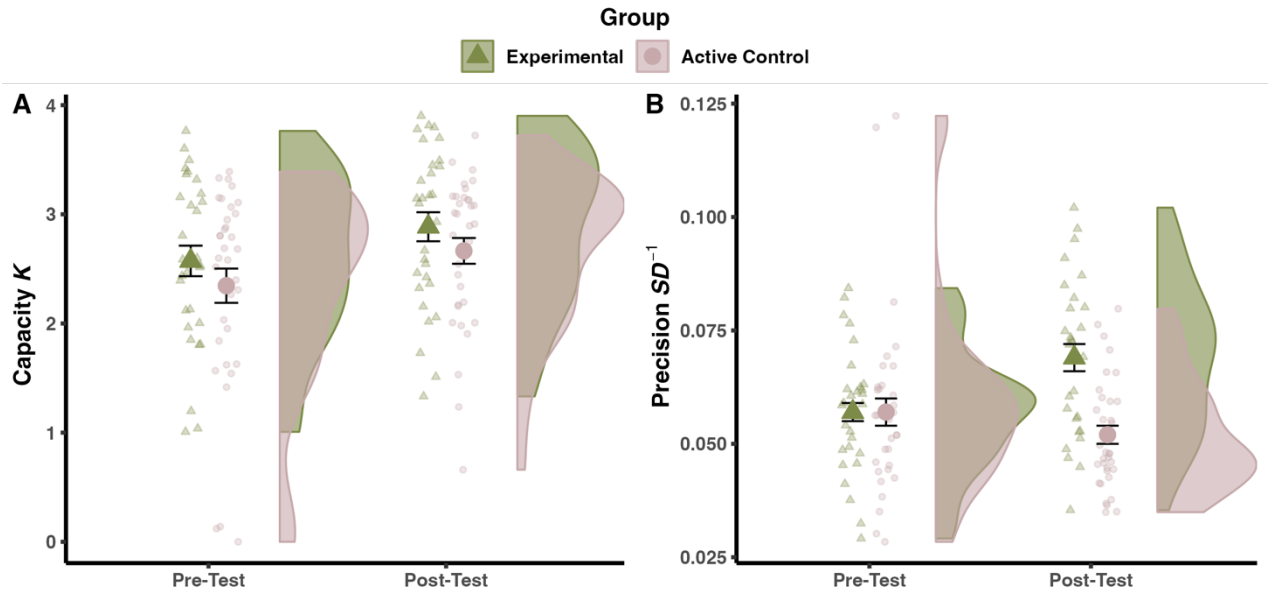
3.3.2.2 Training Effects

Orientation Reproduction. Figure 3.4 illustrates the pre-test to post-test changes in capacity and precision in orientation reproduction. The Time x Group interaction was not significant for capacity, $F(1, 62) < 0.01$, $p = .974$, $\eta^2_G < .01$, $\eta^2_p < .01$, with the absence of the interaction being supported by substantial evidence, $BF_{10} = 1/4.25 \pm 3.26\%$. These results

suggest that training-induced gains cannot be explained by an increase in quantity of representations activated in visual WM.

Figure 3.4

Pre-Post Changes in the Visual WM Training Task on Capacity and Precision



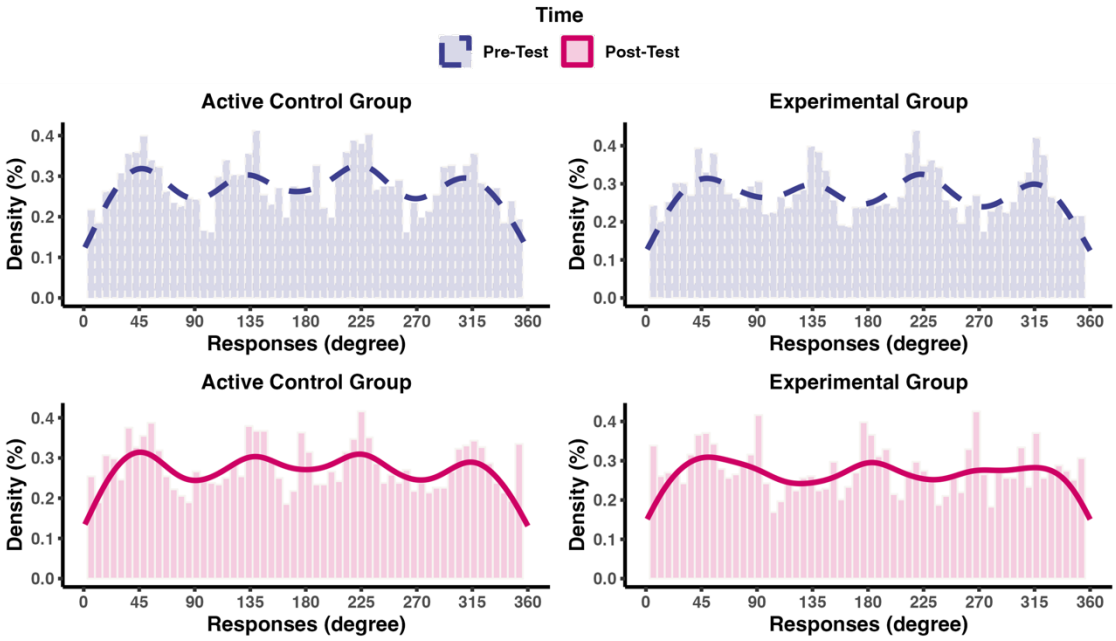
Note. Panel A: Changes in capacity. Panel B: Changes in precision. Left: Small transparent data points show the mean values for each individual. Big solid data points show the mean values at group level, with the error bars representing standard errors. Right: Density distributions of the data for both groups.

For precision, there was a significant Time x Group interaction effect, $F(1, 62) = 25.63$, $p < .001$, $\eta^2_G = .07$, $\eta^2_p = .29$, which was supported by decisive evidence, $BF_{10} > 100 \pm 4.11\%$. In the experimental group, precision significantly increased from pre-test ($M = .06$, $SD = .01$) to post-test ($M = .07$, $SD = .02$), $t(17) = -4.43$, $p < .001$, $\delta t = -1.16$, which was supported by decisive evidence, $BF_{10} > 100 \pm 0.00\%$. In contrast, in the active control group, precision decreased from pre-test ($M = .06$, $SD = .02$) to post-test ($M = .05$, $SD = .01$), $t(21) = 1.99$, $p = .059$, $\delta t = .28$, though the evidence for this decrease was highly ambiguous, $BF_{10} = 1.38 \pm 0.02\%$. Finally, precision was significantly higher in the experimental group than in

the active control group at post-test, $t(28) = 4.36$, $p < .001$, $\delta t = .71$, supported by decisive evidence, $BF_{10} > 100 \pm 0.00\%$. Taken together, we found considerable training-induced gains in visual WM precision in the trained orientation-reproduction task, with large effect sizes for changes from pre-test to post test and for the comparison to the active control group at the post-test. To further explore the differences in changes between the experimental group and the active control group in the orientation-reproduction task (not pre-registered), we examined the distributions of participants' responses at pre-test and post-test. As Figure 3.5 illustrates, we observed a pattern of responses suggesting that, at pre-test, individuals in both groups tended to respond with familiar or canonical orientations, with peaks at 45°, 135°, 225°, and 315°, $\chi^2(7, N = 7680) = 6.30$, $p = .505$, $BF_{10} < 1/100 \pm 0.00\%$. At post-test, however, the distribution of responses differed between the groups, $\chi^2(7, N = 7680) = 44.58$, $p < .001$, with decisive Bayesian evidence, $BF_{10} > 100 \pm 0.00\%$. Specifically, the experimental group showed a larger number of peaks in their response distribution, leading to a flattened density function and suggesting that, after orientation-reproduction training, participants' responses included a larger range of finer differences between orientations. In contrast, the active control showed a similar pattern at pre-test and post-test. These observations may indicate that the experimental group was able to distinguish finer differences in orientations after training.

Figure 3.5

Density of Pre-Post Responses Changes Differs Between Groups



Note. Purple histograms with dashed lines show the density of each response at pre-test, and the pink histograms with solid lines show the density of each response at post-test. Number of bins: 60. Experimental group: n = 30; active control group: n = 34; total responses per participant was 120 each at pre-test and post-test.

Visual Search. For accuracy, the Time x Group interaction was not significant, $F(1, 62) = 1.55$, $p = .218$, $\eta^2_G < .01$, $\eta_p^2 = .02$, with the active control group showing a numerically higher accuracy from pre-test to post-test than the experimental group. However, the evidence was ambiguous, $BF_{10} = 1/2.03 \pm 4.33\%$. For mean RTs, there was a significant Time x Group interaction effect, $F(1, 62) = 9.09$, $p = .004$, $\eta^2_G = .04$, $\eta^2_p = .13$, which was supported by strong evidence, $BF_{10} = 10.95 \pm 2.40\%$. Taken together, participants in the active control group showed larger increases in visual search speed after visual search training than the experimental group without sacrificing accuracy.

3.3.2.3 Transfer Effects

Shape Reproduction. We detected no significant transfer to a task using the same paradigm as the training task but different stimuli. The Time x Group interaction was not significant, $F(1,62) = 1.36$, $p = .249$, $\eta^2_G = .01$, $\eta^2_p = .02$, with, however, capacity decreasing in the experimental group and increasing in the active control group from pre-test to post-test. The evidence for the absence of this interaction was ambiguous, $BF_{10} = 1/2.23 \pm 3.69\%$. For precision, the Time x Group interaction was also non-significant, $F(1,62) = 1.72$, $p = .195$, $\eta^2_G = .01$, $\eta^2_p = .03$, with precision, numerically, slightly improving in the experimental group and remaining stable in the active control group. The evidence supporting the absence of the interaction was again ambiguous, $BF_{10} = 1/1.90 \pm 2.33\%$.

Orientation-Change Detection. Similarly, capacity in a different paradigm but with the same stimuli did not significantly improve after visual WM training. The Time x Group interaction approached significance, $F(1,62) = 3.12$, $p = .082$, $\eta^2_G = .01$, $\eta^2_p = .05$. Numerically, the experimental group performed better at post-test than pre-test, whereas the active control group's performance remained stable. Again, the evidence for the absence of a transfer effect was near-perfectly ambiguous, $BF_{10} = 1/1.05 \pm 2.56\%$. Taken together, there was no transfer to a different type of stimuli or paradigm, with the caveat that the evidence was overall ambiguous.

3.3.3 Summary

We found evidence for improvements in the trained tasks, with the experimental group improving only in precision, but not in capacity, in the trained orientation-reproduction task, and the active control group improving in RTs in the trained visual search task. Therefore, we rejected Hypothesis 1 that training gains reflect increases in capacity, and we concluded that training gains are driven by increased efficiency. As the improvement in precision did not generalise to performance gains in the untrained shape-reproduction task,

we rejected Hypothesis 2 that training gains reflect the acquisition of paradigm-specific expertise, but with the caution that the evidence for the absence of an effect on precision in shape reproduction was ambiguous only. Similarly, there was also no significant effect of orientation-reproduction training on performance in the orientation-change detection task. Therefore, we also rejected Hypothesis 3 that stimulus-specific expertise would transfer to a different paradigm but, again, with the caveat that the Time x Group interaction approached significance, with only ambiguous evidence for the absence of an effect. Therefore, taken together, we found that training gains were stimuli-specific and task-specific, with some ambiguity regarding the potential of these gains in efficiency to generalise to other contexts.

3.4 Discussion

The objective of the study was to identify the mechanisms underlying visual WM training and transfer effects. Specifically, we tested (1) whether training-induced gains after orientation-reproduction training reflect expanded visual WM capacity or enhanced efficiency in using the available capacity by facilitating the acquisition of paradigm-specific or stimulus-specific expertise, and (2) whether such training benefits generalise to other types of stimuli and paradigms. For this purpose, we distinguished training gains in quantity from training gains in quality of visual WM representations and tested transfer effects to an untrained stimulus type (shape reproduction) and paradigm (orientation-change detection).

The results showed that four visual WM training sessions improved the quality of visual WM representations in the trained task but not the quantity. Furthermore, we observed no transfer to different stimuli or a different paradigm. The evidence was ambiguous though, and there was a tendency that the experimental group numerically improved in the orientation-change detection task that used the same stimuli in a different paradigm. Notably, however, if anything, capacity decreased in the experimental group in the shape-reproduction task that uses different stimuli in the same paradigm. Taken together, these findings speak

against broad transfer through expanded capacity, which is consistent with the results from other recent WM training studies which reported limited evidence for transfer (Buschkuehl et al., 2017; De Simoni & von Bastian, 2018; Guye & von Bastian, 2017; Redick et al., 2013).

Instead, these findings suggest that training gains are driven by a more efficient use of the available cognitive capacity (von Bastian & Oberauer, 2014; von Bastian et al., 2022). Furthermore, the lack of transfer effects supports the conclusion that the training-induced efficiency gains were both stimuli-specific and paradigm-specific: neither stimuli-specific expertise nor paradigm-specific expertise were generalisable to the same paradigm with different stimuli or a different paradigm with the same stimuli. More specifically, the untrained shape-reproduction task used the same paradigm as the trained visual WM task but tested the memory of shapes instead of orientations. The lack of transfer to this task suggests that training gains reflect gains in expertise in orientation discrimination which is specific to the stimuli employed in the trained task. Yet, the untrained orientation-change detection task used the same stimuli as the trained visual WM task and also tested memory of orientations, but we still did not observe any transfer. However, different to the trained paradigm, the untrained orientation-change detection task might capitalise on configural information, such as the internal representation of the relationship between all displayed orientations at the maintenance stage (Boduroglu et al., 2009; Buschkuehl et al., 2017). At the same time, at the recall stage, the task requirement to detect only one changed orientation out of all stimuli displayed could possibly reduce the need to focus on the feature precision of each stimulus. This could explain why efficiency gains in the trained task did not generalise to another visual WM paradigm using the same stimuli type.

An alternative, not necessarily mutually exclusive, possibility is that the training gains in the orientation-reproduction task reflect a more refined motor control in reproducing the triangles' orientation. However, the trained orientation-reproduction WM task and the

untrained shape-reproduction task arguably require a similar degree of refined motor control to reproduce the orientation or shape information, respectively, by rotating and clicking the mouse. Hence, if the observed training gains merely reflected better motor control, we should also have observed improvements in the untrained shape-reproduction task which requires similar levels of fine motor control. The observed lack of such improvements renders this possibility unlikely.

The findings of the present study also provide some indications how stimuli-specific and paradigm-specific expertise may operate and interact. Our exploratory inspection of response distributions showed that the experimental group but not the active control group reported a larger number of different orientations after training, suggesting that training in the orientation-reproduction task may have catalysed the development of perceptual expertise allowing for discriminating finer differences in orientations. This is in line with other research showing that visual WM training can boost perceptual processing (Truong et al., 2022). Improved perceptual processing due to stimuli-specific expertise may enhance the perceived perceptual distinctiveness (Olson et al., 2005). Given the premise that the active control group's visual search training involved only little memory (Wolfe & Horowitz, 1998) while sharing similar encoding processing (Kong & Fougnie, 2019), the fact that we observed these precision gains only in the experimental group supports the conclusion that visual WM training-induced gains in efficiency operate at maintenance and recall stage. These stimulus-specific efficiency gains allow for maintaining more precise internal feature representations, and/or discriminating these representations with higher resolution when recalling this feature information.

Developing stimuli-specific, perceptual expertise may also help to use effective paradigm-specific strategies that operate at maintenance and recall stage. Specifically, we found that the experimental group did not only respond a larger number of orientations but

more peaks with canonical orientations after training. Participants may have used canonical orientations as a memory aid for the orientations (e.g., 90, 180, and 270 degrees like the numbers 3, 6 and 9 on a clock face). Increasing the number of available canonical orientations may benefit the effectiveness of such a strategy and increase overall performance. Note that this does not exclude the possibility that both experimental and active control training could have improved sensory discrimination at encoding stage.

3.4.1 Limitations

One major limitation of the current design is that the orientation-change detection task – the untrained paradigm using the same stimuli – did not allow for assessing precision (i.e., the quality of visual WM representations). Consequently, our results cannot fully rule out transfer of gains in the quality of visual WM representations to a different paradigm. Future research with a more fine-grained assessment of the stimulus features is required to identify the mechanisms underlying the transferable gains in quality of visual WM representations.

Another potential limitation of this study is that four training sessions might not be intensive enough to induce transferable training gains in the quality of visual WM representations. Indeed, this possibility is consistent with our results that training gains in the quality of visual WM representations were not detected during training but only at post-test. Furthermore, the spacing of the training sessions may not have optimally supported learning. For example, a design with only one session a week may have allowed for better consolidation of learning effects (e.g., see Lampit et al., 2020). Future research is needed to better understand the optimal intensity and spacing of visual WM training interventions.

Moreover, our training tasks were not adaptive, that is, all participants practised all set sizes irrespective of their individual performance. We chose this design to ensure sufficient measurement of all three set sizes for applying the SMM. However, it might have led to a decrease in motivation. A previous study showed no differences between adaptive and non-

adaptive training both for motivation and training and transfer gains (von Bastian & Eschen, 2016); however, in that study participants still received performance-based feedback. Such feedback likely encourages better engagement with the daily training sessions and reduces attrition, which could be useful especially in an online setting like the current study.

Finally, we did not assess participants' training experience, subjective training gains, or strategies they employed, because we aimed at minimizing the administration time for the benefit of participant retention. However, these data could have added important insights regarding the possible mechanisms underpinning the observed training gains (e.g., see De Simoni & von Bastian, 2018; Guye & von Bastian, 2017). Future research would benefit from including self-report measures for advancing understanding of training-induced change in cognitive performance.

3.4.2 Conclusion

To the best of our knowledge, the findings of the present study are the first to provide evidence from a continuous reproduction task that visual WM training induces stimuli-specific and paradigm-specific gains in the quality but not in the quantity of visual WM representations. These findings support the notion that training enhances cognitive efficiency through the acquisition of expertise but not capacity. A better understanding of how training facilitates a more efficient use of the available visual WM capacity, and how the underlying training benefits are influenced by the characteristics of stimuli and paradigms, will be critical for harnessing the potential benefits of these training benefits.

Chapter 4 – Can training-induced changes in visual working memory be explained by a single model? A model comparison study

Contributions:

Shuangke Jiang (conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, validation, visualization, writing—original draft, writing—review & editing)

Myles Jones (supervision, writing—review & editing)

Claudia von Bastian (conceptualization, funding acquisition, supervision, writing—review & editing)

Findings from this chapter have been presented in the following conferences:

Jiang, S., Jones, M., & von Bastian, C. C. (2022). *Model selection to explain training-induced changes in visual working memory* [Poster]. Psychonomic Society 63rd Annual Meeting, Boston, USA.

Jiang, S., Jones, M., & von Bastian, C. C. (2023). *No single measurement model of visual working memory can explain training-induced change* [Poster]. The 23rd European Society for Cognitive Psychology (ESCoP) Conference, Porto, Portugal.

Work based on this chapter has received 2023 European Society for Cognitive Psychology (ESCoP) Conference Travel funding:

Jiang, S., Jones, M., & von Bastian, C. C. (2023). *No single measurement model of visual working memory can explain training-induced change* [Poster]. The 23rd European Society for Cognitive Psychology (ESCoP) Conference, Porto, Portugal.

Abstract

Computational visual working memory (VWM) models are typically fitted to data from a single testing session. However, VWM performance can change substantially through training. The current study investigated which of four popular VWM measurement models can best account for these changes observed during training. We fitted the Standard Mixture Model (Zhang & Luck, 2008), Swap Model (Bays et al., 2009), Signal Discrimination Model (Oberauer et al., 2021) and Target Confusability Competition Model (Schurgin et al., 2020) to existing data across multiple testing sessions and set size conditions. We first compared these models in the experimental group (n = 30) before, during, and after training for an orientation-reproduction VWM task. We then compared the models in the active control group (n = 34) for their performance of untrained orientation-reproduction task. Furthermore, we compared these models in the untrained shape-reproduction VWM task for both groups. Overall, at the group level, the Standard Mixture Model accounted best for the changes occurring over and after four training sessions in the orientation-reproduction task. However, at the individual level, the preferred model switched across training sessions. The Target Confusability Competition Model fitted best to the data from the shape-reproduction task where training-induced changes were absent; still, similar switching patterns in model preferences as for the orientation-reproduction task were observed across testing sessions. Taken together, these findings speak against the notion that any single current measurement model can fully account for the dynamic changes in VWM performance that were observed in training studies.

Keywords: visual work memory, model comparison, working memory training

4.1 Introduction

To characterise and explain the limited span of visual working memory (VWM) that has been observed empirically, a variety of theoretical models have been developed to enable researchers to test their hypotheses and research questions (Wilken & Ma, 2004; Zhang & Luck, 2008; Bays & Husain, 2008, Oberauer et al., 2017; Schurgin et al., 2020; van den Berg, 2012). However, due to subtle differences in definitions or assumptions made by theorists who use different metaphors to explain the VWM nature, these different theoretical models are often not explicitly distinguishable (Farrell & Lewandowsky, 2018; Oberauer et al., 2018). Without a formal and distinguishable measurement, theoretical models can lead to discrepant, but not mutually exclusive, understandings of VWM.

A particularly heated debate centres on the limits of VWM capacity. Specifically, some researchers argue that there is an upper limit on the quantitative number of discrete representations in VWM, with low memory precision resulting from a lack of discrete representations after reaching the limit (Zhang & Luck, 2008, 2011). Other researchers argue that variable precision of memory underpins the observed limits of capacity (Bays & Husain, 2008; van den Berg et al., 2014; Zokaei et al., 2011). Precision manifests itself at varying levels of noisy representations of feature information. It has also been debated whether the source of memory precision can be modelled by the closeness to the content feature (e.g., orientation), the context feature (e.g., location) or the bindings between content and context (Bays, 2016; Oberauer, 2021; Tomić & Bays, 2022; Williams et al., 2022). Finally, a more recent theoretical notion has challenged all previous views on the source of precision by arguing that it stems from the perceptual similarity of stimuli features (Schurgin et al, 2020).

In light of the current debates in VWM, using an adequate computational measurement model enables researchers to formulate testable hypotheses derived from precise theories, and thus address the ‘theory crisis’ – the poor testability of theories and

hypotheses (Borsboom et al., 2021; Eronen & Bringmann, 2021; Fried & Flake, 2018; Maatman, 2021; Ngiam, 2023; Oberauer & Lewandowsky, 2019; Popov, 2023). Four popular VWM measurement models aim to address these current debates: Standard Mixture Model (SMM; Zhang & Luck, 2008, 2011), Swap Model (SwapM; Bays et al., 2009; Schneegans & Bays, 2016), Signal Discrimination Model (SDM; Oberauer, 2021) and Target Confusability Competition model (TCC; Schurgin et al., 2020). These models have been widely assessed regarding their empirical explanatory adequacy, and applied to test a variety of research questions, typically using the continuous-reproduction paradigm (Heinen et al., 2016; Ovalle Fresa & Rothen, 2019; Sutterer & Awh, 2016; S. Wang et al., 2019; Williams et al., 2022; Zokaei et al., 2011). In this paradigm, individuals are asked to memorise and later reproduce features of stimuli in continuous dimensions (e.g., orientation, shape, colour or location) in order to probe high-resolution contents of VWM (Gorgoraptis et al., 2011; Ma et al., 2014; Wilken & Ma, 2004; Zhang & Luck, 2008; Bays & Husain, 2008;). Performance is assessed by the recall error, which is the difference between the reproduced feature and the originally presented feature of the target item. Using the recall error, these VWM measurement models allow for estimating theoretically essential latent information of VWM representations, and differentiating this information from noise (Oberauer, 2021; Oberauer et al., 2017).

