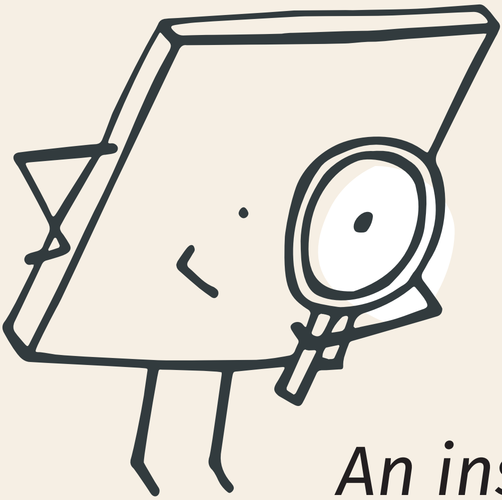
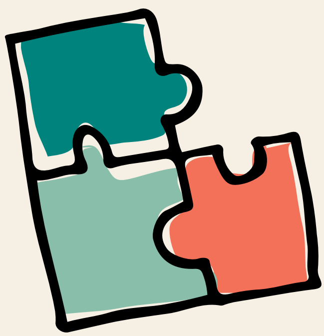
Relying on the external world

Working memory strategies in the context of **memory capacity, external demands,** and **acquired brain injury**



An inspection

by Sanne Böing



Relying on the external world

Working memory strategies
in the context of
memory capacity,
external demands,
and
acquired brain injury

An inspection

by Sanne Böing

Layout: Wendy Bour | Proefschriften.nl
Printed by: Proefschriften.nl
ISBN: 978-90-393-7861-8
DOI: 10.33540/2874

©2025, Sanne Böing
All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form of by any means, electronically, mechanically, by photocopying, recording or otherwise, without the prior written permission of the author.

Relying on the external world:

Working memory strategies
in the context of
memory capacity,
external demands,
and acquired brain injury

Vertrouwen op de buitenwereld:
werkgeheugenstrategieën in de context van geheugencapaciteit,
taakvereisten, en niet-aangeboren hersenletsel
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. ir. W. Hazeleger,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen op

vrijdag 20 juni 2025

des ochtends te

10.15 uur

door

Sanne Böing

geboren op 1 december 1993
te Haarlem

Promotoren:
Prof. dr. S. van der Stigchel
Dr. T.C.W. Nijboer

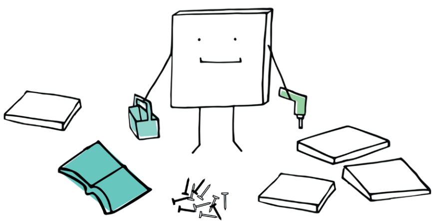
Copromotor:
Dr. A.F. ten Brink

Beoordelingscommissie:
Prof. dr. C.M. van Heugten
Prof. dr. L.J. Kappelle
Prof. dr. C.N.L. Olivers
Dr. I.M. Wiegand
Prof. dr. M.J.E. van Zandvoort

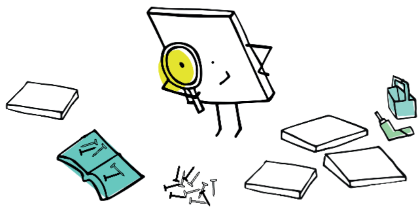
Dit proefschrift werd (mede) mogelijk gemaakt met financiële steun van de European Research Council (ERC) binnen het Horizon 2020 onderzoeks- en innovatieprogramma van de Europese Unie (subsidieovereenkomst 863732).

Preface

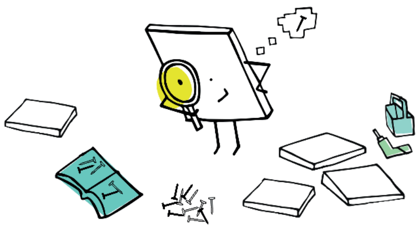
Consider assembling a newly bought wardrobe. After unpacking the boards and taking out your toolbox and screws, you open the instruction leaflet (1). The visual instructions guide your behaviour; you inspect the image of the first screw you need (2), create a mental representation of this screw in your mind's eye (3), use it to search for the screw in the pile of screws (4), and use it for further building (5). You repeat this sequence until your wardrobe is complete (6). In this everyday life activity, there is no need to memorize all the screws, their rotation, and their desired location at once because you can inspect and reinspect the building steps as often as desired.



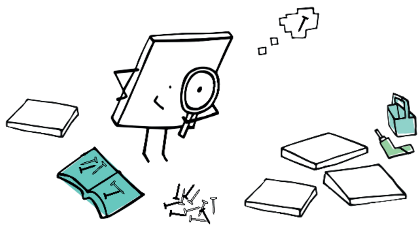
1.



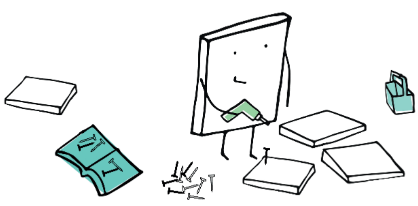
2.



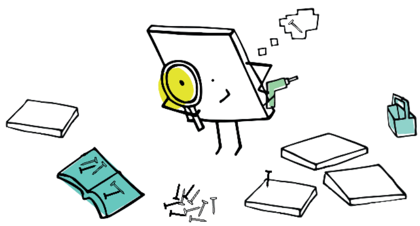
3.



4.

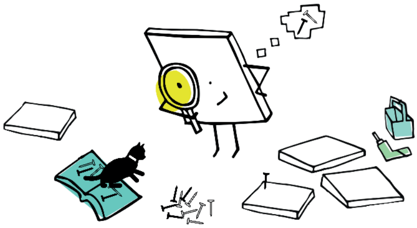


5.



6.

However, the instructions may not be continuously available for you to inspect. At one point your cat may lay down on top of the instruction leaflet, which increases the difficulty with which you can inspect the information to be used. To avoid disturbing your cat, you will try to not memorize one but two screws per inspection, thereby increasing your visual working memory load (7).



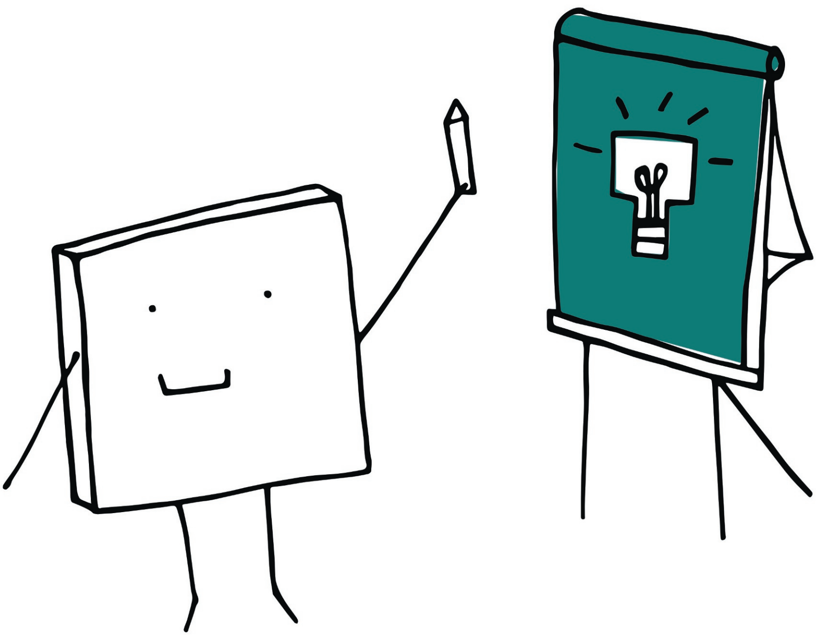
7.

The incentive to increase memory load may become even greater when your cat is grumpy, and you do not want to risk getting scratched each time you try to inspect the instructions. Instead of memorizing two screws, you will increase your memory load to four screws. The interactive and adaptive nature of engaging working memory in natural environments becomes apparent.

Contents

CHAPTER 1

General Introduction



1.1. Working memory

Of all our cognitive functions, memory is perhaps the one that most appeals to our imagination: memory comprises the encoding, storage, recall, recognition and use of relevant information about the world around us and ourselves. Memory is multifaceted, and different systems are involved. Semantic declarative memory allows us to comprehend a language and attribute meaning to objects (e.g., Martin, 2021; Rogers et al., 2003; Winters et al., 2008); episodic memory allows us to create a coherent timeline of our lives, and to relive our happiest but also our darkest moments (e.g., Piolino et al., 2009; Tulving, 2002); short-term memory enables us to dial a phone number that someone just dictated (e.g., Baddeley, 2003; Cowan, 2017); and prospective memory reminds us of upcoming appointments or intentions (e.g., van den Berg et al., 2012). Specifically, the concept of *working* memory refers to a dedicated system that allows to temporarily represent and manipulate information. Working memory serves as an interface between perception, long-term memory, and action. The system is described in a multicomponent model, with separate phonological and visuospatial storage systems, an episodic buffer that integrates information from the modalities and that interacts with long-term memory, and a central executive that flexibly allocates attentional resources (see Figure 1.1; Hitch et al., 2024).

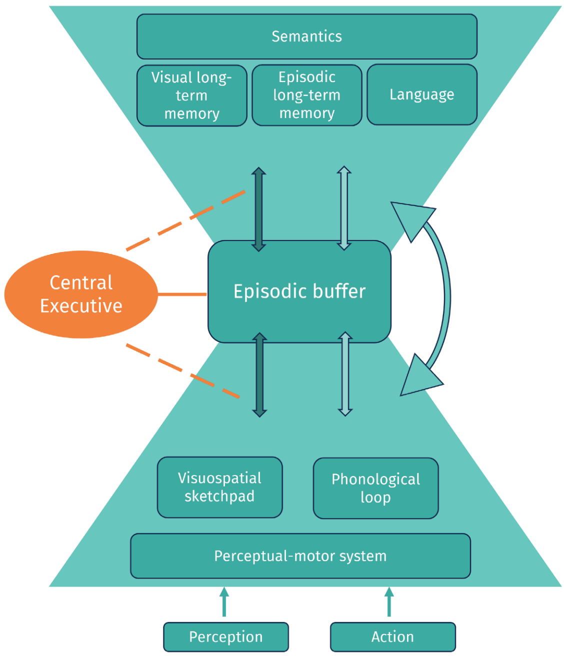


Figure 1.1. The multicomponent model of working memory (Hitch et al., 2024).

Working memory enables us to perform a wide range of behaviours, and is required for complex cognition (Baddeley et al., 2021). The system is considered to play a key role in, among others, decision-making, and problem-solving (Logie et al., 2020), and has been found to correlate with numerous cognitive constructs, such as fluid intelligence (Shelton et al., 2010). Positioned at the intersection of memory and executive function, working memory is not only a core aspect of memory but can also be considered an integral component of executive functions.

1.2. Traditional assessment of working memory

Research on working memory has traditionally been concerned with estimating its maximum capacity (e.g., Conway et al., 2005; Cowan, 2001; Luck & Vogel, 2013). Ranging from Miller's 'magical number' seven (Miller, 1956) to Cowan's four (Cowan, 2001), the limits of working memory are estimated based upon working memory capacity tasks that draw on either the phonological loop (when it concerns verbal material), the visuospatial sketchpad (for non-verbal material), or both. A common feature of capacity tasks is that participants are briefly presented with a varying set size of items that must be stored, manipulated, and recalled after a short retention period. The maximum number of accurately memorized items (i.e., span) serves as an estimation of working memory capacity. To optimize performance on a capacity test, people should try to retain as much information as possible, and thus use their working memory capacity to the fullest.

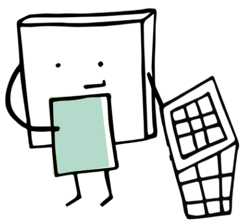
Capacity tasks have provided fundamental insights into the mechanisms of working memory (Luck & Vogel, 2013; Ma et al., 2014), and may be used clinically to discriminate between intact, below average, and impaired capacity, to construct a cognitive profile and subsequently to guide diagnosis in clinical care (e.g., Corsi, 1972; Wechsler, 2012). However, capacity tasks fall short in mimicking actual working memory usage in daily life. Capacity tasks are generally administered in a distraction-free setting with little recruitment of other cognitive functions. In daily life, however, individuals are often faced with a dynamic environment that urges them to simultaneously recruit other cognitive functions in order to successfully complete the memory task at hand. Moreover, the nature of capacity tasks inherently disregards the fact that information often remains available in the external world in everyday life situations. This means that an individual may use the outside world as an external buffer thereby decreasing the need to use their full capacity. Additionally, capacity estimates do not reflect the fact that individuals may exert different memory strategies to arrive at optimal task performance. In other words, capacity tasks are concerned with estimating *memory capacity potential*, rather than testing *whether* and *how* an individual will actually use this potential in more dynamic situations.

1.3. Internal and external memory strategies

Individuals can enhance working memory task performance by using either internal or external strategies. Internal strategies support working memory by optimizing efficient use of working memory capacity, and include visualization techniques (Kaschel et al., 2002; Morrison & Chein, 2011), chunking or grouping (Morrison & Chein, 2011; Norris & Kalm, 2021), and internal rehearsal (Morrison & Chein, 2011; Tan & Ward, 2008). If the individual knows how to exploit such internal memory strategies effectively, the number of reported memory items can increase rapidly, task performance can be improved, and capacity potential may be exploited to the fullest.

On the other hand, external strategies refer to using the outside world as a memory buffer, for example by using calendars, smartwatches, whiteboards, planners, grocery lists or (self-set) cues. The use of such external resources is closely related to the concept of cognitive offloading. Cognitive offloading is described as the use of a physical action (e.g., writing things down, tilting the head to circumvent mental rotation) to reduce the cognitive demand, to overcome capacity limitations (Risko & Gilbert, 2016), and to secure task performance. Externally offloading future intentions by setting cues has, for example, been observed to increase accuracy for short-term future intentions in healthy individuals (Gilbert, 2015a; Gilbert et al., 2023), and writing things down offered support for retro-active memory in younger and older adults (Burnett & Richmond, 2023). Perhaps not so surprising is that increasing the amount of information that has to be remembered increases the likelihood of offloading (Risko & Gilbert, 2016). The use of external memory strategies naturally increases with aging, and exploiting this tendency may especially benefit older adults and memory impaired individuals (Pizzonia & Suhr, 2022). Interestingly, however, people may use offloading even though it does not necessarily benefit performance but mostly serves to safeguard a feeling of security (Risko & Dunn, 2015).

Transferring knowledge in-the-head to knowledge in-the-world is one expression of offloading (Gray & Fu, 2004; Risko & Gilbert, 2016), but one can also decide to not internally load information in the first place, and leave information in the external world for (possible) access later in time (Van der Stigchel, 2020). For example, rather than internalizing the items on the shopping list, one can simply rely on the written note and look up the required information when in the relevant aisle in the supermarket. In such cases, individuals can choose to memorize information to the preferred load and inspect and use the external world when needed. Even with a reduced working memory



capacity, individuals may not be hindered in carrying out such an everyday life activity. At the same time, individuals with a normal working memory capacity may deliberately choose to not use their maximum capacity. In sum, the individual can simply inspect, reinspect, and decide to internalize and act upon information only once needed, thereby circumventing the need for memory use by making a body, head, hand or eye movement.

1.4. Eye movements

In the majority of everyday life activities, humans make three to four eye movements per second (Findlay & Gilchrist, 2003). Saccades, the jerk-like jumpy movements of our eyes, allow us to redirect gaze from one position to another. They are separated by fixations, short periods of time in which the eyes are (mostly) stationary, leading to foveation and high-resolution processing of a novel target. In order for the visual system to reliably encode and detect an object in a scene, gaze must be directed within two degrees visual angle of that object (Nelson & Loftus, 1980). The fixation time needed to consolidate information is found to be as short as 50 ms per item, depending on the item and task characteristics (Vogel et al., 2006). Every additional second of viewing time results in a higher probability of correctly encoding an item for further use (Sahakian et al., 2024), supporting the claim that viewing time predicts the depth of encoding (Koevoet et al., 2023; Somai et al., 2020). Thus, saccade position and viewing times carry implicit information about which information in the outside world is processed, or *sampled*. The oculomotor system therefore offers a unique perspective on the use of the outside world during visual memory tasks: only by virtue of the visual system may we internalize, thus memorize, external visual information. Eye movement characteristics have the potential to reveal how humans use the outside world for just-in-time encoding of the information that is needed at a given point in time, and whether and how individuals choose to deploy their memory capacity potential in interaction with the environment.

1.5. Interactive working memory

1.5.1. Assessing the working memory trade-off between sampling and storing. Interaction with the environment thus allows individuals to underutilize their working memory capacity. The influence of the environmental context on the use of memory has received increased attention in the fields of cognitive engineering and human-system interactions (e.g., Gray & Fu, 2004; Morgan et al., 2009; Waldron et al., 2007), and working memory research (Ballard et al., 1995; Draschkow et al., 2021; Droll & Hayhoe, 2008; Grinschgl, Papenmeier, et al., 2021; Hoogerbrugge, Strauch, Böing, et al., 2024; Kvitelashvili & Kessler, 2024; Melnik et al., 2018; Meyerhoff et al., 2021; Risko & Gilbert, 2016; Sahakian et al., 2023; Somai et al., 2020; Van der Stigchel, 2020). Sampling behaviour – (re)orienting to and (re)inspecting information-to-be-used from the environment once it becomes relevant – is used as an indicator of reliance on the external world. *Sampling* frequency is high when information is relatively easily accessible, and decreases when it is more effortful or costly to access external information (Ballard et al., 1995; Draschkow et al., 2021; Droll & Hayhoe, 2008; Melnik et al., 2018; Sahakian et al., 2023; Somai et al., 2020). A decrease in external sampling implies an increase in storing information internally. This hints at a trade-off between external sampling and internal storing (Figure 1.2), where humans opt for the most cost-efficient alternative (Van der Stigchel, 2020). Interestingly, even when the demands of the environmental context encourage people to shift towards memorization, people may load up less than their maximum capacity or may use a cognitive offloading strategy to avoid full capacity use (e.g., writing things down, creating cues as reminder; Ballard et al., 1995; Draschkow et al., 2021; Gray et al., 2006; Meyerhoff et al., 2021; Morrison & Richmond, 2020; Risko & Dunn, 2015; Sahakian et al., 2023; Somai et al., 2020).

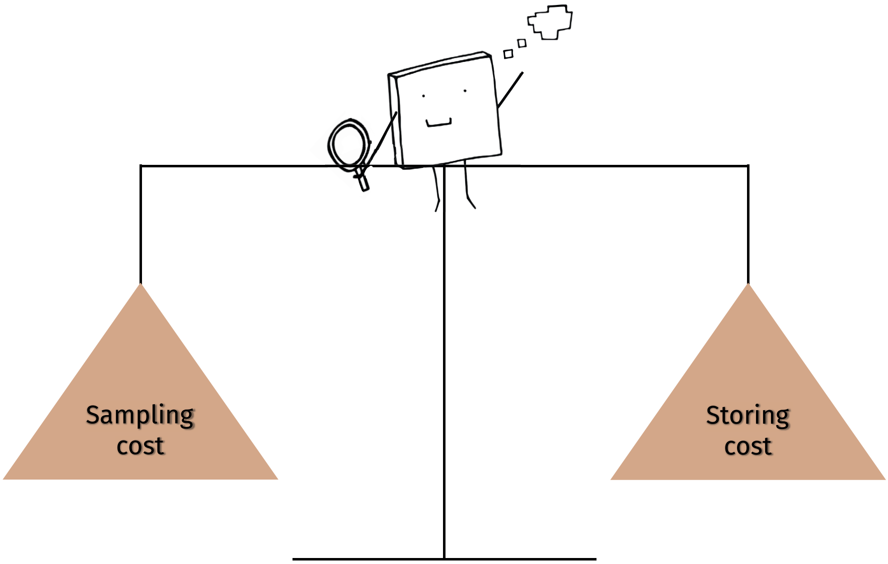


Figure 1.2. The trade-off between external sampling and internal storing. Whenever the cost of sampling or storing increases or decreases, a new equilibrium emerges. If sampling weighs least, this will be the act of choice. Vice versa, if storing weighs least, the individual will memorize information instead of sampling it externally.

Numerous studies manipulated the cost of external sampling, and showed that information availability heavily influences the trade-off between sampling and storing. Variability in the *cost of internal storage*, however, remains largely underexplored. Manipulating the cost of storing may be induced experimentally, for example by using dual-tasks (e.g., Doherty et al., 2019), and although not directly applied to the trade-off framework, such paradigms proved useful in fundamental research assessing the influence of storage costs on memory performance. However, instead of experimentally, increased storage costs also arise naturally. While most studies investigating a trade-off between external offloading and internal memory used healthy adolescents with intact memory functioning, little to no attention is paid to ageing and brain injuries that may result in decreased memory functioning. Neuropsychology may, however, provide insight in how variations in memory storage costs influence the way people, and more specifically patients, choose to deploy their memory. Using the trade-off framework in neuropsychology is not only theoretically interesting, but will also help to better understand patients' memory function.

1.5.2. The working memory trade-off in neuropsychology.

Memory concerns are common in the general ageing population (Ponds et al., 1997). Some degree of memory loss is inherent to getting older and part of a normal ageing process (Brockmole & Logie, 2013; Park et al., 2002; Tulving, 2002). However, neurological pathology, among which neurodegenerative diseases and cerebrovascular accidents, may lead to abnormal levels of memory functionality, or an abnormally steep decrease in memory functionality over the course of time (Berg et al., 2012; Bilgel et al., 2014; Lu et al., 2022). For example, deficits in general memory function, and in working memory specifically, are reported as a characteristic of Alzheimer's disease (Huntley & Howard, 2010; Stopford et al., 2012), and are common after a cerebrovascular accident (CVA, stroke; Kimonides et al., 2018; Lugtmeijer et al., 2021). Working memory problems after stroke influence long-term quality of life, with depressive symptoms as mediator (Kimonides et al., 2018). Moreover, there seems to be a central role for working memory specifically in everyday life activities (Unsworth et al., 2009). Given the importance of working memory in everyday life activities and the accumulating evidence showing that people may underutilize their capacity when the task allows, it is striking that the neuropsychological assessment of working memory still hinges on tasks that are mainly designed to estimate patients' maximum memory capacity. Assessing working memory within the trade-off framework across various patient groups has the potential to fill this hiatus, and may bring us one step closer to estimating memory functioning in daily life situations.

The trade-off evolves around minimizing cognitive effort, thereby dictating an inclination towards the most cost-efficient strategy. Memory deficits may lead to an increased effort to memorize information, as more cognitive resources may need to be allocated to internalize the same amount of information as compared to a (neurologically) healthy individual (Aschenbrenner et al., 2023; Engstrom et al., 2013). In **Chapter 2, 3** and **5**, I test the hypothesis that individuals with memory disorders will rely more on the external world to alleviate their memory burden than neurologically healthy controls, even when sampling is impeded.

The effort expenditure may not only be altered by objectified deficits, but may also be subject to subjective concerns. Intriguingly, what people *believe* about their memory function and their objective ability are frequently incongruent; people often experience memory failure in the absence of impaired memory capacity (Beaudoin & Desrichard, 2011; Mattos et al., 2003; Ponds & Jolles, 1996a). This discrepancy may arise either because capacity tasks are not sensitive enough to detect subtle memory deviations or because there are no objective problems. In the latter case, there are often psychological factors such as depressive symptoms or anxiety that explain (part of) the subjective complaints (Ponds & Jolles, 1996a; Schmand et al., 1997; Steinberg et al., 2013). I hypothesize that (re)sampling is a proxy for an individual's *belief* about their own memory functioning. Sampling could be seen as an act of checking, where the individual may be inclined to reassure themselves about the accuracy of the information to be used. Sampling may then occur as an expression of negative beliefs about one's memory function rather than as a necessity given objective ability. I test how sampling is driven by both objective and subjective components in **Chapter 3.**

Assessing sampling behaviour across various patient groups in a task that allows to use the external world contributes to our fundamental knowledge on working memory usage, and may complement clinical assessment as it could reveal subtle deviations in memory usage – due to either objective problems or subjective complaints – that go unnoticed in capacity tasks. Rather than thinking of memory as a fixed capacity entity that is always fully utilised, I approach working memory as a flexible and interactive system, and hope to elucidate how patients use their memory when they are not forced to use their full capacity.