4.1.1 Standard Mixture Model

The Standard Mixture Model (SMM; Zhang & Luck, 2008, 2011) is a two-component mixture model that distinguishes between the quantity (also called discrete slots or capacity) and quality (precision or resolution) of VWM representations. The SMM assumes that there is an upper limit on the quantitative number of items that can be remembered, estimated as *capacity*. The idea of an upper limit for VWM capacity is consistent with empirical evidence for a fixed number of items that can be remembered (Cowan, 2001; Miller, 1956; Pashler, 1988). At the same time, the quality of VWM representations varies before reaching the

capacity limit. This variation is estimated as *precision*. After reaching the capacity limit, the SMM assumes that the target is either encoded and recalled with a certain precision or not at all. The distinguish roles of quantity and quality of VWM representations that were described in SMM has been supported by neural and behavioural evidence (Alvarez & Cavanagh, 2004; Awh et al., 2007b; Fougnie et al., 2010; Fukuda et al., 2010; Ngiam et al., 2022; Olson & Jiang, 2002; Rouder et al., 2008; Scolari et al., 2008; Souza et al., 2014; Zhang & Luck, 2008).

The SMM specifies two mathematical components, that is, a von Mise distribution approximating a circular normal distribution, and a uniform distribution, denoting response errors for trials where the feature information of the target item was remembered or not:

$$P(x) = (1-g)\frac{1}{2\pi I_0(\kappa)}e^{\kappa \cdot cos(x-x_{\theta t})} + g\frac{1}{2\pi}, \quad (1)$$

where x is the response, $x_{\theta t}$ is the target feature information, κ is the concentration parameter of the von Mises distribution which reflects the precision of representations ($k > 0$; the larger the k, the more concentrate about $x_{\theta t}$), g is the proportion of random guess responses, and $I_0(\kappa)$ is the modified Bessel function of order 0. The parameter for capacity is computed as the product of the probability of remembered information, 1-g, and the set size, n. The parameter for precision is computed as the inverse of the standard deviation (SD^{-1}) of the von Mises distribution, which can be converted from the concentration parameter κ.

4.1.2 Swap Model

The Swap Model (SwapM; Bays et al., 2009; Schneegans & Bays, 2016) is a three-component model. Like the SMM, SwapM also specifies whether the feature information of the target was remembered or not. In addition, SwapM has a third component representing swap (or binding) errors, which refer to mixing up target and non-target information. In contrast to the two-component SMM, SwapM assumes that VWM is limited to a resource that can be shared between all items in a visual scene, and the allocation of such resource is

highly flexible. For example, an item that is prioritized can be represented more precisely in VWM at the cost of decreased precision for other items (Emrich & Lockhart, 2017; Yoo et al., 2018). This theory provides an explanation for binding errors which have been widely observed in empirical research, and attributes them to the source of VWM limits (Schneegans & Bays, 2017; Zokaei et al., 2011).

SwapM decomposes responses errors into three components – one von Mise distribution for the target item information, von Mise distributions for all non-targets, and a uniform distribution for random guess:

$$P(x) = pT\frac{1}{2\pi I_0(\kappa)}e^{\kappa \cdot cos(x-x_{\theta t})} + \frac{pN}{n-1}\cdot\sum_{k}^{n-1}\frac{1}{2\pi I_0(\kappa)}e^{\kappa \cdot cos(x-x_{\theta k})} + pU\frac{1}{2\pi}, \qquad (2)$$

where *x* is the response, $x_{\theta t}$ is the target feature information, $x_{\theta k}$ is the non-target feature information, κ is the concentration parameter of the von Mises distribution which reflects the precision of representations, *n* is the set size, *pT* is the proportion of target information that was remembered, *pN* is the proportion of non-target information that was mistakenly reported, *pU* is the proportion of random guess responses (equivalent to *g* in SMM). *pT, pN, pU* sum to 1.

4.1.3 Signal Discrimination Model

The Signal Discrimination Model (SDM; Oberauer, 2021) assumes that the retrieval of representations in VWM is cue-based (e.g., the location of the probed item). Consequently, SDM quantifies the precision of representations not only on the feature (content) dimension but also on the cue (context) dimension. Thus, SDM distinguishes between three sources of low memory precision. First, similar to the SMM and SwapM, the SDM considers precision on the feature dimension. The degree of activation or reactivation of all features in the testing array is independent of a retrieval cue. Second, the activation of representations in VWM relies on binding between feature information and cue information. The strength of binding implies the extent to which the target is reactivated by a location cue at retrieval. The binding

strength is subject to the similarity of the spatial locations between the target and non-targets. Third, each response option has noise added to the signal strength in both dimensions. These theoretical mechanisms are supported by improved model fits when taking account into mis-bindings between item-feature and cue information (Oberauer et al., 2017; Schneegans & Bays, 2017). Furthermore the empirical patterns of results that swap errors tend to increase when non-target items are spatially close to the target also support this notion (Emrich & Ferber, 2012; Rerko et al., 2014; Souza et al., 2014).

The SDM models the response errors with an activation distribution of von Mises distributions on both the memory feature and cue information centred on the target. A probability function of choosing each response over possible responses is obtained by a signal-detection rule, equivalent to that the activation distribution is normalized by an exponent function:

$$S(x) = (a + c \cdot \sum_i^n exp\,(-s|y_i - y_\theta)) \cdot \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cdot cos(x - x_{\theta t})}, \quad (4)$$

$$P(x) = \frac{exp\,(S(x))}{\sum exp\,(S(x))}, \quad (5)$$

where x is the response, $x_{\theta t}$ is the target feature information, y_i is the location information of each item i of all n items (set size), y_θ is the cue information (the location information of the target). κ and s represent the precision of the memory feature and cue representations, respectively. a and c are the memory strength of the memory feature and binding, respectively.

4.1.4 Target Confusability Competition Model

The TCC model builds on signal-detection accounts of memory. Unlike the other previously mentioned models that measure multiple psychological constructs by distinguishing the different sources of recall errors, the TCC model (TCC; Schurgin et al., 2020) measures error distributions as a function of memory strength. The memory strength represents a formalised psychological similarity scaling function. TCC assumes that the

perceptual similarity of the stimulus features, which is independently measured by an additional perceptual task. The perceptual similarity is fixed across all participants and all memory strengths. Furthermore, noise accumulates and is added to similarity during the maintenance of memory. When the target feature is probed, the familiarity of the target feature gets boosted in the feature space by the amount of memory strength. Through a noisy signal-detection process, the feature information with the strongest maintained similarity will be recalled. Therefore, when the target feature is getting boosted, but the memory strength is weak, a non-target feature, even if far away from the target in the feature space, can sometimes get selected for recall. This is likely due to that the selected non-target's psychological familiarity has resulted the strongest after adding the noise.

In the TCC model, the recall error is modelled as an activation distribution of a Laplace distribution function. A response-selection probability function follows a signal detection rule which is the same as in the SDM (5):

$$S(x) = d' \frac{k}{2} \exp(-k|x - \theta), \quad (6)$$

where x is the response, θ is the mean, k is the scaling rate ($k > 0$; the larger the k, the more spread out about the mean). d' is the memory strength.

4.1.5 Measurement Models in Use

Owing to its limited span and critical involvement in higher cognition, VWM has often been chosen as a target cognitive construct for interventions like cognitive training. Cognitive training involves the repeated practice of cognitive tasks. Individuals' performance is measured before and after training to assess any training-induced cognitive benefits relative to a control group undergoing an alternative intervention. However, whereas large improvements are consistently observed in the trained tasks, the evidence for transfer of these benefits to other cognitive abilities is inconsistent (von Bastian et al., 2022; Melby-Lervag et al., 2016). These inconsistencies may reflect a lack of understanding in theoretical

mechanisms underlying training gains (Redick et al., 2019; Smid et al., 2020; von Bastian & Oberauer, 2014). Measurement models have been used to investigate mechanisms underpinning VWM changes, and thus aids to identify when and why interventions ought (not) to work (Heinen et al., 2016; Jiang et al., 2023; Ovalle Fresa & Rothen, 2019; S. Wang et al., 2019).

Despite the advantages of using a measurement model to quantify and interpret the effects of interventions or experimental conditions, directly using parameters estimated from a chosen model can be problematic (Wilson & Collins, 2019). An extreme case is when the researcher who does not have specific theoretical assumptions and measurement aims runs inferential analyses using multiple existing similar competing models. One might choose a more complex model whose estimates show some significance in the inferential analyses, over another model that does not show any significance but would be equally adequate to explain and more likely to reproduce the empirical observations. This problematic approach is similar to *p*-hacking and reduces the reproducibility. A good practice to avoid this issue is to first evaluate and even compare the overall adequacy of a set of similar models, that is, whether the model explains the empirical findings and whether the explanations are sufficiently parsimonious and substantially plausible (Borsboom et al., 2021).

So far, only a few previous studies have quantitatively compared which possible VWM measurement models most adequately capture cognitive changes (Donkin et al., 2013; Oberauer, 2021; van den Berg et al., 2014; S. Wang et al., 2019; Williams et al., 2022). Yet, theses existing findings revealed inconsistent patterns of model comparison results regarding measuring VWM representations. Moreover, the majority of modelling studies are based on single-session studies (Bays, 2016; Bays & Husain, 2008; Oberauer, 2021; Schurgin et al., 2020; Tomić & Bays, 2022; van den Berg et al., 2014; Williams et al., 2022; Zhang & Luck, 2008). However, intervention research like cognitive training often requires an accurate

quantification of performance changes across multiple testing sessions to make inferences on the training effects and their underlying mechanisms. Therefore, it is important to evaluate how well these models could account for these training-induced changes.

4.1.6 Present Study

To fill this gap, the current study explored which of these four VWM measurement models that have been widely used in empirical, or comparison studies can best account for cognitive changes through training. In particular, we compared the SMM (Zhang & Luck, 2008), SwapM (Bays et al., 2009), SDM (Oberauer, 2021) and TCC (Schurgin et al., 2020) in a multi-session training data set, including different set size conditions. We first compared the models in an experimental group before, during and after training of an orientation-reproduction task for which training-induced improvements were observed (see Chapter 3). We also compared the models in the active control group, who trained a visual search task, fitting data from their performance in an untrained orientation-reproduction task. For the active control group, training-induced improvements were absent but test-retest improvements from pre-test to post-test were still found. Then, we compared these models in both groups fitting data from an untrained shape-reproduction task to explore which measurement model fits best as a baseline for when neither training-induced nor training-unspecific (test/retest) effects are observed. In addition, to explore how well the preferred model can account for dynamic cognitive changes at the individual level, we evaluated to what extent the best-fitting model switched from one testing session to another.

4.2 Method

4.2.1 Sample

The sample included in the present study includes data of all 64 participants who completed the orientation-reproduction task and the shape-reproduction task. Data were originally collected in a visual working training study by Jiang et al. (2023; Chapter 3). The

experimental group (n = 30, 22.73 $\pm$ 3.92 years old) practised an orientation-reproduction task at three different set sizes (2, 4, 6) for four training sessions, while the active control group (n = 34, 21.94 $\pm$ 2.52 years old) practised a visual search task. Performance of both groups in the orientation-reproduction task and the shape-reproduction task were assessed at set size 4 at a pre-test and a post-test before and after training, respectively. For each individual, each condition per testing session consists of 120 observations (trials) for each VWM task.

4.2.2 Orientation-Reproduction Task

Each trial began with a fixation cross displayed centrally for 1000 ms. Next, an array of randomly orientated (0-360°) isosceles triangles was arranged in a circular manner and appeared on the screen for 200 ms, followed by a 1000 ms blank screen. Then, one of the displayed triangles was randomly selected as the target stimulus and presented in a random orientation. Participants were instructed to reproduce the original orientation by rotating the triangle with the computer mouse and click the left mouse-button to record their response.

4.2.3 Shape-Reproduction Task

Following a central fixation cross for 1000 ms, an array of black ring-shaped objects with varying proportions filled in white were distributed on the screen in a circular manner for 200 ms. After a 1000 ms blank screen, one of the displayed objects was randomly selected as the target stimulus. The target stimulus was presented in black colour with a white bar. Participants were instructed to reproduce the original proportion of the white segment by rotating and left clicking the mouse.

4.2.4 Model Comparison

The four VWM measurement models were fitted to each participant's data using the Nelder-Mead simplex method (function fminsearchbnd in Matlab). In general, complex models fit data better than simple models with a cost of having more free parameters.

Therefore, model comparison should rank candidate models balancing the goodness of fit and parsimony. Two common model comparison methods, the Akaike information criterion (AIC; Akaike, 1974) and the Bayesian information criterion (BIC; Schwarz, 2007), were used to measure the relative goodness of fit and parsimony of the four VWM measurement models for each participant, testing session, set size condition and task type. The lower the relative value of AIC or BIC, the better the model was fitted into the data. Because the BIC gives a greater penalty on more complex models and is more appropriate when models are nested (as is the case for the SMM and SwapM), we drew conclusions based on BIC but report both the AIC and BIC values in the supplementary materials (Burnham & Anderson, 2004; Farrell & Lewandowsky, 2018). See relative AIC and BIC values of all models relative to the best fitted model for each participant, task, testing sessions, and set size in Table G1- G5 (Appendix G).

4.3 Results

First, we aimed to select the best fitting model among the four VWM measurement models to capture performance in the trained task when substantial training-induced changes are present at the group level. For this purpose, we fitted these models separately to the orientation-reproduction data from the experimental group at pre-test, post-test, and the four training sessions, and for each set size (set size of 4 at pre-test and post-test; set size of 2, 4, 6 during training). To select the best model when substantial, yet training-unspecific (test/retest), changes are present at the group level, we also fitted the four models to the data from the active control group who practised a visual search task, but their performance in the same orientation-reproduction task was measured during pre-test and post-post at set size 4. In addition, we fitted the models to the data from the shape-reproduction task. This task was trained by neither the experimental group nor the active control group and for which we did not observe any transfer effects (see Chapter 2). According to BIC values, Table 4.1

summarises the percentage of experimental group's data that were best fitted by each model during training. Table 4.2 shows the percentage of both groups' data that were best fitted by each model from pre-test to post-test.

Table 4.1

Percentage of Experimental Group' Data That Were Best Fitted by Each Model During Training

Session	Set Size	Model			
		SMM	SwapM	SDM	TCC
Training 1	2	40.00	20.00	20.00	20.00
	4	63.33	3.33	16.67	16.67
	6	26.67	20.00	33.33	20.00
Training 2	2	40.00	30.00	23.33	6.67
	4	63.33	3.33	13.33	20.00
	6	36.67	10.00	40.00	13.33
Training 3	2	30.00	26.67	23.33	20.00
	4	66.67	0.00	10.00	23.33
	6	33.33	16.67	30.00	20.00
Training 4	2	40.00	20.00	30.00	10.00
	4	83.33	0.00	3.33	13.33
	6	30.00	16.67	23.33	30.00

Note. Based on Bayesian Information Criterion Values (BIC). The lower the BIC value, the better the model was fitted into the data set. The percentage of one person from the experimental group is 3.33%. SMM = Standard Mixture Model; SwapM = Swap Model; SDM = Signal Discrimination Model; TCC = Target Confusability Competition model.

Table 4.2

Percentage of Both Groups' Data That Were Best Fitted by Each Model from Pre-Test to Post-Test.

Session	Model			
	SMM	SwapM	SDM	TCC
Experimental Group (n = 30)				
ORT				
Pre-Test	33.33	23.33	6.67	36.67
Post-Test	50.00	6.67	10.00	33.33
SRT				
Pre-Test	13.33	6.67	0.00	80.00
Post-Test	13.33	3.33	0.00	83.33
Active Control Group (n = 34)				
ORT				
Pre-Test	26.47	11.76	8.82	52.94
Post-Test	35.29	29.41	5.88	29.41
SRT				
Pre-Test	8.82	2.94	0.00	88.24
Post-Test	11.76	0.00	0.00	88.24

Note. Based on Bayesian Information Criterion Values (BIC). The lower the BIC value, the better the model was fitted into the data set. The percentage of one person from the experimental group is 3.33% while the percentage of one person from the active control group is 2.94%. SMM = Standard Mixture Model; SwapM = Swap Model; SDM = Signal Discrimination Model; TCC = Target Confusability Competition model. ORT = Orientation-reproduction task; SRT = Shape-reproduction task.

4.3.1 SMM captured performance during and post training best

For the experimental group's performance in orientation-reproduction task, the SMM showed an overall better model fit for 45.48% of participants relative to the other three models[6] including all six testing sessions and three set size conditions. Especially, at set size 4, SMM and TCC both fitted well at pre-test (the difference was no greater than one

[6] According to AIC values, SMM still came in a close second (33.57%) while the best fitted SDM (39.76%) only marginally fitted better to the training data, rendering less model selection uncertainty.

participant; Table 4.2), but the SMM became more predominant from the first training session (63.33%) onward (Table 4.1). At the post-test, although less participants' data was best fitted by the SMM compared to the fourth training session (83.33%), the SMM was the best fitted to performance of 50% of participants (Table 4.2). This trend may be due to the SMM's ability to explain the training-induced behavioural improvement in quality of VWM specific to this task (Jiang et al., 2023).

In addition, model fits varied for set sizes during training (Table 4.1). Overall, the SMM was the preferred model at lower set sizes (2 and 4) among the other three models. Unlike at set size 4 where the SMM was best fitted to the majority of participants' data (63.33% to 83.33%), only up to 40% of participants preferred SMM at set size 2. At a higher set size (i.e., 6), the SDM fitted better during the first training session. After the first training session, the SMM fitted well equally with SDM. During the last (fourth) training session, SMM fitted well equally with both SDM and TCC. While the evidence is less decisive at higher set size, these patterns of changes in the preferred VWM model suggest that, at least with only little training, the SMM captures performance at lower set sizes best, especially at set size 4.

4.3.2 TCC was largely preferred when substantial changes were absent

Overall, at the pre-test, the TCC model fitted the data best for 64.46 % of participants, from both the experimental group and the active control group, in both the orientation-reproduction task and the shape-reproduction task. Moreover, the TCC model was preferred by over 80% of the participants at both pre-test and post-test in an untrained shape-reproduction task for both groups (Table 4.2). The large preference for this model might reflect the absence of training gain that was transferred to the shape-reproduction task that was reported by Jiang et al. (2023). Indeed, in the orientation-reproduction task that was not trained by the active control group, when changes induced by the test-retest effect were likely

present in this group, the TCC was no longer the best fit, that is, the SMM was preferred (35.29%) over the TCC (29.41%) and SwapM (29.41%). Altogether, it might suggest that TCC can consistently measure limited VWM performance when any substantial change in performance is absent.

4.3.3 No single model can yet capture the dynamic changes in VWM performance

In the previous sections, model fits were interpreted at the group level; however, we observed considerable changes in preferred models at the individual level across testing sessions which we explored further. Table 4.3 summaries the percentage of changes in preferred models across all testing sessions at different set size for individuals from the experimental group completing the orientation-reproduction task. Figure 4.1 illustrates these dynamic changes in the preferred model in the experimental group during orientation-reproduction training. Figure 4.2 demonstrates the dynamic changes in the preferred model from pre-test to post-test for both the experimental group and the active control group in both the orientation-reproduction task and the shape-reproduction task.

Table 4.3

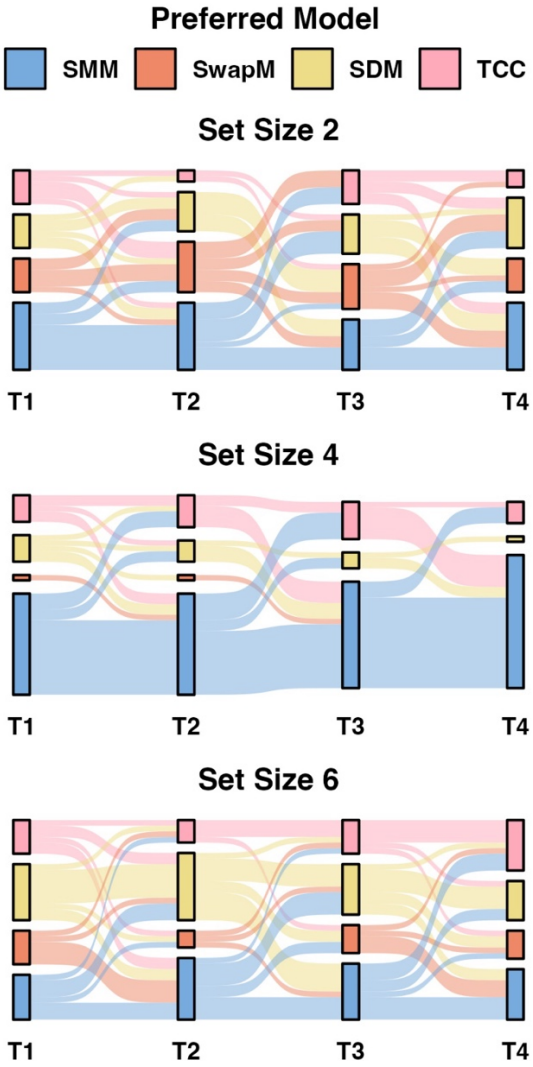
Percentage of Changed Preferred Model for the Experimental Group (n = 30) in the Orientation-Reproduction Task Between Testing Sessions

Set Size	Pre-Test – T1	T1 – T2	T2 – T3	T3 – T4	T4 – Post-test
2	-	53.33	80.00	73.33	-
4	73.33	43.33	50.00	36.67	60.00
6	-	66.67	63.33	56.67	-

Note. The percentage of one person from the experimental group is 3.33%. En dashes represent from one testing session to the next testing session. T1 – T4: 1st training to 4th training session.

Figure 4.1

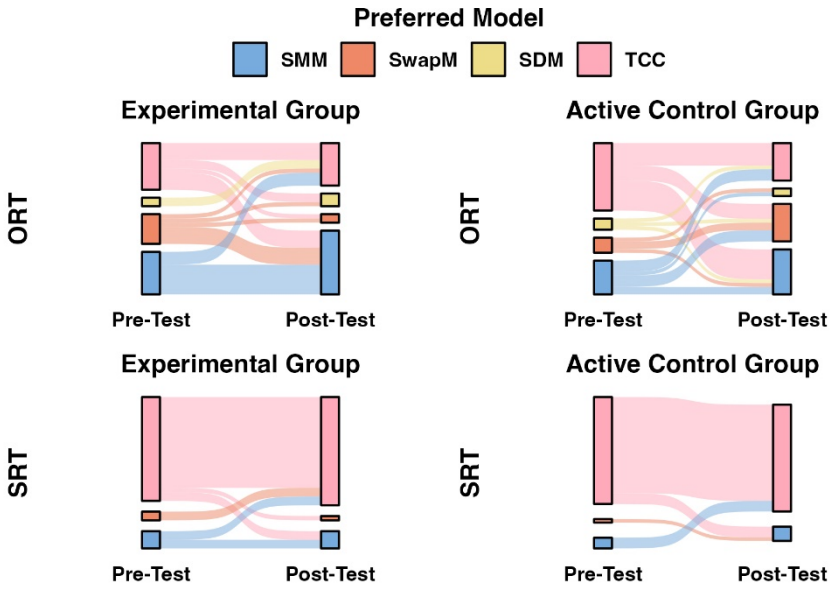
Preferred Model Changes for the Experimental Group (n = 30) During Orientation-Reproduction Training



Note. Sankey nodes represent the proportion of individuals' data was best fitted by the model. Sankey links represent whether the best model fit switched or not from one model to another between two testing sessions or remained unchanged. T1 – T4: 1st training to 4th training.