1.6. Thesis outline

In this thesis, I use the trade-off framework to assess memory use across patient groups with varying levels of memory functioning and healthy controls. Crucially, the paradigm that is used allows participants to store information to their preferred load or to rely on the external world by (re)sampling. **Chapter 2** paves the way for using eye movement behaviour as a proxy of visual working memory usage in individuals with Korsakoff Syndrome, a neuropsychiatric syndrome that is characterized by severe amnesia. **Chapter 3** broadens the scope of our research by testing eye movement behaviour of individuals with a wide range of objective memory capacity impairments and subjective memory complaints who were referred to an outpatient memory clinic. I assess the influence of memory capacity on the tendency to rely on the outside world, and investigate how subjective beliefs further contribute to sampling. From initial group level analyses, I move towards exploring individual differences in strategy use: **Chapter 4** uses secondary data from an online version of the paradigm to create strategy classification labels within a healthy population. I dissociate low-loaders, medium-loaders and high-loaders in a healthy population, and assess whether one of these strategies is better in terms of performance. I further assess the influence of the order in which one encounters situational changes, and check whether people can adaptively shift their eye movement behaviour without affecting performance. Finally, **Chapter 5** builds on the insights from the previous chapter and studies eye movement behaviour in a patient group that is recovering from a cerebrovascular accident (CVA, stroke). In memory rehabilitation, using the external world to support memory (i.e., offloading, sampling) is advocated, but is this actually helpful? I assess spontaneous use of compensation strategies among patients, and test whether a change in sampling behaviour in response to changing external demands has implications for performance. The **General Discussion** summarizes and discusses the implications of my findings from both a theoretical and a neuropsychological perspective. I posit **Debate Boxes** to spark discussion.

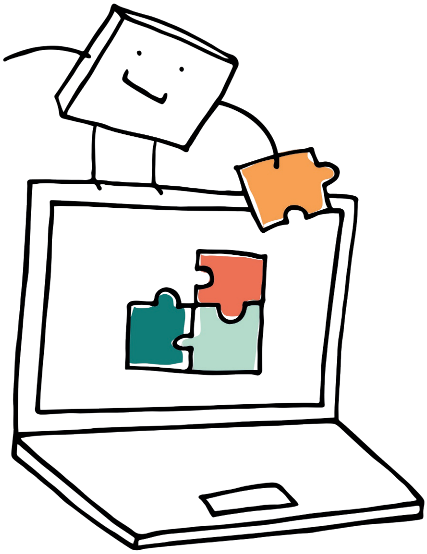
The test materials used in **Chapter 2, 3** and **5** are largely identical, but there are slight study-specific adaptations to the test protocols. The weights assigned to either of the tasks and their outcome variables may vary between chapters. Each chapter explains the materials and variables of main interest, but also refers to their respective Supplementary Materials for more extensive task descriptions. Due to the large overlap in these respective Supplementary Materials, I provide **Supplementary Materials: General** that gives a comprehensive overview of all the tasks used in any of these three chapters. **Supplementary Materials: Chapter-specific** provides the remainder of the Supplementary Materials of the respective chapters that do not overlap.

1.7. Note on jargon

Over the course of four years, wordings and phrasings have slightly changed. Therefore, jargon may differ across chapters while indicating the same thing. The term ‘sampling’ was already introduced, and is interchangeably used with ‘inspecting’ in this dissertation. While ‘offloading’ technically is defined as ‘transferring knowledge out of the head into the world’ and inspecting external information would therefore more accurately be described as ‘non-uploading’, I also use the term ‘offloading’ to indicate sampling behaviour in our populations. ‘Crossings’ and ‘inspections’ indicate the same: only those eye movements that led to sampling of external information. ‘Dwell time’, ‘encoding time’ and ‘inspection time’ all refer to the time spent gazing at the information to be used. ‘Relying on the external world’ is the broader term that encapsulates both the frequency and the duration with which external information is relied upon. ‘Memorizing’ and ‘internal storage’ refer to the activated internal representation of one (or more) item(s) in working memory.

1.8. Let’s puzzle!

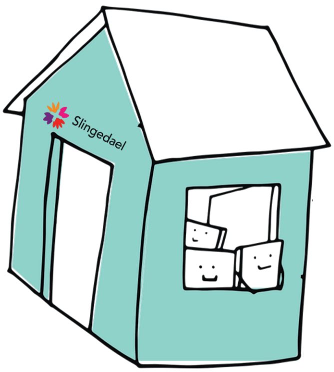
To patients participating in this research, I would always refer to external information to-be-copied as a jigsaw puzzle that they had to rebuild. This puzzle metaphor has not only been central to the task, but also reminded me of researchers’ daily practice. Meticulously inspecting one piece of the puzzle won’t solve the whole thing, but it is a definite requirement before pieces can start falling into their respective place. With this dissertation, I aim to lay some pieces of the puzzle.



CHAPTER 2

Eye movements as proxy for visual working memory usage: Increased reliance on the external world in Korsakoff syndrome

Sanne Böing, Antonia F. Ten Brink, Alex J. Hoogerbrugge, Erik Oudman, Albert Postma, Tanja C.W. Nijboer, & Stefan Van der Stigchel



Published as
Böing, S., Ten Brink, A. F., Hoogerbrugge, A. J., Oudman, E., Postma, A., Nijboer, T. C. W., & Van der Stigchel, S. (2023). Eye movements as proxy for visual working memory usage: Increased reliance on the external world in Korsakoff Syndrome. *Journal of Clinical Medicine*, 12(11). *https://doi.org/10.3390/jcm12113630*

Abstract

In the assessment of visual working memory, estimating the maximum capacity is currently the gold standard. However, traditional tasks disregard that information generally remains available in the external world. Only when to-be-used information is not readily accessible, memory is taxed. Otherwise, people sample information from the environment as a form of cognitive offloading. To investigate how memory deficits impact the trade-off between sampling externally or storing internally, we compared gaze behaviour of individuals with Korsakoff amnesia (n = 24, age range 47–74 years) and healthy controls (n = 27, age range 40–81 years) on a copy task that provoked different strategies by having information freely accessible (facilitating sampling) or introducing a gaze-contingent waiting time (provoking storing). Indeed, patients sampled more often and longer, compared to controls. When sampling became time-consuming, controls reduced sampling and memorized more. Patients also showed reduced and longer sampling in this condition, suggesting an attempt at memorization. Importantly, however, patients sampled disproportionately more often than controls, whilst accuracy dropped. This finding suggests that amnesia patients sample frequently and do not fully compensate for increased sampling costs by memorizing more at once. In other words, Korsakoff amnesia resulted in a heavy reliance on the world as 'external memory'.

Keywords: visual working memory; external memory; acquired brain injury; copy task; eye movements; Korsakoff syndrome; cognitive offloading

1. Introduction

To objectify memory complaints and estimate memory functioning following acquired brain injury, traditionally, patients are asked to memorize an increasing number of briefly presented stimuli and are assessed on how many items have been retained (e.g., Change Detection Task (Luck & Vogel, 2013); Corsi Block-Tapping Task (Corsi, 1972; Kessels et al., 2000); Digit Span Task (Wechsler, 2012)). The maximum storage capacity is then used to dissociate between normative and deviant performance, and subsequently to guide diagnosis and understand patient (dys) functioning in daily situations. However, such tests disregard that in daily life information typically remains available in the external world. We can easily sample information by making eye-movements, using the environment as 'external memory'. Sampling information from the external world is reminiscent of cognitive offloading, where people decide to perform a physical action in order to reduce the internal cognitive effort to carry out a task (Burnett & Richmond, 2023; Gilbert, 2015a; Meyerhoff et al., 2021; Risko & Gilbert, 2016). Similarly, as making an eye movement is easy and quick, people generally tend to sample information instead of memorizing it. Sampling information that is easily accessible in the external world reduces the need to use the maximum VWM storage capacity. Contrarily, when it is difficult or costly to access information in the external world, sampling rates decrease and reliance on internal VWM storage increases (Ballard et al., 1995; Draschkow et al., 2021; Droll & Hayhoe, 2008; Melnik et al., 2018; Sahakian et al., 2023; Somai et al., 2020). This implies a cost-efficient trade-off between sampling and storing. Consequently, the existence of such a trade-off suggests that the maximum storage capacity is often not used in natural behaviour. Capacity scores might therefore not translate to memory functioning in daily life (Van der Stigchel, 2020). A better way to approximate memory functioning in daily life might be by assessing sampling behaviour. The overarching aim of this study was therefore to assess whether eye-movement patterns during the execution of a memory task can serve as a proxy for VWM use in individuals with and without memory impairments. To this end, we compared gaze behaviour of individuals without memory impairments and patients with Korsakoff's amnesia.

Korsakoff's syndrome (KS) is a neuropsychiatric disorder that is caused by thiamine deficiency. Alcohol abuse accounts for 90% of thiamine deficiency (Harper, 1983; Kopelman et al., 2009), but other medical conditions can also lead to KS (Oudman et al., 2021). The syndrome is characterized by severe episodic memory deficits, which are mainly – but not exclusively – expressed as anterograde amnesia: the inability to encode and retrieve new memories. There is general consensus that these long-term declarative memory deficits are part of the cognitive profile of patients with KS (Kopelman et al., 2009). While it was first assumed that working memory was largely spared (see review Kessels & Kopelman, 2012), there is converging evidence suggesting that specific aspects of working memory (capacity) might be impaired in patients with KS (Kessels & Kopelman, 2012; Oudman et al., 2020; Pitel et al., 2008; Van Asselen et al., 2005). As the previously mentioned studies have used different outcome measures to estimate memory capacity, and the results show variability in outcomes of memory capacity (Kessels & Kopelman, 2012; Oudman et al., 2020; Pitel et al., 2008; Van Asselen et al., 2005), straightforward interpretations of capacity scores are difficult. Also, clinical observations point out that patients oftentimes show normal capacity scores when assessed in a test setting but encounter problems when memory is put to use in daily situations. So, rather than assessing *how much* information patients can possibly store, it could be of substantial value to assess *how* patients dynamically employ memory – reflected in their eye-movement behaviour. Previous eye-tracking studies have already provided evidence that VWM usage is low when information is readily available in the outside world, but increases when sampling information becomes costly (Ballard et al., 1995; Draschkow et al., 2021; Droll & Hayhoe, 2008; Melnik et al., 2018; Somai et al., 2020). However, it is currently unclear what happens to the trade-off between sampling and storing when *storage* is more costly or diminished, i.e., in case of memory deficits. Here, we investigate the tendency to sample externally versus storing internally on a copy task based on information availability *and* memory functioning.

Participants were instructed to rebuild an example puzzle as fast and accurately as possible in an empty grid by dragging the pieces of the puzzle to the correct location. If the information-to-be-copied remained available in the outside world, we expected both individuals with and without memory impairments to heavily rely on external sampling. When the cost of sampling increased (i.e., information became less readily available), individuals without memory impairments were expected to shift their strategy towards memorizing information (Ballard et al., 1995; Draschkow et al., 2021; Droll & Hayhoe, 2008; Melnik et al., 2018; Sahakian et al., 2023; Somai et al., 2020). Importantly, however, we expected patients with Korsakoff's amnesia to adhere to the sampling strategy more than healthy controls, because the cost of memorizing as imposed by the individuals' memory condition outweighs the increased cost of sampling. Not only did we expect to find a different trade-off between groups, we also expected the degree of memory deficits to influence the trade-off: the more severe the memory deficit, the more heavily patients would need to rely on sampling over storing. With every extra item that can be memorized (i.e., span increase as measured on traditional memory tasks), people could theoretically load up an extra item per sample, and are therefore expected to rely less on sampling, particularly when information is not readily available and sampling is deemed costly. Regarding the *type* of memory deficits, we would specifically expect this hypothesis to hold for individuals with higher capacity on traditional outcomes of *visual* working memory.

The current study aims to provide a first step in identifying eye-movement markers indicative of subtle changes in memory usage that cannot be captured by means of assessing one's maximum storage load, but that rather occur in dynamic interaction with our environment.

2. Materials and Methods

2.1. Participants

Patients with Korsakoff's syndrome (KS) were recruited via Slingedael Korsakoff Centre of Expertise (see Supplementary Figure S2.1 for a patient flow chart). All patients fulfilled the DSM-V criteria for alcohol induced major neurocognitive disorder (American Psychiatric Association, 2013) and had an extensive history of alcoholism. All patients had severe thiamine deficiency (Wernicke encephalopathy) before onset of KS. None of the patients was in the Wernicke's Encephalopathy phase at the moment of testing, and all were treated according to available guidelines prior to KS diagnosis. Next, age and education matched controls without memory impairments were recruited via various public and university platforms (e.g., Facebook, family members, university intranet, community centres).

We aimed to recruit 25 patients and 25 controls. This number was based upon previous studies, feasibility of including patients, and a power analysis. Previous studies have reported varying sample sizes ranging from 7 (Ballard et al., 1995) to 72 (Melnik et al., 2018). The original trade-off effect has been observed in a group as small as 7 participants (Ballard et al., 1995), and a previous study from our lab has replicated the effect using eye-tracking with 12 participants (Somai et al., 2020). As we expected larger variability in our patient group, we aimed to recruit at least double the amount of participants in either group. Furthermore, recruiting 25 patients was regarded feasible given logistical challenges that come with testing in patient institutions.

Eventually, we were able to include 24 patients (see Supplementary Figure S2.1 for a patient flow chart) and 27 controls. With the current sample size, for a one-tailed t-test with a power of .8, we should be able to reliably detect effects of Cohen's d=.74 (Faul et al., 2009). Effects usually reported in copy task paradigms are similarly large (Draschkow et al., 2021; Sahakian et al., 2023). Moreover, the linear mixed-effects models that we have used, have higher power than t-tests. Therefore, we were confident that our study would have a large enough power. All participants had to speak Dutch and gave written informed consent prior to the start of the experiment. Healthy controls were compensated for their participation with 7EU per hour paid in increments of 30 minutes, and received compensation for travel costs. Patients were not reimbursed for participation.

The project was approved by the Faculty Ethics Review Board of the Faculty of Social and Behavioural Sciences at Utrecht University (protocol numbers 21-0076 and 21-0270) and the local science committee of Slingedael Korsakoff Centre of Expertise. Consent was obtained, and the protocol was carried out in accordance with the Declaration of Helsinki and Utrecht University and Faculty Ethics Review Board requirements.

2.2. Measurements

2.2.1. Experimental computer tasks.

Apparatus. Experimental tasks were run on a Windows 10 Enterprise computer with an Intel Core i7-4790 CPU and 16GB RAM, and displayed on a 27 inch LCD monitor at a resolution of 2560 x 1440 pixels at 100 Hz. Subjects were seated in a dimly lit room and placed their heads in a chin-rest at ~67.5 cm from the monitor. An EyeLink 1000 eye tracker (SR Research Ltd., Canada) was placed in front of the monitor and was used to track the eyes at a sample rate of 1 kHz. Calibration and validation was performed manually with a 9-point grid attempting to achieve a calibration error of less than 2 degrees visual angle (dva).

Copy Task. We used an adapted version of the Copy Task that was developed in our lab (Somai et al., 2020). The aim of the task is to provoke a strategy switch in relying on visual working memory versus sampling information from the outside world. The experiment was programmed in Python 3.7 using the PyQt5 library (Riverbank Computing Limited, 2019) for visual presentation and mouse and keyboard interaction. PyGaze (Dalmaijer et al., 2014) was used to interact with the eye tracker.

Participants were instructed to copy a model puzzle of 6 in a 3 x 3 ‘example’ grid on the left side of the screen to a 3 x 3 empty grid on the right side of the screen. Participants used a computer mouse with their preferred hand to drag one of the 6 items from the right bottom of the screen (the ‘resource’ grid) to the correct cell in the empty grid. The items were adopted from Arnoult (Arnoult, 1956; Figure 2.1A) and consisted of black geometrical shapes that cannot easily be named to measure reliance on VWM instead of verbalisation strategies (Somai et al., 2020).

The Copy Task consisted of two experimental conditions. In the baseline condition, the example grid was always visible (see Figure 2.1B). Therefore, the 'cost' to gather information from the outside world was low: information was freely available. In the experimental condition, a cost was introduced by manipulating when the example grid became visible. The example grid only appeared after fixating the left side of the screen for a total of 2000 ms, during which an hourglass was presented (see Figure 2.1C). This 'gaze-contingent waiting time' was introduced to increase the cost associated with making an eye-movement to sample information from the outside world.

Subjects were instructed to complete each puzzle as quickly and accurately as possible. Whenever an item was placed in the correct location, the background of the cell behind the item turned green for 700ms and the item remained at that location. If the item was incorrectly placed, it disappeared from the location and the background of the cell turned red for 700ms, after which subjects could make another attempt. After placing all six items correctly or after 42 seconds, the trial was ended. If all six items were placed correctly, positive feedback was shown. If subjects failed to correctly place all items within 42 seconds, a message appeared stating that they ran out of time. By introducing a time limit, we aimed to urge subjects to adopt an efficient strategy (Melnik et al., 2018).

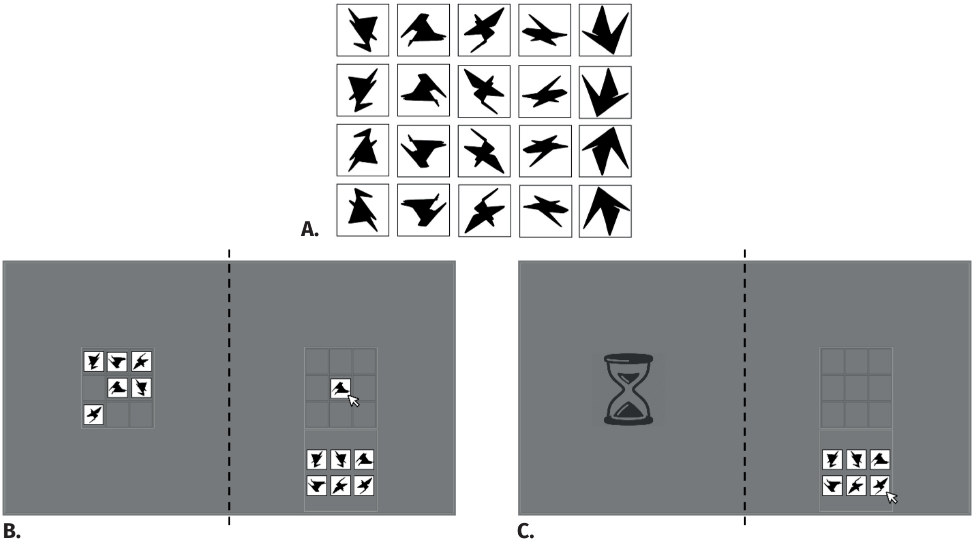


Figure 2.1. **A)** All possible stimuli in the Copy Task. Adopted from Arnoult (1956). An example trial is depicted for the **B)** baseline condition and **C)** high-cost condition of the Copy Task. At the left-hand side of the screen, the example grid is either visible or replaced by an hourglass for 2000 ms (i.e. gaze-contingent occlusion). At the right-hand side of the screen, the empty grid to place the items (top) and the resource grid (bottom) are presented. A trial ended after 42 seconds. Note: the dotted midline is depicted for illustrative purposes and not visible in the experiment.

The Copy Task was divided into two sessions, each session consisting of two blocks. Subjects first performed three practice trials in the baseline condition to get acquainted with the task. Calibration and validation of the eye-tracker were performed after the practice trials. Then, in session one, a baseline block of 15 trials was completed, followed by a high-cost block of 15 trials. In session two, again, a baseline block of 15 trials was completed, followed by a high-cost block of 15 trials, resulting in a total of 30 trials per condition. Although carry-over effects might have played a role (Patrick et al., 2015), we have deliberately chosen for this non-counterbalanced design a priori. The most important consideration was that the gaze-contingency in the high-cost condition is rather complex to understand, especially for patients. We deemed it more most important that the basics of the task were understood first, only to introduce the gaze-contingent waiting time later on.

Before each trial, a drift check (max. 2 degrees visual angle; dva) was performed, and recalibration was performed when deemed necessary. After each block, subjects answered questions on their experience of commitment to and difficulty of the task (not considered in the current analysis). Each session of the Copy Task took 25 – 45 minutes, dependent on calibration time, task speed, and the number and length of breaks.

First, we reported *completion time* and *number of correct placements*, descriptively. Only looking at the completion time would lead to a floor effect for participants who did not complete the trial within 42 seconds (i.e. some participants placed more items correctly than others), and only looking at the number of correct placements would lead to a ceiling effect for participants who completed the trial within 42 seconds (i.e. some participants were faster than others). We, therefore, calculated three performance measures in which the number of correct placements, total attempts (i.e. the sum of the number of correct and incorrect placements), and/or net copy time (i.e. completion time minus the waiting time for the hourglass) were taken into account. *Success rate* reflected the ratio between the number of correct placements and the total attempts. *Speed score* reflected the net copy time divided by the number of correct placements, that is, the net copy time per correctly placed item.

$$Success\ rate\ = \frac{Number\ of\ correct\ placements}{Total\ attempts}$$

$$Speed\ score\ = \frac{Net\ copying\ time}{Number\ of\ correct\ placements}$$

Second, eye-movement outcomes of interest are listed below:

- *Number of crossings.* This refers to the count of *only* those saccades that cross the midline from right to left, thus, which jump from the right (workspace) to the left of the screen (where the example puzzle is located). Crossings capture how often someone looked at the example over the course of the trial.
- *Dwell time per crossing.* This is the total duration of the fixations at the example divided by the number of crossings over the course of the trial. In other words, this score reflects how long someone viewed (i.e., encoded) the example per crossing.
- *Number of crossings per correct placement.* This refers to the count of *only* those saccades that cross the midline from right to left, divided by the number of correct placements. This outcome expresses how often someone needed to inspect the model to place one item correctly.
- *Dwell time per correct placement.* This is the total duration of the fixations at the example, divided by the number of correct placements over the course of a trial. This reflects how much viewing time someone needed to place one item correctly.

Variables were aggregated per participant per condition by mean or median depending on the outcome measure (see Results).

Conceptually, a sampling strategy would translate to a relatively high number of crossings towards the example grid. A memorization strategy would translate to a relatively low number of crossings towards the example grid. Memorization is further expected to translate to longer dwell times per example grid visit to encode more items.

We extracted various other variables (that are not included in the analysis, but serve a descriptive purpose) that can be found in the General Supplementary Materials.

Change Detection Task. Change Detection Tasks are often used in experimental research to assess working memory capacity (Luck & Vogel, 2013). Here, a simplified version of the paradigm from Luck and Vogel was used (Luck & Vogel, 1997; Oudman et al., 2020) (see Figure 2.2). With a varying set size of 2, 3, 4, or 6 items, white bars in different orientations (0°, 30°, 60°, 90°, 120°, 150°) were presented on a black screen for 1000 ms, followed by a gaussian random visual white noise mask for 300 ms. Consecutively, the bars were presented again. One bar was cued by a surrounding red square. The orientation of the cued bar changed in 50% of trials. The orientation of the non-target bars did not change. The participant was instructed to verbally report whether or not the orientation of the cued bar had changed.

Five practice trials were completed prior to the experiment. Here, subjects received feedback on their answers. After practice, 4 blocks of each 20 trials were presented. Every set size was presented 20 times in random order. Here, subjects did not receive feedback on their answers. The task lasted approximately 10 minutes. Eyes were not tracked, only behavioural responses were recorded. Kmax and *d'* were calculated as outcome measures; Kmax is often used in VWM literature (Luck & Vogel, 2013; Magen et al., 2009) and allowed us to compare our findings with previous findings in patients with KS (Oudman et al., 2020). However, *d'* is stated to yield a more robust outcome for visual working memory performance that is less prone to biases in response tendency than Kmax (Williams et al., 2022). Therefore, we used *d'* as capacity score input in further analyses.

$$Kmax = (\text{hit rate} + \text{correct rejection rate} - 1) \times \text{N} \ (\text{N} = \text{memory set size})$$
$$d' = \text{z[p(hits)]} - \text{z[p(falsealarms)]}$$

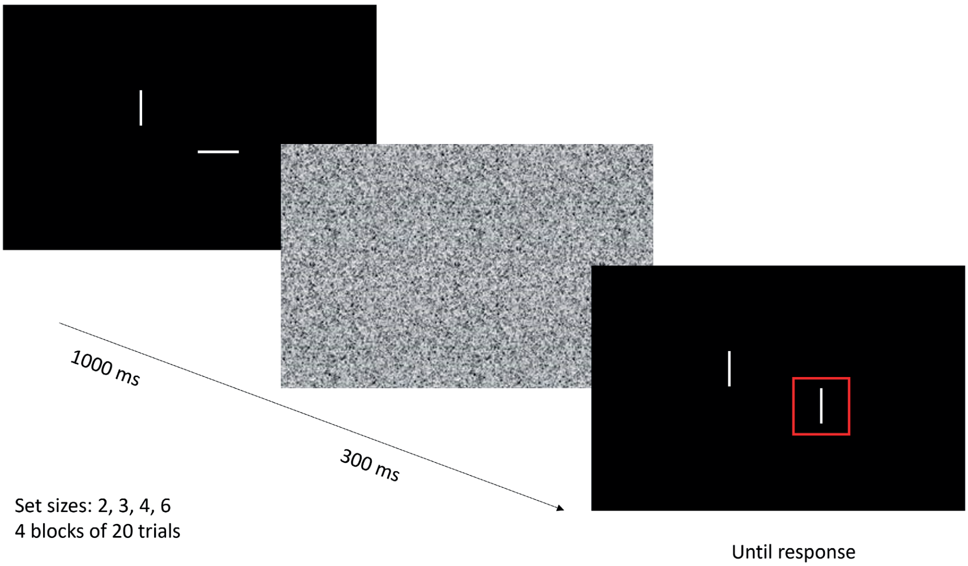


Figure 2.2. Trial overview of the Change Detection Task. Set sizes vary from 2, 3, 4 to 6 white bars in orientations 0°, 30°, 60°, 90°, 120°, 150°.

2.2.2. Neuropsychological tasks. (see Supplementary Materials: General for details).

Location Learning Task (LLT). The standard stimulus set B of the modified Location Learning Task (LLT) was used to assess visuospatial immediate and long-term recall (Kessels et al., 2006, 2014). Primary outcome measures are the learning index (amount of learning over five trials), placement errors (sum of errors over five trials), and the delayed recall score (subtraction of delayed recall placement error minus placement error of fifth trial). A negative score indicates loss of information during retention phase, whereas a positive score indicates a better memory after the retention phase (Kessels et al., 2014).

Rey Auditory Verbal Learning Task. The Rey Auditory Verbal Learning Task (RAVLT; 15 items, Dutch version (Bouma et al., 2012; Saan & Deelman, 1986)) was administered to assess verbal immediate and long-term recall. Outcome measures used are: total number of correct words (range: 0-75) and number of correct words during the delayed recall (range: 0-15). Higher scores indicate better memory capacity.

Digit Span Test (WAIS-IV). We used the Digit Span subtest forward and backward from the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV (Wechsler, 2012)) to assess short-term auditory memory and verbal working memory. The longest sequence that was correctly repeated was used as an outcome measure for maximum capacity (range 2–8 or 2-9, for forward and backward respectively).

Corsi Block Tapping Task. The Corsi Block Tapping Task was used to assess visuospatial working memory (Corsi, 1972; Kessels et al., 2000). We used a digitized version (thus, 2D) of the Corsi Block Tapping Task (Brunetti et al., 2014; Claessen et al., 2015). The forward subtest assesses short-term visuospatial attention; the backward subtest assesses VWM. The longest sequence that was correctly repeated was used as an outcome measure for maximum capacity (forward range 2–9, backward range 2-8).

2.3. Procedure

The test protocol (computer tasks + neuropsychological tasks) was administered with prioritization of tasks with higher importance, while keeping fatigue and physical discomfort (e.g., by keeping the head in the chinrest) at a minimum and taking into account protocols for the delayed assessment of the LLT and RAVLT.
For patients with KS, we divided the test battery in two sessions over separate days (ranging from 1 to 14 days apart, except for one patient who performed the Corsi Block Tapping Task in session 2 and only after 1.5 months). Before the first session, we checked whether patients already had performed some of the neuropsychological tasks as part of standard care or another scientific study that was carried out within six months prior to the experiment. If that was the case,

they were exempt from that task; previously reported scores on those tasks were used in order to prevent unnecessary work load and possible practice effects. Sessions were ended after a maximum of 75 minutes, or when patients became too tired.
For healthy controls, the test protocol was administered in a single visit. The first and second part of the experiment were separated by a break of 10 – 20 minutes. The total administration duration for controls was maximum 3 hours.
Task administration in session 1 comprised (in this order): LLT – direct recall, Copy Task – first session, LLT – delayed recall, Digit Span WAIS IV, and if time allowed: a Fixation and Free viewing task (not taken into account in the current analysis). Task administration in session 2 comprised (in this order): RAVLT – direct recall, Copy Task second session, RAVLT – delayed recall, Corsi Block Tapping Task, and if time allowed: Change Detection Task. See Supplementary Table S2.1 for overview of the test procedure and sessions for controls and patients.

2.4. Data analysis

2.4.1. Pre-processing. Saccades, fixations, and timestamps were extracted with the EyeLink 1000 parser (default EyeLink saccade detection algorithm, SR Research Ltd., Canada). Data pre-processing was implemented in Python 3.10. Data analyses were conducted in R 4.1.2 (R Core Team, 2021).

2.4.2. Group comparisons.

Demographics. Groups (i.e., controls and patients) were matched on age and level of education, since these factors are related to performance on memory tasks (Brockmole & Logie, 2013; Park et al., 2002). Mann-Whitney U tests were performed to assure similarity between groups in terms of age and education. A chi-squared test was performed to check sex distributions across groups.

Dynamic VWM strategy. To analyse differences in VWM strategy across conditions, and to assess whether individuals with memory impairments indeed adhere to the sampling strategy more than those without, we included all trials in a linear mixed-effect model (LMM; Singmann & Kellen, 2019). This approach takes into account missing data and individual differences within groups. LMMs are robust against deviations from normality of the outcome variables (Schielzeth et al., 2020). Several models were generated to analyse the best fit for the data using the lmer function (lme4 package; Bates et al., 2014) in R (R Core Team, 2021). Factors included were Group, Condition, Group*Condition and random slope and intercept for individuals. A likelihood ratio test (ANOVA function of the ltm package; Rizopoulos, 2006) was used for model comparison to investigate which model outperformed the others in explaining the data (a lower AIC/BIC indicating a better fit). χ^2 with $\alpha < .05$ was leading in deciding on the most informative model. After fitting the model, the significance of factors was judged by a value of $p < .05$. The normality of the residuals was visually examined and confirmed for every linear mixed-effects model. Effect sizes were reported as standardized beta-coefficients (β) with a 95% confidence interval. The dependent variables were: *success rate, speed, number of crossings, dwell time per crossing, number of crossings per correct placement*, and dwell time per correct placement. Given that we perform 18 tests (6 models * 3 factors), and using an alpha of 0.05, fewer than *one* of our findings is likely to be a false positive. This should be considered when interpreting the results.

Datasets of all participants were analysed, initially without removal of outliers. In addition, to make sure that findings were not driven by outliers, we removed those participants whose aggregated scores were ≥1.5 times the interquartile range apart from their group median for that specific outcome measure in that specific condition (baseline or high cost). If participants were identified as an outlier in one condition, their data were removed from *both* conditions. After outlier removal, we ran the analyses again. Information on outliers is mentioned in the section of the respective analyses.

2.4.3. Memory functioning and VWM strategy. We expected that the *degree* and *type* of memory deficits *within our patient sample* influences the trade-off between sampling and storing (e.g., lower capacity is expected to relate to more sampling). Therefore, we generated (non-parametric) regression models to predict the *number of crossings per correct placement* and *dwell time per correct placement* in both conditions as a function of memory capacity scores – given age and level of education. These outcome measures were chosen as they reflect both eye-movement sampling behaviour and successful memory employment ('per correct placement'). Each of the capacity scores was included in a separate regression model to predict behaviour on the Copy Task. We ran the models for both conditions separately, as we hypothesized that memory capacity would influence behaviour mostly in a situation where it is beneficial to tax working memory (high cost condition) and not necessarily when information is freely available.

We hypothesized that forward and backward span on the Corsi Block Tapping Task, forward and backward span on the Digit Span Task, and sensitivity (*d'*) on the Change Detection Task (see preregistration: https://osf.io/dbv3g) would be related to sampling measures; for each outcome measure higher scores were expected to result in fewer samples. Other memory task scores (LLT Learning Index and Placement Errors, and RAVLT Total Score) were included in the preregistration for exploratory purposes. Eventually, other than preregistered, we did not take all capacity measures into account. We decided to reduce the number of capacity measures, and with that the number of statistical tests, in order to prevent power issues. We decided to only look at the backward span, and not the forward span, of the Corsi Block Tapping and Digit Span Task; forward scores are clinically mostly interpreted as an (required) attentional span, whereas backward scores are taken as a working memory span. Furthermore, we decided to only analyse one outcome (instead of three) of the LLT (Placement errors).

In clinical neuropsychological practice, raw memory scores are corrected for age and level of education. As memory functioning is to some extend related to these variables (Brockmole & Logie, 2013), and these could confound the influence we attribute to (working) memory capacity scores, level of education and age were included as separate factors in each model. To correct for multiple testing, a Holm-Bonferroni-correction was applied per condition (i.e., low cost, high cost) and dependent variable of the Copy Task (i.e., crossings per correct placement, dwell time per correct placement).

3. Results

3.1. Group comparisons

3.1.1. Demographics. Thirty-two patients with Korsakoff's syndrome (KS; 24 male, *M* = 63.5 years, *SD* = 7.56 years, range 47 – 76) were recruited via Slingedael Korsakoff inpatient Centre of Expertise. One patient dropped out after introduction of the test session. One patient dropped out of the Copy Task after the practice session. Two patients were not able to complete the Copy Task (using a computer mouse) due to motoric impairment. We were unable to track the eyes of another three patients. After checking the medical file, one patient appeared to have suffered a partial stroke. Eventually, twenty-four patients were included (see Table 2.1 for demographic characteristics and see Supplementary Figure S2.1 for a patient flow chart). Patients were without known visual field deficits and had normal or corrected-to-normal vision, except for one patient who had retinal detachment of the left eye. Two patients could not perform the second test session; one deceased and one was bedridden. Due to lowered workload capacity, not all patients were able to complete all the neuropsychological tasks in the available time.

27 controls (10 male, *M* = 58.48 years, *SD* = 8.86 years, range 40-81) were recruited to perform the same test protocol as the patients with KS.

Table 2.1 shows group demographics, obtained scores on neuropsychological assessment, and statistical comparisons between groups. No significant differences between groups were found regarding age ($U = 251$, $p = .171$, $r = -.23$). The level of education differed between groups, where healthy participants had a higher educational level ($M = 5.9$, $SD = 0.92$) than patients with KS ($M = 4.46$, $SD = 1.14$; $U = 536.5$, $p < .001$, $r = .656$). In both groups, however, the level of education was not significantly related to any of our Copy Task outcome measures in both conditions (all $p > .064$). For age, significant correlations were found in both conditions in both groups, but these effects were accounted for as the groups were age-matched. See Supplementary Table S2.2 for statistics on these correlations.

Table 2.1. *Demographic characteristics and scores on the neuropsychological memory tasks per group.*

	Patients with KS			**Healthy controls**			**Test statistic[a]**		
Demographics	***n***	**Mdn (IQR)**	**Range**	***n***	**Mdn (IQR)**	**Range**	**χ^2**	***p***	***d***
Sex	24	16 male		27	10 male		3.357	.067	0.53
							U	***p***	**r**
Age, years	24	64 (8.5)	47 – 74	27	59 (8.5)	40 – 81	251	.170	-.23
Level of education	24	4.5 (1.25)	3 – 7	27	6 (2)	4 - 7	536	<.001**	.66
Time since admission, years	24	3.1 (7.4)	0.1 – 16.9						
Neuropsychological task scores									
Location Learning Task	23			27					
Total displacement score		85.0 (50.5)	45 – 129		31 (25.5)	3 – 75	28	<.0001***	-.91
Learning index (0-1)		0.11 (0.08)	0.03 – 0.3		1.53 (0.4)	0.1 – 1	582	<.0001***	.87
Rey Auditory-Verbal Learning Task	22			27					
Immediate recall: Total correct (0-75)		25 (7.5)	14 – 36		47 (17)	33 – 67	584	<.0001***	.97
Delayed recall: Total correct (0-15)		1 (2)	0 – 4		9 (6)	3 - 14	590	<.0001***	.99
Digit Span Test (WAIS-IV)	24			27					
Forward span (2-9)		5 (1)	4 – 8		6 (1.5)	4 – 9	458	<.01*	.42
Backward span (2-8)		4 (2)	2 – 6		5 (1.5)	2 – 8	483	<.005**	.49
Corsi Block Tapping Task	23			27					
Forward span (2-9)		5 (0)	1 – 8		5 (1)	3 – 8	394	.076	.27
Backward span (2-8)		5 (1)	2 – 7		6 (1)	2 – 7	449	.005**	.45
Change Detection Task	19			27					
Average Kmax score		1.21 (0.67)	0.59 – 1.93		2.17 (0.79)	0.43 – 3.45	450	<.0001***	.75
D-prime		1.29 (0.45)	0.82 – 1.99		2.27 (0.64)	0.63 – 3.8	456	<.0001***	.78

KS = Korsakoff syndrome, sample size n, median Mdn, interquartile range IQR, range (min. – max.). A non-parametric test statistics indicating group differences and effect sizes.: Chi-squared, p-value and d for binomial variable sex, or Mann-Whitney-Wilcoxon U, p-value, and rank-biserial correlation r. *$p \leq .05$, **$p \leq .005$, ***$p \leq .0001$

3.1.2. Dynamic VWM strategy.

Data loss. 1440 trials were planned to be collected over 24 patients (2 sessions x 2 conditions x 15 trials). We removed every first trial in a block from the analysis: this trial served to check whether instructions were retained (additional encouragement was given when needed) and to habituate to the new situation (e.g., transition to gaze-contingent block; -96 trials). Two patients did not complete the second session; one deceased and one was bedridden (-56 trials). Additionally, one patient was not able to finish the gaze-contingent condition in the first session due to a bug in the code (-14 trials). Any reason that could possibly interfere with performance (excessive movement of the participant, forgetfulness of task instructions, apathy, or problems controlling the mouse) was logged, and corresponding trials (71 trials) were removed from further analysis. 8 trials were removed because the eye-tracker lost signal. To summarize, 245 trials were excluded leaving 1195 trials for analysis.

1620 trials were planned to be collected over 27 healthy controls (2 sessions x 2 conditions x 15 trials). Again, we removed every first trial in a block from analysis (-108). Due to e.g., coaching or movement, 4 additional trials needed to be excluded from analysis. Although a drift check was implemented, some trials had started with a drift check above the 2 dva threshold. If the error exceeded a 5 dva threshold, the trials were excluded to make sure that this would not confound our definition of a crossing (see Supplementary Results Chapter 2 for drift check descriptives). For one participant, this meant that almost none of the trials in the second session were valid. We therefore excluded the whole second session of this participant. In sum, we excluded 45 trials because of exceeding the drift check threshold. Finally, 1463 trials were left for analysis.

Table 2.2 displays outcomes of interest on the Copy Task for both groups, split per condition. Per participant, outcome measures were aggregated by mean over trials per condition except for time-based outcome measures, which were aggregated by median. Group scores (i.e. medians) were then calculated.

Table 2.2. *Outcomes on the Copy Task for patients with Korsakoff syndrome (KS) and healthy controls split on conditions (baseline, high cost).*

	Patients with KS			Healthy controls		
Copy task scores	**n**	**Mdn (IQR)**	**Range**	**n**	**Mdn (IQR)**	**Range**
Completion time, s	24			27		
Baseline		31.85 (13.97)	17.46 – 42		17.75 (4.2)	12.1 – 24.95
High cost		42 (0.001)	33.68 – 42		32.33 (7.75)	24.05 – 42
Net copying time, s	24			27		
Baseline		31.85 (13.97)	17.46 – 42		17.75 (4.2)	12.1 – 24.95
High cost		32.34 (4.96)	21.65 – 37.77		24 (4.66)	18.54 – 36
Correct placements	24			27		
Baseline		5.84 (0.45)	3.22 – 6		6 (0)	5.89 – 6
High cost		3.54 (1.33)	1.78 – 6		5.85 (0.47)	4.07 – 6
Success rate	20			25		
Baseline		0.97 (0.08)	0.86 – 1		0.97 (0.03)	0.88 – 1
High cost		0.82 (0.22)	0.37 – 0.99		0.91 (0.11)	0.64 – 0.97
Speed score, s	24			27		
Baseline		5.62 (2.4)	3 – 14.17		3.04 (0.77)	2.1 – 4.41
High cost		11.95 (6.02)	3.68 – 20.77		4.15 (0.99)	3.28 – 10.84
Number of crossings	24			27		
Baseline		11.7 (4.68)	8.07 – 18.78		9.54 (2.57)	5.29 – 13.07
High cost		3.82 (1.9)	2.04 – 6.3		3.79 (2.12)	1.82 – 6.96
Dwell time per crossing, s	24			27		
Baseline		0.49 (0.12)	3.56 – 1.06		0.38 (0.08)	0.28 – 0.51
High cost		1.22 (1.03)	0.54 – 4.22		1.21 (0.94)	0.56 – 5.62
Number of crossings per correct placement	24			27		
Baseline		2.14 (1.1)	1.35 – 4.65		1.59 (0.43)	0.88 – 2.24
High cost		1.2 (0.31)	0.75 – 2.29		0.77 (0.4)	0.30 – 1.4
Dwell time per correct placement, s	24			27		
Baseline		1.01 (0.69)	0.56 – 3.23		0.60 (0.16)	0.41 – 0.98
High cost		1.49 (1.03)	0.54 – 3.05		0.83 (0.24)	0.55 – 3.03

Valid datasets n, median Mdn, interquartile range (IQR), and range (min. - max.)

Behavioural performance. Completion time is depicted in Figure 2.3A. In the baseline condition, all controls and almost all patients were able to complete trials within time. When introducing the gaze-contingent waiting time in the high cost condition, most controls were still able to complete the puzzle, but patients struggled to place all six items on the correct location within time.
Figure 2.3B shows the number of correct placements within a trial, which shows lower values for patients, especially in the high cost condition.

As a measure of how effective people were in placing items correctly without making errors, success rate (i.e. the number of correct placements divided by the total number of attempts) was calculated (see Figure 2.3C). A linear mixed-effect model was fit to success rate to analyse the influence of group and condition, while controlling for individual differences. There was no main effect of group (t = -1.53, p = .133, β = -0.05 [-0.12, 0.01]), but a main effect of condition (t = -3.37, p = .002, β =

-0.24 [-0.38, -0.1]) was found, where success rate was lower in the high cost than baseline condition. Additionally, there was an interaction effect between group and condition (t = -2.76, p = .009, β = -0.23 [-0.4,-0.07]), with patients performing disproportionately worse than controls in the high cost condition. Two outliers (2 healthy controls) were detected for success rate. After outlier exclusion, the linear mixed-effects model was run again. A main effect of group on success rate appeared (t = -2.12, p=.037, β = -.07 [-0.13, -0.01]), where controls outperformed patients. The effects of condition and the interaction remained the same (see Supplementary Table S2.3 for results before and after outlier removal).

Speed score is depicted in Figure 2.3D. A linear mixed-effect model was fit to speed score to analyse the influence of group and condition. The model showed a main effect of group (t = 5.57, p < .001, β = 0.03 [0.19, 0.4]), with patients being slower than controls. A main effect of condition was also present (t = 2.97, p = .005 β = 0.15 [0.05, 0.5]); participants took longer to place one item correctly in the high cost condition as compared to the baseline. In the high cost condition, patients became disproportionately slower than controls as indicated by an interaction effect (t = 4.65, p < .001, β = 0.28 [0.16, 0.41]).

For speed score five outliers were detected (3 healthy controls, 2 patients). After outlier removal, no differences were found in the results. See Supplementary Table S2.3 for statistical results before and after outlier removal.

Sampling behaviour. The previous analyses showed that patients had more difficulty completing the task compared to controls: more mistakes were made and they were slower. But how did participants arrive at their performance? The next question was whether or not patients show the same eye-movement behaviour as controls across conditions, and whether patients with memory impairment indeed adhered to a sampling strategy more than controls.

Both the number of crossings (Figure 2.4A) and the dwell time per crossing (Figure 2.4B) were significantly predicted by group (t = 3.47, p = .001, β = 0.25 [0.11, 0.39]; t = 2.92, p = .004, β = 0.05 [0.02, 0.08], respectively) and condition (t=-12.1, p < .001, β = -0.65 [-0.75, -0.54]; t = 4.97, p < .001, β = 0.45 [0.27, 0.63], respectively). In general, patients sampled more and dwelled longer than controls. Both groups reduced sampling and dwelled longer when sampling cost was high compared to when sampling cost was low. An interaction effect was only found for the number of crossings: patients made fewer crossings in the high cost condition (t = -3.44, p = .001, β = -0.22 [-0.34, -0.09]). This could (at least partly) be explained by the fact that they had more difficulty performing the task (being slower and less accurate, and therefore having less time within the trial to make a crossing). No outliers were detected for number of crossings. Seven outliers were detected for dwell time per crossing (3 healthy controls, 4 patients). Then, the new fit yielded a main effect of group (t = 2.25, p = .025, β = 0.04 [0.01, 0.08]) and condition (t = 6.96, p <.001, β = 0.44 [0.32, 0.57]), which aligns with the findings before outlier removal. See Supplementary Table S2.3 for results before and after outlier removal.

When looking at sampling behaviour with respect to placing one item correctly (Figure 2.4C, 2.4D), the same pattern was observed: patients made significantly more encoding crossings than controls (t = 4.08, p < .001, β =0.38 [0.2, 0.56]) and both groups made fewer crossings in the high cost condition compared to the no cost condition (t = -8.65, p < .001, β = -0.47 [-0.57, -0.36]). Looking at the absolute values, the results show that controls were able to retain multiple items per crossing (<1 crossing per correct placement) in the high cost condition, whereas patients still needed 1 crossing or more. Patients also dwelled longer to correctly place one item than controls (t = 4.61, p < .001, β = 0.28 [0.16, 0.4]). In the high cost condition, both groups dwelled longer to place one item correctly (t = 2.13, p = .039, β = 0.16 [0.01, 0.31]). No interaction effects were present for crossings per correct placement nor dwell time per correct placement (t = -1.59, p = .12, β = -0.1 [-0.23,0.02], and t = 0.97, p = .338, β = 0.09 [-0.09, 0.26] respectively). For number of crossings per correct two outliers were detected (2 patients). After removal of the outliers, the same effects were found as before removal. Finally, for dwell time per correct six outliers were detected (3 healthy controls, 3 patients). The results of the new model fit show a main effect of group, but the effect of condition vanished: here, the effect of condition was driven by the outliers. Participants did not sample longer per correctly placed item in the high cost condition. Nonetheless, the interaction effect held. Again, see Supplementary Table S2.3 for results before and after outlier removal.

Note on multiple testing. We correct for multiple testing by taking the least significant finding with a grain of salt, as this might reflect a false positive. This concerns the effect of condition found for dwell time per correct placement (t = 2.13, p = .039, β = 0.16 [0.01, 0.31]).

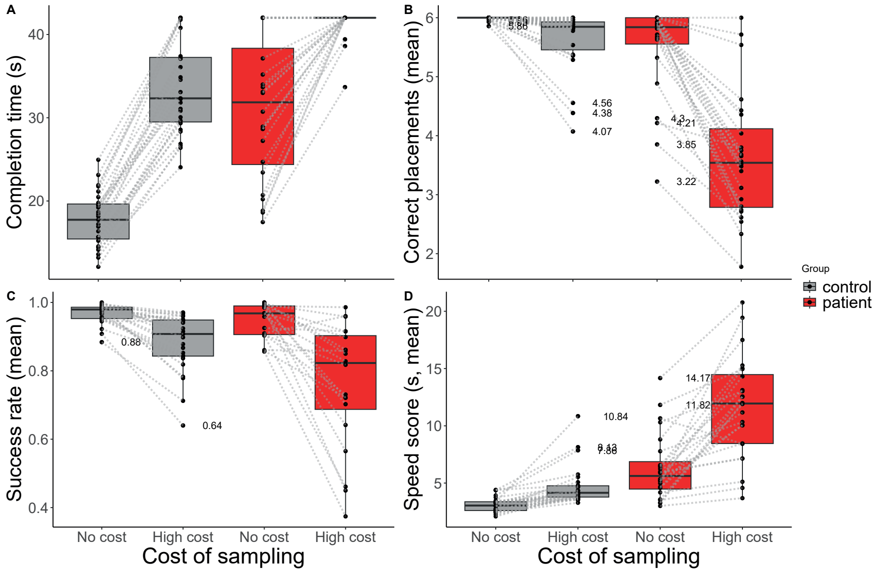


Figure 2.3. Performance scores. **A)** Completion time (mdn, 42s time limit). When introducing a high cost gaze-contingent waiting time, patients failed to complete the puzzle within time. **B)** Mean correct placements (maximum 6) per trial, **C)** mean success rate, and **D)** mean speed score for controls (grey) and patients (red) across conditions (no cost, high cost). Black dots and grey lines represent scores for individual participants. Outlier values (median ± 1.5*interquartile range) are labelled.