Figure 4.2

Preferred Model Changes from Pre-Test to Post-Test



Note. Sankey nodes represent the proportion of individuals' data was best fitted by the model. Sankey links represent whether the best model fit switched or not from one model to another between two testing sessions or remained unchanged. ORT = orientation-reproduction task; SRT = shape-reproduction task.

As shown in Table 4.3 and Figure 4.1, between training sessions, the lower and higher set size (i.e., 2 and 6 respectively) overall showed more switches in model preference compared to that at set size 4. Thereby, this overall pattern, again, suggested that the SMM accounted better for performance changes at set size 4. Furthermore, the percentage of switches in model preference dramatically increased from the pre-test to the first training session (73.33%), and from the last training session to the post-test (60%). During training, different set sizes were randomly presented, whereas only set size 4 was presented at pre-test and post-test, which may have possibly led participants to adjust their strategies. Therefore, the increase in the switches in model preference between these sessions might reflect a need for different VWM models when individuals obtain training benefits through the acquisition of strategies.

In addition, from the pre-test to post-test only, both groups showed a similar pattern of changes in model preference between the two types of VWM task. Specifically, more changes were observed in the orientation-reproduction task (60%) than in the shape-reproduction task (23.33%, see Figure 4.3). Similarly, 70.58% of the active control group switched their model preference from pre-test to post-test in the orientation-reproduction task after training a different task (visual search), but only 20% of this group changed their preferred model in the shape-reproduction tasks. These patterns suggest that different models present better fits for different types of stimuli.

Taken together, the overall pattern at the individual level suggests that, though SMM overall explains the performance in the orientation-reproduction task best at the group level, no single model can yet capture the dynamic changes in VWM performance during training. As discussed in Chapter 3, these changes could be due to acquisition of expertise or strategies by individuals to deal with different conditions including set size and stimulus type.

4.4 Discussion

The goal of the current study was to identify which possible VWM measurement models could better capture VWM performance when intervention-induced changes are present or not, across multiple testing sessions, without compromising parsimony. In a series of comparisons, we evaluated four VWM measurement models: the SMM (Zhang & Luck, 2008), SwapM (Bays et al., 2009), SDM (Oberauer, 2021) and TCC (Schurgin et al., 2020). Overall, at the group level, we found that the SMM was the simplest and best-fitting model for our training data from the experimental group who showed substantial training-induced improvement in the quality of VWM performance from the pre-test to post-test. In contrast, when substantial cognitive changes were absent, the TCC fitted the data better. Furthermore, at the individual level, we found that participants were prone to switching their VWM model preference across testing sessions and conditions, suggesting that none of the current

measurement models can fully account for the dynamic, substantial changes in VWM performance. These switches in model preference for the orientation-reproduction task data were present even in the active control group who did not train for this task but showed test-retest practice effects, suggesting that model preference switches are not specific to training-induced changes. In addition, changes in model preference were much less pronounced in the data from the untrained shape-reproduction task; hence, the stimulus type seems to also impact which VWM measurement model fits best.

The pattern of our results that the SMM (Zhang & Luck, 2008) was favoured when substantial changes were present whereas the TCC (Schurgin et al., 2020) was preferred when changes were absent, provides three major implications for the debates on the nature of VWM representations as well as explanations on cognitive changes through intervention. Firstly, estimated parameters (capacity and precision) regarding the quantity and quality of VWM representations are dissociable if substantial changes have been induced. This idea is in line with previous intervention research where selective improvement in quantity or quality were observed (Ovalle Fresa & Rothen, 2019; Sutterer & Awh, 2016; S. Wang et al., 2019). Secondly, precision is limited that can be improved through intervention, such as brain stimulation (Heinen et al., 2016) or cognitive training (Jiang et al., 2023; Ovalle Fresa & Rothen, 2019). Thirdly, although, as suggested by the TCC model, perceptual familiarity can be a dominant source of imprecise memory, our results suggest it is not a main contributor to the observed substantial cognitive changes. To explain these substantial changes in VWM, it appears to be that a simple, generic view of changeable precision by SMM for the remembered feature information is sufficient, in comparison to more complex analysis of the limited precision by SwapM and SDM.

Our observation of the changes in model preference for data from the same task also warrants a cautious note for researchers who use computational models like these four VWM

models to examine training-induced changes. In principle, the choice of a measurement model often depends on researchers' measurement aims and theoretical assumptions. However, effects of study design, such as the number of testing sessions, conditions, and stimulus type, are typically neglected in modelling research but should also be considered when choosing a measurement model. More importantly, the fluctuations in model preference at the individual level could reflect dynamic changes in strategy use, which can be acquired through training or test-retest practice, as well as adjustments to different conditions, and type of stimulus.

In addition, together with the stimulus-specific training-induced benefits observed in Chapter 3, the pattern at the pre-test that the SMM and TCC fitted the data from the orientation-reproduction task best, whereas the TCC fared best only for the data from the shape-reproduction task, suggests a possible effect of stimulus type on how stimulus information is stored in memory. Indeed, processing orientation information and shape information in VWM involves different neural networks (Faillenot et al., 1997). Furthermore, the memory strength for the shape feature may rely more on the subjective similarity of shapes (like colours), while orientations benefit less from their similarity. Such a stimulus-specific effect is in line with the finding that substantial stimulus-specific variations in memory precision can distort the decision drawn from model comparison (Pratte et al., 2017).

4.4.1 Limitations and Future Research

One major limitation is that we determined the best-fitting model based on absolute minimum BIC values. However, for the data of some individuals, who probably tried out different strategies during the same session, might show minor differences in preferring different models. As a result, the BIC values for non-winning models might only be marginally bigger than that for the winning model, thus leading to ambiguous model comparison results for these individuals. Together with a relatively small sample size, this

uncertainty at the individual level might confound the overall pattern of our results. Future studies should include a bigger sample to enable accounting for these individual differences. A Bayesian hierarchical framework can be promising to address these limitations (Frischkorn & Popov, 2023); however, no feasible computational modelling existed at the time of the present data.

Another limitation is that the current comparison study focused only on un-cued, partial-report continuous-reproduction paradigms. Although we accounted for the stimulus-type (orientation vs shape), the possibility of paradigm-specific preferences on models cannot be excluded. Future research could test its generation to a different type of paradigm, for example, a pre-cued paradigm where participants are informed of the location of the target, or a whole-report paradigm where participants need to report all stimuli's feature instead of one probed item as the target. The former might lead to better fits of the cue-based binding models such as the SDM, while the latter might lead to better fits of swap-error featured models like the SwapM.

4.4.2 Conclusion

To the best of our knowledge, the current study is the first to quantitatively compare all four popular VWM measurement models together, aiming to provide explanations for cognitive changes through interventions like cognitive training. The overall pattern of results at the group level supports that SMM can best account for substantial cognitive changes. In contrast, when substantial changes are absent, TCC can capture VWM performance better. However, at the individual level, none of the current models can fully account for the changes induced by training or test-retest effects.

Chapter 5 – General Discussion

Contributions:

Shuangke Jiang (conceptualization, writing – original draft and editing)

Myles Jones (supervision and review)

Claudia von Bastian (supervision, review, and editing)

5.1 Thesis main findings

This thesis had three overarching goals addressing the replication and theory crises in the field of WM enhancement. The first goal was to empirically assess the replicability and robustness of the effects of tDCS on improving WM. The second goal was to investigate the mechanisms underlying cognitive changes by empirically testing predictions based on a capacity-efficiency theoretical model of training and transfer effects (von Bastian, 2021). The third goal was to evaluate a set of four existing VWM measurement models that were widely employed in the continuous-reproduction VWM paradigm, but their theoretical assumptions have been hotly debated.

As shown in Chapter 2, we did not observe any benefits of single-session, anodal parietal or prefrontal tDCS on VWM capacity and precision. In particular, we failed to replicate the selective, large effect of parietal tDCS on increasing VWM capacity at a big set size that was reported by Wang et al. (2019). In Chapter 3, we found stimuli-specific and paradigm-specific training gains in precision but not in the capacity of VWM representations. The pattern of our results supports that training gains are driven by a more efficient use of available cognitive capacity instead of expanding the existing capacity. In Chapter 4, we found that the Standard Mixture Model (Zhang & Luck, 2008) was the simplest and best-fitting model for our training data of the experimental group who showed substantial training-induced improvement while the Target Confusability Competition model was the simplest and best-fitting model when substantial cognitive changes were absent in a different group and a different WM task at the group level. However, at the individual level, none of the current models could fully account for the dynamic changes induced by training or training-unspecific effects.

5.2 Theoretical implications

The overall pattern of our empirical findings in Chapter 2 and Chapter 3 suggested that WM can be enhanced through interventions like WM training but not necessarily through a single session of tDCS. We investigated two possible pathways to enhance WM that were described in the capacity-efficiency theoretical models (von Bastian et al., 2021), and measured by the Standard Mixture Model (Zhang & Luck, 2008). Our findings speak against the first proposed pathway – through expanding capacity as a general resource – by showing the absence of changes in capacity estimates after both tDCS and training interventions. On the other hand, our findings support the second pathway – through enhancing efficiency in using existing capacity – by showing training-induced, but not tDCS-induced, enhancement in precision estimates.

Also, investigating the intervention-induced changes in WM can aid better understanding of WM with the examples of our two empirical studies in Chapter 2 and Chapter 3. Data from both studies was best fitted by the measurement Standard Mixture Model which stems from the *discrete-slot* class of explanatory models. This class of models assumes a fixed upper limit on the quantitative number of items that can be remembered. After reaching the capacity limit, the information is stored in an *all-or-none* manner, that is, with a certain precision or not at all. The pattern of absent changes in capacity in both studies supports the notion of the upper limit on capacity which is less likely to be altered by both tDCS and WM training. Although by WM training only, the quality can still be improved.

Nevertheless, the findings in Chapter 4 have warranted a cautious note for researchers, who use computational models like the four measurement models, Standard Mixture Model (Zhang & Luck, 2008, 2011), Swap Model (Bays et al., 2009; Schneegans & Bays, 2016), Signal Discrimination Model (Oberauer, 2021) and Target Confusability Competition model (Schurgin et al., 2020), to examine VWM changes induced by interventions like cognitive

training. Before any change occurs, perceptual familiarity may be a dominant source of imprecise memory, as assumed in the Target Confusability Competition model, but not a main contributor to substantial cognitive changes. When substantial changes are induced, the Standard Mixture Model is sufficient to explain these substantial changes in VWM. In the Standard Mixture Model, a simple generic view of limited precision of the remembered feature information is assumed. Nonetheless, different computational models and theoretical explanations are needed to fully account for dynamic individuals' changes in cognitive status. Therefore, our current theoretical-computational understanding of VWM still needs further refining, especially in WM enhancement research where cognitive changes are more likely to be induced.

5.4 Practical implications

Our findings in Chapter 2 highlighted the importance of replication studies investigating the robust tDCS effect on improving VWM. Often weak or inflated statistically positive effects of tDCS were reported with a small sample size in the field. These claims have motivated its therapeutic application to vulnerable populations like older adults with mild cognitive impairment or Alzheimer's disease. Accounting for possible caveats from the original study (Wang et al., 2019), our pre-registered replication included a bigger sample size, completed counterbalancing, a larger number of trials and a more challenging version of the task. The substantial evidence for a null effect that was observed in our replication study suggests that previously reported positive effects of tDCS and its therapeutic application need to be interpreted with caution without large-scale and multi-centre clinical trials. Thus, the replication approach from Chapter 2 directly addresses the replication crisis in the field and highlights the importance of pre-registering the difference between the original and replication studies.

Furthermore, our work in Chapter 3 has shown a need for shifting focus in the field from '*Can cognitive training improve WM or not*' onto '*Why cognitive training can (not) improve WM*', to explain the inconsistent benefits through training. Owing to the development of theoretical explanatory models of training and transfer effects (e.g., the capacity-efficiency model; von Bastian et al., 2021), we are able to test theoretically informed hypotheses about possible training gains. Our findings have further drawn attention, from the capacity-driven mechanism, to investigating the efficiency-driven mechanism of training-induced gains through the acquisition of expertise or strategies.

Last but not least, this thesis is characterised by investigating two types of interventions (i.e., tDCS and WM training) for studying WM enhancement. Our approach to combining cognitive enhancement research with computational measurement models of VWM in both Chapter 2 and Chapter 3 has strengthened the links between our hypotheses and the theory of VWM and its enhancement. This approach has contributed to not only addressing the theory crisis in the field but also facilitating the generalisability and replicability of interventions.

5.5 Limitations and future research

Despite these important implications, several general limitations need to be considered. One limitation of the employed tDCS technique in Chapter 2 – the low spatial resolution of conventional tDCS devices with two electrodes – can bring challenges to the assessment of the robust and replicable effects of tDCS. Conventional tDCS often results in widespread, diffuse modulation over the entire cortical surface owing to the large size and separation of electrode pads (Datta et al., 2009). The low spatial resolution translates into low precision when targeting brain regions corresponding to specific cognitive functions, preventing drawing consistent conclusions about the optimal parameters and causal nature of areas implicated in cognitive processes (Reinhart & Woodman, 2015). Together with individual differences in head anatomy and morphology of brain structures (e.g., sulci and gyri), this

may further make replications more challenging. Indeed, the effects of tDCS are highly variable, which also is sensitive to inter-individual differences in the tDCS-induced electric fields in the cortex (Laakso et al., 2019). In Chapter 2, we aimed to conceptually replicate the effects of such a conventional tDCS as reported in the original study by Wang et al. (2019). The low spatial resolution of tDCS and/or individual differences in head and brain anatomy may contribute to differences in the findings of tDCS studies. Future research should consider improving the spatial resolution and specificity of the stimulated head model. One way to improve the precision of tDCS is to use high-definition tDCS to modulate the exact neural networks that serve for the specific cognitive functions like VWM, augmenting computational models that simulate individual electric fields (Wischnewski et al., 2021).

In addition to these methodological challenges in replicating the tDCS effects, one single replication like in Chapter 2 does not disprove the general effect of tDCS (Simons, 2014). One single replication study can add information about the reliability of the originally reported effects by providing accumulative evidence to the field, until it reaches a consistent conclusion. This single replication from a different lab has inevitable differences like the characteristics of recruited participants and lab environment, which could potentially contribute to the inconsistent findings. Therefore, accounting for robust effects of tDCS across samples and even experimental settings is of great significance for future replication studies. One good, though effortful, practice to examine true tDCS effects is direct replications by multiple laboratories, which accounts for the reliability and generalisability through internal meta-analyses with a good control of the methodological differences (Elliott et al., 2021; Pavlov et al., 2021; Simons, 2014; Strzelczyk et al., 2023).

Another limitation is that we focused on behavioural data alone. In Chapter 3, we speculated that training gains are a result of the acquisition of stimuli-specific expertise (i.e., better sensory discrimination), possibly operating at the perceptual encoding stage. Moreover,

such expertise might boost the precision of VWM representations at the maintenance and recall stages. Confirming these hypotheses and gaining a better understanding of the stages at which these training-induced benefits occur will facilitate the implementation of more effective and transferable training interventions. To achieve this goal, future studies could combine the behavioural evidence with VWM-related EEG signals, such as contralateral delay activity, theta (4–8 Hz) and alpha (8–12 Hz) oscillations.

Another limitation in Chapter 3 was that we did not measure directly which strategies participants used, which could also impact the interpretation of our findings reported in Chapter 4. Although we inspected the use of more cardinal orientations after training an orientation reproduction task, we were not able to directly differentiate strategies that were used by individuals between the trained, and untrained tasks with a different stimulus or paradigm type. This could limit our further understanding of the efficiency-driven training gains that were observed in Chapter 3, that is, whether these training gains are a product of acquired strategies that can be transferred to a different stimulus or paradigm type. At the same time, we were not able to identify the changes of strategy use during training, which could potentially provide evidential explanations on why a different measurement model is needed by individuals from one training session to another in Chapter 4.

A general limitation of model comparisons is the inevitably arbitrariness of model selection criteria. We chose the overall optimal measurement model when training-induced was present among existing well-documented models at the group level based on Bayesian information criterion (BIC). BIC is theoretically motivated by the notion of a 'true' model among model candidates, so that, with increasing sample size, the probability of the 'true' model being selected also increases (Aho et al., 2014; Vrieze, 2012). We used the BIC and assumed that one of the possible models could so far provide the best explanation on training-induced changes in Chapter 4. Although we only chose our model candidates as they are

widely used and applied to the continuous-reproduction paradigm, our study cannot exclude the possibility that other models also or better account for these cognitive changes. However, it is known that "*Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful.*" (Box & Draper, 1987, p. 424). One solution for future research to balance it could be conducting factorial comparisons by combining the competing factors from all four models to increase the possible model space (van den Berg et al., 2014; Oberauer 2021).

Moreover, another general limitation in terms of the choice of model comparison method exists. Our measurement aims were to evaluate which model would best detect training-induced changes in VWM which could possibly be generalised to different samples of data, like from different training programmes and even other intervention studies. A more conservative model comparison method like BIC suits our aims better (Evans, 2019). However, researchers who have different measurement aims might need a different model comparison method that could lead to different conclusions. For example, for those who are interested in which model can best predict the source of recall errors in VWM and believe the reality is more complex than the current model candidates, a more liberal method like Akaike information criterion (AIC) may be preferred. AIC enables detection of small effects that are often compromised by BIC. As a result, a more complex model might be preferred. Indeed, when based on AIC (as reported in supplementary materials in Chapter 4), overall, a more complex model, SDM, would fit the same data sample best. Still, it is an open question how to address such model selection uncertainty due to the choice of model comparison methods in the field (Burnham & Anderson, 2004; Yang, 2005). Future work could at least pre-register their assumptions and measurement aims in order to avoid similar confusions that we encountered.

5.6 Conclusion

Notwithstanding the above limitations, this thesis demonstrated a case study of addressing the replication and theory crisis in the field of WM enhancement. We highlighted the importance of pre-registered replication studies and theory-informed hypothesis testing. In contrast to previously reported positive effects, substantial evidence from our study favoured the absence of any cognitive benefits after tDCS regardless of stimulation site and task difficulty. Considering the complexity of tDCS parameters and setups, our null findings highlight the critical importance of conducting replications for building a robust and informative body of evidence on the effectiveness of non-invasive brain stimulation on WM enhancement. Also, this thesis supports the notion that training enhances cognitive efficiency through the acquisition of expertise or strategies, but training is unlikely to expand capacity *per se*. A better understanding of how training facilitates a more efficient use of the available VWM capacity, and how the underlying training benefits are influenced by the characteristics of stimuli and paradigms, will be critical for harnessing the potential training benefits. Last but not least, this thesis demonstrated that none of the current existing VWM measurement models meet the need for capturing dynamic cognitive changes as observed through WM interventions. A new or improved VWM measurement model with an explanatory theoretical framework (e.g., efficiency-driven mechanisms) that also fully accounts for the efficiency-driven WM training benefits, is required.

Appendix A

Post Stimulation Ratings – Session 1 / session 2

Participant ID:_____

Please rate your "current" levels of pain, attention and fatigue.

Minimal Pain Maximal Pain

I----------I----------I----------I----------I----------I----------I

1 2 3 4 5 6 7

Poorest Attention Maximal Attention

I----------I----------I----------I----------I----------I----------I

1 2 3 4 5 6 7

Minimal Fatigue Maximal Fatigue

I----------I----------I----------I----------I----------I----------I

1 2 3 4 5 6 7

Appendix B

tDCS Adverse Effects Questionnaire:

Session:

ID:

Do you experience any of the following symptoms or side-effects?	Enter a value (1–4) in the space below (1, absent; 2, mild; 3, moderate; 4, severe)	If present: Is this related to tDCS? (1, none; 2, remote; 3, possible; 4, probable; 5, definite)
Headache		
Neck pain		
Scalp pain		
Tingling		
Itching		
Burning sensation		
Skin redness		
Sleepiness		
Trouble concentrating		
Acute mood change		
Others (specify)		

Appendix C

For each of the sessions you have completed you received either real or sham* tDCS.

Please circle the appropriate answer.

Which tDCS do you think you received during session 1? Real / Sham / I am not sure

Which tDCS do you think you received during session 2? Real / Sham / I am not sure

Which tDCS do you think you received during session 3? Real / Sham / I am not sure

* The purpose of sham stimulation was to act as a control (placebo), ensuring participants experienced a similar itching feeling that receded over the first few seconds of active stimulation. Sham stimulation lasted for a few seconds from the initial time of the stimulation.

Appendix D

Table D1

Individual Head Model

Participant	NI	TT	C
1	38.00	35.00	61.50
2	34.00	37.00	55.00
3	34.00	37.00	56.50
4	32.00	32.50	54.00
5	36.00	35.00	56.00
6	37.00	37.00	59.00
7	35.00	33.50	55.00
8	36.00	34.00	58.00
9	35.00	34.00	55.00
10	36.00	36.50	56.00
11	33.00	36.00	55.50
12	35.00	34.50	57.80
13	34.00	36.50	56.50
14	36.50	37.50	58.00
15	37.00	36.50	60.50
16	32.50	32.50	54.00
17	33.00	36.00	57.00
18	32.00	33.00	54.00
19	36.00	37.50	58.00
20	38.00	37.00	60.00
21	35.50	35.50	57.00
22	33.00	33.00	54.00
23	35.50	35.00	58.00
24	33.50	35.00	57.50
25	32.00	35.00	53.00
26	34.00	37.00	55.00
27	37.00	36.00	56.50
28	35.00	35.00	55.00
29	34.00	35.00	56.00
30	32.00	34.00	51.50
31	35.00	34.00	57.00
32	32.00	34.00	54.50
33	36.00	37.00	56.00
34	32.00	34.00	53.00
35	34.00	35.00	52.00
36	32.00	35.00	53.50
37	35.00	38.00	56.50
38	34.00	35.00	56.00
39	36.00	37.00	58.00
40	35.00	37.00	55.50
41	32.00	34.00	53.50
42	37.00	35.00	59.00
43	36.00	35.00	56.50
44	33.00	35.00	54.00
45	36.00	36.00	56.50
46	34.00	35.00	53.50
47	35.00	36.00	58.00

Table D1

Individual Head Model

Participant	NI	TT	C
48	34.00	33.00	55.00

Note. NI = length between nasion and inion (cm): TT = length from left tragus to right tragus (cm); C = circumference length (cm).