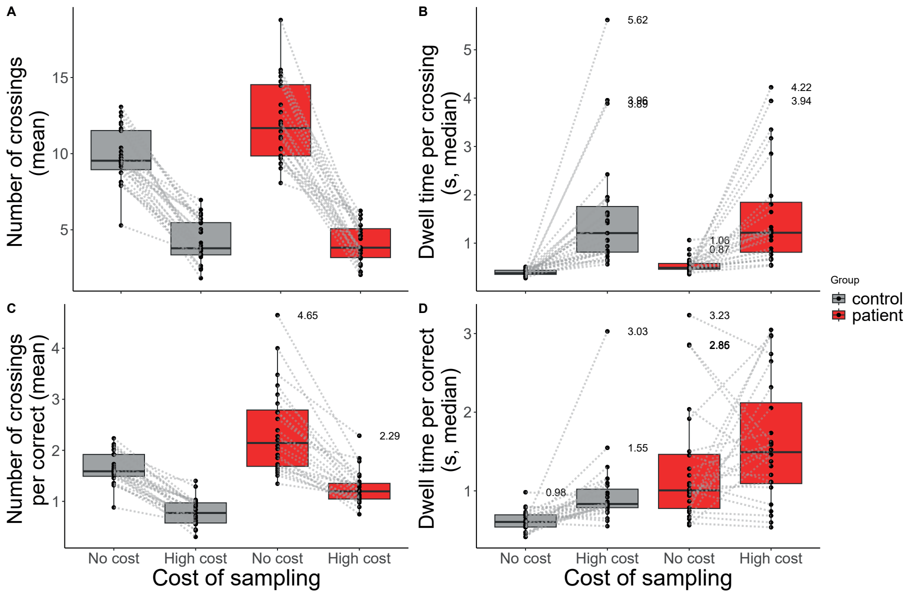


Figure 2.4. Eye-movement measures as indicator for sampling behaviour. **A)** Mean number of crossings within a trial, **B)** median dwell time per crossing, **C)** mean number of crossings needed to make one correct placement, and **D)** median dwell time needed to make one correct placement for controls (grey) and patients (red) across conditions (no cost, high cost). Black dots and grey lines represent outcomes of individual participants. Outlier values (median ±1.5 * interquartile range) are indicated.

3.2. Memory functioning and dynamic VWM use

To explore whether the degree and type of memory deficits have an influence on sampling behaviour within the patient sample, regression models were generated to predict the number of crossings per correct placement and dwell time per correct placement in both conditions as a function of memory capacity scores – given age and level of education. Table 2.3 shows regression estimates and uncorrected and Holm-Bonferroni corrected (for variables of interest) p-values for each model.

Opposed to what was expected in the baseline condition specifically, the uncorrected raw p-values show that some of the capacity scores on traditional neuropsychological working memory assessment (Digit Span, Corsi) related (with a medium effect size) to sampling behaviour. A higher capacity on the Digit Span yielded fewer crossings to place one item correctly (p = .034, β = -0.45 [-0.87, -0.04]). A higher capacity on the Corsi yielded fewer crossings (p = .019 β = -0.49 [-0.89, -0.09]) and shorter dwell times (p = .038, β = -0.41 [-0.8,-0.03]) to place one item correctly. So, the higher the memory capacity, the lower the number of crossings and the lower the dwell time that were needed to place one item correctly.

In the condition with the gaze-contingent waiting time (high cost condition), there was one predictor: *d'*. The higher the *d'* – indicating a better visual working memory performance – the lower the number of crossings per correctly placed item (p = .023, β = -0.49 [-0.9, -0.08]), but the higher the dwell time (p = .011, β = 0.62 [0.17, 1.08]). Contrary to our expectations, there were no significant relations between the other memory capacity measures and sampling behaviour in the high cost condition (all p > .351).

After correcting for multiple comparisons, none of the relations in either the baseline or high cost condition remained significant. Therefore, the general conclusion is that degree and type of memory deficits do not predict sampling behaviour (sampling nor dwelling) in neither of the conditions.

Table 2.3. *Unstandardized coefficient estimates, raw p-values and Holm-Bonferroni adjusted p-values for factors (fixed covariates level of education and age, and capacity score of interest) within the regression models to predict sampling behaviour for the patients with Korsakoff's Syndrome on the Copy Task (crossings per correct placement, dwell time per correct placement) split on condition (baseline, high cost).*

	No. of crossings per correct placement						**Dwell time per correct placement**					
	Baseline			**High cost**			**Baseline**			**High cost**		
Digit Span – BW span	**Est.**	**Raw *p***	**Holm**	**Est.**	**Raw *p***	**Holm**	**Est.**	**Raw *p***	**Holm**	**Est.**	**Raw *p***	**Holm**
N=24												
Education	-0.042	.782		0.059	.377		0.087	.506		0.120	.464	
Age	0.026	.298		0.014	.215		0.035	.106		-0.031	.251	
Digit Span BW	-0.349	**.034***	.306	-0.028	.681	1	-0.261	.061.	.427	-0.033	.845	1
Corsi – BW span												
N=23												
Education	-0.127	.413		0.053	.445		0.015	.911		0.093	.588	
Age	0.03	.234		0.013	.249		0.039	.077.		-0.029	.293	
Corsi Span BW	-0.402	**.019***	.19	-0.061	.393	1	-0.300	**.038***	.306	-0.043	.808	1
D'												
N=19												
Education	-0.165	.440		-0.065	.190		0.053	.681		0.088	.609	
Age	0.027	.394		0.017	.030 *		0.034	.085		-0.009	.724	
D'	-0.643	.338	1	-0.377	**.023 ***	.207	-0.399	.334	1	1.541	**.011 ***	.11
LLT - displacement errors												
N=23												
Education	-0.072	.664		0.044	.516		0.079	.564		0.11	.503	
Age	0.028	.341		0.018	.134		0.031	.203		-0.026	.372	
displacement errors	-0.00002	.998	1	0.001	.744	1	-0.003	.638	1	0.006	.351	1
RAVLT – total score												
N=22												
Education	-0.070	.714		0.024	.756		0.025	.875		0.046	.809	
Age	0.027	.424		0.022	.115		0.043	.119		-0.018	.585	
Total score	-0.007	.843	1	0.009	.500	1	0.02	.483	1	0.03	.386	1

Note. Digit Span – BW span = backward span on the WAIS IV Digit Span, Corsi – BW span = backward span on the Corsi Block Tapping Test, *D'* = dprime on the Change Detection Task, LLT – Placement Errors = total number of placement errors on the Location Learning Task, RAVLT – total score = number of correctly reproduced words over five trials from the Raven Auditory Verbal Learning Task. *p<.05.

4. Discussion

In neuropsychological assessment of visual working memory (VWM), estimating the maximum capacity is currently the gold standard. However, previous studies have shown that if possible, people rather fall back onto (i.e. sample from) information in the external world instead of memorizing it (Ballard et al., 1995; Draschkow et al., 2021; Droll & Hayhoe, 2008; Melnik et al., 2018; Somai et al., 2020). Only when sampling is impeded, people decrease the amount of inspecting behaviour and instead memorize more information at once (Ballard et al., 1995; Draschkow et al., 2021; Droll & Hayhoe, 2008; Melnik et al., 2018; Somai et al., 2020). We hypothesized that when memory is impaired, an even more pronounced reliance on external sampling would occur. We assessed whether eye-movements (used for external sampling) that are made during the execution of a memory task can serve as a proxy for VWM use in a group of healthy controls and patients with Korsakoff's amnesia.

Our dynamic working memory task yielded eye-movement behaviour in healthy controls in line with the expectations: controls sampled often when possible, and sampled less often when information was less readily available. In the latter situation, they increased dwell time on the model. This behaviour is in line with that observed in previous studies using a copy task that manipulated the availability of information (Ballard et al., 1995; Draschkow et al., 2021; Droll & Hayhoe, 2008; Melnik et al., 2018; Sahakian et al., 2023; Somai et al., 2020). It shows that whether or not information was available provoked different eye-movement patterns in our healthy population.

Further, our results indicate that patients with Korsakoff's syndrome (KS) relied more on sampling – and thus on the external world as a memory buffer – than controls. This difference between groups was already observed when information was freely available in the external world. While executing our Copy Task, patients inspected the example on average 2.14 times to place one item correctly, whereas controls only looked 1.59 times. The values in our study indicate that both patients and controls inspected the example more than once in order to place one item correctly and thus often reinspected the example before making a placement. This reinspection behaviour conceptually replicates earlier findings, where results showed that people, when given the opportunity, will not load up more than roughly one item in VWM per inspection (Sahakian et al., 2023; Somai et al., 2020).

When information was less readily available in the external world (i.e., which we manipulated by introducing waiting time whenever the participant viewed the example), patients and controls adapted their behaviour: both groups sampled less often as compared to when the information was freely available, but the encoding time per sample increased. We interpreted this as an attempt to memorize more information at once. Nonetheless, the waiting time, which induced the shift in strategy from sampling to memorization, came at a cost. Participants made more errors and were slower. In patients with KS, this cost was most profound: patients had difficulty completing the trial within time and obtained lower performance scores than controls. So, although patients dynamically adapt their strategy when confronted with less accessible information - as reflected in their eye-movement behaviour - they fail to do so as effectively as controls. Furthermore, in order to successfully place one item correctly, patients needed to sample more often (1.2 times) than controls (0.77 times). This aligns with the expectation that patients would adhere to a sampling strategy more than controls, even when sampling was costly.

The increased reliance on the external world could be explained by deficits in working memory. Indeed, patients with KS performed worse than controls on all (but one) classical tasks that assessed memory subdomains, which confirms their impaired memory ability relative to controls and aligns with earlier findings of compromised (working) memory in patients with KS (Kessels & Kopelman, 2012; Oudman et al., 2020; Van Asselen et al., 2005). This supports the idea that impaired memory ability causes increased sampling: patients who have difficulty encoding or retrieving information need to sample multiple times (and, importantly, more often than controls) to strengthen the memory trace before being able to make a correct placement. It is therefore tempting to attribute a heavier reliance on external sampling to memory problems solely. However, if these memory problems were to underly sampling behaviour exclusively, we would expect individuals performing at the low end of the capacity spectrum to rely most strongly on the external world. Interestingly, however, we found that (lower) capacity scores on memory subdomains were not predictive of (lesser) sampling – and thus externalization – behaviour in patients with KS. The absence of this correlation adds to the mixed findings regarding the relation between memory capacity and sampling behaviour, where some studies find correlations while others do not (Meyerhoff et al., 2021; Morrison & Richmond, 2020; Risko & Dunn, 2015). These inconsistencies might partly be explained by different approaches in the assessment of reliance on the external world, ranging from offering the possibility to directly sample from the external world, to demanding a more active and thought-through role of the participant (intended offloading, writing). Furthermore, previous studies (Meyerhoff et al., 2021; Morrison & Richmond, 2020; Risko & Dunn, 2015) used a different operationalization of working memory capacity. For example, Meyerhoff and colleagues (2021) used the Corsi Block Tapping Task forward span to estimate VWM capacity. To be able to compare our results with those found in the study of Meyerhoff and colleagues, we conducted additional analyses with inclusion of the forward span (both verbal and visual, see Supplementary Table S2.4), which showed that the forward span neither did predict sampling frequency or duration. Thus, in our population objective outcomes of memory capacity do not relate to the frequency of sampling. It is possible, however, that this relation is not observed because there is no linear relationship. Theoretically, it could be the case that there is some sort of threshold of memory functioning that is needed to not heavily rely on sampling, and people will continuously sample when this threshold is not reached. Furthermore, stimuli that were used to estimate capacity in traditional tasks have different visual features than the stimuli used in our Copy Task. Possibly, estimating capacity by means of memorizing a sequence of the currently used stimuli would yield different results. Still, we argue that patients with KS should be able to load up at least two items at once: none of the patients in our sample had a capacity score <2 on any of the classical neuropsychological tests. Yet, they sampled multiple times to correctly place one item, even when sampling costs were high. This argues against the idea that mere memory ability is at the core of sampling behaviour. If not ability, what then causes these heightened levels of sampling – both when information is freely available and when it is not – in patients as compared to controls?

The fact that we did not find a relation between the currently administered memory capacity tasks and sampling behaviour on our copy task could be because these tasks might measure different constructs of memory. Earlier studies that adopted copying tasks interpreted frequent external sampling as putting little reliance on internal VWM (Ballard et al., 1995; Draschkow et al., 2021; Somai et al., 2020). Revisits (sampling more than once per correctly placed item), subsequently, could then be interpreted as an expression of non-successful encoding at the first inspection. However, recently it was found that (re)visiting behaviour does not necessarily mean that VWM content is completely put to use before taking another look at the example (Sahakian et al., 2023). Rather, it could be argued that sampling behaviour serves some sort of soothing behaviour to increase one's confidence in their memory strength. This idea would fit with the study of Morrison and Richmond (2020) who suggested that the subjective estimation of one's memory capacity influences sampling behaviour to a larger extent than objective memory capacity. The findings of both Sahakian and colleagues (2023) and Morrison and Richmond (2020) point out that the frequency of sampling is not inherently a proxy for the amount of information that is stored in VWM, which would explain why pure capacity scores are not predictive of the amount of sampling.

Plausibly, sampling behaviour does not reflect the (in)ability to use memory, but reluctance to use memory as a consequence of higher costs to internally storing information. With impaired memory, internally storing information, even for only one or two items, is likely associated with high effort, and sampling would be regarded a more cost-efficient strategy even when sampling costs are large. With non-impaired memory, the effort associated with retaining multiple items per sample would be lower, and internally storing information would be regarded the more cost-efficient strategy when sampling costs became large. In a healthy population, choosing externalization over internal storage has been found to indeed depend on perceived reduction of effort (Risko & Dunn, 2015). Offloading (in this case writing down a sequence of letters) was perceived a higher effort than internal storage for small set sizes. This pattern flipped with increasing set sizes (Risko & Dunn, 2015). Observations in our healthy population can therefore be aligned with the idea of reducing perceived effort: when we introduced the waiting time, sampling might have been perceived more effortful by controls than memorizing a small number of shapes. For patients with KS, the increased cost of sampling did potentially not outweigh the cost associated to memorization. Therefore, heightened sampling could be a reflection of increased reluctance to use internal memory storage in order to minimize perceived effort in patients with KS.

The decision to offload or memorize is not only dependent on effort, but also on the desire to be accurate: in a previous study where accuracy was at stake, participants were more inclined to fall back onto the external world to support memory, even when this would not necessarily lead to better performance as compared to using only memory (Risko & Dunn, 2015). Sampling (here, reinspecting the example) could in this case be seen as an expression of checking behaviour. Our participants were instructed to perform as accurately and quickly as possible, but they were not punished for errors nor slowness other than receiving feedback. Errorless performance was therefore possibly not deemed to be as important. When checking was easy, people tended to revisit the example (>1 sample per correctly placed item), but when sampling was impeded, checking – and thereby assuring accuracy – might not have been not worthwhile anymore. Actually, when sampling costs increased, it could be seen as a strategy shift to make more attempts (albeit faulty) to avoid sampling, and to 'squeeze' out more information from memory at the expense of accuracy (Sahakian et al., 2023). Thus, sampling behaviour can vary depending on whether effort-minimization or time-accuracy expenditure is prioritized.

So, sampling behaviour is likely to be the end-product of (perceived) working memory ability (Morrison & Richmond, 2020), effort minimization (Risko & Dunn, 2015), and task demands (speed and/or accuracy; Risko & Dunn, 2015; Sahakian et al., 2023). Additionally, our copy task did probably not only tax working memory in order to complete the puzzle as fast and accurately as possible; our task called upon a certain level of executive functioning to monitor what puzzle pieces had already been placed and to keep a structured workflow. Note that participants exerted control over the visibility of information: a gaze-contingent waiting time required them to wait for 2 seconds, after which they could decide how long they would inspect – and thus encode information from – the example puzzle. During the experiment, we observed that patients with KS needed more guidance in the task instructions. We suspect that some patients had difficulty understanding how to exert control over the gaze-contingent appearance of the example puzzle. This would fit with the frequent report of executive deficits in patients with KS (Brand, 2007; Maharasingam et al., 2013). Potentially, patients may have wanted to sample from the example more frequently, but lacked the full understanding on how to accomplish this. Indeed, our data (see Supplementary Figure S2.2) shows that patients actually moved their eyes towards the side of the screen with the example more often, but only a part of these crossings remained fixated long enough to reveal the example puzzle. Then, when patients finally waited long enough to make the example appear, they could have been inclined to directly place the stimulus they encoded, failing to oversee the consequence of having to wait again to make the example reappear. This somewhat impulsive eye-movement behaviour can be supported by the fact that disinhibitory control is often observed in patients with KS (Gerridzen et al., 2018; Moya et al., 2021). In the acute phase of Wernicke's encephalopathy, which precedes the development of Korsakoff's syndrome, oculomotor symptoms such as nystagmus are often observed (Wernicke, 1981, in (Kopelman et al., 2009)) and some of these may remain present in the chronic phase. Yet, as the outcome measures we used are rather crude, we do not believe that these would be influenced by nystagmus. Patients with KS display only subtle impairments in recognizing and naming real world objects (letters) with degraded perceptual clarity or common objects (e.g., animals) from atypical perspectives (Kasse et al., 2019). Also, spatial perception is not hampered (Kasse et al., 2019). Furthermore, if patients would have had difficulty with perceiving the stimuli, they would have performed already worse in the baseline condition and differences between conditions could not been explained by it.

Although a reduced understanding of task instructions could partly explain our results, we are confident that patients clearly understood the task manipulation. We base this upon the observation that they did perform a strategy shift: patients either decreased the amount of sampling and memorized more, or made more placement attempts (albeit faulty) in order to avoid sampling. Still, executive deficits could have contributed to their impaired performance on the task. For example, cognitive flexibility is associated with better performance on jigsaw puzzles (Fissler et al., 2018), which are to some extent similar to our Copy Task.

Patients with KS have shown deficits in this cognitive domain, where they are slower when rule switching is required, are worse at inhibition of previously learned rules, and show more perseveration errors (Oscar-Berman et al., 2004). Likewise, it is possible that patients had more difficulty switching between blocks, that they stuck to their previous sampling strategy, and/or that they made perseveration errors in our task (e.g., placing the same stimulus at the same wrong location multiple times) leading to worse performance. As we do not have quantitative neuropsychological data on these functions, we cannot rule out the possibility that they have influenced our results. Additionally, differences in psychomotor and information processing speed (Fissler et al., 2018; Welch et al., 1997), and apathy (Arts et al., 2017) or other clinical manifestations of the syndrome such as depression (Gerridzen et al., 2018; Takahashi et al., 2021) can partly explain the finding that patients took longer than controls. We acknowledge that there are multiple facets that may influence behaviour on our task. Despite the difficulty disentangling the factors that contribute to visual working memory usage, we argue that exactly because of this, our task approaches more naturalistic VWM usage than mere memory tasks. After all, in daily life, the way that we deal with information is also the result of the complex interplay between several cognitive factors and the task at hand.

Although it is too early to directly translate our findings to a clinical implementation, we can speculate about the potential clinical value of a dynamic task such as ours. Diagnostically, a task such as ours offers a possibility to detect differences in working memory usage in a more dynamic environment than the classical working memory paradigms. It might allow to reveal different strategies that are put to use, and facilitate detection of switching abilities of the patient. Future research should elucidate how eye-movement markers on these dynamic tasks predict functioning in (instrumental) activities of daily living. Once established, this could give insights in the extent to which patients are able to function independently, which might help assigning patients to the care facility that is most adapted to their level of functioning.

Patients with KS often reside in clinical institutions that are tailored to the needs of this population (Kopelman et al., 2009). One aspect that puts a burden on caretakers is the need to constantly remind patients with KS of important appointments or agreements, such as taking one's medicines. To enlighten this burden, rehabilitation implementations evolve around finding solutions that fit patients' memory functioning. Using 'external memory' in the form of notebooks or calendars has been described as among the most common in supporting people with memory deficits (Sohlberg et al., 2007). More specific to patients with KS, errorless and/or procedural learning in (instrumental) activities of daily living were investigated as novel approaches, as implicit memory is relatively spared in KS (Oudman et al., 2013, 2015). Other developments are aimed at using technologies to support memory in patients with KS (de Joode et al., 2013; Smits et al., 2022). With regards to these memory aids, a future direction might be to assess whether patients' inclination to rely on the outside world is linked to the ability to effectively use these memory aids, e.g., whether patients benefit from sampling from a smartwatch (which is constantly available around the wrist) versus sampling from a notebook (which is not always in the same room as the patient).

To conclude, our results offer a framework to think more thoroughly about how dynamic tasks such as ours could be used to combine diagnostics and rehabilitation.

Limitations. Several limitations need to be considered when interpreting our results and making future recommendations. The dynamic nature of the task resulted in a higher complexity as compared to neuropsychological tasks targeting cognitive functions in isolation, thereby involving other cognitive functions apart from working memory (see also Fissler et al., 2018). We specifically designed the test battery to get an as broad as possible memory profile, but this came at the (foreseen) cost of excluding measures for other cognitive domains. Although we addressed how other cognitive factors might have potentially influenced our results, we cannot substantiate these by objective measures.

Second, we have performed multiple analyses on the relation between capacity scores and sampling behaviour. No effects were present after correcting for multiple statistical tests. This might have been due to limited power, which could be resolved in the future by including larger sample sizes, or by reducing the amount of statistical tests by using, for example, compound scores for memory functioning or sampling behaviour.

Furthermore, our experimental paradigm comes with several practical limitations. The requirement to use a computer mouse excluded severely motorically impaired patients from participating, and might have led to slower performance in participants who had little experience in using a computer mouse. Furthermore, using an eye-tracker in patient populations comes with general limitations relating to the inability to hold position for an extended period, oculomotor deficits and/or droopy eyelids, the tendency to move the head, reinstating calibration and validation cycles, and so on. This could have led to an inclusion bias (e.g., non-compliant patients could not be calibrated or produced datasets with signal loss, and could therefore not be included).

Although we have successfully gathered eye-tracking data for our study, the technical challenges that comes with eye-tracking should definitely be taken into account when using such tasks in clinical settings. It requires profound knowledge of the apparatus and familiarity with the type of data to be able to use it for diagnostical and rehabilitation purposes. While eye-tracking paradigms can yield rich datasets and valuable knowledge, they should be finetuned to the patient population, and task administration time, exclusion criteria for participation and prospected outcomes should be weighed against the investment to implement such paradigms. Paradigms that do not involve eye-tracking but measure sampling differently (e.g., by mouse movements such as in Meyerhoff et al., 2021; Sahakian et al., 2023) might offer solace, although future research should elucidate whether and how these different outcome measures (hand vs. eye movements) are directly interchangeable.

5. Conclusion

Differences in performance and sampling behaviour between patients with KS and healthy controls could be driven by several factors. Although we cannot (yet) pinpoint the (most pronounced) underlying factor causing sampling behaviour, assessing sampling behaviour clearly yields additional value on a clinical level as to how patients dynamically use information in situations that demand memory usage. We conclude that Korsakoff's amnesia evokes a relatively heavy reliance on external sampling, even when sampling is costly. Naturalistic eye-movement markers can serve as a proxy for these subtle changes in memory usage that are not captured by assessing one's maximum storage capacity, but that rather occur in dynamic interaction with the environment.

Supplementary Materials: Supporting information can be downloaded from https:// www. mdpi.com/article/10.3390/jcm12113630/s1, and can be found in the Supplementary Materials of this dissertation.

Author Contributions: Sanne Böing: conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, visualization, writing – original draft preparation, writing – review and editing. **Antonia F. ten Brink:** conceptualization, methodology, resources, software, supervision, writing – original draft preparation, writing – review and editing. **Alex J. Hoogerbrugge**: conceptualization, methodology, resources, software, writing – review and editing. **Erik Oudman:** conceptualization, resources, writing – review and editing. **Albert Postma:** conceptualization, resources, writing – review and editing. **Tanja C.W. Nijboer**: conceptualization, methodology, supervision, writing – review and editing. **Stefan Van der Stigchel:** conceptualization, funding acquisition,

methodology, supervision, writing – review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 863732).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the Faculty Ethics Review Board of the Faculty of Social and Behavioural Sciences at Utrecht University (protocol numbers 21-0076 and 21-0270) and the local science committee of Slingedael Korsakoff Centre of Expertise.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent was obtained from the patient(s) to publish this paper.

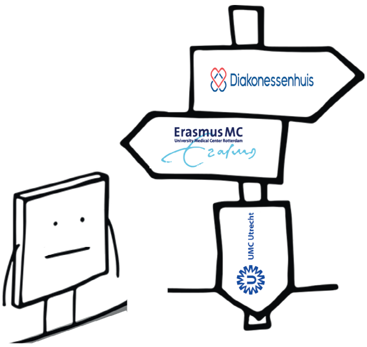
Data Availability Statement: The data presented in this study are openly available in Open Science Framework at https://osf.io/83nsw.

Declaration of Interest: The authors declare no conflict of interest.