Appendix E

Table E1

Differences in Relative Akaike Information Criterion Values (ΔAIC)

Participant	Sham			PPC			DLPFC		
	Set size								
	2	4	6	2	4	6	2	4	6
1	-2.00	2.72	-2.00	-2.00	10.61	1.10	-2.01	7.92	5.65
2	-2.00	12.18	11.40	-2.00	4.04	3.65	-2.00	6.36	5.64
3	-2.06	-2.00	0.11	-2.25	-1.34	3.03	-2.00	-2.00	0.48
4	2.17	3.14	-0.27	-0.52	-2.00	14.15	-2.00	0.08	-0.35
5	-2.00	-1.74	-1.20	-2.00	-1.96	1.78	-2.00	-0.88	4.65
6	-2.00	-1.69	2.61	-2.00	2.10	-0.53	-2.00	2.07	-0.69
7	-2.00	1.13	-2.00	2.93	-1.84	-1.14	2.64	1.67	-1.80
8	-1.55	0.94	-0.99	-1.14	1.00	5.69	-2.00	0.02	-2.00
9	2.21	2.19	-1.14	-1.45	2.37	-2.00	-2.00	2.76	-0.72
10	-2.00	-1.81	-0.93	-2.00	-0.69	8.61	5.36	0.97	-1.04
11	-1.87	3.82	-1.70	-2.00	-0.83	-1.65	-2.00	3.94	-1.75
12	-2.00	-1.79	-1.99	-2.00	-0.71	18.69	-2.07	-1.33	-1.68
13	1.96	0.53	5.94	-1.66	6.71	0.11	1.20	4.83	1.38
14	-2.00	-1.96	-1.08	-1.16	-0.91	-1.46	-2.00	-0.32	-0.34
15	-2.00	10.32	10.30	-2.00	0.94	16.73	0.77	7.78	28.49
16	4.75	-2.00	-1.91	2.24	-2.00	0.40	-0.29	-2.00	6.84
17	-1.71	-2.00	-2.00	-2.00	7.14	1.83	-1.73	-1.89	-1.94
18	0.91	1.93	1.06	-2.00	0.89	1.25	0.22	5.11	7.25
19	-2.00	-2.00	-1.32	-2.00	1.43	-1.84	-2.00	0.19	-1.49
20	-2.00	-2.00	-2.00	-2.00	2.02	0.78	-2.00	1.52	-0.87
21	-2.00	-0.26	-0.17	-2.00	-1.22	5.03	-2.00	2.58	-2.00
22	-1.81	-1.83	-2.00	3.18	-2.00	-2.00	-2.00	6.07	1.65
23	-2.00	0.18	-0.10	2.60	1.10	1.61	-2.00	6.06	3.12
24	-2.00	3.73	-0.10	-2.00	-2.00	-2.00	-2.01	-1.37	12.07
25	-2.00	0.47	1.92	-2.00	7.94	4.18	-2.00	1.71	3.35
26	-2.01	7.04	5.67	-2.00	-2.00	7.59	-2.00	-0.96	1.29
27	-2.00	8.34	12.27	-2.00	-0.92	9.94	-2.00	0.22	-2.00
28	7.34	6.74	2.95	-2.00	-1.02	4.87	-2.00	-2.00	0.27
29	-2.00	-2.00	-2.00	-2.00	-2.00	1.03	-2.00	-0.05	-2.00
30	-1.23	-0.39	7.61	-1.96	10.18	5.79	-2.00	2.80	7.68
31	-1.12	-1.44	0.54	-1.47	-1.73	5.46	-2.00	-0.02	-0.41
32	-2.00	0.17	-1.98	-2.00	-2.00	-1.94	-2.00	-0.71	1.12
33	-1.35	-2.00	-2.00	-2.00	7.70	-1.82	-2.00	1.19	-1.07
34	-2.00	-1.44	-2.00	-2.00	-2.00	7.48	-1.99	-2.00	1.42
35	2.55	-1.97	-1.57	2.31	-0.89	2.84	-2.00	-1.77	-2.00
36	3.97	-1.34	-1.88	-2.00	9.42	2.22	1.35	1.32	-0.96
37	-0.39	7.62	-1.14	-2.06	0.76	2.46	-0.75	2.48	0.92
38	-0.38	3.43	1.23	-2.00	-1.83	-2.00	9.43	-0.48	1.72
39	-2.01	0.62	0.10	-2.11	-2.00	-1.08	-2.00	-2.00	-0.77
40	-2.00	-2.00	-2.00	-2.01	-0.44	-2.00	-2.00	4.28	3.66
41	-1.86	-1.35	-1.78	-2.00	-0.34	0.73	3.27	-2.00	1.97
42	-2.00	-0.75	-1.86	-2.01	9.60	3.78	-2.00	4.18	5.34
43	-2.00	1.03	2.22	3.27	-1.69	0.85	-1.93	-1.88	2.38
44	-2.00	-2.00	-1.33	-2.02	-1.77	-0.78	-1.14	-1.74	-2.00
45	-1.98	11.99	2.68	-2.12	0.41	9.60	-2.00	-0.79	4.17

Table E1

Differences in Relative Akaike Information Criterion Values (ΔAIC)

Participant	Sham			PPC			DLPFC		
	Set size								
	2	4	6	2	4	6	2	4	6
46	-2.00	3.93	-0.46	4.81	8.39	2.08	3.88	1.25	-1.90
47	-2.00	3.27	-2.00	-2.01	-1.66	-1.96	-2.00	-2.00	2.28
48	-1.24	-2.00	1.03	-2.00	0.91	-2.00	1.25	-1.38	-1.92

Note. Negative values favours SMM while positive values favours SM. The lower value represents a better model fit. SMM, standard mixture model; SM, swap model.

Table E2

Differences in Relative Bayesian Information Criterion Values (ΔBIC)

Participant	Sham			PPC			DLPFC		
	Set size								
	2	4	6	2	4	6	2	4	6
1	-6.63	-1.91	-6.63	-6.63	5.98	-3.53	-6.63	3.30	1.02
2	-6.63	7.55	6.78	-6.63	-0.59	-0.98	-6.63	1.73	1.01
3	-6.69	-6.63	-4.52	-6.88	-5.96	-1.60	-6.63	-6.63	-4.15
4	-2.46	-1.48	-4.90	-5.15	-6.63	9.53	-6.63	-4.54	-4.97
5	-6.63	-6.36	-5.83	-6.63	-6.58	-2.84	-6.63	-5.50	0.02
6	-6.63	-6.31	-2.02	-6.63	-2.52	-5.15	-6.63	-2.56	-5.32
7	-6.63	-3.50	-6.63	-1.69	-6.47	-5.77	-1.98	-2.96	-6.43
8	-6.17	-3.68	-5.62	-5.76	-3.63	1.06	-6.63	-4.61	-6.63
9	-2.41	-2.43	-5.77	-6.08	-2.25	-6.63	-6.63	-1.87	-5.34
10	-6.63	-6.44	-5.56	-6.63	-5.32	3.99	0.74	-3.66	-5.66
11	-6.49	-0.80	-6.32	-6.63	-5.46	-6.28	-6.63	-0.68	-6.37
12	-6.63	-6.41	-6.62	-6.63	-5.33	14.07	-6.70	-5.95	-6.31
13	-2.67	-4.10	1.32	-6.28	2.09	-4.52	-3.43	0.21	-3.25
14	-6.63	-6.59	-5.70	-5.79	-5.54	-6.09	-6.63	-4.94	-4.96
15	-6.63	5.69	5.67	-6.63	-3.69	12.10	-3.86	3.15	23.87
16	0.13	-6.63	-6.53	-2.38	-6.63	-4.23	-4.92	-6.63	2.22
17	-6.33	-6.62	-6.63	-6.63	2.51	-2.80	-6.36	-6.51	-6.56
18	-3.72	-2.70	-3.57	-6.63	-3.74	-3.37	-4.40	0.48	2.62
19	-6.63	-6.63	-5.94	-6.63	-3.19	-6.47	-6.63	-4.43	-6.11
20	-6.63	-6.62	-6.62	-6.63	-2.61	-3.84	-6.63	-3.10	-5.50
21	-6.63	-4.88	-4.80	-6.63	-5.84	0.41	-6.63	-2.04	-6.63
22	-6.43	-6.46	-6.63	-1.44	-6.63	-6.63	-6.63	1.44	-2.97
23	-6.63	-4.44	-4.72	-2.03	-3.52	-3.01	-6.63	1.44	-1.50
24	-6.63	-0.89	-4.72	-6.63	-6.63	-6.63	-6.63	-5.99	7.45
25	-6.63	-4.15	-2.70	-6.63	3.31	-0.44	-6.63	-2.92	-1.28
26	-6.63	2.41	1.04	-6.63	-6.63	2.97	-6.63	-5.58	-3.34
27	-6.63	3.72	7.65	-6.63	-5.54	5.31	-6.63	-4.41	-6.63
28	2.71	2.11	-1.67	-6.63	-5.65	0.25	-6.63	-6.63	-4.36
29	-6.63	-6.63	-6.63	-6.63	-6.62	-3.60	-6.63	-4.68	-6.62
30	-5.85	-5.02	2.98	-6.59	5.56	1.17	-6.63	-1.83	3.06
31	-5.75	-6.06	-4.09	-6.10	-6.36	0.84	-6.63	-4.64	-5.03
32	-6.63	-4.46	-6.60	-6.63	-6.63	-6.57	-6.63	-5.33	-3.51
33	-5.98	-6.63	-6.63	-6.63	3.07	-6.44	-6.63	-3.44	-5.69
34	-6.63	-6.07	-6.63	-6.63	-6.63	2.86	-6.62	-6.62	-3.21
35	-2.07	-6.60	-6.20	-2.32	-5.52	-1.79	-6.62	-6.39	-6.63
36	-0.66	-5.97	-6.50	-6.63	4.80	-2.41	-3.27	-3.30	-5.58
37	-5.01	3.00	-5.76	-6.68	-3.86	-2.17	-5.37	-2.14	-3.70
38	-5.01	-1.19	-3.40	-6.63	-6.46	-6.63	4.81	-5.11	-2.91
39	-6.63	-4.00	-4.53	-6.74	-6.63	-5.70	-6.63	-6.63	-5.40
40	-6.63	-6.63	-6.63	-6.63	-5.06	-6.63	-6.63	-0.35	-0.96
41	-6.49	-5.97	-6.40	-6.63	-4.97	-3.90	-1.36	-6.63	-2.66
42	-6.63	-5.37	-6.48	-6.64	4.97	-0.84	-6.63	-0.44	0.71
43	-6.63	-3.60	-2.41	-1.35	-6.32	-3.78	-6.56	-6.51	-2.24
44	-6.63	-6.63	-5.95	-6.65	-6.40	-5.40	-5.76	-6.36	-6.63
45	-6.60	7.36	-1.95	-6.75	-4.21	4.97	-6.63	-5.41	-0.45

Table E2

Differences in Relative Bayesian Information Criterion Values (ΔBIC)

Participant	Sham			PPC			DLPFC		
	Set size								
	2	4	6	2	4	6	2	4	6
46	-6.63	-0.69	-5.08	0.19	3.77	-2.54	-0.75	-3.37	-6.53
47	-6.63	-1.36	-6.63	-6.63	-6.29	-6.58	-6.63	-6.63	-2.34
48	-5.87	-6.63	-3.60	-6.63	-3.72	-6.63	-3.38	-6.01	-6.54

Note. Negative values favours SMM while positive values favours SM. The lower value represents a better model fit. SMM, standard mixture model; SM, swap model.

Appendix F

Supplementary analyses for Chapter 2 – Does transcranial direct current stimulation enhance visual working memory? A replication study

Blindness and post-stimulation rating

To ensure participants were blind to the condition, we examined their self-reported guesses about which stimulation condition they thought they had received by running one-sample Wilcoxon tests, as the data violated the assumption of normality. For the sham condition, guessing accuracy was 52.10%, which was not significantly different from chance guessing of 50%, $V = 612.50$, $p = 0.78$, $r = 0.04$, $BF_{01} = 6.14 \pm 0.06\%$. This result confirms that overall participants were not able to differentiate the sham stimulation from active stimulations, indicating our blinding was effective in preventing placebo effects.

We also investigated the level of pain, attention, and fatigue after each stimulation. We found that only pain level after DLPFC stimulation ($M = 1.83$, $SD = 1.10$) was significantly higher than that after sham ($M = 1.33$, $SD = 0.72$), $t(47) = 3.07$, Bonferroni-adjusted $p = .011$, Cohen's $d = 0.44$, $BF_{10} = 9.29 \pm 0.00\%$. However, this did not affect their overall cognitive performance. Note that there was no significant difference between stimulation conditions regarding the levels of attention (Bonferroni-adjusted p ranged 0.067 to 1.000) and fatigue (Bonferroni-adjusted p values were 1.000).

Adverse tDCS Effects

Overall, the current tDCS setup did not lead to any severe adverse effect, which indicates the safety of the montage and paradigms (Table F1). The most common adverse events reported in the total of 144 sessions (3 sessions for each participant) that were related to tDCS were tingling (58.3% of all experimental sessions) and itching (45.9%) sensations, followed by skin redness (30.5%)

Table F1

Adverse Effects Self-Reported After Stimulation and Sham Sessions

Adverse effect	In general (%)	tDCS-related (%)
Tingling sensation	61.8	58.3
Itching sensation	52.1	45.9
Skin redness	34.7	30.5
Headache	15.3	10.4
Scalp pain	13.2	12.5
Neck pain	5.6	2.1
Concentration problems	14.6	7.6
Sleepiness	36.1	13.2
Acute mood change	4.9	3.5

Note. Percentages refer to proportions of 144 sessions (3 sessions for each participant).

Appendix G

Table G1

Akaike Information Criterion Values Relative to The Winning Model for The Experimental Group in Orientation Reproduction Task

ID	Session													
	Pre-Test	T1			T2			T3			T4			Post-Test
	Set Size													
	4	2	4	6	2	4	6	2	4	6	2	4	6	4
	Standard Mixture Model													
1	0.00	0.00	11.13	6.89	0.81	11.90	5.23	0.00	0.00	13.48	10.17	3.09	1.17	2.70
5	3.90	2.06	11.07	3.14	1.37	2.58	0.00	3.54	1.46	9.28	15.96	2.80	1.95	5.20
8	2.98	0.00	2.54	3.31	8.25	8.71	0.00	0.00	2.29	1.10	10.48	2.06	1.10	0.00
10	0.00	0.14	0.00	5.38	0.00	0.00	0.00	1.79	0.95	0.80	0.25	0.00	1.12	0.00
11	0.00	5.04	8.05	0.32	2.26	3.20	2.11	0.00	0.00	0.00	3.83	0.00	5.27	0.48
12	5.51	0.00	0.00	8.65	0.00	1.75	0.00	4.91	0.00	0.00	1.38	0.00	0.00	2.98
14	7.78	1.47	18.05	10.22	16.62	6.93	0.72	5.94	0.00	2.73	0.90	12.97	2.11	0.63
17	2.08	1.01	7.34	2.41	1.36	1.34	15.37	0.00	2.44	0.66	0.00	0.00	3.57	0.00
18	3.00	2.36	7.58	10.28	0.00	2.71	2.76	1.98	0.00	0.00	12.18	0.25	0.00	6.53
19	0.00	0.00	0.00	0.00	0.00	0.00	3.59	0.00	0.00	0.00	6.95	0.00	0.00	0.00
20	3.65	3.21	8.12	0.00	2.31	17.40	0.00	0.00	14.51	0.00	0.00	9.70	7.43	0.00
21	1.99	1.31	0.55	1.59	4.51	2.83	0.00	4.17	0.00	4.68	0.00	0.00	2.29	0.83
22	3.24	0.00	1.82	5.76	0.63	12.46	0.00	5.33	6.64	1.14	0.00	0.27	3.21	0.98
23	0.83	3.39	8.15	2.79	0.00	6.70	0.44	0.00	0.00	7.21	0.00	0.00	3.47	0.00
24	10.01	3.21	5.68	0.00	3.37	5.32	0.92	7.45	0.18	1.90	1.23	0.00	0.00	0.00
25	4.45	3.90	2.04	0.00	0.72	1.29	2.37	0.90	3.53	8.60	0.00	0.00	0.00	2.94
26	0.00	0.00	0.67	0.00	0.00	0.00	11.56	13.14	0.82	4.80	0.00	0.00	0.09	0.80
27	7.45	6.67	10.18	3.54	4.89	12.50	0.00	12.73	0.00	1.79	7.87	0.69	1.99	30.54
29	7.79	16.75	0.00	6.26	5.44	0.00	15.36	6.91	1.30	7.22	20.92	6.43	0.00	8.61
30	16.18	4.23	1.06	1.46	3.04	0.00	1.64	5.58	0.00	7.66	12.85	0.83	21.51	5.14
32	0.00	0.00	0.07	0.00	0.00	0.30	1.26	14.78	0.94	0.00	0.00	7.23	0.00	0.85
33	6.44	0.17	3.11	1.91	9.46	9.27	5.75	1.87	3.12	1.88	2.02	4.50	6.93	1.32
34	0.44	0.00	3.70	3.75	0.00	0.00	0.00	1.29	11.04	0.00	0.00	2.24	4.98	7.08
36	0.00	0.00	0.00	0.00	0.00	0.00	9.04	0.00	4.53	4.46	0.00	0.00	0.00	0.07
37	0.00	0.00	1.73	0.00	0.00	1.81	0.00	4.62	0.00	0.00	0.00	0.47	1.99	0.00
39	0.00	0.00	4.00	6.85	8.65	0.00	2.43	3.50	0.00	2.27	0.00	0.00	0.73	2.21
40	23.31	7.10	0.00	18.65	7.84	0.00	4.08	8.28	1.13	3.46	5.33	0.00	1.94	21.01
42	5.78	3.34	0.00	3.77	3.93	0.00	0.87	3.87	4.07	0.00	0.40	0.27	1.36	2.29
44	1.93	0.53	1.22	4.66	0.00	0.00	15.25	0.00	9.31	0.00	6.69	3.25	1.43	9.73
46	42.20	31.74	4.43	13.71	8.40	1.49	12.69	0.92	1.85	1.30	14.68	0.00	13.60	0.00
Sum	160.94	97.62	122.30	125.31	93.86	110.48	113.44	113.49	70.09	86.43	134.08	57.05	89.23	112.92
	Swap Model													
1	1.00	0.54	13.13	5.54	0.00	13.90	6.75	2.00	2.00	0.00	3.45	5.09	0.00	4.64
5	0.00	2.50	11.46	0.00	0.00	4.57	1.47	5.54	1.53	2.91	4.99	3.48	2.80	6.78

Table G1

Akaike Information Criterion Values Relative to The Winning Model for The Experimental Group in Orientation Reproduction Task

ID	Session													
	Pre-Test	T1			T2			T3			T4			Post-Test
	Set Size													
	4	2	4	6	2	4	6	2	4	6	2	4	6	4
8	0.00	2.00	4.54	0.30	5.84	4.52	0.03	2.00	3.38	0.81	3.53	3.99	0.00	2.00
10	1.31	0.00	1.95	7.29	1.31	2.00	2.00	3.79	0.00	2.46	2.15	1.29	3.10	0.45
11	2.00	6.68	8.04	2.20	0.00	5.20	0.00	2.00	2.00	1.46	0.02	2.00	2.97	2.09
12	4.75	2.00	2.00	8.62	1.39	3.75	0.63	6.91	2.00	2.00	0.00	2.00	0.42	3.77
14	1.80	0.00	20.05	8.14	3.77	8.93	0.88	0.00	2.00	4.73	2.35	11.81	0.77	1.85
17	0.00	0.00	5.18	1.91	0.94	3.34	13.85	1.54	4.44	2.40	2.00	2.00	4.27	1.64
18	5.00	1.87	0.00	5.35	0.62	4.43	2.88	0.53	1.54	0.65	0.06	0.00	1.99	0.00
19	2.00	2.00	2.00	2.00	1.51	2.00	2.48	1.59	2.00	2.00	0.00	2.00	2.00	2.00
20	4.02	5.21	4.94	2.00	4.05	19.40	1.99	2.00	13.63	2.00	2.00	11.70	6.29	2.00
21	0.38	2.53	1.66	0.00	6.51	1.22	1.15	3.06	2.00	6.02	1.69	2.00	0.00	2.83
22	0.00	2.00	2.73	0.00	0.72	9.45	1.84	7.31	8.54	0.00	2.00	1.63	5.21	0.00
23	0.00	3.94	9.31	4.79	2.00	0.00	2.43	2.00	2.00	2.53	1.99	0.16	0.04	2.00
24	0.00	1.56	2.39	1.83	0.00	6.05	1.42	0.00	2.18	0.00	2.83	2.00	2.00	1.97
25	6.45	1.88	3.58	1.08	0.75	3.12	3.15	0.27	5.51	0.00	1.72	1.29	1.86	2.91
26	1.62	1.98	2.67	0.69	0.77	2.00	6.60	5.20	2.82	0.00	0.12	2.00	0.00	1.47
27	0.00	6.13	5.27	0.00	0.00	8.56	1.85	6.12	2.00	3.47	0.00	2.69	0.18	0.00
29	5.58	0.00	2.00	3.40	0.00	2.00	9.95	5.02	3.30	8.12	8.75	8.43	2.00	9.97
30	10.82	2.55	1.28	1.90	1.68	2.00	1.15	0.00	1.99	8.11	11.03	2.63	13.04	1.90
32	0.77	2.00	2.07	1.94	1.99	0.00	2.73	0.00	2.94	0.67	1.62	8.14	2.00	1.63
33	0.00	0.00	5.11	3.79	8.89	11.27	6.79	0.00	5.12	2.30	0.00	6.50	6.71	0.00
34	2.44	0.93	3.39	3.52	1.33	2.00	2.00	3.29	12.49	0.58	2.00	2.94	5.24	2.94
36	1.36	0.94	2.00	1.73	2.00	2.00	0.00	1.86	4.02	4.80	2.00	2.00	2.00	1.73
37	1.28	1.51	3.67	1.17	1.14	2.05	2.00	6.62	0.36	1.87	1.71	0.00	3.91	1.93
39	2.00	1.73	6.00	3.83	0.00	2.00	0.98	0.00	1.94	2.85	0.52	2.00	2.69	3.56
40	0.00	4.66	0.46	12.37	0.92	2.00	3.18	0.00	3.13	3.94	2.90	2.00	2.34	8.16
42	0.00	0.00	2.00	4.13	0.00	2.00	0.00	5.87	6.07	2.00	1.78	0.00	2.99	2.46
44	1.20	0.00	3.22	2.69	1.85	2.00	9.12	2.00	8.98	2.00	1.30	3.55	0.00	4.60
46	8.08	0.00	5.05	0.00	0.00	2.93	2.12	1.84	3.25	1.77	6.05	1.99	2.79	2.00
Sum	63.86	57.15	137.16	92.21	49.97	134.71	91.40	78.35	113.16	72.45	70.55	97.31	79.62	79.26
	Signal Discrimination Model													
1	1.00	0.85	0.00	0.00	2.10	0.93	0.00	1.40	3.29	0.64	0.00	0.00	0.95	0.00
5	2.65	0.00	0.00	1.90	1.99	0.00	0.42	6.34	0.00	0.00	0.00	0.00	1.34	0.00
8	1.51	2.02	0.00	0.00	0.00	0.00	0.63	3.41	0.00	0.00	0.00	0.00	0.29	5.11
10	3.43	3.39	5.48	7.21	1.89	2.02	4.07	5.58	1.34	3.71	2.80	2.82	5.01	0.07
11	2.46	0.51	0.00	0.46	0.10	0.00	4.02	3.60	1.84	0.81	0.00	1.23	0.00	0.00

Table G1

Akaike Information Criterion Values Relative to The Winning Model for The Experimental Group in Orientation Reproduction Task