CHAPTER 3

Inspecting the external world: Memory capacity, but not memory self-efficacy, predicts offloading in working memory

Sanne Böing, Antonia F. Ten Brink, Carla Ruis, Zoë A. Schielen, Esther van den Berg, J. Matthijs Biesbroek, Tanja C.W. Nijboer, & Stefan Van der Stigchel



Published as
Böing, S., Brink, A. F. T., Ruis, C., Schielen, Z. A., van den Berg, E., Biesbroek, J. M., Nijboer, T. C. W., & Van der Stigchel, S. (2025). Inspecting the external world: Memory capacity, but not memory self-efficacy, predicts offloading in working memory. *Journal of Clinical and Experimental Neuropsychology.* https://doi.org/10.1080/13803395.2024.2447263

Abstract

Individuals with memory impairments may need to rely often on the external world (i.e., offloading). By memorizing only a fraction of the items at hand, and repeatedly looking back to the remainder of items (i.e., inspecting), they can avoid frailty or effortful memory use. However, individuals with subjective concerns may also prefer to rely on the external world even though their capacity is intact. Crucially, capacity assessment fails to recognize offloading strategies, while inspection assessment may reveal how people choose to deploy memory in everyday life. To disentangle the relative contributions of memory capacity and memory self-efficacy to offloading behaviour, we recruited 29 individuals who were referred to a memory clinic and 38 age-matched individuals. We assessed memory capacity using neuropsychological measures, and memory self-efficacy using questionnaires. Inspection behaviour was assessed in a Copy Task that allowed participants to store information to their preferred load or to rely on the external world. Referred individuals had lower capacity scores and lower memory self-efficacy. They inspected as often as controls, but used longer inspections and performed worse. Across all subjects, memory capacity – but not memory self-efficacy – explained inspection frequency and duration, with higher capacity associated with fewer and shorter inspections. Capacity measures thus translate to how people choose to deploy their memory in tasks that do not force full capacity use. However, people generally avoided remembering more than two items per inspection, and thus avoided using their full capacity. Inspection behaviour was not further explained by memory self-efficacy, suggesting that inspections are not a sensitive measure of constraints experienced in everyday life. Although we provide support for the predictive value of capacity tasks in tasks with more degrees of freedom, capacity tasks overlook offloading behaviour that individuals may employ to avoid using their full memory capacity in everyday life.

Keywords: offloading; sampling; working memory; metamemory; neuropsychological assessment

1. Introduction

Memory complaints are common in the general ageing population (Ponds et al., 1997). Although some degree of memory loss is inherent to getting older, subjectively experienced memory problems are a precursor to cognitive impairment and may be indicative of underlying pathology (Drag & Bieliauskas, 2010; Jessen et al., 2010; Saykin et al., 2006; Steinberg et al., 2013). Concerns about memory functioning may therefore warrant a referral to a memory clinic. To discriminate between intact, below average, and impaired memory capacity, the referred individual is asked to encode, maintain and report as much information as possible. The resulting score is used to construct a cognitive profile and subsequently to guide diagnosis. However, memory capacity scores that are obtained in a clinical setting could fall within the normal range even though the individual may report subjective memory complaints in daily life (Beaudoin & Desrichard, 2011). This discrepancy may be due in part to the fact that traditional tasks force people to use a particular strategy (i.e., to memorize as much as possible). Yet, when people can choose whether or not to load memory to maximum capacity, they are likely to minimize the internal cognitive effort involved in performing a task and rely on information from the external world (Burnett & Richmond, 2023; Gilbert, 2015a; Meyerhoff et al., 2021; Risko & Dunn, 2015). In other words, they choose to use an offloading strategy (e.g., writing things down, creating cues as reminder; Ballard et al., 1995; Böing et al., 2023, in press; Draschkow et al., 2021; Gray et al., 2006; Meyerhoff et al., 2021; Morrison & Richmond, 2020; Risko & Dunn, 2015; Sahakian et al., 2023; Somai et al., 2020). Offloading may not only minimize effort, but also support accurate task completion in healthy individuals (Burnett & Richmond, 2023; Gilbert, 2015a; Gilbert et al., 2020, 2023). The use of external memory strategies is frequently reported among older adults (although there are mixed findings for clinical samples) and its usage is even found to increase with increasing age (Pizzonia & Suhr, 2022). Traditional capacity tasks disregard this element of choice in the employment of memory strategies, and, consequently, do not necessarily capture the actual use of memory in everyday life. This means that subtle deviations (e.g., increased reliance on external strategies) leading to the subjective experience of memory failure might go unnoticed in memory assessment. Therefore, rather than thinking of memory as a fixed capacity entity that is always fully utilised, we should consider how one uses their memory. In this study, we approach working memory as a system that people use differently depending on the accessibility of information, their maximum memory capacity, and their expectations of how their memory will function.

The act of memorizing information is highly dependent on the accessibility of information. This becomes apparent in (visual) working memory paradigms that allow individuals to choose how much information they internalize in working memory and how often they fall back onto external information. Sampling behaviour – the act of (re)orienting to and (re)inspecting information-to-be-used from the environment once it becomes relevant – is used as an indicator of such reliance on the external world; sampling is shown to occur often when information is relatively easily accessible, and to decrease when it is more effortful to access external information (Ballard et al., 1995; Böing et al., 2023; Draschkow et al., 2021; Droll & Hayhoe, 2008; Melnik et al., 2018; Sahakian et al., 2023; Somai et al., 2020). Given that the visual environment is generally stable, this implies a strong preference for external sampling in activities of daily living (e.g., looking back and forth at a grocery list rather than learning it by heart). Moreover, this reliance on sampling from the external world is even stronger when it is difficult to memorize information; people with impaired memory adhere to sampling rather than using working memory, even when sampling becomes costly (Böing et al., 2023). The extent to which people rely on external sampling versus internal working memory storage thus appears to depend on the interplay between information accessibility and working memory capacity.

Although lower levels of working memory functioning are to some extent associated with increased reliance on the external world (Meyerhoff et al., 2021; Morrison & Richmond, 2020; Risko & Dunn, 2015), there is no robust linear relationship between memory capacity and sampling behaviour (Böing et al., 2023). Even healthy individuals who are able to remember multiple items (i.e., capacity of two or more items) show frequent inspecting. They only memorize up to two items at a time when information remains accessible in the external world. In fact, reinspecting (more than 1 inspection per item) is often observed. This reinspecting behaviour has recently been interpreted as an expression of strengthening memory traces before acting on them (e.g., reaching an action threshold; Sahakian et al., 2023). In other words, people may have some residual information in working memory, but are not confident enough to use it, and therefore decide to inspect again. Along this line, we hypothesize that (re)inspecting is a proxy for an individual's belief about their own memory functioning. Speculating reinspecting to be an act of reassurance about the accuracy of the representation of the information to be used (i.e., checking oneself), we expect that individuals with negative beliefs about their memory functioning or self-reported memory failures may engage in reinspecting behaviour more often than individuals with more positive expectations about their memory functioning. Such beliefs can be captured by measures of memory self-efficacy, where low levels of memory self-efficacy indicate uncertainty or negative beliefs about memory functioning. Negative beliefs about memory functioning may be co-occurring with impaired memory capacity, but may also exist in the absence of impaired memory capacity (Ponds & Jolles, 1996a). Further, the construct of memory self-efficacy is related, but not synonymous, to subjective cognitive decline. Subjective cognitive decline regards the perceived decline in memory function within a person over time, whereas memory self-efficacy refers to the subjective judgement of one's memory functioning at a certain point in time. Even though a person may perceive cognitive decline in their memory function over time, they may still consider their memory functioning at the later timepoint to be adequate (thus, having a sufficient level of memory self-efficacy). On the other hand, perceived cognitive decline may be experienced by the individual to such an extent that it lowers the level of memory self-efficacy. Moreover, the level of memory self-efficacy within an individual may be low but stable over time, thus without subjective cognitive decline. In sum, experiencing perceived decline can, but does not necessarily, lower the level of memory self-efficacy, and memory self-efficacy is not necessarily congruent with objective functioning.

Low levels of memory self-efficacy may lead to greater reliance on the external world, even when this is not necessary given the objectively intact memory capacity. This over-reliance on the external world could, in turn, be experienced as a memory failure by the individual, strengthening the drive to obtain a referral to a memory clinic. Crucially, capacity assessment does not capture memory concerns, while subtle deviations in inspection behaviour may be a result of both capacity limitations and underlying memory uncertainty. Assessment of inspection behaviour may, therefore, bridge the gap between clinically objectifiable deficits and the subjective experience of memory decline or failure for which the individual is referred to the hospital, and can help to integrate the co-occurring effects of both objective and subjective aspects of memory functioning. Further, inspection behaviour may serve as an objective measure of external memory strategies, called for by Pizzonia and Suhr (2022).

In an attempt to disentangle the relative contributions of information accessibility, memory capacity, and memory self-efficacy to reliance on the external world, we assessed inspection behaviour of individuals with different levels of memory capacity (as determined by objective metrics in the verbal and visuospatial domain for short- and longer term maintenance) and different levels of self-reported memory complaints, on a Copy Task that either facilitated inspecting or encouraged memorizing by varying the availability of external information. To this aim, we recruited individuals who had been referred to a memory clinic, as well as age-matched individuals who had not been referred to a memory clinic. This allowed us to include individuals with a wide range of subjective memory complaints and objective memory capacity impairments, resulting in memory profiles with different combinations of subjective and objective performance. As age and the level of education are known to be associated with performance on memory tasks (Brockmole & Logie, 2013; Park et al., 2002), the non-referred group was matched to the referred group on these characteristics. We compared memory use in the two groups across two conditions that differed in the cost (low or high) of accessing information from the external world. Both referred and non-referred individuals were expected to reduce inspecting behaviour when information was less readily available (Böing et al., 2023). In addition, the referred individuals were expected to rely more on the external world (due to higher effort to store information or higher levels of memory complaints; Hurt et al., 2012) than non-referred individuals, even when information would not be readily available (Böing et al., 2023). For both referred and non-referred individuals, lower levels of memory capacity and higher levels of subjective memory complaints were expected to predict increased inspection frequency. As depression has been found to be associated with decreased memory performance and subjective memory complaints (Johansson et al., 1997; Schmand et al., 1996; Turvey et al., 2000), we also explored this attribute as a potential (confounding) factor influencing inspection behaviour. Assessing inspection behaviour and its underlying attributes may be an elegant and much needed way to approximate memory use in daily life, and may serve as a starting point to increase our understanding of patients' objective, subjective and interactive memory functioning.

2. Materials and Methods

2.1. Participants

Individuals referred for memory assessment were recruited via the outpatient memory clinics of the University Medical Centre Utrecht (UMCU), the Erasmus MC University Medical Centre Rotterdam, and Diakonessenhuis Hospital. These clinics have different specializations, and the types of referrals vary accordingly. The memory outpatient clinic of the neurology department of the UMCU sees a heterogeneous group of adults of all ages who experience memory problems due to, for example, neurodegenerative diseases, traumatic brain injury, an as yet unknown cause, or as a result of psychological factors. The memory clinic of the gerontology department of the UMCU specifically focuses on older adults (>65 years). MCI and dementia are regularly diagnosed. The route of referral is similar for both clinics of the UMCU: individuals may have initiated a referral themselves or are referred by their general practitioner or by other clinicians within the hospital (e.g., endocrinology, nephrology) who suspect cognitive decline. The outpatient memory clinic at the Erasmus MC specializes in Alzheimer's disease but also diagnoses other types of dementia. At the memory clinic of the neurology department of the Diakonessenhuis Hospital, a heterogeneous group of adults of all ages are seen; patients are mainly referred by general practitioners and either a brief cognitive screening tool or extensive neuropsychological testing is used, depending on the differential diagnosis and complexity of the case. Note that despite a referral to any of the clinics, a medical diagnosis may not be made after assessment.

The eligibility of referred individuals was based on the judgment of a neuropsychologist and/or a multidisciplinary team within the outpatient memory clinic. To be eligible for participation, referred individuals had to either self-report memory complaints, have objective memory impairment based on neuropsychological assessment, or have memory impairment observed by a clinician. Referred individuals had to be between 18 and 85 years old, speak Dutch fluently, and be able to give consent. They were excluded if there was evidence of visuospatial neglect, deficits in visual perception, aphasia, or if motor impairments prevented the use of a computer mouse.

Partners or family members accompanying the referred person were actively approached to act as matched controls. In addition, age- and education-matched controls were recruited via various public and university platforms (e.g., social media, family members, university intranet, community centres).

We recruited two groups (i.e., referred to the memory clinic and matched controls) with the aim of having at least 25 participants in each group. These numbers were determined by considering previous studies that have tested sampling behaviour, and a power analysis. The original trade-off effect on sampling versus storing has been observed in a group of only 7 participants (no mention of effect size; Ballard et al., 1995), which was replicated by Somai and colleagues (2020) in a group of 12 participants (only unstandardized β coefficients for linear mixed-effect models mentioned). As we expected greater variability in our target groups due to the heterogeneity of referral reasons and a wider age range, we aimed to recruit at least twice as many participants in each group. A previous study from our research group showed that this number was sufficient to detect differences in eye movement behaviour between patients with Korsakoff syndrome and controls (detected effect sizes β in the range of 0.05 – 0.38; Böing et al., 2023).

All participants gave written informed consent prior to the start of the experiment. Participants were compensated for their participation with 7EU per hour paid in increments of 30 minutes, and received compensation for travel costs.

We included 29 referred individuals (see Supplementary Figure S3.1 for a flowchart) and 38 non-referred controls. With the current sample size, for a one-tailed non-parametric Wilcoxon-Mann-Whitney t-test ($\alpha = .05$) with a power of .8, we should be able to reliably detect effects of Cohen's d=0.63 (Faul et al., 2009). Effects commonly reported in comparable paradigms are similarly large (Draschkow et al., 2021; Sahakian et al., 2023). Furthermore, the linear mixed-effects models we used have higher power than t-tests. Therefore, we were confident that our study would have a sufficient power.

The project was approved by the Faculty Ethics Review Board of the Faculty of Social and Behavioural Sciences at Utrecht University (protocol numbers 21-0076 and 21-0269). The protocol was conducted in accordance with the Declaration of Helsinki.

2.2. Procedure

Parts of the methods section are similar to those described in our previous study (Böing et al., 2023).

After participants agreed to participate, they received an online link to fill out questionnaires (see 'Questionnaires' for a description) at home in the period 14 to 1 day(s) before their test session. Individuals that already completed the Hospital Anxiety and Depression Scale in the outpatient memory clinic were exempt from filling in this questionnaire online. Questionnaires were administered to characterize the referred and non-referred group, and to rule out depression as a potential confound.

At the university testing facility, the rest of the test protocol (see 'Experimental computer tasks' and 'Neuropsychological tasks' for a description) was administered in a single visit. The first and second session of the experiment were separated by a break of 10 to 20 minutes, and the total test duration was a maximum of 3 hours. All tasks that were administered were memory tasks. These memory tasks were included to get an idea of the memory capacity across the groups, but also with the aim of integrating them into a memory compound score (see Analysis) that takes into account memory capacity in both the verbal and visual domains for both short-term and longer term delays. Task administration in session 1 comprised (in this order): Location Learning Task – direct recall, Copy Task – first session, Location Learning Task – delayed recall, Digit Span WAIS IV, and if time allowed: a Fixation and Free viewing task (not taken into account in the current analysis). Task administration in session 2 comprised (in this order): Rey Auditory-Verbal Learning Task – direct recall, Copy Task – second session, Rey Auditory-Verbal Learning Task – delayed recall, Corsi Block Tapping Task, and if time allowed: Change Detection Task.

At the end of the test protocol, the Metamemory In Adulthood questionnaire was administered. This was the case only for a subset of participants as the questionnaire was added later to the test protocol. This questionnaire was added to get an extra measure on beliefs about one's memory function, and was used in the calculation of the subjective memory compound score (see Analysis). See Supplementary Table S3.1 for a schematic overview of the test procedure.

Before their visit, we checked whether individuals that were referred to the hospital had already performed some of the neuropsychological tasks as part of standard care. If this were the case, they were exempt from that task; previously reported scores on those tasks were used in order to prevent unnecessary workload and avoid potential practice effects. Practice effects can occur after short time intervals between testing sessions, and can last up to 7 years (Calamia et al., 2012). Therefore, we should be wary with 'overtesting' people. As a rule of thumb, task administration in the hospital had to be within a period of six months before their visit to the study site to remain valid. The six month rule was based upon clinical practice where six months is believed to be long enough to have general task effects to wear off. The majority of referred individuals was scheduled within three months after their visit to the clinician. The period between assessment in the clinic and assessment in the research facility was sometimes less than three months, but never more than six months. It is improbable that substantial cognitive changes have occurred within this time frame. Almost all referred individuals were exempt from the Digit Span Test and Rey Auditory-Verbal Learning Test (see below) as these tasks are commonly used in both screening or extensive neuropsychological testing.

2.3. Measurements

2.3.1. Experimental computer tasks.

Apparatus. Experimental tasks were run on a Windows 10 Enterprise computer with an Intel Core i7-4790 CPU and 16GB RAM, and displayed on a 27 inch LCD monitor at a resolution of 2560 x 1440 pixels at 100 Hz. An EyeLink 1000 eye tracker (SR Research Ltd., Canada) was placed at the desktop to track the eyes at a sample rate of 1 kHz. Participants were seated with their heads in a chin-rest at ~67.5 cm from the monitor, and the lights were dimmed during administration of the experimental tasks. Eye-tracker calibration and validation were performed manually with a 9-point grid attempting to achieve a calibration error of less than 2 degrees of visual angle (dva).

Copy Task. Identical to our previous study (Böing et al., 2023), we adapted a Copy Task that was originally used in our research group (Somai et al., 2020) to better fit our participant population. The task aimed to provoke a strategy switch in relying on internal visual working memory versus inspecting information from the outside world. The experiment was programmed in Python 3.7 using the PyQt5 library (Riverbank Computing Limited, 2019) for visual presentation and mouse and keyboard interaction. PyGaze (Dalmaijer et al., 2014) was used to interact with the eye tracker.

A model puzzle consisting of 6 items in a 3 x 3 example grid was shown at the left-hand side of the screen (see Figure 3.1). At the right-hand side of the screen, a 3 x 3 empty grid was presented, with a 2 x 3 resource grid presented below. The resource grid only contained items that were needed to copy the model; no distractors were present. Items were adopted from Arnoult (Arnoult, 1956; Figure 3.1A) and consisted of black geometrical shapes that could not easily be named to measure reliance on VWM instead of verbalisation strategies (Somai et al., 2020).

The task consisted of two experimental conditions. In the baseline or 'low-cost' condition, the example grid was visible throughout the trial (Figure 3.1B). In this way, the 'cost' to gather information from the outside world was low. In the experimental 'high-cost' condition, we raised the cost to inspect information from the external world by introducing a gaze-contingent waiting time: the example appeared after fixating the left side of the screen for a total of 2000 ms. During the waiting time an hourglass was presented (Figure 3.1C). If participants looked back to the right during the waiting interval, the delay-clock would pause, and would restart as soon as the eyes were redirected to the hourglass again, so that gaze-contingent waiting always was 2000 ms, and never more. Once the example became visible, it remained on screen until the participant would move their eyes towards the right side of the screen after which it would disappear.

Participants were instructed to rebuild the model puzzle as quickly and accurately as possible by dragging items from the resource grid to the empty grid using a computer mouse. Participants received direct feedback: if an item was placed incorrectly, the item disappeared and the background of the cell turned red for 700ms, after which subjects could make another attempt. If the item was placed correctly, the background of the cell turned green for 700ms and the item remained fixed. A trial ended after correct placement of six items, or when the time-limit of 42 seconds had passed. The time-limit of 42 seconds was based on the study of Somai and colleagues (2020) in which high-cost conditions with 200, 1500 and 3000 ms delays were used. The authors observed maximum completion times of 30 seconds for placing six items in either of the three variations. As we tested older adults and patients with potential cognitive decline, we anticipated our subjects to need more time. We therefore complemented the maximum observed completion time of Somai and colleagues by adding the gaze-contingent delay of 2000 ms for each item that had to be placed in the high-cost condition.

In case someone would inspect once per item (which seems plausible from Somai et al., 2020), this would result in an additional 12 seconds. The choice to impose a time-limit at all was made because we wanted to have some control over the maximum task administration time, as we were bound to a larger protocol with limited testing time. After successful completion of a trial, positive feedback was shown (a thumbs up symbol). If subjects failed to correctly place all items within the time-limit, they were shown feedback that they ran out of time. By introducing the time-limit, we encouraged subjects to adopt a time-efficient strategy (Melnik et al., 2018). There was no specific incentive to increase the importance of accuracy as compared to speed or vice versa. Faster trial completion would yield faster task completion, serving as an incentive to increase working pace, but trying to be accurate also serves faster task completion, as making mistakes may also lead to increased completion times. This speed-accuracy trade-off is taken into account by the analysis of task performance (see 'Performance measures' below).

We administered two sessions of the Copy Task, each session consisting of two blocks. First, three practice trials were performed in the low-cost condition to get familiar with the task. Calibration and validation of the eye-tracker were performed after the practice trials. Both sessions started with a low-cost block of 15 trials, followed by a high-cost block of 15 trials, resulting in a total of 30 trials per condition when combining data of the two sessions. This block design could have led to carry-over effects (Patrick et al., 2015), but we have deliberately chosen for this non-counterbalanced design a priori. We opted for this to make sure that our participants (especially older adults and/or cognitively impaired individuals) understood the basics of the task before being introduced to the more complex gaze-contingent high-cost condition.

A drift check (max. 2 dva) was performed before each trial, and recalibration was performed when deemed necessary. After each block, participants answered questions on their experience of commitment to and difficulty of the task (not considered in the current analysis). Each session of the Copy Task took 25 to 45 minutes, dependent on the calibration time, the participants' work pace, and the number and length of breaks.

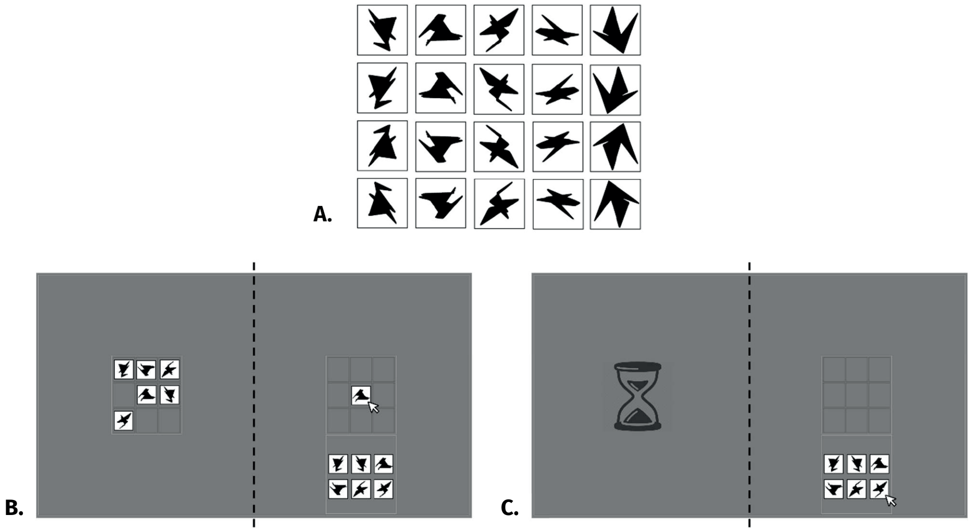


Figure 3.1. **A)** All possible stimuli in the Copy Task. Adopted from Arnoult (1956). An example trial is depicted for the **B)** low-cost condition and **C)** high-cost condition of the Copy Task. At the left-hand side of the screen, the example grid is either visible or replaced by an hourglass for 2000 ms (i.e., gaze-contingent occlusion). At the right-hand side of the screen, the empty grid to place the items (top) and the resource grid (bottom) are presented. A trial ended after 42 seconds. Note: the dotted midline is depicted for illustrative purposes and not visible in the experiment. The Copy Task layout is adopted and adjusted from Somai and colleagues (2020), and Böing and colleagues (2023).

Performance measures. We defined and calculated several outcome measures to describe between-group performance on the Copy Task (see Supplementary Materials: General). For between-group analysis, we calculated the linear integrated speed-accuracy score (LISAS; Vandierendonck, 2017) per individual per condition (low-cost, high-cost) as:

$$LISAS = RTij + \ PEij \ x \frac{S_{RT}}{S_{PE}}$$

where RT_{ij} (reaction time) denotes the trial *i* net copying time (completion time minus hourglass waiting time) divided by the number of correct placements for individual j. The reaction time data was log transformed to account for skewness associated with time measures. PE_{ij} refers to the proportion of errors on trial *i* and equals 1 minus the number of correct placements divided by the total attempts in that trial. S_{RT} denotes the individual *j*'s overall net copying time standard deviation, and S_{PE} is the individual *j*'s overall *PE* standard deviation. Standard deviations were calculated for individual j by collapsing all trials without split on condition (Vandierendonck, 2017). The LISAS was chosen as it combines two outcomes of performance (accuracy and speed) and weighs their importance equally. Lower LISAS reflects better (i.e., more accurate and faster) performance.

Eye-movement measures. We defined and calculated several outcome measures to describe between-group inspection behaviour on the Copy Task (see Supplementary Materials: General). For the between-group analysis, the number of inspections per correct placement was chosen as it reflects eye movement inspection behaviour regardless of overall performance (i.e., 'per correct placement'). Dwell time per correct was analysed as well.

Change Detection Task. (see Supplementary Materials: General for details).
To assess visual working memory capacity in a traditional lab paradigm, we used a simplified version of the Change Detection Task from Luck and Vogel (Luck & Vogel, 1997; Oudman et al., 2020). Participants completed 80 trials in which they verbally reported whether or not they detected a change in the orientation of one bar amongst 2, 3, 4, or 6 bars before and after a white noise mask was presented. *D'* (dprime) was calculated as capacity outcome measure. *D'* is stated to yield a robust outcome for visual working memory performance that is less prone to biases in response tendency than, for example, Kmax (Williams et al., 2022).

$$d' = \mathrm{z[p(hits)] - z[p(falsealarms)]}$$

2.3.2. Neuropsychological tasks. (see Supplementary Materials: General for details).
The neuropsychological tasks that are described below all have a similar task instruction: to memorize and report back as much as possible. Therefore, they are all grafted on obtaining a maximum capacity score. We transform (part of) these capacity scores into one memory compound score that takes into account both verbal and visual (working) memory performance (see 2.4.3. Objective memory capacity).

Location Learning Task. To assess visuospatial immediate and long-term recall the standard stimulus set B of the modified Location Learning Task was used (Kessels et al., 2006, 2014). From this task displacement errors (sum of errors over five trials) can be calculated. A higher number of displacement errors indicates worse memory performance (Kessels et al., 2014). Further, a learning index can be derived, and a delayed recall score can be obtained after prompting the individual to place as many items as possible after ~30 minutes. Only the displacement errors are used in the memory compound score, because these reflect short-term encoding success. Higher displacement error scores indicate worse performance. This score is reversed in pre-processing of the data to ensure that higher numbers reflect better performance. The delayed recall scores reflect longer-term retrieval processes, which are only of secondary interest in the current study.

Rey Auditory Verbal Learning Task. To assess verbal immediate and long-term recall, the Rey Auditory Verbal Learning Task (15 items, Dutch version; Bouma et al., 2012; Saan & Deelman, 1986) was administered. The outcome measure used here is the total number of correctly recalled words over the course of five trials (range: 0-75). Higher scores reflect better memory function. A delayed recall score is obtained after prompting the individual to recall as many words as possible after ~25 minutes. Only the direct recall score is used in the memory compound score, because it reflects short-term encoding success. The delayed recall scores reflect longer-term retrieval processes, which are only of secondary interest in the current study.

Digit Span (WAIS-IV). The Digit Span subtask forward and backward from the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV; Wechsler, 2012) were administered to assess short-term auditory memory and verbal working memory. The longest sequence that was correctly repeated was used as an outcome measure for maximum capacity (span range 2–8 or 2-9, for forward and backward respectively). As such, higher scores indicate better performance. Deviating from our protocol, some patients completed the Digit Span task from the WAIS-III as these were part of the standard administration in the hospital. WAIS-III has a different item score system than WAIS-IV, and therefore yields a different classification of scale scores. However, this has no implications for raw span scores, and therefore, the span scores obtained from the hospital could be used without conversion problems.

Corsi Block Tapping Task. A digitized version (2D) of the Corsi Block Tapping Task was used to assess visuospatial working memory (Brunetti et al., 2014; Claessen et al., 2015; Corsi, 1972; Kessels et al., 2000). The forward subtask assesses short-term visuospatial attention; the backward subtask assesses VWM. To quantify maximum capacity, the span of the longest sequence that was correctly repeated was used (forward range 2–9, backward range 2-8). Higher scores indicate better performance.

2.3.3. Questionnaires

Memory complaints. Participants were asked whether they experienced memory problems (yes/no). This answer was used to categorize participants with and without subjective memory problems. As this question is inclusive but fairly unspecific (Abdulrab & Heun, 2008), we included the Cognitive Failures Questionnaire and the Metamemory In Adulthood questionnaire to obtain a better idea about subjective memory experience.

Cognitive Failures Questionnaire. As a measure of subjective cognitive functioning in the broader term, the Dutch Cognitive Failure Questionnaire is a 25-item questionnaire inquiring about the frequency with which participants experience small mistakes in daily life, on a 5-point scale, globally targeting attention and memory (Broadbent et al., 1982; Ponds et al., 2006), for example: "Do you find to forget whether you've turned off a light or a fire or locked the door?". Items 2, 6, 16, 17, 18, 23, and 24 together make up for a subscale 'absentmindedness' (Ponds et al., 2006) containing items about memory. We used this subscale as an outcome of self-reported memory failure occurrences; it was used in the memory self-efficacy compound.

3

Fatigue. We used the 4-statement Dutch Verkorte Vermoeidheidsvragenlijst to assess experienced fatigue over the previous two weeks (Alberts et al., 1997; Bleijenberg et al., 2009). One of the statements is: "I feel tired". On a 7-point scale, participants were asked to indicate to what extent the statement held true, where higher scores indicate more fatigue. One statement was rephrased ("I feel fit"), so that lower scores indicated more fatigue, and needed to be reversed in scoring. Total scores range from 4 to 28, and a score ≥ 18 indicates severe fatigue. These are reported as a group descriptive.

Hospital Anxiety and Depression Scale. The Dutch Hospital Anxiety and Depression Scale is a 14-item self-report questionnaire that is often administered in clinical care as a screener to assess complaints of anxiety (7 items) and depression (7 items), without focusing on physical complaints (Spinhoven et al., 1997). Scores can be interpreted per subdomain. Scores within the range of 0–7 to indicate no anxiety or depression, 8–10 to indicate possible anxiety or depression, and scores of 11–21 to indicate probable anxiety or depression (Jungen et al., 2019). Note that these results alone are not used to make a clinical diagnosis, but rather serve as an indicator of the presence of distress (Spinhoven et al., 1997). The depression scale is taken into consideration to account for the potential influence of depression on task performance.

Metamemory in Adulthood. The abridged version of the Dutch Metamemory in Adulthood questionnaire was adapted from Ponds and Jolles (1996). It consists of 58 items that inquire about memory and attention, and an additional 16 items that ask about strategies people apply to support memory in daily life. Participants indicated the extent to which they agree with the statement on a 5-point scale. Several scale scores can be computed: Task, Capacity, Change, Anxiety, Achievement, Locus, External Strategies, and Internal Strategies. A memory self-efficacy score – the outcome of self-reported memory functioning – was derived from the Capacity, Change, and Anxiety subscale together. This score was used in the memory self-efficacy compound.

2.4. Pre-processing

2.4.1. Referral. We created a binary variable 'Referral' to indicate whether or not the individual was referred, independently of the outcome of their assessment at the outpatient clinic.

2.4.2. Inspection behaviour. Saccades, fixations, and timestamps were extracted using the EyeLink 1000 parser (default EyeLink saccade detection algorithm, SR Research Ltd., Canada). Data pre-processing was implemented using Python 3.10. Every first trial in each block was removed from analysis: this trial served to check whether the instructions had been retained (additional instructions were given when needed) and to habituate the participant to the new situation (e.g., from low-cost to high-cost). If additional instructions (on mouse use, task, posture) were provided, the trial number was logged and the invalid trial was excluded from analysis. This was the case for 43 of the trials in the group of referred individuals and for 14 of the trials in the group of non-referred controls (see 3.2.1. Data Loss). Variables were calculated as described in 'Measurements'. Data analyses were conducted using R 4.1.2 (R Core Team, 2017).

2.4.3. Objective memory capacity. To get an estimate of objective memory capacity across groups, raw capacity scores (x) were transformed to z-scores for separate tasks. To bundle these into a single memory domain compound score, we averaged the z-scores of the separate capacity tasks to get a single value for general memory performance. Z was denoted as:

$$z = \frac{x - \mu}{\sigma}$$

where x denotes the raw score of the individual, μ refers to the mean for the complete sample, and σ refers to the standard deviation within the complete group. The 'general memory' z-compound is then calculated by summing all available memory z-scores for the individual and dividing it by the number of tasks administered. The scores used in calculation of the objective memory compound score are: Rey Auditory Verbal Learning Task – direct recall (over five trials), Location Learning Task – displacement errors (over five trials), Digit Span forward span, Digit Span backward span, Corsi Block Tapping Task forward span, Corsi Block Tapping Task backward span, and dprime. Note that the delayed recall scores are not taken into account in the compound score. We decided not to do this, as we could not assure that the delay period was equally long for all the participants; the Copy Task often took too long, and the delayed recall may have only taken place after 45 minutes, which is almost twice the time window that is used in clinical care and valid interpretation of the score. We therefore consider the delayed recall scores for the Rey Auditory Verbal Learning Task and Location Learning Task with a grain of salt, but descriptively report them nonetheless.

2.4.4. Memory self-efficacy. The subscale 'absentmindedness' of the Cognitive Failure Questionnaire was taken as an outcome of subjective memory failure. Further, a memory self-efficacy score can be derived from the Capacity, Change and Anxiety subscale of the Metamemory In Adulthood questionnaire. This memory self-efficacy score and the subscale Absentmindedness were transformed into a memory self-efficacy compound z-score. The Metamemory In Adulthood questionnaire was added to the protocol later (as a result of advancing insights), so we only have this data for a smaller part of the participants (n=18 for referred individuals, n=15 for non-referred controls).

3

2.5. Data analyses

2.5.1. Group characteristics. To assure similarity between groups in terms of age and education, Mann-Whitney U tests were performed. A chi-squared test was performed to compare sex distributions between groups. Scores on neuropsychological tasks and questionnaires were reported to characterize groups, and chi-squared tests and proportion z-tests were performed to test group differences.

2.5.2. Inspection strategies and performance across groups. For descriptive purposes, we reported inspection behaviour and performance across the referred and non-referred group. For each individual, we aggregated outcome measures by the mean over trials per condition (low-cost, high-cost), except for time-based outcome measures which were aggregated by the median. Group scores (i.e., medians) were then calculated from these individual values.

To assess group differences in inspection behaviour and performance, we ran linear mixed-effects models (LMM; Singmann & Kellen, 2019) on either of the outcomes across both conditions. Factors included in the LMM were Referral, Condition, Referral*Condition, and random intercept and slope to control for individual differences. After fitting the model, the significance of factors was judged using an alpha of 0.05. The normality of the residuals was visually examined and confirmed for every linear mixed-effects model. Effect sizes were reported as standardised beta-coefficients (β) with a 95% confidence interval. LMM were chosen over mere ANOVAs because of their robustness against deviations from normality of the outcome variables, and because they control for missing data and individual differences (Schielzeth et al., 2020).

Initially, datasets of all participants were analysed without the removal of outliers. To rule out the possibility that our findings were solely driven by outliers, we removed participants whose aggregated scores were ≥1.5 times the interquartile range apart from the total group median for that specific outcome measure in that specific condition (low-cost or high-cost). When identified as outlier in either of the conditions, data of this participant were removed from both conditions. After outlier exclusion, the analyses were run again. Information on the effect of outliers is mentioned in the section of the respective analyses.

2.5.3. Inspecting behaviour based on memory capacity and memory self-efficacy. To investigate the effects of objective and subjective memory functioning on inspection behaviour and related performance, we included the memory capacity compound z-score and memory self-efficacy compound z-score in regression models to predict the number of inspections per correct placement and dwell time per correct placement (as measures of inspection behaviour), and LISAS (as measure of performance). Condition (low-cost, high-cost) was also included as explanatory factor in the model. Age, level of education, and depression score were included as covariates in all models. The significance of factors was judged using an alpha of 0.05.

2.5.4. Predictive value of memory capacity subtasks. To explore the predictive value of memory capacity in verbal or visual working and/or long term memory on inspection behaviour and related performance, we ran (non-parametric) regression models to predict LISAS, and number of crossings per correct placement and dwell time per correct placement in both conditions as a function of z-transformed memory capacity task scores, with covariates age and education. Each of the capacity tasks was included in a separate regression model. The significance of factors was judged using an alpha of 0.05. Results are reported in the Supplementary Results Chapter 3.

3. Results

3.1. Group characteristics

We approached 66 referred individuals through the outpatient memory clinics. Thirty-seven were interested in participation and were invited to the testing facility. Six of these cancelled their appointment without wanting to reschedule, and two test sessions were prematurely ended because the participant was not able to complete the copy task. Eventually, we were able to obtain a valid dataset (with copy task completion being the lead criterium) of 29 referred individuals (see Table 3.1 for demographic characteristics; see Supplementary Figure S3.1 for a patient flow chart; see Supplementary Figure S3.2 for information on suspected neurological aetiology). All individuals were without known visual field defects and had normal or corrected-to-normal visual acuity.

Forty-eight non-referred individuals were recruited as control group. Four dropped out, four were not tested on the copy task due to technical problems, and one participant did not meet our inclusion criteria. For one of the participants, we were unable to track the eyes. Eventually, we obtained a valid dataset (with copy task completion being the lead criterium) of 38 non-referred individuals (see Table 3.1 for demographic characteristics and see Supplementary Figure S3.3 for a control flow chart).

Group characteristics, scores on neuropsychological assessment and questionnaires, and statistical comparisons between groups are displayed in Table 3.1. Note that the level of education is characterized according to the classification of Verhage (1964, 1965), that is commonly used in Dutch clinical care, and classifies the level of education (ranging from 1 to 7) based on the number of education years.

Table 3.1. *Demographic characteristics, scores on memory capacity tasks, and questionnaires, split per group (i.e., individuals referred or not referred to a memory clinic), medians (IQR) or frequencies are depicted.*

	Individuals referred to memory clinic			**Non-referred matched controls**			**Test Statistic [a]**
	n	**n (%)/ Mdn (IQR)**	**Range**	***n***	**n (%)/ Mdn (IQR)**	**Range**	
Demographics							
Sex, male	29	17 (58.6%)		38	15 (39.5%)		χ^2 =1.710, p=.191, d=0.324
Age in years	29	67 (10)	37-80	38	60 (11.8)	40-81	U=404, *p*=.063, *r*=-0.27
Level of education	29	6 (1)	4-7	38	6 (1.75)	4-7	U=589, *p*=.621, *r*=0.07
Suspected neurological aetiology							
yes		17 (58.6%)					
no		11 (37.9%)					
ambiguous diagnosis		1 (3.4%)					
Do you experience memory problems? yes		26 (89.7%)			9 (23.7%)		χ^2= 26.108, *p* <.001**
Fatigue, % severe fatigue		11 (37.9%)			6 (15.8%)		χ^2=3.169, *p*=.076
Anxiety							
Not present (score 0-7)		19 (65.5%)			31 (81.6%)		
Potential (score 8-10)		5 (17.2%)			7 (18.4%)		
Likely (score ≥11)		5 (17.2%)			0 (0%)		
Depression							
Not present (score 0-7)		24 (82.8%)			36 (94.7%)		
Potential (score 8-10)		4 (13.8%)			2 (5.3%)		
Likely (score ≥11)		1 (3.4%)			0 (0%)		
HADS Total score		10 (9)	2 – 22		5 (7.75)	0 – 19	U =280, *p* <.001**, *r*=-0.49

	Individuals referred to memory clinic			Non-referred matched controls			Test Statistic [a]
	n	n (%)/ Mdn (IQR)	Range	*n*	n (%)/ Mdn (IQR)	Range	
Neuropsychological task scores							
Location learning task	29			38			
Total displacement score [b]		49 (35)	5 – 150		27.5 (26)	0 – 75	U=262, *p* <.001***, *r*=-0.53
Learning index (0–1)		0.29 (0.21)	0.05 – 1		0.52 (0.42)	0.10 – 1	U=772, p=.005**, *r*=0.40
Delayed recall: Placement errors	28	3.5 (10.5)	0 - 39	37	1 (4)	0 – 19	U=312, p=.005**, *r*=-0.40
Rey auditory–verbal learning task	28			38			
Immediate recall: Total correct (0–75) [b]		36.5 (15.2)	13 – 51		42 (18)	30 – 67	U=778, *p* <.001***, *r*=0.52
Delayed recall: Total correct (0–15)		4.5 (5)	0 – 13		8 (6)	3 – 14	U=796, *p* <.001***, *r*=0.54
Digit span (WAIS-III/IV)	29			38			
Forward span (2–9) [b]		5 (1)	4 – 7		6 (2)	4 – 9	U=777, p=.002**, *r*= 0.41
Backward span (2–8) [b]		4 (1)	2 – 7		5 (2)	2 – 8	U=708, p=.041*, *r*=0.28
Corsi block-tapping task	29						
Forward span (2–9) [b]		5 (2)	2 – 7	38	5 (1)	3 – 8	U=642, p=.229, *r*=0.16
Backward span (2–8) [b]		5 (2)	2 – 7	37	6 (1)	2 – 7	U=626, p=.234, *r*=0.17
Change detection paradigm	23			36			
d' [b]		1.79 (1.17)	0.82 – 3.36		2.23 (0.99)	0.25 – 3.8	U=498, p=.197, *r*=0.20
Impairment within memory domainc							
% impaired		5 (17.25%)			2 (5.3%)		
% below average		15 (51.7%)			5 (13.2%)		
% within normal range		9 (31.05%)			31 (81.5%)		
Memory capacity compound, z	29	-0.36 (1.07)	-1.53 – 0.69	38	0.24 (0.79)	-0.88 – 1.64	U=846, *p* <.001***, *r*=0.54

Table 3.1. *(continued) Demographic characteristics, scores on memory capacity tasks, and questionnaires, split per group (i.e., individuals referred or not referred to a memory clinic), medians (IQR) or frequencies are depicted.*

	Individuals referred to memory clinic			**Non-referred matched controls**			**Test Statistic [a]**
	n	**n (%)/ Mdn (IQR)**	**Range**	***n***	**n (%)/ Mdn (IQR)**	**Range**	
Memory questionnaires							
Cognitive Failure Questionnaire	29			38			
Total score (max. 125)		63 (21)	38 – 93		54.5 (13.5)	25 – 109	U=344, p=.009*, *r*=-0.38
Scale absentmindedness (max. 35) [e]		17 (6)	10 – 25		14 (4.75)	7 – 30	U=308, p=.002**, *r*=-0.44
Metamemory In Adulthood	18			15			
Scale Anxiety (lower is better) [d]		3.33 (0.58)	2.17 – 3.83		2.67 (1.04)	1.25 – 4	U=69.5, p=.019*, *r*=-0.49
Scale Capacity (higher is better)		2.54 (0.7)	1.75 – 3.67		3.25 (0.58)	2.58 – 4.5	U=226, p=.001***, *r*=0.67
Scale Change (higher is better)		2.35 (0.65)	1.5 – 3.2		3.2 (0.65)	2.11 – 4.3	U=230, *p* <.001***, *r*=0.70
Sum score Memory Self-Efficacy [e]		2.62 (0.7)	2.11 – 3.32		3.27 (0.57)	2.26 – 4.42	U=228, *p* <.001***, *r*=0.69
Memory self-efficacy compound, z	29	-0.28 (1.11)	-1.55 – 1.17	38	0.25 (0.89)	-2.98 – 1.79	U=804, p=.001***, *r*=0.46

n = sample size, Mdn = median, IQR = interquartile range, range (min.–max.). Sample size may differ per outcome variable. [a] Non-parametric test statistics indicating group differences and effect sizes: chi-squared, p-value, and d for binomial variables, or Mann–Whitney–Wilcoxon U, p-value, and rank-biserial correlation r for continuous data. [b] Capacity scores used in memory capacity compound z-score; Location Learning Task displacement errors are reversed so that higher scores indicate better performance on all capacity tasks. [c] Impaired: a score <2nd percentile on ≥ 2 sub tasksb (without *d'*); Below average: a score <2nd percentile on 1 sub taskb (without *d'*) and/or a score between 2nd - 9th percentile on ≥ 2 sub tasksb (without *d'*); Within normal range: does not fit criteria for impairment or below average. [d] Anxiety scale is reversed in calculation of the Memory Self-Efficacy sum score, so that higher scores indicate better subjective memory experience. [e] Scores used in calculation of memory self-efficacy compound z-score; higher scores indicate better subjective memory experience. * $p \leq 0.05$, ** $p \leq 0.005$, *** $p \leq 0.001$.

3.2. Inspection strategies and performance across groups

3.2.1. Data loss. Across 29 participants in the referred group, 1740 trials were planned to be collected. All first trials of each block were removed to assure task comprehension (116 trials). Fourteen trials were lost due to technical issues. Any reason that could possibly interfere with performance (excessive movement of the participant, forgetting the task instructions, problems controlling the mouse) was logged, and the corresponding trials (43 trials) were removed from further analysis. These included trials in which the eye-tracker lost signal. Despite the implementation of a drift check, some trials were started with a drift check above the 2 degrees visual angle threshold. When exceeding 5 degrees visual angle, trials were excluded (14 trials). Finally, 1553 trials were left for analysis.

Across 38 participants in the non-referred group, 2280 trials were planned to be collected. Again, all first trials of each block were removed (152 trials). In the non-referred group, one participant did not complete the second session of the Copy Task (minus 28 trials). Trials that were invalid due to signal loss, excessive movement of the participant, forgetting the task instructions, or problems controlling the mouse were removed (14 trials). For one participant, we were urged to exclude the entire second session because the majority of trials exceeded the drift check threshold. In total, 45 trials needed to be excluded because of exceeding the drift check threshold. In this group, 2041 trials were left for analysis.

3.2.2. Descriptive values. Group scores for inspection behaviour and performance across conditions (low-cost and high-cost) were calculated and reported in Table 3.2. We confirmed that there was no differential effect (interaction) of session number across groups on our outcome measures of interest (in bold, Table 3.2) to ensure that pooling the conditions across sessions was a valid practice and outcome measures would not be confounded by differences in session effects between groups.

Table 3.2. *Group scores (referred, non-referred) for outcomes of performance and inspection behaviour across conditions (low-cost and high-cost). Variables in bold are used in subsequent analyses.*

Copy Task Scores	Individuals referred to memory clinic			Non-referred matched controls		
	n	Mdn (IQR)	Range	*n*	Mdn (IQR)	Range
Completion time, s	29			38		
Low-cost [a]		21.5 (11)	13.3 – 42		18.5 (5.95)	12.1 – 33.4
High-cost		38.2 (6.42)	30.8 – 42		33.8 (10)	24 – 42
Net copying time, s	29			38		
Low-cost [a]		21.5 (11)	13.3 – 42		18.5 (5.95)	12.1 – 33.4
High-cost		28 (5.45)	22.4 – 32.33		23.8 (5)	18.5 – 36
Correct placements (0-6)	29			38		
Low-cost		6 (0.15)	4.39 – 6		6 (0.03)	5.57 – 6
High-cost		5.46 (1.52)	3.07 – 6		5.83 (0.57)	2.23 – 6
Success rate (0-1)	27			36		
Low-cost		0.97 (0.03)	0.78 – 1		0.97 (0.04)	0.84 – 1
High-cost		0.87 (0.06)	0.74 – 0.98		0.91 (0.01)	0.50 – 0.98
Speed score, s	29			38		
Low-cost		3.7 (2.18)	2.35 – 11.3		3.25 (1.18)	2.09 – 6.73
High-cost		5.24 (3.16)	3.75 – 11.3		4.16 (1.23)	3.28 – 15.2
LISAS				36		
Low-cost	26	1.38 (0.63)	0.94 – 2.52		1.22 (0.43)	0.77 – 2.14
High-cost	27	2.03 (0.54)	1.49 – 2.75		1.67 (0.42)	1.26 – 3.71
Number of crossings	29			38		
Low-cost		9.89 (3.29)	5.96 – 18.6		9.62 (2.52)	5.29 – 14.1
High-cost		4.73 (1.16)	2.93 – 6.18		4.15 (1.87)	1.82 – 6.96
Dwell time per crossing, s	29			38		
Low-cost		0.47 (0.14)	0.27 – 1.06		0.39 (0.11)	0.26 – 0.57
High-cost		1.17 (0.66)	0.56 – 3.73		1.13 (0.79)	0.56 – 5.62
Number of inspections per correct placement	29			38		
Low-cost		1.88 (0.67)	0.99 – 5.44		1.61 (0.42)	0.88 – 2.9
High-cost		1.05 (0.55)	0.49 – 1.66		0.84 (0.37)	0.30 – 1.9
Dwell time per correct placement, s	29			38		
Low-cost		0.74 (0.38)	0.44 – 2.61		0.62 (0.18)	0.36 – 1.11
High-cost		1.14 (0.64)	0.56 – 2.8		0.84 (0.26)	0.55 – 3.03

Valid datasets *n*, median Mdn, interquartile range (IQR), and range (min.–max.). [a] Completion time and net copying time in the low-cost condition are the same. In the high-cost condition, the net copying time is the completion time minus the hourglass waiting time.