ID	Session													
	Pre-Test	T1			T2			T3			T4			Post-Test
	Set Size													
	4	2	4	6	2	4	6	2	4	6	2	4	6	4
12	0.00	0.61	0.84	0.00	5.69	0.00	1.31	0.00	3.51	5.02	3.24	5.13	3.46	0.00
14	0.00	1.24	0.00	0.00	0.00	0.00	0.00	4.51	1.01	0.00	0.00	0.00	0.00	0.00
17	3.30	0.62	0.00	0.00	0.00	0.00	0.00	0.68	2.77	0.00	4.69	0.70	4.07	5.84
18	6.32	0.00	5.37	0.00	0.65	0.00	0.00	0.00	7.91	1.93	0.00	0.59	4.05	1.91
19	6.90	3.13	0.40	5.87	4.95	3.80	0.00	0.93	3.76	3.69	6.29	7.59	5.46	4.18
20	0.37	1.49	0.00	1.94	0.00	0.00	3.40	1.91	0.00	1.91	2.61	0.00	0.00	1.30
21	0.00	3.04	0.00	1.56	5.27	0.00	3.34	0.00	3.67	2.64	3.72	2.35	2.05	4.14
22	1.60	6.15	0.00	0.82	0.00	0.00	5.27	0.00	0.00	1.74	4.67	0.00	4.05	0.17
23	0.57	0.00	0.00	0.56	3.54	5.72	0.00	7.39	1.52	0.00	1.36	4.69	0.00	3.84
24	2.12	3.51	0.00	2.16	0.46	0.00	0.00	2.57	0.00	2.30	0.00	3.37	3.92	2.89
25	5.53	0.00	0.00	2.59	2.04	3.87	5.27	0.00	0.00	0.60	0.63	0.49	0.21	0.00
26	5.84	6.59	0.00	4.27	2.93	0.13	0.00	0.00	0.10	3.02	3.19	3.45	4.50	0.00
27	5.00	7.86	0.00	0.03	3.38	0.00	3.09	0.00	7.55	3.69	4.52	0.00	1.25	5.02
29	2.50	3.89	1.88	0.00	4.89	0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.91	0.00
30	0.00	0.00	0.00	3.95	0.00	7.94	0.00	0.56	3.80	0.00	0.00	0.00	0.00	0.00
32	4.90	3.60	0.00	6.07	2.39	4.08	0.00	11.77	0.00	1.65	1.02	3.87	2.22	0.00
33	2.79	4.10	0.00	0.00	0.00	0.00	0.00	1.76	0.00	0.00	6.12	0.00	0.00	4.24
34	4.67	2.10	0.00	0.00	3.67	5.61	3.62	3.97	0.00	2.77	1.33	0.00	6.80	2.47
36	4.98	0.76	2.98	1.85	3.76	4.07	0.28	2.87	0.00	0.00	5.20	0.82	6.56	0.00
37	0.28	1.65	0.00	3.83	5.53	0.00	3.85	7.44	4.76	2.36	4.53	5.10	5.66	3.11
39	5.11	1.36	0.00	0.00	3.96	1.61	0.00	5.04	5.03	1.68	0.72	5.49	3.02	4.38
40	1.68	0.00	6.13	0.00	0.00	7.04	0.00	9.08	0.00	0.00	0.00	1.10	0.00	0.00
42	3.30	3.65	3.03	0.00	1.56	0.58	3.06	6.00	0.00	3.35	0.00	0.00	4.33	0.00
44	0.00	2.47	0.00	0.00	1.28	3.69	0.00	3.57	0.00	1.35	0.00	0.00	1.71	0.00
46	0.00	0.74	0.00	1.59	0.49	0.00	0.00	0.49	0.00	2.37	0.00	2.71	0.00	4.22
Sum	78.79	65.32	26.10	46.65	62.50	51.27	41.60	90.86	51.86	47.20	56.64	51.50	71.83	52.90
	Target Confusability Competition Model													
1	11.26	9.17	36.20	12.31	9.80	0.00	20.60	7.28	20.54	16.95	18.69	24.40	5.17	5.01
5	3.73	4.76	35.30	2.93	8.75	14.40	2.57	0.00	23.66	7.45	21.88	32.48	0.00	13.40
8	2.87	0.50	6.19	7.93	4.97	2.55	6.42	5.42	4.32	14.68	7.78	2.66	6.85	8.66
10	0.91	5.50	10.96	0.00	0.98	17.72	0.28	0.00	3.23	0.00	0.00	9.27	0.00	2.06
11	5.88	0.00	2.59	0.00	16.26	15.35	6.97	32.08	11.40	5.95	12.95	9.32	7.43	10.96
12	7.35	6.20	7.54	15.80	19.49	13.03	7.52	8.43	16.88	13.83	15.54	23.11	21.44	17.46
14	13.20	13.27	30.13	21.65	20.50	19.68	14.77	19.98	13.38	13.39	24.75	10.11	6.42	6.84
17	9.83	5.87	20.62	0.83	15.43	15.96	24.61	1.82	0.00	4.34	9.37	11.81	0.00	15.54

Table G1

Akaike Information Criterion Values Relative to The Winning Model for The Experimental Group in Orientation Reproduction Task

ID	Session													
	Pre-Test	T1			T2			T3			T4			Post-Test
	Set Size													
	4	2	4	6	2	4	6	2	4	6	2	4	6	4
18	0.00	4.85	28.75	12.00	5.51	21.02	1.79	16.33	42.21	9.66	18.13	15.21	1.10	6.66
19	17.74	13.82	23.88	16.60	18.06	20.10	18.08	9.09	18.31	14.15	30.93	21.87	11.58	6.24
20	0.00	0.00	10.10	8.28	16.86	20.13	14.70	3.75	30.32	23.34	11.79	14.64	28.29	5.33
21	12.07	0.00	16.79	2.96	0.00	6.08	1.16	2.51	12.67	0.00	21.86	24.28	2.75	0.00
22	14.08	9.82	18.00	14.36	7.34	26.66	3.39	6.57	28.31	4.79	5.54	13.35	0.00	8.92
23	4.99	16.66	21.86	0.00	11.63	40.18	10.52	14.76	14.01	10.92	12.43	31.72	13.51	8.33
24	9.36	0.00	24.19	1.40	7.61	11.86	1.74	16.77	5.18	2.49	12.62	10.94	3.71	8.19
25	0.00	11.30	20.79	2.36	0.00	0.00	0.00	6.96	7.81	9.07	0.04	8.13	2.96	3.40
26	12.35	27.28	22.38	13.30	21.09	22.63	24.37	25.99	0.00	25.45	10.11	13.84	10.86	16.97
27	13.04	0.00	19.85	2.67	1.81	28.58	0.09	17.22	34.87	0.00	13.44	19.98	0.00	28.80
29	0.00	23.02	26.38	11.20	8.95	13.09	17.81	6.76	3.33	7.64	18.08	20.97	12.22	13.21
30	14.61	6.12	7.02	0.00	2.91	26.56	1.69	11.90	0.66	12.28	15.19	14.10	26.63	7.44
32	36.48	9.70	7.88	13.96	13.29	13.50	4.94	34.61	18.58	2.01	12.56	0.00	8.43	24.67
33	21.91	24.70	26.80	8.63	8.04	28.47	29.49	20.22	26.26	23.32	12.04	17.48	8.00	36.86
34	0.00	5.36	11.12	1.35	3.90	25.14	2.03	0.00	19.78	3.49	0.32	17.43	0.00	0.00
36	6.52	0.17	7.83	0.55	4.65	13.95	10.87	11.28	8.79	4.98	17.46	4.62	17.93	6.80
37	1.89	4.42	5.81	1.18	13.60	7.28	2.38	0.00	21.74	2.49	9.54	31.80	0.00	1.05
39	9.21	4.72	3.78	8.38	18.49	24.12	1.04	11.98	19.02	0.00	11.30	52.78	0.00	0.00
40	37.66	10.26	28.29	23.61	16.99	10.78	11.12	23.18	14.62	9.56	30.20	21.17	5.00	38.97
42	7.34	10.08	21.34	10.35	11.07	1.88	7.33	0.00	0.86	0.40	2.50	18.64	0.00	4.87
44	3.04	6.03	21.93	7.33	8.70	9.01	19.00	8.59	22.42	5.13	11.68	10.31	8.46	23.80
46	41.60	34.65	2.76	10.90	21.00	3.93	13.52	0.00	6.16	0.00	16.45	0.96	15.12	1.29
Sum	318.92	268.24	527.05	232.82	317.68	473.63	280.77	323.50	449.31	247.75	405.15	507.38	223.85	331.73

Note. N = 30. This table includes IDs that were included for analysis. The value for winning model is 0. The smaller the value, the better the fit. Only set size of 4 was assessed at Pre-Test and Post-Test. T1 – T4: 1st training to 4th training session.

Table G2

Bayesian Information Criterion Values Relative to The Winning Model for The Experimental Group in Orientation Reproduction Task

ID	Session													
	Pre-Test	T1			T2			T3			T4			Post-Test
	Set Size													
	4	2	4	6	2	4	6	2	4	6	2	4	6	4
	Standard Mixture Model													
1	0.00	0.00	3.36	5.92	0.32	15.79	4.26	0.00	0.00	12.99	9.20	0.00	0.69	0.47
5	2.96	1.09	3.30	2.66	0.89	0.00	0.00	4.02	0.00	8.31	14.99	0.00	2.44	0.00
8	2.90	0.00	0.24	2.53	7.28	10.05	0.00	0.00	1.86	0.13	9.52	3.28	0.61	0.00
10	1.88	0.00	0.00	5.87	0.00	0.00	0.21	2.28	1.60	1.28	0.74	0.00	1.61	0.73
11	0.00	5.53	9.35	0.80	1.78	0.00	1.63	0.00	0.00	0.00	3.32	0.00	4.30	0.00
12	0.95	0.00	0.00	7.68	0.00	0.00	0.00	3.94	0.00	0.00	0.89	0.00	0.00	0.00
14	3.20	0.99	10.28	9.25	15.65	0.00	0.00	5.46	0.00	1.76	0.00	6.75	1.14	0.00
17	0.00	0.52	0.00	2.06	0.39	0.00	14.40	0.00	6.32	0.00	0.00	0.00	4.05	0.00
18	5.78	1.39	3.69	9.31	0.00	0.00	1.79	1.01	0.00	0.00	11.64	0.00	0.00	3.74
19	0.00	0.00	0.00	0.00	0.00	0.00	2.62	0.00	0.00	0.00	6.46	0.00	0.00	0.00
20	6.44	3.69	1.91	0.00	1.34	9.63	0.00	0.00	6.73	0.00	0.00	1.93	6.46	0.00
21	0.00	1.79	0.00	1.10	4.99	0.63	0.00	3.20	0.00	5.17	0.00	0.00	1.80	3.61
22	0.45	0.00	0.00	5.28	0.00	4.69	0.00	4.36	0.00	0.66	0.00	0.00	3.69	0.00
23	0.00	2.42	0.38	3.27	0.00	2.82	0.00	0.00	0.00	6.24	0.00	0.00	2.94	0.00
24	7.22	3.69	0.00	0.00	2.89	0.00	0.00	6.96	0.00	1.42	0.26	0.00	0.00	0.00
25	7.24	2.93	0.00	0.00	1.20	5.17	2.86	0.15	0.00	8.12	0.44	0.00	0.00	2.33
26	0.00	0.00	0.00	0.00	0.00	0.00	10.59	12.17	4.70	4.32	0.00	0.00	0.00	0.00
27	4.66	7.15	2.41	3.06	4.40	4.73	0.40	11.76	0.00	2.27	7.38	0.00	2.47	27.76
29	10.57	16.27	0.00	5.29	4.95	0.00	14.39	5.94	1.86	6.25	19.95	0.00	0.00	3.04
30	10.61	3.26	0.00	1.94	2.07	0.00	0.67	5.10	3.23	6.69	11.88	0.00	20.54	0.49
32	0.00	0.00	0.00	0.00	0.00	0.00	0.29	14.29	0.00	0.00	0.00	11.11	0.00	0.00
33	3.65	0.00	0.00	0.94	8.49	1.50	4.78	1.39	0.00	0.91	1.54	0.00	5.96	0.00
34	3.23	0.00	0.00	2.89	0.00	0.00	0.00	1.77	3.27	0.00	0.17	0.00	5.46	9.87
36	0.00	0.32	0.00	0.00	0.00	0.00	8.55	0.00	0.00	3.49	0.00	0.00	0.00	0.00
37	0.90	0.00	0.00	0.00	0.00	0.00	0.00	5.10	0.00	0.00	0.00	0.00	2.47	1.74
39	0.00	0.00	4.10	5.88	8.16	0.00	1.88	3.01	0.00	2.75	0.00	0.00	1.22	5.00
40	20.52	6.13	0.00	17.68	6.87	0.00	3.11	7.80	0.00	2.49	4.36	0.00	0.97	15.44
42	2.99	2.85	0.00	2.80	3.45	2.01	0.39	4.35	7.10	0.08	0.00	0.00	1.85	0.21
44	1.67	0.04	0.00	3.69	0.00	0.00	14.28	0.00	1.54	0.00	5.72	0.00	0.94	4.16
46	36.63	31.26	5.55	13.23	7.91	1.44	11.72	1.41	0.00	1.79	13.71	2.92	12.63	1.50
Sum	134.45	91.33	44.57	113.13	83.04	58.45	98.81	105.46	38.21	77.13	122.17	26.00	84.25	80.07
	Swap Model													
1	3.79	1.03	9.24	5.06	0.00	21.68	6.26	2.48	5.89	0.00	2.96	5.89	0.00	5.21
5	1.85	2.01	7.58	0.00	0.00	5.88	1.96	6.51	3.96	2.42	4.51	4.57	3.77	4.36
8	2.71	2.48	6.13	0.00	5.35	9.74	0.51	2.48	6.84	0.33	3.05	9.10	0.00	4.79
10	5.98	0.35	5.83	8.26	1.79	5.89	2.69	4.76	4.54	3.43	3.12	5.18	4.07	3.97
11	4.79	7.65	13.22	3.17	0.00	5.89	0.00	2.48	5.89	1.94	0.00	5.89	2.49	4.39
12	2.98	2.48	5.89	8.14	1.88	5.89	1.11	6.43	5.89	2.48	0.00	5.89	0.90	3.58
14	0.00	0.00	16.17	7.66	3.29	5.89	0.64	0.00	5.89	4.25	1.94	9.47	0.28	4.00
17	0.71	0.00	1.72	2.05	0.46	5.89	13.36	2.03	12.21	2.22	2.48	5.89	5.24	4.43
18	10.57	1.39	0.00	4.86	1.10	5.61	2.39	0.04	5.43	1.13	0.00	3.64	2.48	0.00
19	4.79	2.48	5.89	2.48	1.99	5.89	1.99	2.08	5.89	2.48	0.00	5.89	2.48	4.79
20	9.59	6.18	2.61	2.48	3.56	15.51	2.48	2.48	9.74	2.48	2.48	7.81	5.80	4.79
21	1.17	3.50	4.99	0.00	7.48	2.91	1.64	2.58	5.89	6.99	2.18	5.89	0.00	8.40
22	0.00	2.48	4.80	0.00	0.57	5.57	2.32	6.83	5.79	0.00	2.48	5.25	6.18	1.81
23	1.95	3.45	5.42	5.76	2.48	0.00	2.47	2.48	5.89	2.04	2.47	4.04	0.00	4.79
24	0.00	2.53	0.60	2.31	0.00	4.62	0.99	0.00	5.89	0.00	2.35	5.89	2.48	4.75
25	12.02	1.40	5.43	1.57	1.72	10.89	4.12	0.00	5.87	0.00	2.65	5.17	2.35	5.09
26	4.41	2.47	5.89	1.18	1.26	5.89	6.12	4.71	10.59	0.00	0.61	5.89	0.39	3.45
27	0.00	7.10	1.39	0.00	0.00	4.68	2.73	5.63	5.89	4.44	0.00	5.89	1.15	0.00
29	11.16	0.00	5.89	2.92	0.00	5.89	9.47	4.53	7.74	7.64	8.26	5.89	2.48	7.18

Table G2

Bayesian Information Criterion Values Relative to The Winning Model for The Experimental Group in Orientation Reproduction Task

ID	Session													
	Pre-Test	T1			T2			T3			T4			Post-Test
	Set Size													
	4	2	4	6	2	4	6	2	4	6	2	4	6	4
30	8.04	2.07	4.10	2.87	1.19	5.89	0.66	0.00	9.11	7.63	10.55	5.68	12.56	0.03
32	3.55	2.48	5.89	2.43	2.48	3.58	2.24	0.00	5.89	1.16	2.10	15.91	2.48	3.57
33	0.00	0.32	5.89	3.30	8.40	7.39	6.31	0.00	5.89	1.82	0.00	5.89	6.23	1.47
34	8.02	1.41	3.58	3.14	1.82	5.89	2.48	4.26	8.61	1.07	2.65	4.59	6.21	8.51
36	4.15	1.74	5.89	2.21	2.48	5.89	0.00	2.35	3.38	4.31	2.48	5.89	2.48	4.44
37	4.96	1.99	5.83	1.66	1.62	4.13	2.48	7.59	4.24	2.36	2.19	3.41	4.88	6.46
39	4.79	2.22	9.99	3.34	0.00	5.89	0.91	0.00	5.83	3.82	1.00	5.89	3.66	9.14
40	0.00	4.17	4.35	11.88	0.43	5.89	2.69	0.00	5.89	3.45	2.42	5.89	1.86	5.37
42	0.00	0.00	5.89	3.64	0.00	7.89	0.00	6.84	12.98	2.56	1.87	3.62	3.96	3.17
44	3.73	0.00	5.89	2.21	2.33	5.89	8.63	2.48	5.09	2.48	0.81	4.19	0.00	1.81
46	5.29	0.00	10.06	0.00	0.00	6.77	1.64	2.81	5.28	2.74	5.56	8.80	2.30	6.28
Sum	120.99	65.41	176.01	94.58	53.70	199.26	91.31	84.87	197.86	77.69	73.19	182.84	89.19	130.03
	Signal Discrimination Model													
1	6.57	1.82	0.00	0.00	2.58	12.59	0.00	2.37	11.06	1.13	0.00	4.69	1.43	3.35
5	7.28	0.00	0.00	2.38	2.47	5.19	1.38	7.79	6.31	0.00	0.00	4.98	2.80	0.38
8	7.00	2.99	5.47	0.19	0.00	9.11	1.60	4.38	7.34	0.00	0.00	9.00	0.78	10.69
10	10.88	4.23	13.25	8.66	2.86	9.80	5.25	7.04	9.77	5.16	4.25	10.59	6.46	6.38
11	8.03	1.97	9.07	1.92	0.58	4.58	4.50	4.57	9.61	1.78	0.47	9.00	0.00	5.09
12	1.02	1.58	8.61	0.00	6.66	6.03	2.28	0.00	11.28	5.99	3.73	12.90	4.43	2.60
14	0.99	1.72	0.00	0.00	0.00	0.84	0.25	4.99	8.78	0.00	0.07	1.55	0.00	4.94
17	6.79	1.11	0.43	0.62	0.00	6.43	0.00	1.65	14.42	0.31	5.66	8.47	5.52	11.42
18	14.68	0.00	9.25	0.00	1.62	5.07	0.00	0.00	15.69	2.90	0.43	8.11	5.02	4.69
19	12.47	4.10	8.17	6.84	5.92	11.58	0.00	1.90	11.54	4.66	6.78	15.37	6.43	9.76
20	8.73	2.94	1.56	2.91	0.00	0.00	4.37	2.88	0.00	2.88	3.58	0.00	0.00	6.88
21	3.58	4.50	7.22	2.05	6.73	5.57	4.31	0.00	11.44	4.10	4.69	10.12	2.53	12.50
22	4.39	7.12	5.95	1.30	0.34	0.00	6.24	0.00	1.13	2.22	5.64	7.50	5.50	4.77
23	5.32	0.00	0.00	2.01	4.51	9.60	0.53	8.36	9.29	0.00	2.33	12.47	0.44	9.42
24	4.91	4.97	2.10	3.13	0.94	2.45	0.05	3.06	7.59	2.78	0.00	11.14	4.89	8.46
25	13.90	0.00	5.73	3.56	3.50	15.53	6.72	0.21	4.24	1.09	2.04	8.26	1.18	4.97
26	11.42	7.56	7.10	5.24	3.90	7.90	0.00	0.00	11.76	3.50	4.15	11.22	5.38	4.77
27	7.79	9.31	0.00	0.52	3.86	0.00	4.46	0.00	15.32	5.14	5.01	7.08	2.70	7.81
29	10.86	4.37	9.65	0.00	5.37	7.96	0.00	0.00	8.33	0.00	0.00	1.34	1.88	0.00
30	0.00	0.00	6.71	5.41	0.00	15.71	0.00	1.04	14.80	0.00	0.00	6.94	0.00	0.92
32	10.47	4.57	7.70	7.04	3.36	11.55	0.00	12.26	6.83	2.62	1.99	15.53	3.19	4.72
33	5.58	4.90	4.66	0.00	0.00	0.00	0.00	2.24	4.65	0.00	6.60	3.28	0.00	8.50
34	13.03	3.07	4.07	0.10	4.64	13.38	4.59	5.42	0.00	3.74	2.47	5.53	8.26	10.84
36	10.55	2.05	10.75	2.82	4.73	11.84	0.76	3.84	3.24	0.00	6.17	8.59	7.53	5.50
37	6.75	2.62	6.04	4.80	6.50	5.97	4.82	8.90	12.53	3.33	5.50	12.40	7.12	10.43
39	10.69	2.33	7.88	0.00	4.44	9.38	0.42	5.53	12.80	3.13	1.69	13.27	4.48	12.74
40	4.47	0.00	13.90	0.00	0.00	14.81	0.00	9.57	6.65	0.00	0.00	8.87	0.00	0.00
42	6.08	4.13	10.80	0.00	2.05	10.36	3.54	7.45	10.80	4.40	0.57	7.50	5.79	3.50
44	5.32	2.95	6.55	0.00	2.25	11.46	0.00	4.54	0.00	2.32	0.00	4.52	2.19	0.00
46	0.00	1.22	8.90	2.07	0.98	7.73	0.00	1.94	5.92	3.83	0.00	13.40	0.00	11.29
Sum	219.55	88.13	181.53	63.57	80.78	232.40	56.07	111.92	253.14	66.99	73.82	253.62	95.94	187.30
	Target Confusability Competition Model													
1	8.47	8.69	24.54	10.86	8.83	0.00	19.14	6.79	16.65	15.98	17.23	17.43	4.20	0.00
5	0.00	3.31	23.64	1.96	7.78	7.94	2.09	0.00	18.32	5.99	20.43	25.80	0.00	5.41
8	0.00	0.02	0.00	6.66	3.51	0.00	5.93	4.94	0.00	13.22	6.33	0.00	5.88	5.88
10	0.00	4.88	7.08	0.00	0.49	13.83	0.00	0.00	0.00	0.00	0.00	5.39	0.00	0.00
11	3.09	0.00	0.00	0.00	15.29	8.27	6.00	31.59	7.52	5.47	11.96	5.44	5.98	7.69
12	0.00	5.72	3.66	14.34	19.01	7.40	7.04	6.98	12.99	13.35	14.57	19.22	20.95	11.69
14	5.83	12.30	18.47	20.19	19.05	8.86	13.57	19.01	9.49	11.94	23.37	0.00	4.97	3.42

Table G2

Bayesian Information Criterion Values Relative to The Winning Model for The Experimental Group in Orientation Reproduction Task

ID	Session													
	Pre-Test	T1			T2			T3			T4			Post-Test
	Set Size													
	4	2	4	6	2	4	6	2	4	6	2	4	6	4
17	4.96	4.90	9.39	0.00	13.98	10.73	23.15	1.34	0.00	3.20	8.88	7.92	0.00	12.75
18	0.00	3.40	20.98	10.54	5.03	14.42	0.33	14.88	38.32	9.18	17.11	11.08	0.61	1.08
19	14.96	13.33	19.99	16.11	17.58	16.21	16.62	8.61	14.43	13.66	29.96	17.98	11.09	3.46
20	0.00	0.00	0.00	7.79	15.40	8.47	14.22	3.27	18.66	22.86	11.31	2.98	26.83	2.54
21	7.29	0.00	12.35	1.99	0.00	0.00	0.67	1.06	8.78	0.00	21.37	20.39	1.78	0.00
22	8.51	9.34	12.29	13.39	6.23	15.01	2.90	5.12	17.79	3.82	5.05	9.19	0.00	5.16
23	1.37	15.21	10.20	0.00	11.14	32.41	9.59	14.28	10.12	9.46	11.94	27.84	12.50	5.54
24	3.78	0.00	14.63	0.91	6.64	2.65	0.33	15.80	1.11	1.52	11.16	7.05	3.23	5.40
25	0.00	9.84	14.86	1.88	0.00	0.00	0.00	5.72	0.40	8.10	0.00	4.24	2.48	0.00
26	9.56	26.80	17.82	12.81	20.60	18.74	22.92	24.53	0.00	24.48	9.63	9.95	10.28	13.38
27	7.46	0.00	8.19	1.70	0.84	16.92	0.00	15.76	30.98	0.00	12.46	15.41	0.00	23.22
29	0.00	22.05	22.50	9.74	7.98	9.21	16.35	5.30	0.00	6.18	16.62	10.65	11.73	4.85
30	6.25	4.67	2.08	0.00	1.46	22.67	0.24	10.93	0.00	10.82	13.74	9.38	25.18	0.00
32	33.70	9.22	3.92	13.48	12.80	9.31	3.48	33.64	13.76	1.52	12.07	0.00	7.94	21.04
33	16.34	24.05	19.80	7.18	6.59	16.81	28.03	19.25	19.26	21.86	11.07	9.09	6.54	32.75
34	0.00	4.87	3.53	0.00	3.42	21.25	1.55	0.00	8.12	3.00	0.00	11.30	0.00	0.00
36	3.73	0.00	3.95	0.07	4.16	10.06	9.90	10.79	0.38	3.53	16.97	0.74	17.44	3.95
37	0.00	3.93	0.20	0.69	13.11	1.59	1.90	0.00	17.86	2.01	9.06	27.44	0.00	0.00
39	6.43	4.24	0.00	6.92	17.52	20.24	0.00	11.01	15.13	0.00	10.81	48.89	0.00	0.00
40	32.09	8.80	24.40	22.16	15.53	6.89	9.66	22.21	9.60	8.10	28.75	17.28	3.54	30.61
42	1.76	9.11	17.45	8.90	10.11	0.00	6.36	0.00	0.00	0.00	1.62	14.49	0.00	0.00
44	0.00	5.06	16.82	5.88	8.22	5.13	17.55	8.11	10.76	4.65	10.22	3.18	7.49	15.44
46	33.24	33.68	0.00	9.93	20.03	0.00	12.07	0.00	0.42	0.00	15.00	0.00	13.66	0.00
Sum	208.80	247.41	332.74	206.09	292.32	305.02	251.59	300.92	300.85	223.90	378.69	359.74	204.32	215.25

Note. N = 30. This table includes IDs that were included for analysis. The value for winning model is 0. The smaller the value, the better the fit. Only set size of 4 was assessed at Pre-Test and Post-Test. T1 – T4: 1st training to 4th training session.