3.2.3. Inspection behaviour analysis. A linear mixed-effect model was fit to predict the number of inspections per correct placement by referral (referred, non-referred) and condition (low-cost, high-cost). A main effect of condition was found (t = -11.178, p < .001, β = -0.87 [-1.02, -0.72]), with more model inspections in the low-cost condition as compared to the high-cost condition. No effect of referral was found (p = .069), nor was there an interaction effect (p = .5): referred individuals inspected the model just as often as non-referred controls to place one item correctly. Figure 3.2 visualizes findings for the inspection behaviour analysis.

The same factors were included in a model with dwell time per correct placements as dependent variable. Again, a main effect of condition was found (t = 4.15, p < .001, β = 0.38 [0.20, 0.56]), showing that inspection durations of the model increased in the high-cost condition. Here, a main effect of referral was found (t = 2.87, p = .005, β = 0.35 [0.11, 0.59]), showing that referred individuals took more time to inspect the model for one correct placement compared to non-referred controls. No interaction effect between referral and condition was found (p = .8).

Sensitivity analyses. After outlier removal, the effect of condition was still present for both the number of inspections as well as the dwell time per correct placement (both p < .001). The effect of referral on the number of inspections remained insignificant for the number of inspections (p = .06), and vanished for dwell time per correctly placed item (p = .09).

3.2.4. Performance analysis. A linear mixed-effect model was fit to analyse the influence of referral (referred, non-referred) and condition (low-cost, high-cost) on speed-accuracy performance (LISAS). The model yielded a main effect of condition (t = 9.19, p < .001, β = 0.53 [0.41, 0.64]) with performance decreasing with high-cost inspecting as compared to low-cost inspecting. A main effect of referral (t = 2.66, p = .01, β = 0.26 [0.07, 0.46]) was present, indicating that referred individuals performed worse than non-referred individuals. No interaction between referral and condition was found (p = .7). Figure 3.3 visualizes findings for the performance analysis.

Sensitivity analysis. When running the same models after outlier removal, the effects of condition and referral persisted (p < .001, p < .005, respectively), and again, no interaction was found (p = .35).

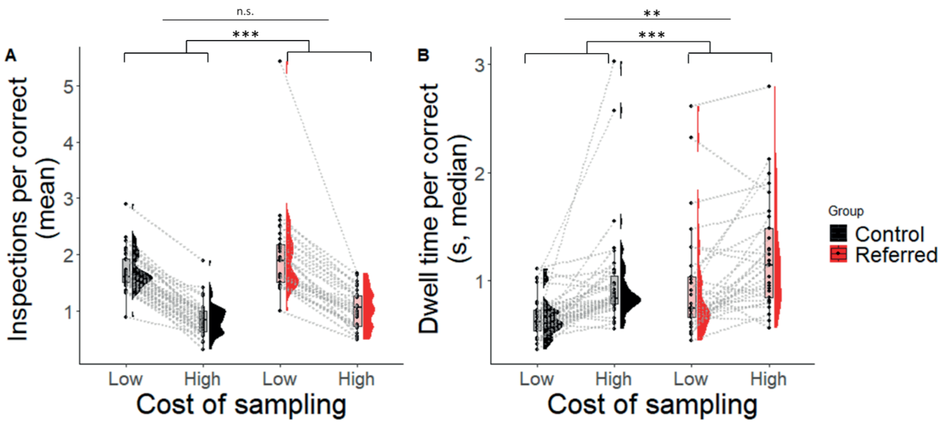


Figure 3.2. Eye-movement measures as indicator for inspection behaviour. **A)** Mean number of inspections needed to make one correct placement, **B)** Median dwell time in seconds per correct placement for non-referred controls (black) and referred individuals (red) across conditions (low-cost, high-cost). Black dots and grey lines represent outcomes of individual participants. Asterisks indicate significant effects. n.s = non-significant, **p* ≤ .05, ***p* ≤ .01,*** *p* ≤ .001,

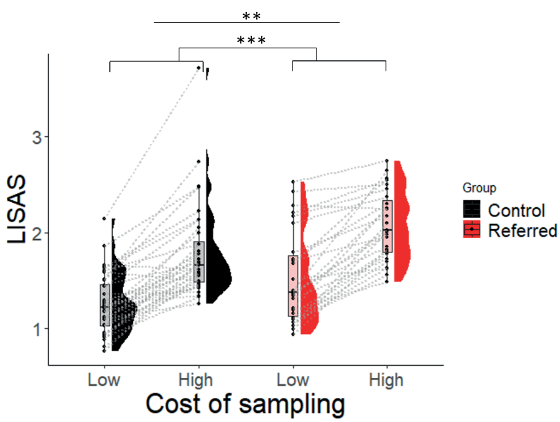


Figure 3.3. Performance outcome expressed in LISAS for non-referred controls (black) and referred individuals (red) across conditions (low-cost, high-cost). Black dots and grey lines represent outcomes of individual participants. Lower LISAS indicates better performance. Asterisks indicate significant effects. **p* ≤ .05, ***p* ≤ .01,*** *p* ≤ .001

3.3. Inspection behaviour based on memory capacity and memory self-efficacy

Referred and non-referred individuals did not differ in the number of inspections per correct placement, and the difference between groups for the dwell time needed to place one item correctly was mostly driven by outliers. This indicates that the distinction between groups is not so clear-cut, which can be attributed to the fact there is large overlap between groups (see Supplementary Figure S3.4). Some referred individuals showed no objective memory impairments, while some non-referred individuals did show objective memory impairments. Although the groups statistically differed on both objective memory capacity and memory self-efficacy (see Table 3.1), referral as a sole factor appears not to be sensitive enough to explain inspection behaviour. To investigate the effects of objective memory functioning and memory self-efficacy on inspection behaviour and related performance, we tested the predictive value of the memory capacity compound and the memory self-efficacy compound. As there were no strong indicators of severe depression in our sample (see Table 3.1), we decided to exclude this covariate from the model to reach higher power. The level of education and age were included as covariate.

3

The number of inspections per correct placement was predicted by condition ($t = -12.8$, $p < .001$, $\beta = -0.9$), and the memory capacity compound score ($t = -4.56$, $p < .001$, $\beta = -0.36$), but no effect of memory self-efficacy was found ($p = .28$). Interaction effects were absent (all $p > .3$). Dwell time per correct placement was influenced by condition ($t = 4.34$, $p < .001$, $\beta = 0.32$), and memory capacity ($t = -3.83$, $p < .001$, $\beta = -0.32$). Again, no effect of the subjective component was found ($p = .74$), and no interaction effects were apparent (all $p > .05$). Finally, LISAS was significantly predicted by condition ($t = 8.6$, $p < .001$, $\beta = 0.54$), memory capacity ($t = -5.12$, $p < .001$, $\beta = -0.36$), but not by memory self-efficacy ($p = .67$). The effects of condition and memory capacity held under nonparametric tests. Figure 3.4, 3.5 and 3.6 visualize the observed effects per outcome variable. Covariates are not taken into account in these figures.

In summary, higher capacity was associated with fewer and shorter inspections per correctly placed item, and better performance, regardless of whether or not information was readily available. Memory self-efficacy was associated with neither of these outcomes. The same conclusions were drawn when running the analyses with only the CFQ as a measure of memory self-efficacy (to make sure that the hiatus in number of valid data sets of people completing the MIA would not bias the results for this component; see Supplementary Results Chapter 3 for elaboration).

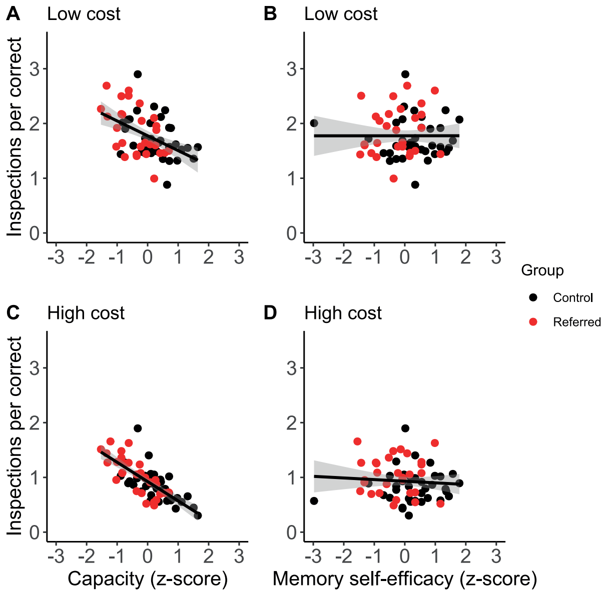


Figure 3.4. The relation between the number of inspections per correct placement and the capacity compound z-score **(A,C)**, and memory self-efficacy compound z-score **(B,D)** in the low cost and high-cost condition. Non-referred controls are depicted in black, individuals referred to the outpatient memory clinic are depicted in red. A smoothed linear coefficient is added in black with confidence intervals in grey. Covariates are not taken into account in this figure.

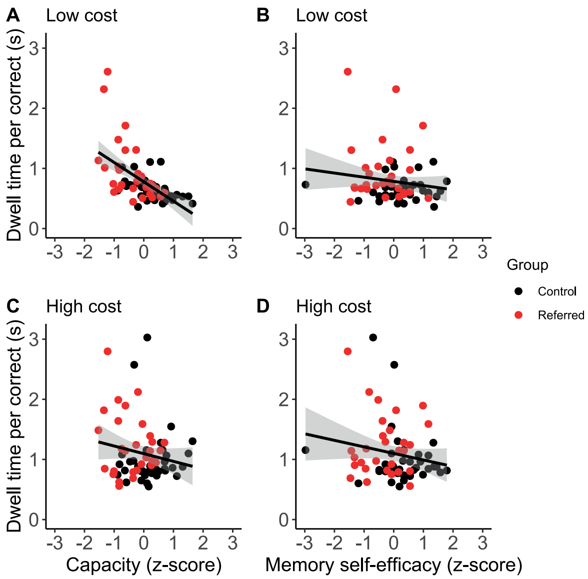


Figure 3.5. The relation between dwell time per correct placement and the capacity compound z-score **(A,C)**, and memory self-efficacy compound z-score **(B,D)** in the low cost and high-cost condition. Non-referred controls are depicted in black, individuals referred to the outpatient memory clinic are depicted in red. A smoothed linear coefficient is added in black with confidence intervals in grey. Covariates are not taken into account in this figure.

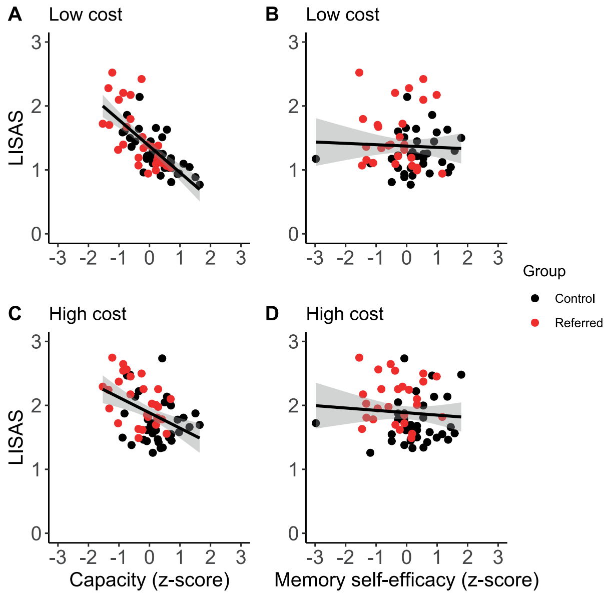


Figure 3.6. The relation between LISAS and the capacity compound z-score **(A,C)**, and memory self-efficacy compound z-score **(B,D)** in the low cost and high cost condition. Non-referred controls are depicted in black, individuals referred to the outpatient memory clinic are depicted in red. A smoothed linear coefficient is added in black with confidence intervals in grey. Covariates are not taken into account in this figure.

3.4. Predictive value of memory capacity subtasks and level of memory functioning

The results of the non-parametric regression models to explore the number of crossings per correct placement and dwell time per correct placement in both conditions as a function of raw memory capacity subtask scores, with covariates age and level of education, are reported in Supplementary Table S3.2. We have analysed all subtasks (also the delayed recall scores) to explore any relation between memory subprocesses (e.g., short-term encoding versus long-term retrieval) and inspection behaviour. After correcting for multiple tests, we found that verbal attentional span (Digit Span forward), visual working memory capacity (*d'*), and verbal encoding (Rey Auditory-Verbal Learning Task total score) were related to inspection frequency and duration when information was freely available. When inspecting information became more costly, the other subtasks also started to exert their influence on inspection frequency; we found that all but two (Location Learning Task displacement errors and Location Learning Task delayed recall) subtests were predictive of inspection frequency in the high-cost condition, implying that higher memory capacity on each of these subtasks resulted in fewer inspections needed to place one item correctly. Interestingly, none were related to dwell time in the high-cost condition. Although interpreting these results is premature due to the relatively small sample size, we cautiously infer that there may be a benefit of both verbal and visual attentional and working memory span, resulting in fewer inspections and shorter inspection duration in stable visual environments, and that one may benefit further from higher capacity in situations where information is less readily available and memorization is prompted. Individuals with higher capacity rely less on the outside world.

To get a gist about clinical value, we further visualized inspections per correct as a function of level of memory performance category (intact, below average, impaired; see Supplementary Figure 3.5). The impaired group (performance < 2nd percentile on two or more subtasks, American Psychiatric Association, 2013; Hendriks et al., 2014, 2020) inspected significantly more than the intact and below average group (performance < 2nd percentile on one subtask (and/)or below the 9th percentile on two or more subtasks), but the intact and below average group did not differ from each other.

4. Discussion

Relying on the external world by (re)inspecting (i.e., 'sampling' or 'offloading') visual information alleviates the need to load internal memory to its full capacity. Since visual information remains available in the external world, it is unlikely that the full capacity of visual memory will be used in everyday life if given the option not to. However, to clinically objectify memory complaints that have warranted a referral for cognitive assessment, it is precisely this capacity characteristic that is examined, and it is generally not taken into account that the individual can exploit their environment as a support system. Crucially, this means that current memory assessment fails to incorporate the possibility that people may choose to memorize information at the preferred rather than the maximum load, and instead (re)inspect information from the external world. To complement the clinical approximation of memory use in everyday life, we assessed memory capacity, subjective experience of memory functioning (i.e., memory self-efficacy), and inspection behaviour in individuals who were either referred to an outpatient clinic for cognitive assessment or not.

As expected, compared to the non-referred group, the referred group had a lower memory capacity as measured with standard neuropsychological memory tests, and lower levels of memory self-efficacy as measured with memory self-efficacy questionnaires. When participants were asked to copy an example puzzle to an empty grid, there was no difference in inspection frequency: both groups inspected the example puzzle equally often to place one item correctly. Only the degree of availability of information showed to influence inspection frequency: both groups made fewer (re)inspections when they had to wait every time they wanted to inspect the example puzzle (high cost condition) as compared to when the information was continuously available (low cost condition), replicating results from previous studies (Draschkow et al., 2021; Sahakian et al., 2023; Somai et al., 2020). Interestingly, the referred group did not inspect more often than the non-referred group in either of the conditions, while an effect of memory impairment was previously shown in a previous study where a group of severely memory-impaired individuals with Korsakoff syndrome inspected more often as compared to age-matched healthy controls (Böing et al., 2023). The absence of a group difference in the current sample can partly be explained by the heterogeneity of both the referred and non-referred group: some individuals were referred to a memory clinic but performed only slightly below average, others were not referred but showed impaired memory performance. As we aimed for a group that varied in both objective memory impairments and subjective memory complaints, it was no surprise that groups overlapped regarding memory functioning, and that there was no clear-cut impaired versus non-impaired difference. Our main aim was to include all individuals within one model and investigate the independent effects of objective memory capacity and subjective memory functioning, rather than referral. We found that people with a lower memory capacity inspected more frequently as compared to those with a higher memory capacity. As a subsequent exploration, we interpreted the raw capacity span scores against appropriate norm scores (controlled for age and education), to check the effect of clinical memory impairment. Those who would be classified as clinically impaired, inspected more frequently as compared to those whose performance was below average or intact. These results align with our previous study showing distinctly different behaviour for memory-impaired individuals than healthy controls (Böing et al., 2023). This effect thus only arises with more profound memory deficits and shows that inspection frequency is not a sensitive measure to map subtle memory deficiencies; measuring inspection frequency only distinguishes two subgroups (impaired vs. non-impaired) rather than three (impaired vs. below average vs. intact, tested with standard neuropsychological capacity assessments). Future studies with a larger sample size would allow to dissociate subgroups based on inspection behaviour not only in terms of the level of memory function, but also in terms of clinical status (e.g., mild cognitive impairment, dementia, Parkinson's, or presence of psychological factors).

Although there was no difference between the referred and non-referred groups regarding how often they inspected information, there was a difference in inspection duration. The referred group showed increased dwell times, indicating a potential necessity for longer encoding times or a slower evidence (here, confidence) accumulation to reach an action threshold (Lee et al., 2023; Sahakian et al., 2023), although it should be noted that this effect disappeared when outliers were removed. The referred group also showed worse performance than the non-referred group in terms of task speed-accuracy, which may be a reflection of the longer information uptake, but which could also arise because more errors were made, or slower information processing speed in general. In summary, the referred group did not rely more often but potentially longer on the outside world, and showed weakened performance as compared to the non-referred group.

As memory capacity varies on a continuous rather than dichotomous scale, we were particularly interested to see whether and how the objective memory capacity span and memory self-efficacy would influence inspection frequency, regardless of referral or clinical status. Surprisingly, we found that memory capacity, but not memory self-efficacy, was related to both inspection frequency and dwell time. Higher memory capacity related to fewer and shorter (re)inspections to place one item correctly. These results show that the standard neuropsychological memory capacity tasks used in clinical care generalize to behaviour in a more free setting where participants can choose whether or not to load memory. In other words, the amount of information one is able to memorize translates to how often someone falls back onto the external world, and consequently, how much information someone will memorize in day-to-day memory use. This relation is clinically implied but not so often directly tested, although there have been some earlier attempts in which capacity scores assessed with specific neuropsychological measures (e.g. Corsi Forward Span) were related to different types of cognitive offloading (Meyerhoff et al., 2021). Cognitive offloading is a broad concept and refers to any type of behaviour that is executed to decrease the effort associated with cognitive processes (e.g., using a calendar to support prospective memory, but also tilting one's head to avoid mental rotation). We consider inspection behaviour as a subordinate of cognitive offloading behaviour. Our results add to the offloading literature by taking several clinically relevant memory capacity subtasks into account and testing them as an integrated memory concept as well as displaying their individual predictive value on inspection behaviour.

One remark on estimating memory capacity is that the standard neuropsychological task outcomes are based on a single measurement in time, making them susceptible to measurement errors and attentional lapses. Composing a capacity score that combines performance on multiple memory subtasks (such as in the present study, and in, e.g., Morrison & Richmond, 2020) partly accounts for such momentary influences, but cannot fully eliminate them. Surely it would be more robust to extract an average capacity span from repeated trials for each subtask (as Meyerhoff and colleagues (2021) did for the Corsi Block Tapping Task Forward Span), but this is time-costly and burdensome for patients in a clinical setting. Other momentary influences relate to motivational and personal aspects of task completion, such as the desire to be accurate or certain rather than quick, or the drive to perform at a personally set maximum or a 'just' sufficient level. Consider also that undergoing assessment as part of a scientific study comes with a different incentive, mindset and setting than completing these for the sake of diagnosis in a clinical setting.

Apart from using an integrated memory concept rather than single capacity measures, the current results also add another clinical perspective to the offloading literature: while most offloading studies are based upon healthy student populations, we extend our findings to adults with memory impairments (including Böing et al., 2023). Importantly, the various lines of research on strategic use of the external world emphasize the wide variety of memory support strategies, ranging from either the trade-off in memorizing versus sampling in working memory (Böing et al., 2023; Draschkow et al., 2021; Hoogerbrugge, Strauch, Böing, et al., 2024; Sahakian et al., 2023; Somai et al., 2020; Van der Stigchel, 2020) to more conscious and deliberate cognitive offloading (e.g., writing things down to aid memory, placing a cue for oneself to remember intentions; see Gilbert et al. (2023) and Risko & Gilbert (2016) for a review). With regards to the relation between memory capacity and any type of offloading, findings are mixed, showing that they are subject to specific characteristics of the memory task being used. For example, prospective memory (i.e., remembering an intention to act out in the future) requires a different allocation of subsystems and operates on a different timescale than making sure to remember and dial the correct number from an appointment note, and memory capacity may differentially affect these processes. It is therefore difficult to generalize findings of the current study on visual working memory capacity, but also those of the aforementioned studies, to a general concept of real-world memory usage. Nevertheless, our findings emphasize the need to take memory strategies into consideration when trying to approximate freedom-of-choice memory use alongside memory capacity. The relevance of such considerations is subserved by our current observation that, although higher memory capacity relates to fewer inspections, people still avoid maximum capacity usage, and prefer to memorize one up to three items maximum in working memory. Even when information is not always readily available (high cost condition), people often take some degree of time or effort expenditure for granted (in waiting, annoyance, physical exertion) to avoid full memory capacity use. What's more, increasing the amount of information to be remembered increases the likelihood of offloading (Risko & Gilbert, 2016). This adaptive behaviour is missed in the regular assessment of memory capacity.

It is intriguing that subjective and objective capacity measures are frequently discrepant; people may experience memory failure in the absence of impaired memory capacity (Beaudoin & Desrichard, 2011; Mattos et al., 2003; Ponds & Jolles, 1996a). In our results, similarly, the lack of a relation between subjective memory performance on the one hand, and capacity and inspection behaviour on the other, illustrates that memory self-(in)efficacy does not translate to maximum capacity nor actual visual working memory usage. Whereas we expected people with lower confidence to check more often, memory self-efficacy was not associated with the frequency with which one relied on the external world. This contrasts with Gilbert (2015b), who posed that both task-specific as well as domain-general metacognitive confidence (i.e., like our quantification of memory self-efficacy) explain offloading. The incongruency between their and our findings likely arises due to different operationalizations of offloading, with the study of Gilbert (2015b) focusing on intention offloading (hinting at prospective memory) and ours on working memory. In a task that engages working memory in a way that is more similar to ours, Grinschgl and colleagues (2021) manipulated domain-general confidence by facilitating fake feedback, and found, like us, no effect of confidence on offloading behaviour. Following the same line as Hertzog and colleagues (1987, as cited in Beaudoin & Desrichard, 2011), the authors propose a distinction between metacognitive beliefs, i.e., a more generic gist about one's memory, and metacognitive experiences, reflecting confidence about performance on a specific task (or trial) that had just been completed. We captured metacognitive beliefs, but not metacognitive experiences across the test procedure (e.g., self-efficacy after a capacity task, versus after an (un)successful copy task trial), while the latter might have exerted an effect on inspection strategy in the way that Gilbert (2015b) described. This line of reasoning suggests that memory confidence fluctuates depending on prior experience, task-specific characteristics and the moment in time one is asked to judge confidence. Further diving into the waters of memory uncertainty, Sahakian and colleagues (2023) showed that, even within trials, people were not keen on acting upon content that they were not confident enough about, although they had some residual information in working memory. This aligns with the observation that people may use offloading even though it does not necessarily benefit performance but mostly serves to safeguard a feeling of security (Risko & Dunn, 2015). Reinspecting could then be seen as an act to accumulate confidence, and the threshold of certainty that needs to be reached would then be described as an action threshold (Sahakian et al., 2023). We expected our measure of memory self-efficacy to express the individual's general action threshold: some people would be more confident than others, and would therefore need less reassurance, resulting in fewer inspections. Yet, individual differences in memory self-efficacy could not account for inspection frequency (nor duration). If one wants to make claims about reinspecting – or offloading in the broader sense – as an expression of reassurance behaviour, it might be valuable to test people with specific tendencies as observed in, for example, individuals with obsessive-compulsive disorder (Karadag et al., 2005; Tolin et al., 2001), or high levels of performance failure anxiety. A further peril of judging one's own memory (that seeps through to the memory self-efficacy construct and also into the reason for referral) is that the general public often falsely attributes cognitive failures to memory dysfunction, while they are often the result of deficits in other domains (e.g., executive function, attention, motivation; Hendriks et al., 2014). We cannot rule out the influence of these factors within our sample.