Table G3

Information Criterion Values Relative to The Winning Model for The Active Control Group in Orientation Reproduction Task

ID	Pre-Test		Post-Test	
	AIC	BIC	AIC	BIC
	Standard Mixture Model			
49	9.40	3.82	32.63	29.85
50	0.22	0.93	7.11	4.32
51	2.57	1.82	0.00	0.00
52	3.82	3.16	5.37	0.00
53	9.31	3.73	4.57	6.18
54	2.77	3.11	5.85	3.06
55	0.20	0.13	1.58	0.00
56	0.00	0.66	0.00	0.00
57	1.52	4.31	0.00	0.00
61	4.66	5.44	0.00	2.24
62	1.81	0.00	0.00	0.10
64	11.20	5.62	0.00	0.00
67	2.46	1.75	0.02	1.45
69	6.46	3.67	0.00	0.00
70	17.19	14.40	6.78	1.20
71	1.18	3.26	0.00	0.84
72	1.41	4.20	1.30	0.00
74	0.74	0.00	11.28	5.71
76	3.83	2.68	8.17	5.39
79	1.51	0.00	4.88	0.85
80	6.24	3.61	2.21	4.78
81	0.00	0.00	0.00	0.00
82	5.84	3.26	0.00	0.00
83	5.28	1.94	19.86	15.44
84	0.00	2.56	3.22	6.00
86	2.17	0.00	5.95	1.21
89	2.00	4.79	1.20	3.99
90	25.55	22.76	13.37	10.58
91	0.00	0.00	15.92	13.13
92	0.00	0.00	7.07	9.86
93	0.00	0.00	2.44	0.56
94	0.00	0.29	4.58	0.00
96	2.62	0.00	4.19	0.00
97	0.00	1.35	3.47	0.69
Sum	131.95	103.25	173.02	127.43
	Swap Model			
49	5.07	2.28	0.00	0.00
50	0.00	3.50	0.00	0.00
51	4.53	6.58	2.00	4.79
52	1.44	3.56	6.03	3.45
53	6.72	3.93	5.16	9.56
54	3.58	6.69	0.00	0.00
55	2.20	4.91	0.00	1.21
56	1.35	4.79	1.98	4.76

Table G3

Information Criterion Values Relative to The Winning Model for The Active Control Group in Orientation Reproduction Task

ID	Pre-Test		Post-Test	
	AIC	BIC	AIC	BIC
57	3.81	9.38	2.00	4.79
61	0.30	3.87	0.95	5.98
62	0.00	0.98	1.87	4.76
64	6.38	3.59	1.78	4.56
67	0.00	2.08	0.00	4.22
69	0.00	0.00	1.79	4.58
70	0.00	0.00	6.00	3.22
71	0.55	5.42	2.00	5.63
72	4.08	9.66	3.01	4.50
74	0.00	2.05	9.51	6.72
76	5.04	6.67	0.00	0.00
79	0.00	1.28	1.24	0.00
80	3.08	3.23	3.66	9.02
81	1.30	4.09	2.00	4.79
82	4.01	4.22	0.96	3.75
83	0.56	0.00	1.63	0.00
84	1.21	6.56	4.92	10.50
86	0.00	0.62	1.95	0.00
89	4.00	9.57	2.55	8.13
90	0.00	0.00	0.00	0.00
91	0.87	3.65	0.00	0.00
92	2.00	4.79	8.32	13.90
93	1.52	4.31	2.61	3.51
94	1.58	4.66	6.58	4.79
96	2.65	2.82	3.97	2.57
97	1.98	6.11	0.00	0.00
Sum	69.79	135.87	84.49	133.68
	Signal Discrimination Model			
49	0.00	0.00	10.42	13.21
50	3.83	10.12	2.68	5.46
51	0.00	4.83	3.73	9.30
52	0.00	4.91	0.00	0.20
53	0.00	0.00	0.00	7.19
54	0.00	5.91	3.58	6.36
55	0.00	5.50	1.92	5.92
56	2.85	9.08	4.48	10.06
57	5.55	13.91	4.72	10.29
61	0.00	6.36	0.86	8.68
62	2.22	5.98	2.37	8.05
64	0.00	0.00	1.75	7.32
67	0.52	5.39	4.11	11.12
69	2.89	5.68	5.83	11.41
70	6.84	9.63	0.00	0.00
71	0.00	7.65	3.49	9.91
72	5.43	13.79	0.00	4.27

Table G3

Information Criterion Values Relative to The Winning Model for The Active Control Group in Orientation Reproduction Task

ID	Pre-Test		Post-Test	
	AIC	BIC	AIC	BIC
74	3.64	8.48	0.00	0.00
76	0.00	4.42	4.46	7.25
79	2.26	6.33	0.00	1.54
80	0.00	2.94	0.00	8.14
81	2.16	7.74	1.14	6.72
82	0.00	2.99	2.81	8.38
83	0.00	2.23	0.00	1.16
84	2.97	11.10	1.06	9.43
86	0.75	4.15	0.00	0.84
89	5.56	13.93	4.05	12.42
90	4.27	7.06	3.11	5.90
91	2.62	8.19	9.08	11.87
92	3.89	9.47	8.14	16.50
93	2.97	8.54	0.00	3.69
94	4.49	10.36	0.00	1.00
96	0.00	2.95	0.00	1.39
97	0.89	7.82	2.32	5.11
Sum	66.62	227.47	86.12	230.08
	Target Confusability Competition Model			
49	21.49	13.13	55.56	49.98
50	2.07	0.00	6.98	1.40
51	3.53	0.00	2.96	0.17
52	3.45	0.00	13.24	5.08
53	20.56	12.19	1.18	0.00
54	2.46	0.00	10.24	4.67
55	2.86	0.00	8.74	4.37
56	2.13	0.00	10.22	7.44
57	0.00	0.00	11.43	8.64
61	2.00	0.00	0.55	0.00
62	18.71	14.12	2.68	0.00
64	17.83	9.47	14.86	12.07
67	3.50	0.00	1.35	0.00
69	6.30	0.73	12.86	10.07
70	27.44	21.86	22.21	13.84
71	0.71	0.00	1.95	0.00
72	0.00	0.00	5.18	1.10
74	15.56	12.04	18.21	9.85
76	3.94	0.00	19.93	14.36
79	4.59	0.30	6.96	0.14
80	5.42	0.00	0.22	0.00
81	13.94	11.15	6.96	4.17
82	5.37	0.00	7.29	4.51
83	10.30	4.17	17.03	9.83
84	0.23	0.00	0.00	0.00
86	7.72	2.76	8.48	0.96

Table G3

Information Criterion Values Relative to The Winning Model for The Active Control Group in Orientation Reproduction Task

ID	Pre-Test		Post-Test	
	AIC	BIC	AIC	BIC
89	0.00	0.00	0.00	0.00
90	31.21	25.64	27.97	22.40
91	17.70	14.91	16.59	11.02
92	4.16	1.37	0.00	0.00
93	6.31	3.52	4.67	0.00
94	2.49	0.00	8.63	1.26
96	29.54	24.13	27.79	20.81
97	1.44	0.00	9.55	3.97
Sum	294.96	171.48	362.47	222.11

Note. N = 34. Performance was measured at set size 4. This table includes IDs that were included for analysis. The value for winning model is 0. The smaller the value, the better the fit. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion

Table G4

Information Criterion Values Relative to The Winning Model for The Experimental Group in Shape Reproduction Task

ID	Pre-Test		Post-Test	
	AIC	BIC	AIC	BIC
	Standard Mixture Model			
1	3.61	6.40	0.00	0.00
5	5.57	5.54	1.11	2.76
8	0.00	2.15	6.34	9.12
10	1.61	4.40	2.85	5.63
11	0.36	0.37	3.55	6.28
12	0.00	0.84	4.67	7.46
14	2.52	4.19	11.87	14.66
17	3.25	0.46	0.00	2.58
18	0.00	0.00	4.23	7.02
19	11.00	13.79	0.00	0.00
20	0.93	2.00	9.60	12.39
21	0.52	3.25	3.79	6.58
22	0.00	0.03	4.60	7.39
23	3.74	6.52	4.00	6.78
24	0.59	3.38	1.25	4.04
25	3.71	0.92	5.44	4.89
26	0.00	0.90	4.35	7.13
27	4.44	7.23	10.81	13.59
29	2.50	2.59	0.95	3.74
30	4.29	2.47	3.54	6.33
32	6.36	9.14	2.68	0.00
33	0.89	3.68	10.99	13.77
34	2.95	2.95	2.86	5.65
36	0.00	0.00	0.00	0.00
37	5.47	8.26	0.01	2.80
39	2.71	5.49	5.58	8.37
40	0.00	0.00	0.00	0.00
42	1.70	4.49	1.70	3.96
44	3.90	6.69	4.34	1.55
46	0.00	0.00	2.49	5.28
Sum	72.61	108.14	113.59	169.76
	Swap Model			
1	2.57	8.14	1.91	4.70
5	0.00	2.76	0.00	4.44
8	2.00	6.93	8.34	13.91
10	2.63	8.21	4.62	10.19
11	0.46	3.26	0.30	5.82
12	1.86	5.49	5.95	11.53
14	1.14	5.59	13.87	19.45
17	0.00	0.00	2.00	7.36
18	1.51	4.30	4.48	10.05
19	13.00	18.58	2.00	4.79
20	0.00	3.86	5.08	10.65
21	1.04	6.56	5.79	11.36

Table G4

Information Criterion Values Relative to The Winning Model for The Experimental Group in Shape Reproduction Task

ID	Pre-Test		Post-Test	
	AIC	BIC	AIC	BIC
22	0.91	3.73	6.60	12.18
23	2.35	7.92	4.65	10.23
24	2.42	8.00	3.25	8.83
25	0.00	0.00	0.00	2.24
26	2.00	5.69	5.23	10.80
27	5.98	11.55	11.91	17.48
29	4.50	7.37	2.95	8.52
30	4.16	5.14	0.13	5.71
32	8.36	13.93	0.00	0.10
33	2.85	8.42	12.99	18.56
34	0.00	2.79	4.07	9.65
36	2.00	4.79	1.95	4.74
37	0.39	5.96	1.74	7.32
39	4.71	10.28	6.07	11.64
40	2.00	4.79	2.00	4.79
42	3.70	9.28	0.00	5.06
44	3.67	9.24	0.00	0.00
46	1.65	4.44	4.35	9.93
Sum	77.85	197.01	122.23	262.02
	Signal Discrimination Model			
1	0.93	9.29	3.93	9.50
5	1.07	6.62	1.34	8.56
8	2.92	10.64	8.81	17.18
10	4.80	13.16	5.07	13.43
11	0.00	5.59	0.00	8.30
12	2.62	9.04	4.60	12.97
14	0.00	7.24	11.57	19.93
17	0.87	3.66	2.86	11.01
18	3.64	9.21	2.62	10.99
19	7.83	16.19	3.42	8.99
20	0.03	6.68	5.60	13.96
21	0.00	8.31	7.04	15.40
22	3.64	9.25	7.57	15.93
23	3.93	12.29	4.42	12.78
24	4.31	12.67	5.30	13.66
25	1.99	4.78	2.21	7.23
26	3.95	10.42	5.09	13.45
27	7.36	15.72	10.08	18.45
29	0.00	5.66	1.73	10.09
30	0.00	3.76	1.09	9.45
32	7.78	16.14	0.79	3.68
33	1.16	9.52	11.29	19.66
34	2.91	8.49	5.00	13.36
36	2.66	8.24	1.50	7.07
37	2.32	10.68	1.19	9.55

Table G4

Information Criterion Values Relative to The Winning Model for The Experimental Group in Shape Reproduction Task

ID	Pre-Test		Post-Test	
	AIC	BIC	AIC	BIC
39	6.47	14.83	7.79	16.15
40	1.42	7.00	3.46	9.04
42	4.93	13.30	1.05	8.89
44	5.62	13.98	4.49	7.27
46	4.43	10.01	5.83	14.19
Sum	89.59	292.38	136.72	360.14
	Target Confusability Competition Model			
1	0.00	0.00	2.79	0.00
5	2.81	0.00	1.14	0.00
8	0.64	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00
11	2.77	0.00	0.06	0.00
12	1.94	0.00	0.00	0.00
14	1.12	0.00	0.00	0.00
17	7.68	2.11	0.21	0.00
18	4.24	1.45	0.00	0.00
19	0.00	0.00	10.38	7.60
20	1.71	0.00	0.00	0.00
21	0.06	0.00	0.00	0.00
22	2.76	0.00	0.00	0.00
23	0.00	0.00	0.00	0.00
24	0.00	0.00	0.00	0.00
25	10.74	5.16	3.33	0.00
26	1.89	0.00	0.00	0.00
27	0.00	0.00	0.00	0.00
29	2.70	0.00	0.00	0.00
30	4.60	0.00	0.00	0.00
32	0.00	0.00	16.92	11.45
33	0.00	0.00	0.00	0.00
34	2.79	0.00	0.00	0.00
36	9.60	6.81	4.45	1.66
37	0.00	0.00	0.00	0.00
39	0.00	0.00	0.00	0.00
40	7.57	4.78	18.50	15.71
42	0.00	0.00	0.52	0.00
44	0.00	0.00	12.06	6.49
46	6.64	3.85	0.00	0.00
Sum	72.25	24.17	70.35	42.90

Note. N = 30. Performance was measured at set size 4. This table includes IDs that were included for analysis. The value for winning model is 0. The smaller the value, the better the fit. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion

Table G5

Information Criterion Values Relative to The Winning Model for The Active Control Group in Shape Reproduction Task

ID	Pre-Test		Post-Test	
	AIC	BIC	AIC	BIC
		Standard Mixture Model		
49	2.17	0.00	11.98	14.77
50	0.00	0.00	4.43	7.22
51	0.00	2.51	5.14	7.93
52	2.13	4.91	6.06	8.85
53	2.23	5.01	5.29	8.08
54	10.06	7.27	2.02	0.00
55	7.59	10.38	1.91	4.70
56	9.32	12.11	2.75	5.54
57	1.43	4.22	1.73	4.51
61	9.42	10.20	6.48	9.27
62	5.62	8.41	0.02	2.81
64	7.59	10.37	0.00	1.38
67	0.00	0.55	4.28	7.07
69	5.63	8.42	8.89	6.45
70	2.26	5.05	2.29	2.37
71	0.46	1.39	2.75	5.54
72	6.67	9.45	3.88	6.67
74	7.44	10.23	10.05	12.84
76	3.33	6.12	5.24	3.51
79	6.04	8.82	10.28	13.06
80	11.42	7.96	4.74	7.53
81	5.27	6.81	0.00	0.00
82	9.75	7.73	0.41	0.49
83	9.76	10.75	3.91	6.70
84	0.00	2.69	5.68	8.47
86	5.67	8.46	0.00	0.00
89	2.27	5.06	1.62	4.40
90	0.52	3.31	0.00	1.69
91	0.94	0.00	21.50	16.90
92	0.00	1.36	2.69	5.48
93	4.28	7.07	2.39	5.18
94	4.84	7.63	4.15	2.67
96	4.18	6.96	1.83	0.00
97	2.15	4.94	3.41	6.20
Sum	150.46	206.18	147.80	198.25
		Swap Model		
49	0.00	0.62	13.85	19.42
50	1.50	4.28	6.43	12.01
51	2.00	7.30	7.14	12.72
52	4.13	9.70	6.98	12.55
53	4.23	9.80	7.29	12.87
54	0.00	0.00	0.00	0.76
55	5.76	11.33	3.55	9.13
56	11.32	16.90	4.04	9.62

Table G5

Information Criterion Values Relative to The Winning Model for The Active Control Group in Shape Reproduction Task

ID	Pre-Test		Post-Test	
	AIC	BIC	AIC	BIC
57	3.43	9.00	1.92	7.49
61	1.75	5.31	8.38	13.96
62	7.62	13.20	0.03	5.61
64	8.66	14.24	2.00	6.17
67	1.27	4.60	5.31	10.88
69	5.82	11.39	4.96	5.32
70	2.28	7.85	3.92	6.79
71	0.00	3.72	4.30	9.88
72	8.67	14.24	5.88	11.46
74	8.78	14.35	12.05	17.63
76	5.33	10.90	1.02	2.07
79	4.07	9.65	9.89	15.47
80	6.13	5.46	2.44	8.02
81	3.44	7.77	2.00	4.79
82	5.83	6.60	0.17	3.03
83	9.20	12.98	5.91	11.48
84	1.90	7.38	7.68	13.26
86	7.67	13.25	1.37	4.16
89	4.27	9.85	3.40	8.97
90	2.52	8.10	1.09	5.56
91	0.00	1.84	11.66	9.85
92	2.00	6.15	4.69	10.27
93	3.17	8.75	0.69	6.26
94	6.84	12.41	0.00	1.31
96	2.58	8.15	0.00	0.96
97	4.15	9.73	5.41	10.98
Sum	146.32	296.82	155.46	300.69
	Signal Discrimination Model			
49	0.03	3.44	13.39	21.76
50	3.45	9.02	7.80	16.16
51	3.81	11.90	8.55	16.92
52	5.40	13.77	5.57	13.93
53	4.89	13.25	7.95	16.31
54	2.61	5.40	1.67	5.22
55	5.49	13.85	3.55	11.91
56	10.69	19.05	3.30	11.66
57	5.50	13.86	4.41	12.77
61	0.00	6.35	9.95	18.31
62	5.11	13.47	2.65	11.01
64	6.78	15.15	3.75	10.71
67	2.72	8.84	6.42	14.78
69	4.31	12.67	0.00	3.14
70	3.04	11.40	0.00	5.65
71	1.22	7.72	6.64	15.01
72	9.39	17.76	7.22	15.58

Table G5

Information Criterion Values Relative to The Winning Model for The Active Control Group in Shape Reproduction Task

ID	Pre-Test		Post-Test	
	AIC	BIC	AIC	BIC
74	8.34	16.70	9.75	18.12
76	4.80	13.16	0.00	3.84
79	5.44	13.80	6.13	14.49
80	0.00	2.12	1.24	9.61
81	0.00	7.12	4.30	9.88
82	0.00	3.56	0.00	5.65
83	0.00	6.56	2.60	10.96
84	2.55	10.81	8.47	16.83
86	8.68	17.05	2.20	7.78
89	6.21	14.57	5.07	13.44
90	3.71	12.08	1.82	9.08
91	1.71	6.34	0.00	0.98
92	3.85	10.79	6.34	14.70
93	5.22	13.58	2.56	10.92
94	8.26	16.62	0.86	4.96
96	4.16	12.52	0.28	4.03
97	5.08	13.44	5.09	13.46
Sum	142.44	387.71	149.55	389.55
	Target Confusability Competition Model			
49	5.10	0.14	0.00	0.00
50	4.84	2.06	0.00	0.00
51	0.27	0.00	0.00	0.00
52	0.00	0.00	0.00	0.00
53	0.00	0.00	0.00	0.00
54	9.08	3.51	18.56	13.75
55	0.00	0.00	0.00	0.00
56	0.00	0.00	0.00	0.00
57	0.00	0.00	0.00	0.00
61	2.02	0.00	0.00	0.00
62	0.00	0.00	0.00	0.00
64	0.00	0.00	1.41	0.00
67	2.24	0.00	0.00	0.00
69	0.00	0.00	5.22	0.00
70	0.00	0.00	2.71	0.00
71	1.85	0.00	0.00	0.00
72	0.00	0.00	0.00	0.00
74	0.00	0.00	0.00	0.00
76	0.00	0.00	4.52	0.00
79	0.00	0.00	0.00	0.00
80	6.25	0.00	0.00	0.00
81	1.24	0.00	6.12	3.33
82	4.81	0.00	2.71	0.00
83	1.80	0.00	0.00	0.00
84	0.09	0.00	0.00	0.00
86	0.00	0.00	6.10	3.31

Table G5

Information Criterion Values Relative to The Winning Model for The Active Control Group in Shape Reproduction Task

ID	Pre-Test		Post-Test	
	AIC	BIC	AIC	BIC
89	0.00	0.00	0.00	0.00
90	0.00	0.00	1.10	0.00
91	4.72	0.99	7.39	0.00
92	1.43	0.00	0.00	0.00
93	0.00	0.00	0.00	0.00
94	0.00	0.00	4.27	0.00
96	0.00	0.00	5.48	0.86
97	0.00	0.00	0.00	0.00
Sum	45.75	6.69	65.58	21.26

Note. N = 34. Performance was measured at set size 4. This table includes IDs that were included for analysis. The value for winning model is 0. The smaller the value, the better the fit. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion

References

Agam, Y., & Sekuler, R. (2007). Interactions between working memory and visual perception: An ERP/EEG study. *NeuroImage*, *36*(3), 933–942.

Aho, K., Derryberry, D., & Peterson, T. (2014). Model selection for ecologists: The worldviews of AIC and BIC. *Ecology*, *95*(3), 631–636. https://doi.org/10.1890/13-1452.1

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Alcauter, S., Lin, W., Keith Smith, J., Short, S. J., Goldman, B. D., Steven Reznick, J., Gilmore, J. H., & Gao, W. (2014). Development of thalamocortical connectivity during infancy and its cognitive correlations. *Journal of Neuroscience*, *34*(27), 9067–9075. https://doi.org/10.1523/JNEUROSCI.0796-14.2014

Algina, J., Keselman, H. J., & Penfield, R. D. (2005). An alternative to Cohen's standardized mean difference effect size: A robust parameter and confidence interval in the two independent groups case. *Psychological Methods*, *10*(3), 317–328. https://doi.org/10.1037/1082-989X.10.3.317

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*(1), 20–29. https://doi.org/10.1016/j.jecp.2009.11.003

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111. https://doi.org/10.1111/j.0963-7214.2004.01502006.x

Andrews, S. C., Hoy, K. E., Enticott, P. G., Daskalakis, Z. J., & Fitzgerald, P. B. (2011). Improving working memory: The effect of combining cognitive activity and anodal transcranial direct current stimulation to the left dorsolateral prefrontal cortex. *Brain*

Stimulation, *4*(2), 84–89. https://doi.org/10.1016/j.brs.2010.06.004

Arciniega, H., Gözenman, F., Jones, K. T., Stephens, J. A., Berryhill, M. E., Popa-Wagner, A., Tamás Kincses, Z., & Li, D. (2018). Frontoparietal tDCS benefits visual working memory in older adults with low working memory capacity. *Frontiers in Aging Neuroscience*, *10*(57). https://doi.org/10.3389/fnagi.2018.00057

Awh, E., Barton, B., & Vogel, E. K. (2007a). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*(7), 622–628. https://doi.org/10.1111/j.1467-9280.2007.01949.x

Awh, E., Barton, B., & Vogel, E. K. (2007b). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*(7), 622–628. https://doi.org/10.1111/j.1467-9280.2007.01949.x

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612–637. https://doi.org/10.1037/0033-2909.128.4.612

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin*, *130*(4), 553–573. https://doi.org/10.1037/0033-2909.130.4.553

Baumert, A., Buchholz, N., Zinkernagel, A., Clarke, P., MacLeod, C., Osinsky, R., & Schmitt, M. (2020). Causal underpinnings of working memory and Stroop interference control: Testing the effects of anodal and cathodal tDCS over the left DLPFC. *Cognitive, Affective and Behavioral Neuroscience*, *20*(1), 34–48. https://doi.org/10.3758/s13415-019-00726-y

Bavelier, D., Shawn Green, C., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: Learning to learn and action video games. *Annual Review of Neuroscience*, *35*(June 2016), 391–416. https://doi.org/10.1146/annurev-neuro-060909-152832

Bays, P. M. (2016). Evaluating and excluding swap errors in analogue tests of working memory. *Scientific Reports*, *6*(19203). https://doi.org/10.1038/srep19203

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 7.1. https://doi.org/10.1167/9.10.7

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854. https://doi.org/10.1126/science.1158023

Belleville, S., Gilbert, B., Fontaine, F., Gagnon, L., Ménard, E., & Gauthier, S. (2006). Improvement of episodic memory in persons with mild cognitive impairment and healthy older adults: Evidence from a cognitive intervention program. *Dementia and Geriatric Cognitive Disorders*, *22*(5–6), 486–499. https://doi.org/10.1159/000096316

Belleville, S., Mellah, S., De Boysson, C., Demonet, J. F., & Bier, B. (2014). The pattern and loci of training-induced brain changes in healthy older adults are predicted by the nature of the intervention. *PLoS ONE*, *9*(8). https://doi.org/10.1371/journal.pone.0102710

Bikson, M., Reato, D., & Rahman, A. (2012). Cellular and network effects of transcranial direct current stimulation. In *Transcranial Brain Stimulation* (pp. 55–92). https://doi.org/10.1201/b14174-5

Boduroglu, A., Shah, P., & Nisbett, R. E. (2009). Cultural differences in allocation of attention in visual information processing. *Journal of Cross-Cultural Psychology*, *40*(3), 349–360. https://doi.org/10.1177/0022022108331005

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science*, *8*(4), 445–454. https://doi.org/10.1177/1745691613491271

Borsboom, D., van der Maas, H. L. J., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory

construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science*, *16*(4), 756–766. https://doi.org/10.1177/1745691620969647

Box, G. E. P., & Draper, N. R. (1987). Empirical model-building and response surfaces. In *Empirical model-building and response surfaces.* John Wiley & Sons.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, *138*(4), 487–502. https://doi.org/10.1037/a0016797

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., Grange, J. A., Perugini, M., Spies, J. R., & van 't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, *50*(1), 217–224. https://doi.org/10.1016/j.jesp.2013.10.005

Brunoni, A. R., Nitsche, M. A., Bolognini, N., Bikson, M., Wagner, T., Merabet, L., Edwards, D. J., Valero-Cabre, A., Rotenberg, A., Pascual-Leone, A., Ferrucci, R., Priori, A., Boggio, P. S., & Fregni, F. (2012). Clinical research with transcranial direct current stimulation (tDCS): Challenges and future directions. *Brain Stimulation*, *5*(3), 175–195. https://doi.org/10.1016/j.brs.2011.03.002

Brunoni, A. R., & Vanderhasselt, M. A. (2014). Working memory improvement with non-invasive brain stimulation of the dorsolateral prefrontal cortex: A systematic review and meta-analysis. *Brain and Cognition*, *86*(1), 1–9. https://doi.org/10.1016/j.bandc.2014.01.008

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644

Buschkuehl, M., Jaeggi, S. M., Mueller, S. T., Shah, P., & Jonides, J. (2017). Training change

detection leads to substantial task-specific improvement. *Journal of Cognitive Enhancement*, *1*(4), 419–433. https://doi.org/10.1007/s41465-017-0055-y

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Carpenter, J., Sherman, M. T., Kievit, R. A., Seth, A. K., Lau, H., & Fleming, S. M. (2019). Domain-general enhancements of metacognitive ability through adaptive training. *Journal of Experimental Psychology: General*, *148*(1), 51–64. https://doi.org/10.1037/xge0000505

Chen, J., Wang, Z., Chen, Q., Fu, Y., & Zheng, K. (2022). Transcranial direct current stimulation enhances cognitive function in patients with mild cognitive impairment and early/mid Alzheimer's disease: A systematic review and meta-analysis. *Brain Sciences*, *12*(5). https://doi.org/10.3390/brainsci12050562

Constantinidis, C., & Klingberg, T. (2016). The neuroscience of working memory capacity and training. *Nature Reviews Neuroscience*, *17*(7), 438–449. https://doi.org/10.1038/nrn.2016.43

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552. https://doi.org/10.1016/j.tics.2003.10.005

Cosmo, C., DiBiasi, M., Lima, V., Grecco, L. C., Muszkat, M., Philip, N. S., & de Sena, E. P. (2020). A systematic review of transcranial direct current stimulation effects in attention-deficit/hyperactivity disorder. *Journal of Affective Disorders*, *276*, 1–13. https://doi.org/10.1016/j.jad.2020.06.054

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

Craik, F. I. M., & Bialystok, E. (2006). Cognition through the lifespan: Mechanisms of change. *Trends in Cognitive Sciences*, *10*(3), 131–138. https://doi.org/10.1016/j.tics.2006.01.007

Croxson, P. L., Kyriazis, D. A., & Baxter, M. G. (2011). Cholinergic modulation of a specific memory function of prefrontal cortex. *Nature Neuroscience*, *14*. https://doi.org/10.1038/nn.2971

Curby, K. M., & Gauthier, I. (2007). A visual short-term memory advantage for faces. *Psychonomic Bulletin & Review*, *14*(4), 620–628.

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, *7*(9), 415–423. https://doi.org/10.1016/S1364-6613(03)00197-9

Dai, M., Li, Y., Gan, S., & Du, F. (2019). The reliability of estimating visual working memory capacity. *Scientific Reports*, *9*(1), 1–8. https://doi.org/10.1038/s41598-019-39044-1

Datta, A., Bansal, V., Diaz, J., Patel, J., Reato, D., & Bikson, M. (2009). Gyri-precise head model of transcranial direct current stimulation: Improved spatial focality using a ring electrode versus conventional rectangular pad. *Brain Stimulation*, *2*(4), 201–207. https://doi.org/10.1016/j.brs.2009.03.005

Dawood, A. Bin, Dickinson, A., Aytemur, A., Howarth, C., Milne, E., & Jones, M. (2019). Investigating the effects of tDCS on Visual Orientation Discrimination Task Performance: 'The possible influence of placebo.' *Journal of Cognitive Enhancement*, *4*, 235–249. https://doi.org/https://doi.org/10.1007/s41465-019-00154-3

De Simoni, C., & von Bastian, C. C. (2018). Working memory updating and binding training:

Bayesian evidence supporting the absence of transfer. *Journal of Experimental Psychology: General*, *147*(6), 829–858. https://doi.org/10.1037/xge0000453

Dedoncker, J., Brunoni, A. R., Baeken, C., & Vanderhasselt, M.-A. (2016). A systematic review and meta-analysis of the effects of transcranial direct current stimulation (tDCS) over the dorsolateral prefrontal cortex in healthy and neuropsychiatric samples: influence of stimulation parameters. *Brain Stimulation*, *9*, 501–517. https://doi.org/10.1016/j.brs.2016.04.006

Donkin, C., Nosofsky, R. M., Gold, J. M., & Shiffrin, R. M. (2013). Discrete-Slots Models of Visual Working-Memory Response Times. *Psychol Rev*, *120*(4), 873–902. https://doi.org/10.1037/a0034247.

Dumont, R., Majerus, S., & Hansenne, M. (2021). Transcranial direct current stimulation (tDCS) over the intraparietal sulcus does not influence working memory performance. *Psychologica Belgica*, *61*(1), 200–211. https://doi.org/10.5334/PB.534

Elliott, E. M., Morey, C. C., AuBuchon, A. M., Cowan, N., Jarrold, C., Adams, E. J., Attwood, M., Bayram, B., Beeler-Duden, S., Blakstvedt, T. Y., Büttner, G., Castelain, T., Cave, S., Crepaldi, D., Fredriksen, E., Glass, B. A., Graves, A. J., Guitard, D., Hoehl, S., … Voracek, M. (2021). Multilab Direct Replication of Flavell, Beach, and Chinsky (1966): Spontaneous Verbal Rehearsal in a Memory Task as a Function of Age. *Advances in Methods and Practices in Psychological Science*, *4*(2). https://doi.org/10.1177/25152459211018187

Emrich, S. M., & Ferber, S. (2012). Competition increases binding errors in visual working memory. *Journal of Vision*, *12*(4), 1–16. https://doi.org/10.1167/12.4.12

Emrich, S. M., & Lockhart, H. A. (2017). Attention mediates the flexible allocation of visual working memory resources. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(7), 1454–1465. https://doi.org/10.1037/xhp0000398.supp

Engle, R. W., Laughlin, J. E., Tuholski, S. W., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331. https://doi.org/10.1037//0096-3445.128.3.309

Eronen, M. I., & Bringmann, L. F. (2021). The theory crisis in psychology: How to move forward. *Perspectives on Psychological Science*, *16*(4), 779–788. https://doi.org/10.1177/1745691620970586

Evans, N. J. (2019). Assessing the practical differences between model selection methods in inferences about choice response time tasks. *Psychonomic Bulletin and Review*, *26*(4), 1070–1098. https://doi.org/10.3758/s13423-018-01563-9

Faillenot, I., Sakata, H., Costes, N., Decety, J., & Jeannerod, M. (1997). Visual working memory for shape and 3D-orientation: a PET study. *NeuroReport*, *8*(4). https://journals.lww.com/neuroreport/fulltext/1997/03030/visual_working_memory_for_shape_and.10.aspx

Farrell, S., & Lewandowsky, S. (2018). Computational modeling of cognition and behavior. In *Computational modeling of cognition and behavior.* Cambridge University Press. https://doi.org/10.1017/9781316272503

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. http://www.gpower.hhu.de/fileadmin/redaktion/Fakultaeten/Mathematisch-Naturwissenschaftliche_Fakultaet/Psychologie/AAP/gpower/GPower3-BRM-Paper.pdf

Fitch, A., Smith, H., Guillory, S. B., & Kaldy, Z. (2016). Off to a good start: The early development of the neural substrates underlying visual working memory. *Frontiers in Systems Neuroscience*, *10*(AUG). https://doi.org/10.3389/fnsys.2016.00068

Foroughi, C. K., Monfort, S. S., Paczynski, M., McKnight, P. E., & Greenwood, P. M. (2016). Placebo effects in cognitive training. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(27), 7470–7474. https://doi.org/10.1073/pnas.1601243113

Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision*, *10*(12), 27. https://doi.org/10.1167/10.12.27.

Fregni, F., Boggio, P. S., Nitsche, M., Bermpohl, F., Antal, A., Feredoes, E., Marcolin, M. A., Rigonatti, S. P., Silva, M. T. A., Paulus, W., & Pascual-Leone, A. (2005). Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory. *Experimental Brain Research*, *166*(1), 23–30. https://doi.org/10.1007/s00221-005-2334-6

Fried, E. I., & Flake, J. K. (2018, March). Measurement matters. *Association for Psychological Science*. https://www.psychologicalscience.org/observer/measurement-matters

Frischkorn, G. T., & Popov, V. (2023). A tutorial for estimating mixture models for visual working memory tasks in brms: Introducing the Bayesian Measurement Modeling (bmm) package for R. In *PsyArXiv*. https://doi.org/https://doi.org/10.31234/osf.io/umt57.

Frost, A., Moussaoui, S., Kaur, J., Aziz, S., Fukuda, K., & Niemeier, M. (2021). Is the n-back task a measure of unstructured working memory capacity? Towards understanding its connection to other working memory tasks. *Acta Psychologica*, *219*, 103398. https://doi.org/10.1016/J.ACTPSY.2021.103398

Froudist-Walsh, S., López-Barroso, D., José Torres-Prioris, M., Croxson, P. L., & Berthier, M. L. (2018). Plasticity in the working memory system: Life span changes and response to injury. *Neuroscientist*, *24*(3), 261–276. https://doi.org/10.1177/1073858417717210

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin and Review*, *17*(5), 673–679. https://doi.org/10.3758/17.5.673

Gathercole, S. E., Dunning, D. L., Holmes, J., & Norris, D. (2019). Working memory training involves learning new skills. *Journal of Memory and Language*, *105*(December 2018), 19–42. https://doi.org/10.1016/j.jml.2018.10.003

Goldthorpe, R. A., Rapley, J. M., & Violante, I. R. (2020). A systematic review of non-invasive brain stimulation applications to memory in healthy aging. *Frontiers in Neurology*, *11*(575075). https://doi.org/10.3389/FNEUR.2020.575075/BIBTEX

Gorgoraptis, N., Catalao, R. F. G., Bays, P. M., & Husain, M. (2011). Dynamic updating of working memory resources for visual objects. *Journal of Neuroscience*, *31*(23), 8502–8511. https://doi.org/10.1523/JNEUROSCI.0208-11.2011

Guest, O., & Martin, A. E. (2021). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*, *16*(4), 789–802. https://doi.org/10.1177/1745691620970585

Guye, S., & von Bastian, C. C. (2017). Working memory training in older adults: Bayesian evidence supporting the absence of transfer. *Psychology and Aging*, *32*(8), 732–746. https://doi.org/10.1037/pag0000206

Halsey, L. G., Curran-Everett, D., Vowler, S. L., & Drummond, G. B. (2015). The fickle P value generates irreproducible results. *Nature Methods*, *12*(3), 179–185. https://doi.org/10.1038/nmeth.3288

Hedges, L. V, & Schauer, J. M. (2019). More than one replication study is needed for unambiguous tests of replication. *Journal of Educational and Behavioral Statistics*, *44*, 543–570. https://doi.org/10.3102/1076998619852953

Heinen, K., Sagliano, L., Candini, M., Husain, M., Cappelletti, M., & Zokaei, N. (2016).

Cathodal transcranial direct current stimulation over posterior parietal cortex enhances distinct aspects of visual working memory. *Neuropsychologia*, *87*, 35–42. https://doi.org/10.1016/j.neuropsychologia.2016.04.028

Hill, A. T., Fitzgerald, P. B., & Hoy, K. E. (2016). Effects of anodal transcranial direct current stimulation on working memory: A systematic review and meta-analysis of findings from healthy and neuropsychiatric populations. *Brain Stimulation*, *9*(2), 197–208. https://doi.org/10.1016/j.brs.2015.10.006

Horvath, J. C., Forte, J. D., & Carter, O. (2015). Quantitative review finds no evidence of cognitive effects in healthy populations from single-session transcranial direct current stimulation (tDCS). *Brain Stimulation*, *8*(3), 535–550. https://doi.org/10.1016/j.brs.2015.01.400

Hoy, K. E., Emonson, M. R. L., Arnold, S. L., Thomson, R. H., Daskalakis, Z. J., & Fitzgerald, P. B. (2013). Testing the limits: Investigating the effect of tDCS dose on working memory enhancement in healthy controls. *Neuropsychologia*, *51*, 1777–1784. https://doi.org/10.1016/j.neuropsychologia.2013.05.018

Hsu, T.-Y., Tseng, P., Liang, W.-K., Cheng, S.-K., & Juan, C.-H. (2014). Transcranial direct current stimulation over right posterior parietal cortex changes prestimulus alpha oscillation in visual short-term memory task. *NeuroImage*, *98*, 306–313. https://doi.org/10.1016/j.neuroimage.2014.04.069

Hsu, W.-Y., Ku, Y., Zanto, T. P., & Gazzaley, A. (2015). Effects of noninvasive brain stimulation on cognitive function in healthy aging and Alzheimer's disease: A systematic review and meta-analysis. *Neurobiology of Aging*, *36*(8), 2348–2359. https://doi.org/10.1016/j.neurobiolaging.2015.04.016

Ikkai, A., & Curtis, C. E. (2011). Common neural mechanisms supporting spatial working memory, attention and motor intention. *Neuropsychologia*, *49*(6), 1428–1434.

https://doi.org/10.1016/j.neuropsychologia.2010.12.020

Jackson, M. P., Rahman, A., Lafon, B., Kronberg, G., Ling, D., Parra, L. C., & Bikson, M. (2016). Animal models of transcranial direct current stimulation: Methods and mechanisms. *Clinical Neurophysiology*, *127*(11), 3425–3454. https://doi.org/10.1016/j.clinph.2016.08.016

Jacobson, L., Koslowsky, M., & Lavidor, M. (2012). TDCS polarity effects in motor and cognitive domains: A meta-analytical review. *Experimental Brain Research*, *216*(1), 1–10. https://doi.org/10.1007/s00221-011-2891-9

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(19), 6829–6833. https://doi.org/10.1073/pnas.0801268105

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2012). Cogmed and working memory training–Current challenges and the search for underlying mechanisms. *Journal of Applied Research in Memory and Cognition*, *1*(3), 211–213. https://doi.org/10.1016/j.jarmac.2012.07.002

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, *18*(4), 394–412. https://doi.org/10.1080/09658211003702171

Jiang, S., Jones, M., & von Bastian, C. C. (2023). Mechanisms of cognitive change: Training improves the quality but not the quantity of visual working memory representations. *Journal of Cognition*, *6*(1), 1–20. https://doi.org/10.5334/joc.306

Johnson, E. L., Arciniega, H., Jones, K. T., Kilgore-Gomez, A., & Berryhill, M. E. (2022). Individual predictors and electrophysiological signatures of working memory enhancement in aging. *NeuroImage*, *250*(January), 118939.

https://doi.org/10.1016/j.neuroimage.2022.118939

Jones, K. T., & Berryhill, M. E. (2012). Parietal contributions to visual working memory depend on task difficulty. *Frontiers in Psychiatry*, *3*(SEP), 1–11. https://doi.org/10.3389/fpsyt.2012.00081

Jung, Y. J., Kim, J. H., & Im, C. H. (2013). COMETS: A MATLAB toolbox for simulating local electric fields generated by transcranial direct current stimulation (tDCS). *Biomedical Engineering Letters*, *3*(1), 39–46. https://doi.org/10.1007/s13534-013-0087-x

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the N-back task: A question of construct validity. *Journal of Experimental Psychology: Learning Memory and Cognition*, *33*(3), 615–622. https://doi.org/10.1037/0278-7393.33.3.615

Karbach, J., & Verhaeghen, P. (2015). Making working memory work: A meta-analysis of executive control and working memory training in younger and older adults. *Psychol Sci*, *25*(11), 2027–2037. https://doi.org/10.1177/0956797614548725

Karthikeyan, R., Smoot, M. R., & Mehta, R. K. (2021). Anodal tDCS augments and preserves working memory beyond time-on-task deficits. *Scientific Reports*, *11*(1), 1–11. https://doi.org/10.1038/s41598-021-98636-y

Kassambara, A. (2021). *rstatix: Pipe-Friendly framework for basic statistical tests*. https://cran.r-project.org/package=rstatix

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., … Nosek, B. A. (2018). Many labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, *1*(4), 443–

490. https://doi.org/10.1177/2515245918810225

Klem, G. H., Lüders, H. O., Jasper, H. H., & Elger, C. (1999). The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. *Electroencephalography and Clinical Neurophysiology*.

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, *14*(7), 317–324. https://doi.org/10.1016/j.tics.2010.05.002

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, *24*(6), 781–791. https://doi.org/10.1076/jcen.24.6.781.8395

Kong, G., & Fougnie, D. (2019). Visual search within working memory. *Journal of Experimental Psychology: General*, *148*(10), 1688–1700. https://doi.org/10.1037/xge0000555

Kvarven, A., Strømland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour*, *4*(4), 423–434. https://doi.org/10.1038/s41562-019-0787-z

Kyllonen, P. C., Christal, R. E., Woltz, D., Ackerman, P., & Jensen, A. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*(4), 389–433. https://doi.org/https://doi.org/10.1016/S0160-2896(05)80012-1

Laakso, I., Mikkonen, M., Koyama, S., Hirata, A., & Tanaka, S. (2019). Can electric fields explain inter-individual variability in transcranial direct current stimulation of the motor cortex? *Scientific Reports*, *9*(1), 1–10. https://doi.org/10.1038/s41598-018-37226-x

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*(NOV), 863. https://doi.org/10.3389/fpsyg.2013.00863

Lampit, A., Gavelin, H. M., Sabates, J., Launder, N. H., Hallock, H., Finke, C., Krohn, S., &

Peeters, G. (2020). Computerized cognitive training in cognitively healthy older adults: A systematic review and network meta-analysis. *MedRxiv*. https://doi.org/10.1101/2020.10.07.20208306

Lara, A. H., & Wallis, J. D. (2015). The role of prefrontal cortex in working memory: A mini review. *Frontiers in Systems Neuroscience*, *9*(173). https://doi.org/10.3389/fnsys.2015.00173

Lawrence, M. A. (2016). *ez: Easy analysis and visualization of factorial experiments*. https://cran.r-project.org/package=ez

Lerche, V., Voss, A., & Nagler, M. (2017). How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. *Behavior Research Methods*, *49*(2), 513–537. https://doi.org/10.3758/s13428-016-0740-2

Li, C. H., He, X., Wang, Y. J., Hu, Z., & Guo, C. Y. (2017). Visual working memory capacity can be increased by training on distractor filtering efficiency. *Frontiers in Psychology*, *8*(FEB), 1–12. https://doi.org/10.3389/fpsyg.2017.00196

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400. https://doi.org/10.1016/j.tics.2013.06.006

Ly, A., Verhagen, J., & Wagenmakers, E. J. (2016). Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*, *72*, 19–32. https://doi.org/10.1016/j.jmp.2015.06.004

Ma, W. J., Husain, M., Bays, P. M., Ma, W. J., Husain, M., Bays, P. M., Ji Ma, W., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356. https://doi.org/10.1038/nn.3655

Maatman, F. O. (2021). *Psychology's theory crisis, and why formal modelling cannot solve*

it. https://doi.org/10.31234/osf.io/puqvs

Mair, P., & Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*, *52*, 464–488.

Mancuso, L. E., Ilieva, I. P., Hamilton, R. H., & Farah, M. J. (2016). Does transcranial direct current stimulation improve healthy working memory?: A meta-analytic review. *Journal of Cognitive Neuroscience*, *28*(8), 1063–1089. https://doi.org/10.1162/jocn

Martinussen, R., Hayden, J., Hogg-Johnson, S., & Tannock, R. (2005). A meta-analysis of working memory impairments in children with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*(4), 377–384. https://doi.org/10.1097/01.CHI.0000153228.72591.73

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis?: What does "failure to replicate" really mean? *American Psychologist*, *70*(6), 487–498. https://doi.org/10.1037/a0039400

McNab, F., Varrone, A., Farde, L., Jucaite, A., Bystritsky, P., Forssberg, H., & Klingberg, T. (2009). Changes in cortical dopamine D1 receptor binding associated with cognitive training. *Science*, *323*(5915), 800–802. https://doi.org/10.1126/science.1166102

Medina, J., & Cason, S. (2017a). No evidential value in samples of transcranial direct current stimulation (tDCS) studies of cognition and working memory in healthy populations. *Cortex*, *94*, 131–141. https://doi.org/10.1016/j.cortex.2017.06.021

Medina, J., & Cason, S. (2017b). No evidential value in samples of transcranial direct current stimulation (tDCS) studies of cognition and working memory in healthy populations. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *94*, 131–141. https://doi.org/10.1016/J.CORTEX.2017.06.021

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270–291.

https://doi.org/10.1037/a0028228

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer": Evidence from a meta-analytic review. *Perspectives on Psychological Science*, *11*(4), 512–534. https://doi.org/10.1177/1745691616635612

Meule, A. (2017). Reporting and interpreting working memory performance in n-back tasks. *Frontiers in Psychology*, *8*(352). https://doi.org/10.1002/hup.1248

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge University Press. https://doi.org/https://doi.org/10.1017/CBO9781139174909

Morey, R. D., & Rouder, J. N. (2021). *BayesFactor: Computation of Bayes factors for common designs*. https://cran.r-project.org/package=BayesFactor

Moriya, J. (2019). Visual-working-memory training improves both quantity and quality. *Journal of Cognitive Enhancement*, *3*, 221–232. https://doi.org/10.1007/s41465-018-00120-5

Morra, S., & Borella, E. (2015). Working memory training: From metaphors to models. *Frontiers in Psychology*, *6*(August), 1–4. https://doi.org/10.3389/fpsyg.2015.01097

Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise

and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin and Review*, *18*(1), 46–60. https://doi.org/10.3758/s13423-010-0034-0

Navarro, D. (2015). *Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.6)*. https://learningstatisticswithr.com

Ngiam, W. X. Q. (2023). Mapping visual working memory models to a theoretical framework. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-023-02356-5

Ngiam, W. X. Q., Foster, J. J., Adam, K. C. S., & Awh, E. (2022). Distinguishing guesses from fuzzy memories: Further evidence for item limits in visual working memory. *Attention, Perception, and Psychophysics*, 1695–1709. https://doi.org/10.3758/s13414-022-02631-y

Nikolin, S., Martin, D., Loo, C. K., & Boonstra, T. W. (2018). Effects of tDCS dosage on working memory in healthy participants. *Brain Stimulation*, *11*, 518–527. https://doi.org/10.1016/j.brs.2018.01.003

Nitsche, M. A., & Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulatio. *Journal of Physiology*, *527*(3), 633–639.

Nord, C. L., & Jonathan, R. P. (2015). Non-invasive direct current brain stimulation for depression : the evidence behind the hype. *Acnr*, *15*(5), 9–11.

Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLoS Biology*, *18*(3), 1–8. https://doi.org/10.1371/journal.pbio.3000691

Oberauer, K. (2021). Measurement models for visual working memory—A factorial model comparison. *Psychological Review*, *September*(27). https://doi.org/10.1037/rev0000328

Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin and Review*, *26*(5), 1596–1618. https://doi.org/10.3758/s13423-

019-01645-2

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885–958. https://doi.org/10.1037/bul0000153

Oberauer, K., Stoneking, C., Wabersich, D., & Lin, H. Y. (2017). Hierarchical Bayesian measurement models for continuous reproduction of visual features from working memory. *Journal of Vision*, *17*(5), 1–27. https://doi.org/10.1167/17.5.11

Oberauer, K., Süß, H. M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, *36*(6), 641–652. https://doi.org/10.1016/j.intell.2008.01.007

Ohn, S. H., Park, C.-I., Yoo, W.-K., Ko, M.-H., Pil Choi, K., Kim, G.-M., Taek Lee, Y., Kim, Y.-H., & toYun-Hee Kim, C. (2008). Time-dependent e¡ect of transcranial direct current stimulation on the enhancement of working memory. *Neuroreport*, *19*(43–47). https://pdfs.semanticscholar.org/e70a/6221d2365bfa6e7a3dd492c2e55aabba6463.pdf

Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience*, *7*(1), 75–79. https://doi.org/10.1038/nn1165

Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the "strong-object" hypothesis. *Perception and Psychophysics*, *64*(7), 1055–1067. https://doi.org/10.3758/BF03194756

Olson, I. R., Jiang, Y., & Moore, K. S. (2005). Associative learning improves visual working memory performance. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(5), 889–900. https://doi.org/10.1037/0096-1523.31.5.889

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(aac4716). https://doi.org/10.1126/science.aac4716

Ovalle Fresa, R., & Rothen, N. (2019). Training enhances fidelity of color representations in visual long-term memory. *J Cogn Enhanc*, *3*(3), 315–327.

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*(2), 299–320. https://doi.org/10.1037/0882-7974.17.2.299

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*(4), 369–378. https://doi.org/10.3758/BF03210419

Pasternak, T., Lui, L. L., & Spinelli, P. M. (2015). Unilateral prefrontal lesions impair memory-guided comparisons of contralateral visual motion. *Journal of Neuroscience*, *35*(18), 7095–7105. https://doi.org/10.1523/JNEUROSCI.5265-14.2015

Pavlov, Y. G., Adamian, N., Appelhoff, S., Arvaneh, M., Benwell, C. S. Y., Beste, C., Bland, A. R., Bradford, D. E., Bublatzky, F., Busch, N. A., Clayson, P. E., Cruse, D., Czeszumski, A., Dreber, A., Dumas, G., Ehinger, B., Ganis, G., He, X., Hinojosa, J. A., … Mushtaq, F. (2021). #EEGManyLabs: Investigating the replicability of influential EEG experiments. *Cortex*, *144*, 213–229. https://doi.org/10.1016/j.cortex.2021.03.013

Pelletier, S. J., Lagace, M., St-Amour, I., Arsenault, D., Cisbani, G., Chabrat, A., Fecteau, S., Levesque, M., & Cicchetti, F. (2015). The morphological and molecular changes of brain cells exposed to direct current electric field stimulation. *International Journal of Neuropsychopharmacology*, *18*(5), pyu090–pyu090. https://doi.org/10.1093/ijnp/pyu090

Popov, V. (2023). *If god handed us the ground-truth theory of memory, how would we recognize it?* https://doi.org/10.31234/osf.io/ay5cm

Pratte, M. S., Park, Y. E., Rademaker, R. L., & Tong, F. (2017). Accounting for stimulus-specific variation in precision reveals a discrete capacity limit in visual working

memory. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(1), 6–17. https://doi.org/10.1037/xhp0000302

R Core Team. (2022). *R: A language and environment for statistical computing* (4.1.3). R Foundation for Statistical Computing. https://www.r-project.org/

Redick, T. S. (2019). The hype cycle of working memory training. *Current Directions in Psychological Science*, *28*(5), 423–429. https://doi.org/10.1177/0963721419848668

Redick, T. S., & Lindsey, D. R. B. (2013). Complex span and n-back measures of working memory: A meta-analysis. *Psychonomic Bulletin and Review*, *20*(6), 1102–1113. https://doi.org/10.3758/s13423-013-0453-9

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, *142*(2), 359–379. https://doi.org/10.1037/a0029082

Redick, T. S., Shipstead, Z., Wiemers, E. A., Melby-lervåg, M., & Hulme, C. (2015). What's working in working memory training? An educational perspective. *Educ Psychol Rev*, *27*(4), 617–633. https://doi.org/10.1007/s10648-015-9314-6.What

Reinhart, R. M. G., & Woodman, G. F. (2015). The surprising temporal specificity of direct-current stimulation. *Trends in Neurosciences*, *38*(8), 459–461. https://doi.org/10.1016/j.tins.2015.05.009

Rerko, L., Oberauer, K., & Lin, H.-Y. (2014). Spatial transposition gradients in visual working memory. In *The Quarterly Journal of Experimental Psychology* (Vol. 67, Issue 1, pp. 3–15). Taylor & Francis. https://doi.org/10.1080/17470218.2013.789543

Robison, M. K., McGuirk, W. P., & Unsworth, N. (2017). No evidence for enhancements to visual working memory with transcranial direct current stimulation to prefrontal or posterior parietal cortices. *Behavioral Neuroscience*, *131*(4), 277–288.

https://doi.org/10.1037/bne0000202

Roche, N., Geiger, M., & Bussel, B. (2015). Mechanisms underlying transcranial direct current stimulation in rehabilitation. *Annals of Physical and Rehabilitation Medicine*, *58*, 214–219. https://doi.org/10.1016/j.rehab.2015.04.009

Rose, E. J., & Ebmeier, K. P. (2006). Pattern of impaired working memory during major depression. *Journal of Affective Disorders*, *90*(2–3), 149–161. https://doi.org/10.1016/j.jad.2005.11.003

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 of the United States of America*, *105*(16), 5975–5979. https://doi.org/10.1073/pnas.0711295105

Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin and Review*, *18*(2), 324–330. https://doi.org/10.3758/s13423-011-0055-3

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(2012), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Ruffini, G., Wendling, F., Merlet, I., Molaee-Ardekani, B., Mekonnen, A., Salvador, R., Soria-Frisch, A., Grau, C., Dunne, S., & Miranda, P. C. (2013). Transcranial current brain stimulation (tCS): Models and technologies. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *21*(3), 333–345. https://doi.org/10.1109/TNSRE.2012.2200046

Salehinejad, M. A., Nejati, V., Mosayebi-Samani, M., Mohammadi, A., Wischnewski, M., Kuo, M. F., Avenanti, A., Vicario, C. M., & Nitsche, M. A. (2020). Transcranial direct current stimulation in ADHD: A systematic review of efficacy, safety, and protocol-

induced electrical field modeling results. *Neuroscience Bulletin*, *36*(10), 1191–1212. https://doi.org/10.1007/s12264-020-00501-x

Saucedo-Marquez, C. M., Zhang, X., Swinnen, S. P., Meesen, R., & Wenderoth, N. (2013). Task-Specific effect of transcranial direct current stimulation on motor learning. *Frontiers in Human Neuroscience*, *7*(333), 1–12. https://doi.org/10.3389/fnhum.2013.00333

Sawaguchi, T., & Goldman-Rakic, P. S. (1991). D1 dopamine receptors in prefrontal cortex: Involvement in working memory. *Science*, *4996*(251), 947–950. https://doi.org/10.1126/science.1825731

Schneegans, S., & Bays, P. M. (2016). No fixed item limit in visuospatial working memory. *Cortex*, *83*, 181–193. https://doi.org/10.1016/j.cortex.2016.07.021

Schneegans, S., & Bays, P. M. (2017). Neural architecture for feature binding in visual working memory. *Journal of Neuroscience*, *37*(14), 3913–3925. https://doi.org/10.1523/JNEUROSCI.3493-16.2017

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. *Nature Human Behaviour*, *4*(11), 1156–1172. https://doi.org/10.1038/s41562-020-00938-0

Schwarz, G. (2007). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Scolari, M., Vogel, E. K., & Awh, E. (2008). Perceptual expertise enhances the resolution but not the number of representations in working memory. *Psychonomic Bulletin and Review*, *15*(1), 215–222. https://doi.org/10.3758/PBR.15.1.215

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*(4), 628–654. https://doi.org/10.1037/a0027473

Siegert, A., Diedrich, L., & Antal, A. (2021). New methods, old brains—A systematic review

on the effects of tDCS on the cognition of elderly people. *Frontiers in Human Neuroscience*, *15*(October). https://doi.org/10.3389/fnhum.2021.730134

Simons, D. J. (2014). The Value of Direct Replication. *Perspectives on Psychological Science*, *9*(1), 76–80. https://doi.org/10.1177/1745691613514755

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest*, *17*(3), 103–186. https://doi.org/10.1177/1529100616661983

Smid, C. R., Karbach, J., & Steinbeis, N. (2020). Toward a science of effective cognitive training. *Current Directions in Psychological Science*, *29*(6), 531–537. https://doi.org/10.1177/0963721420951599

Souza, A. S., Rerko, L., Lin, H. Y., & Oberauer, K. (2014). Focused attention improves working memory: implications for flexible-resource and discrete-capacity models. *Attention, Perception, and Psychophysics*, *76*(7), 2080–2102. https://doi.org/10.3758/s13414-014-0687-2

Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, *78*(2), 364–375. https://doi.org/10.1016/j.neuron.2013.01.039

Störmer, V. S., Passow, S., Biesenack, J., & Li, S. C. (2012). Dopaminergic and cholinergic modulations of visual-spatial attention and working memory: Insights from molecular genetic research and implications for adult cognitive development. *Developmental Psychology*, *48*(3), 875–889. https://doi.org/10.1037/a0026198

Strzelczyk, D., Clayson, P. E., Sigurdardottir, H. M., Mushtaq, F., Yuri, G., Devillez, H., Lukashevich, A., Rocha, H. A., Chung, Y. H., Ortego, M., Stoermer, V. S., Alanis, J. C. G., Löffler, C., Biel, A. L., Tretow, A., Xu, W., Hämäläinen, J., Lu, Z., Choi, Y. M., …

Langer, N. (2023). *memory capacity : a multi-site registered replication*. *August*, 1–35.

Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision*, *13*(10), 1–8. https://doi.org/10.1167/13.10.9

Sutterer, D. W., & Awh, E. (2016). Retrieval practice enhances the accessibility but not the quality of memory. *Psychonomic Bulletin and Review*, *23*(3), 831–841. https://doi.org/10.3758/s13423-015-0937-x

Teng, C., & Kravitz, D. J. (2019). Visual working memory directly alters perception. *Nature Human Behaviour*, *3*, 827–836. https://doi.org/10.1038/s41562-019-0640-4

The Improved Clinical Effectiveness through Behavioural Research Group. (2006). Designing theoretically-informed implementation interventions. *Implementation Science*, *1*(1), 1–8. https://doi.org/10.1186/1748-5908-1-4

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*, 751–754. https://doi.org/10.1038/nature02466

Tomić, I., & Bays, P. M. (2022). Perceptual similarity judgments do not predict the distribution of errors in working memory. In *bioRxiv*. https://doi.org/10.1101/2022.05.24.493200

Truong, J., Buschkuehl, M., Smith-Peirce, R. N., Carrillo, A. A., Seitz, A. R., & Jaeggi, S. M. (2022). Change-detection training and its effects on visual processing skills. *Scientific Reports*, *12*(1), 1–14. https://doi.org/10.1038/s41598-022-15649-x

Tseng, P., Hsu, T. Y., Chang, C. F., Tzeng, O. J. L., Hung, D. L., Muggleton, N. G., Walsh, V., Liang, W. K., Cheng, S. K., & Juan, C. H. (2012). Unleashing potential: Transcranial direct current stimulation over the right posterior parietal cortex improves change detection in low-performing individuals. *Journal of Neuroscience*, *32*(31), 10554–10561. https://doi.org/10.1523/JNEUROSCI.0362-12.2012

Tsukada, H., Nishiyama, S., Fukumoto, D., Sato, K., Kakiuchi, T., & Domino, E. F. (2005). Chronic NMDA Antagonism Impairs Working Memory, Decreases Extracellular Dopamine, and Increases D1 Receptor Binding in Prefrontal Cortex of Conscious Monkeys Chronic NMDA Antagonism Impairs Working Memory, Decreases Extracellular Dopamine, and Increases D 1 Receptor Binding in Prefrontal Cortex of Conscious Monkeys. *Article in Neuropsychopharmacology*, *30*, 1861–1869. https://doi.org/10.1038/sj.npp.1300732

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*(1), 124–149. https://doi.org/10.1037/a0035234

Van Den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(22), 8780–8785. https://doi.org/10.1073/pnas.1117465109

von Bastian, C. C., Belleville, S., Reinhartz, A., & Strobach, T. (2023). Reply to 'Efficiency and capacity mechanisms can coexist in cognitive training.' *Nature Reviews Psychology*, *2*(2), 128. https://doi.org/10.1038/s44159-022-00147-8

von Bastian, C. C., Belleville, S., Udale, R. C., Reinhartz, A., Essounni, M., & Strobach, T. (2022). Mechanisms underlying training-induced cognitive change. *Nature Reviews Psychology*, *1*(1), 30–41. https://doi.org/10.1038/s44159-021-00001-3

von Bastian, C. C., & Eschen, A. (2016). Does working memory training have to be adaptive? *Psychological Research*, *80*(2), 181–194. https://doi.org/10.1007/s00426-015-0655-z

von Bastian, C. C., Guye, S., & De Simoni, C. (2019). How strong is the evidence for the effectiveness of working memory training? In *Cognitive and Working Memory Training* (pp. 58–76). https://doi.org/10.1093/oso/9780199974467.003.0004

von Bastian, C. C., Locher, A., & Ruflin, M. (2013). Tatool: A Java-based open-source programming framework for psychological studies. *Behavior Research Methods*, *45*(1), 108–115. https://doi.org/10.3758/s13428-012-0224-y

von Bastian, C. C., & Oberauer, K. (2014). Effects and mechanisms of working memory training: A Review. *Psychological Research*, *78*(6), 803–820. https://doi.org/10.1007/s00426-013-0524-6

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*(2), 228–243. https://doi.org/10.1037/a0027127

Wang, K., & Qian, J. (2021). Training with high perceptual difficulty improves the capacity and fidelity of internal representation in VWM. *Psychological Research*, *85*(6), 2408–2419. https://doi.org/10.1007/s00426-020-01404-2

Wang, S., Itthipuripat, S., & Ku, Y. (2019). Electrical stimulation over human posterior parietal cortex selectively enhances the capacity of visual short-term memory. *The Journal of Neuroscience*, *39*(3), 528–536. https://doi.org/10.1523/JNEUROSCI.1959-18.2018

Wetzels, R., & Wagenmakers, E. J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin and Review*, *19*(6), 1057–1064. https://doi.org/10.3758/s13423-012-0295-x

Wiecki, T. V., & Frank, M. J. (2013). A computational model of inhibitory control in frontal cortex and basal ganglia. *Psychological Review*, *120*(2), 329–355. https://doi.org/10.1037/a0031542

Wilhelm, O., Hildebrandt, A., Oberauer, K., & Deyoung, C. G. (2013). *What is working memory capacity, and how can we measure it?*

https://doi.org/10.3389/fpsyg.2013.00433

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 1120–1135. https://doi.org/10.1167/4.12.11

Williams, J. R., Robinson, M. M., & Brady, T. F. (2022). There is no theory-free measure of "swaps" in visual working memory experiments. *Computational Brain and Behavior*, *2008*. https://doi.org/10.1007/s42113-022-00150-5

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *ELife*, *8*. https://doi.org/10.7554/eLife.49547

Wischnewski, M., Mantell, K. E., & Opitz, A. (2021). Identifying regions in prefrontal cortex related to working memory improvement: A novel meta-analytic method using electric field modeling. *Neuroscience and Biobehavioral Reviews*, *130*(July), 147–161. https://doi.org/10.1016/j.neubiorev.2021.08.017

Wolfe, J. M., & Horowitz, T. S. (1998). Visual search has no memory. *Nature*, *394*, 575–577.

Woodham, R., Rimmer, R. M., Mutz, J., & Fu, C. H. Y. (2021). Is tDCS a potential first line treatment for major depression? *International Review of Psychiatry*, *33*(3), 250–265. https://doi.org/10.1080/09540261.2021.1879030

Woodman, G. F., Vogel, E. K., & Luck, S. J. (2001). Visual search remains efficient when visual working memory is full. *Psychological Science*, *12*(3), 219–224.

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*(7080), 91–95. https://doi.org/10.1038/nature04262

Yang, Y. (2005). Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation. *Biometrika*, *92*(4), 937–950. https://doi.org/10.1093/biomet/92.4.937

Yoo, A. H., Klyszejko, Z., Curtis, C. E., & Ma, W. J. (2018). Strategic allocation of working memory resource. *Scientific Reports*, *8*(1), 1–8. https://doi.org/10.1038/s41598-018-

34282-1

Yuen, K. K. . (1974). The two-sample trimmed t for unequal population variances. *Biometrika*, *61*(1), 165–170. http://www.jstor.com/stable/2334299

Zaehle, T., Sandmann, P., Thorne, J. D., Jäncke, L., & Herrmann, C. S. (2011). Transcranial direct current stimulation of the prefrontal cortex modulates working memory performance: combined behavioural and electrophysiological evidence. In *BMC Neuroscience* (Vol. 12). https://doi.org/10.1186/1471-2202-12-2

Zerr, P., Gayet, S., van den Esschert, F., Kappen, M., Olah, Z., & Van der Stigchel, S. (2021). The development of retro-cue benefits with extensive practice: Implications for capacity estimation and attentional states in visual working memory. *Memory and Cognition*, *49*(5), 1036–1049. https://doi.org/10.3758/s13421-021-01138-5

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860

Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological Science*, *22*(11), 1434–1441. https://doi.org/10.1177/0956797611417006

Živanović, M., Paunović, D., Konstantinović, U., Vulić, K., Bjekić, J., & Filipović, S. R. (2021). The effects of offline and online prefrontal vs parietal transcranial direct current stimulation (tDCS) on verbal and spatial working memory. *Neurobiology of Learning and Memory*, *179*(December 2020). https://doi.org/10.1016/j.nlm.2021.107398

Zokaei, N., Gorgoraptis, N., Bahrami, B., Bays, P. M., & Husain, M. (2011). Precision of working memory for visual motion sequences and transparent motion surfaces. *Journal of Vision*, *11*(14), 1–18. https://doi.org/10.1167/11.14.2