The heterogeneity in tasks, groups, cognitive profiles and personal characteristics adds complexity to disentangling factors that influence memory usage, but at the same time reflects the complexity with which the clinician is faced when assessing cognitive functioning in the clinic. We therefore embrace this complexity when trying to approximate memory usage in daily life and underline that memory assessment is multifaceted. Different tasks may evoke different self-perceived memory ability, capacity estimates may vary across the verbal and visual modalities, the ease with which external stimuli are internalized may vary because of stimulus characteristics, semantic labelling, and familiarity (e.g., shopping items versus abstract geometric shapes), and how one chooses to use either the internal memory load or the external world is differentially balanced based on both internal and external demands. Further, the ability to learn over the course of trials and draw from long-term memory (e.g., stronger memories for repeated stimuli) may differentially influence individuals. When adding degrees of freedom to a task, as in everyday life, other cognitive domains may further start to interact with how memory is engaged. For example, our task not only required working memory, but also aspects of executive functioning (attention, planning, monitoring). As we have tested memory functioning and not executive functioning, we can only make an attempt to attribute our findings to working memory. Yet, one should be wary that this does not nearly explain the additional cognitive processes that may play a role when interacting with the environment and engaging working memory in everyday life. Taken together, approaching memory functioning by only examining memory capacity does not do justice to the many layers of memory usage: individuals – with and without memory capacity constraints – may employ a variety of compensatory fallbacks dependent on the task at hand, explaining why we do see differences between referred and non-referred individuals on tasks that force maximum capacity use, but lesser so on a task where people have more degrees of freedom and can choose how many items they memorize at once.

To put this in a clinical perspective: memory complaints, as repeatedly shown, do not necessarily translate to memory capacity measures, and memory capacity does only to a certain extent translate to actual memory use when given the choice. Although forced capacity tasks appear to be more sensitive to categorize individuals' performance as clinically impaired, below average or intact as compared to our freedom-of-choice Copy Task, they fail to capture the workaround that people use to prevent maximum capacity usage. When given some wiggle room, those with lower memory capacity can compensate by increasing their reliance on the external world. These individuals might need to inspect information somewhat more often or longer to use information correctly, but this may be the relatively 'cheap' price one pays to work around capacity limits.

Acknowledgements: The authors thank T. Hamers and B. de Zwart for helping with data collection. We thank A.J. Hoogerbrugge for programming the Copy Task, and supporting with data pre-processing.

Supplementary Materials: Additional descriptions, tables, figures and analyses that were reported in the manuscript can be found in the Supplementary Materials of this dissertation.

Author Contributions: Sanne Böing: conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, visualization, writing – original draft, writing – review and editing. **Antonia F. Ten Brink :** conceptualization, methodology, resources, software, supervision, writing – review and editing. **Zoë A. Schielen:** conceptualization, methodology, resources, writing – review and editing. **Carla Ruis:** conceptualization, methodology, resources, writing – review and editing. **Esther van den Berg :** conceptualization, methodology, resources, writing – review and editing. **J. Matthijs Biesbroek:** resources, writing – review and editing. **Tanja C.W. Nijboer:** conceptualization, methodology, supervision, writing – review and editing. **Stefan Van der Stigchel:** conceptualization, funding acquisition, methodology, supervision, writing – review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 863732).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the Faculty Ethics Review Board of the Faculty of Social and Behavioural Sciences at Utrecht University (protocol numbers protocol numbers 21-0076 and 21-0269).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent was obtained from the patients to use their data in this paper.

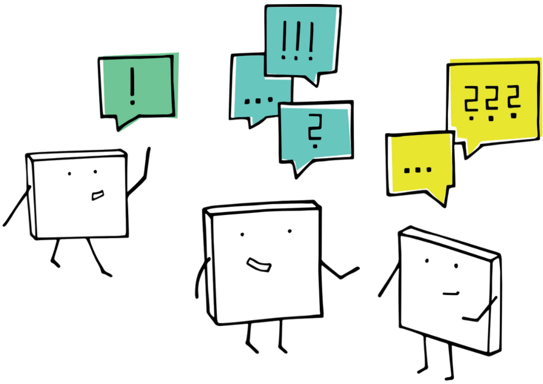
Data Availability Statement: The data presented in this study are openly available in Open Science Framework at https://osf.io/ys67b/.

Declaration of Interest The authors declare no conflict of interest.

CHAPTER 4

Relying on the external world: Individuals variably use low- and medium-loading, but rarely high-loading, strategies when engaging visual working memory

Sanne Böing, Beleke de Zwart, Antonia F. Ten Brink, Tanja C.W. Nijboer, & Stefan Van der Stigchel



Published as

Böing, S., Ten Brink, A. F., de Zwart, B., Nijboer, T. C. W., & Van der Stigchel, S. (2025). Relying on the external world: Individuals variably use low- and medium-loading, but rarely high-loading, strategies when engaging visual working memory. *Visual Cognition*. https://doi.org/10.1080/13506285.2025.2484523

Abstract

In naturalistic environments, individuals typically rely on external sampling rather than fully utilizing their visual working memory capacity. When external sampling becomes more costly people memorize more information (i.e., loading). To investigate individual differences in sampling versus loading strategies, healthy participants (n=88) performed a copying task under low-cost (immediate accessibility) and high-cost (delayed accessibility) conditions, which were counterbalanced to test the effect of prior experience. We categorized participants as low-loaders (sampling >1 per item), medium-loaders (loading ≥1 items per inspection), and high-loaders (loading ≥3 items per inspection), and found that both sampling cost and prior experience affected sampling frequency and category. Crucially, low- and medium-loading strategies were common, but individuals seldom exhibited a high-loading strategy that approached working memory capacity limits. Despite individual differences in sampling, participants flexibly adapted sampling frequency to changing task demands: neither sampling frequency nor adjustments in sampling frequency across conditions impacted performance. This suggests that while individuals have distinct working memory strategies, they can adjust these as task demands change without affecting performance. We propose that a trade-off between effort, performance goals, and prior experience determines the current sampling strategy, with individual variation in their preferred approach.

Keywords: offloading; memory strategies; individual differences; sampling; copy task

Abbreviations:

LISAS	Linear Integrated Speed-Accuracy Score
LMM	Linear Mixed-effects Model

1. Introduction

Research on visual working memory has traditionally been concerned with estimating its maximum capacity (e.g., Luck & Vogel, 2013). To optimize performance on a capacity test, people should try to memorize as much information as possible, which means they should use their memory capacity to the fullest. Although these capacity tasks have provided fundamental insights into the mechanisms of working memory (Luck & Vogel, 2013; Ma et al., 2014) and have been useful in compiling cognitive profiles and subsequently guide diagnoses in clinical care (e.g., Corsi, 1972), they fall short in mimicking actual working memory usage in daily life, and do not grasp the wide variety of working memory strategies that individuals might use to support task performance.

The influence of the environmental context on the use of memory has received increased attention in the fields of cognitive engineering and human-system interactions (e.g., Gray & Fu, 2004; Morgan et al., 2009; Waldron et al., 2007), working memory (Ballard et al., 1995; Draschkow et al., 2021; Droll & Hayhoe, 2008; Grinschgl, Papenmeier, et al., 2021; Hoogerbrugge, Strauch, Böing, et al., 2024; Kvitelashvili & Kessler, 2024; Melnik et al., 2018; Meyerhoff et al., 2021; Risko & Gilbert, 2016; Sahakian et al., 2023; Somai et al., 2020; Van der Stigchel, 2020), and neuropsychology (Böing et al., 2023, 2025). Interestingly, results in these fields indicate that, even when the demands of the environmental context encourage people to shift towards memorization, people still do not employ their full capacity, and often keep relying on the outside world. To not fully tax memory, people may load up less than their maximum capacity or may use a cognitive offloading strategy (e.g., writing things down, creating cues as reminder; Ballard et al., 1995; Böing et al., 2023, 2025; Draschkow et al., 2021; Gray et al., 2006; Meyerhoff et al., 2021; Morrison & Richmond, 2020; Risko & Dunn, 2015; Sahakian et al., 2023; Somai et al., 2020). To illustrate these findings in the context of a natural environment, consider assembling a piece of furniture. The instruction leaflet is available for one to rely on. Hence, there is no need to memorize all the screws, their rotation, and their desired location at once because one can inspect and reinspect the building steps depicted on the leaflet as often as desired. In such cases, one can afford to rely on the external world by (re)inspecting information in a just-in-time manner once the need arises (Hoogerbrugge et al., 2023; Hoogerbrugge, Strauch, Böing, et al., 2024). Only when this ‘sampling’ is impeded (say, one has to walk a few metres to one’s computer screen for a digital leaflet), people load up more information per iteration, and shift from relying on the external world to relying on internal memory (Ballard et al., 1995; Böing et al., 2023, 2025; Draschkow et al., 2021; Fu & Gray, 2000; Sahakian et al., 2023; Somai et al., 2020). This shows the interactive and adaptive nature of engaging working memory in natural environments.

The tendency to rely on the outside world has been observed on a group level, but there may be large individual differences regarding the degree to which one relies on either the external world or internal memory, and how much one switches between those when task demands change. The dynamic allocation of external (sampling) and internal memory resources depends upon many different factors, including memory capacity (Böing et al., 2023, 2025; Meyerhoff et al., 2021; Morrison & Richmond, 2020; Risko & Dunn, 2015), general beliefs about or experienced successes with one's memory functioning (e.g., Böing et al., 2025; Gilbert, 2015b), the perceived importance of accuracy and/or quick task completion (Risko & Dunn, 2015; Sahakian et al., 2023), the goal of effort minimization (Kvitelashvili & Kessler, 2024; Risko & Dunn, 2015), and, as mentioned before, task setup and information availability (Ballard et al., 1995; Draschkow et al., 2021; Gray et al., 2006; Sahakian et al., 2023; Somai et al., 2020). Task properties impose hard constraints on how to carry out a task, while other (personal) factors make up soft constraints (Gray et al., 2006). The unique combination of factors for a given individual and situation likely determines which strategy is applied. Yet, in both experimental working memory paradigms and clinical assessment of working memory, individual variability in working memory strategies have often been neglected (Logie, 2023). As opposed to looking at aggregate scores of performance, assessing individual variability in a range of working memory tasks can yield theoretical as well as clinical advancements in our understanding of working memory use (Logie, 2023).

The limited body of literature that addressed individual variation in strategy deployment did so within the context of capacity tasks. For example, Morrison and colleagues (2016) explored how verbal working memory tasks with different demands prompted inter-task and individual variability in strategy deployment. They found that internal strategies (e.g., rehearsing, grouping, or visualization) were not homogenously applied across all participants nor consistent within participants across various tasks. Similarly, the deployment of external strategies for memorizing verbal information (Morrison & Richmond, 2020) or future intentions (Meyerhoff et al., 2021; Risko & Dunn, 2015), such as writing something down or placing reminders when given the choice, also varied between individuals. With regards to visuospatial working memory, many different strategies have been described to support task performance, among which chunking, holistic encoding, and visuospatial rehearsal, which usage is likely dependent on the task, the individual, and even on the individual trial (Gonthier, 2021). Although these studies substantiate the idea that there is indeed variation in individual approaches to working memory strategies, these findings are, yet again, based upon paradigms where the goal is to memorize and report as much information as possible, and where it is disregarded that information generally remains available in the outside world. As situations often allow for reliance on the external world, (re)inspecting information in a just-in-time manner might be the strategy of choice for the majority of people, but also in these situations individual differences will likely be present.

To take on a perspective on visual working memory that is not focused on how much information people can retain, but rather how one uses working memory to interact with external information, we should consider both situational demands and, specifically, individual approaches to these situational demands.

Our primary aim was to characterize individual sampling versus loading strategies and their effects on visual working memory task performance. We used secondary data that were collected during a copying task with changing task demands (Sahakian et al., 2023) that conceptually replicated previous copying task studies. Supplemental to previous studies that aimed to engage working memory in more natural environment, the present study attempts to identify underlying strategies that were adopted by the individual. Participants were instructed to rebuild an example puzzle as fast and accurately as possible in a condition in which the example was easily accessible versus a condition in which it was more time-consuming to access the example. Other than in traditional working memory paradigms, this Copy Task did not force using one's maximum capacity in order to successfully complete the task (e.g., memorizing as much as possible at once; full-loading), but rather allowed the individual to opt for their preferred load. Additional to the group level analysis that was done before (Sahakian et al., 2023), we here attribute meaning to individual behaviour and describe varying strategies. With the assumption that people differ in their natural inclination to either sample or store information when no strategies are forced, we expected some participants to heavily rely on the external world (i.e., to sample often; low-loaders), and some to rely more on their internal memory (medium to high-loaders). Furthermore, when information was not readily available anymore, we expected that people would adjust their strategy to the new situation in a way specific to the individual. Potentially, some people might stick to sampling (low-loading), while others start memorizing more (medium to high-loading). It is possible that behavioural adjustments to changing situational demands may come at a cost of task performance: we hypothesize that larger changes in sampling behaviour come with switch costs (thus, decreased performance) as one needs to discontinue their current behaviour and adjust to the new situation. To assess this idea, we analyse whether initial sampling frequency influences performance, and whether changes in sampling frequency across conditions influence performance across conditions.

As a secondary aim, we considered how the order of the encountered situational demands (thus, prior experience) plays a role in opting for a specific strategy. When people start with a situation in which information remains continuously available, they might adopt a strategy of relying on the external world without memorizing much. When people are initially confronted with a situation in which information can only be sampled after a waiting time, they might adopt a strategy that is biased towards memorization instead of sampling. The context of the first situation then primes people to keep using their initial strategy even if the context changes: an earlier study showed that using a specific source to support memory (here, the internet) predicted future reliance on the same source (Storm et al., 2017). Vice versa, Patrick and colleagues (2015) found, in a task similar to ours, that exposure to only one trial in which information was hard to retrieve already was sufficient to prime people towards a memorizing strategy. This implies that people's perseverance with previous behaviour may result in carry-over effects that are likely to influence the individual's strategy of choice. However, Risko & Dunn (2015) did not find order effects on the decision to offload. Therefore, we took a closer look at whether and how the order of conditions (i.e., experience) in the current copy task primed individuals' strategy initiation and continuation.

Revealing individual differences in working memory strategies in response to changing situational demands contributes to our knowledge of working memory functioning as a dynamic system rather than a fixed entity that is always put to use in an identical manner.

2. Materials and methods

2.1. Participants

We used an existing open access dataset (Sahakian et al., 2023) for analysis. The authors recruited their participants through the platform Prolific. Participants could reside anywhere in the world, but had to be fluent in English, and were excluded if they participated in an earlier study of the same authors. 88 participants provided viable datasets for analyses. The majority of participants resided in Europe. Participants were only asked about gender and age categories, due to which demographic information is limited to proportions of categories. Of 88 participants, 53.4% were male, 45.5% were female, and 1.14% identified as non-binary. 62.5% of participants were in the age category 18 – 25 years old, 22.7% were 26 – 33 years old, 6.8% were 34 – 41 years old, and 8% were over 42 years old.

All participants gave written informed consent prior to the start of the online experiment. Participation was compensated with (the equivalent of) £6.25 upon completion of the task. The study was in accordance with the Declaration of Helsinki, and was approved by the Ethics Committee of the Faculty of Social and Behavioural Sciences of Utrecht University (21-0297).

2.2. Task

A previously developed copying task (Somai et al., 2020) was adapted such that it was suitable for online task administration (Sahakian et al., 2023). Results obtained from the online paradigm showed to yield qualitatively similar results as earlier conducted lab-based studies (Draschkow et al., 2021; Sahakian et al., 2023; Somai et al., 2020), confirming the reliability of the task. Different from the previously published group-level analyses and results, we used data derived from the online study to obtain information about individual differences in the allocation of external or internal memory resources across different situational demands. An extensive overview of the experiment, procedure and pre-processing of the data can be found in the original paper (Sahakian et al., 2023). Here, we provide an overview of relevant task features and variables.

The layout of the Copy Task is depicted in Figure 4.1A. Participants were instructed to copy 6 items from the 4 x 4 example grid (model; shown on the left side of the screen) to a 4 x 4 grid (workspace; located on the inner right side of the screen) as fast and accurately as possible. Items were dragged from the 4 x 4 resources grid that was located to the far right of the screen. Stimuli consisted of 20 polygons (Figure 4.1C; Arnoult, 1956), and 20 colours were added that were selected from the HSLuv (www.hsluv.org) colour space with 90% saturation and 65% luminance, resulting in 400 unique stimuli. For each trial in the experiment, a random selection *without* replacement of four shapes and four colours was used to create 16 unique stimuli. From these 16 stimuli, 6 were randomly selected *with* replacement, and randomly positioned in the *model* for each trial (Figure 4.1B; Sahakian et al., 2023). Whenever participants dragged an item over the workspace, the closest cell in the grid was highlighted in yellow. If the item was released at the correct cell, it would automatically align to the centre of that cell. Conversely, if the item was released at an incorrect location, it would return to its initial location in the resources. Note that stimuli remained available in the resource grid, even if they had already been correctly placed. As the experiment was conducted online, a cursor-directed aperture was incorporated in the experimental trials to extract sampling behaviour (Anwyl-Irvine et al., 2022). This technique involves covering the display with an opaque black overlay whilst leaving only a circular area around the cursor transparent, being just large enough to allow full visibility of the model at once (Figure 4.1B and Video at osf.io/w7zag). The transparency of the aperture followed a Gaussian function: it was fully transparent at the centre and gradually less transparent towards the edges.

The Copy Task consisted of two experimental conditions, each consisting of 24 trials (48 trials in total). In the low-cost condition, the moving speed of the aperture was smoothly aligned with the pace of the mouse movement. In the high-cost condition, the aperture moved with a reduced speed (approximately 1.67 seconds) when crossing the midline from the workspace to the model (across the dark grey border) making sampling more time-consuming. There was no delay when participants moved the cursor from model to workspace (see Video at osf.io/3z8xn). The conditions were blocked, as this allowed individuals to engage in a consistent strategy (Janssen & Gray, 2012; Patrick et al., 2015). The order of blocks was counterbalanced, which allowed dividing the groups based on the order in which they encountered the experimental conditions (i.e., situational demands). When the low-cost condition was followed by the high-cost condition, we referred to this order as ‘low-cost first. When the high-cost condition was completed first, and the low-cost condition thereafter, we referred to this order as ‘high-cost first.

In part of the trials (59.2%), participants were interrupted during a trial, and had to answer a two-alternative forced choice (2-AFC) question (“probe questions”). This study does not focus on 2-AFC data (the reader is referred to the original paper of Sahakian et al. (2023) for further details), but we consider the trials valid to be included in the current analysis.

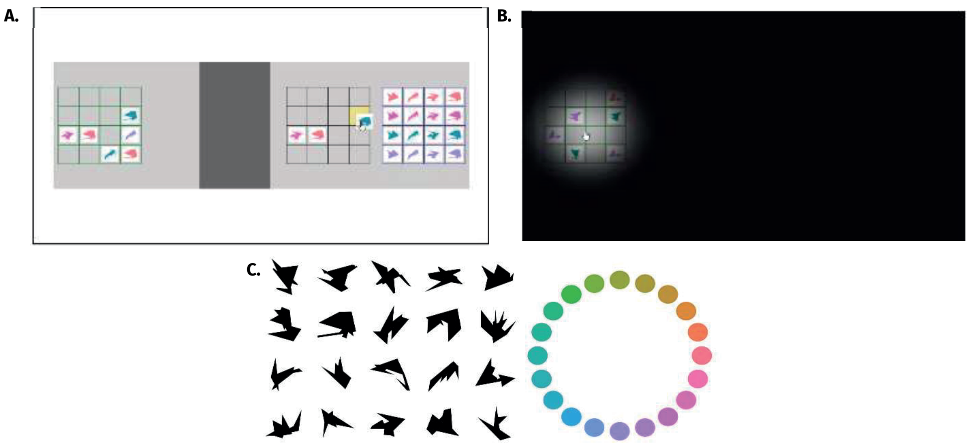


Figure 4.1. Experimental layout. **A)** illustrates the experimental layout, consisting of the 4 x 4 example grid (model) shown on the left side of the screen, and the 4 x 4 grid (workspace) depicted on the inner right side and the resources grid (Resources) depicted on the outer right side. **B)** shows a still from an experimental trial, as seen by the participants. A cursor-directed opaque black overlay covered the display as depicted in Panel A, allowing for full visibility of the Model at once. **C)** The stimuli set (20 polygons x 20 colors; 400 unique stimuli (Sahakian et al., 2023)). Figures adapted with permission.

2.3. Outcome measures

To describe individual differences in strategy adoption, we derived several outcome measures from the task. First, we computed how often a participant inspected the example model (i.e. crossed the cursor from left to right), averaged over trials, resulting in the mean number of model inspections. Second, we computed how long someone viewed (i.e., encoded) the example model per inspection. This was calculated by dividing the total duration of model inspections in seconds by the number of inspections per trial, and consecutively taking the median *dwell time at the model per inspection*. Conceptually, a sampling strategy translates to a high number of model inspections with shorter dwell times per inspection, whereas a memorization strategy translates to a low number of model inspections with longer dwell times per inspection. Therefore, participants were classified based upon the number of model inspections. These classification labels are solely introduced for descriptive purposes, and may serve as a rule of thumb to give some insight on the individual's general sampling tendency. Participants who made on average more than one inspection per correctly placed item were classified as low-loaders. For example, such participants may decide to first memorize the item, find the item in the resource grid, then reinspect the model to memorize the location, and only then place the item in the empty grid. Participants who correctly remembered one or more item(s) per inspection (i.e., made one or fewer than one inspection per correctly placed item) were classified as medium-loaders. Those who correctly remembered three or more items per inspection (i.e., made 0.33 inspections or fewer per correctly paced item) were classified as high-loaders (also see Analysis).

Participants were instructed to rebuild the example puzzle as quickly and accurately as possible. Although we thereby do not emphasize either outcome, we cannot guarantee that individuals have an equal attribution of importance for either speed or accuracy. This leaves room for individual differences in motivation to either perform the task without errors, or 'quick but dirty'. To control for potential individual differences in motivation and to deal with the presence of a speed-accuracy trade-off, performance was assessed by calculating a *linear integrated speed-accuracy score* (LISAS; Vandierendonck, n.d., 2017, 2021), depicted in Equation 1.

$$LISAS = RTij + \ PEij \ x \frac{S_{RT}}{S_{PE}} \qquad (1)$$

Here, RT_{ij} refers to the completion time on trial *i* divided by the number of correct placements on trial *i* for individual *j*. In the high-cost condition, the aperture delay was first subtracted from the completion time. The speed data was log transformed to account for skewness associated with time measures. PE_{ij} refers to the proportion error on trial *i* (1 minus the number of correct placements divided by the total attempts). We calculated the standard deviations S_{RT} and S_{PE} per individual over both conditions collapsed (Vandierendonck, n.d., 2017, 2021). A lower LISAS reflects better performance.

To investigate the degree to which each participant adapted their sampling strategy between conditions, we divided the number of inspections in the high-cost condition by those in the low-cost condition (‘change factor number of model inspections’). The higher the change factor number of model inspections, the larger the adaptation from the low-cost condition to the high-cost condition. The same was done for the LISAS (‘change factor LISAS’), to see how performance changed for the individual. A higher change factor LISAS reflected a larger drop in performance from the low-cost to the high-cost condition.

2.4. Pre-processing

The data were retrieved from https://osf.io/pkxdc (Sahakian et al., 2023). In the current study, we used the same filter and exclusion criteria as in the study of Sahakian et al. (2023); no additional participants were excluded from the analysis. The trials with probe questions (n = 1569, 59.2%) and without probe questions did not differ regarding the number of model inspections, dwell times per inspection, and the number of errors (Sahakian et al., 2023). Therefore, trials with and without probe questions were treated similarly. The main analyses were performed on all data (n = 4219 trials). The level of significance was set to an alpha of 0.05. To make sure that findings were not driven by outliers, we ran sensitivity analyses after removing those trials with scores ≥1.5 times the interquartile range apart from the group median for that specific outcome measure (i.e., number of model inspections or LISAS) in that specific condition (i.e., low-cost or high-cost) per order of condition (i.e., low-cost first, high-cost first). Information on outlier removal and sensitivity analyses can be found in the Supplementary Results Chapter 4.

2.5. Analysis

2.5.1 Individual differences in strategy and effects of situational demands and prior experience. Statistical analyses were performed in R 4.1.2 (R Core Team, 2021).

First, we provided the number of model inspections for each condition (low-cost, high-cost) and both presentation orders (low-cost first, high-cost first), and we ran a non-parametric Kendall Rank correlation between the number of model inspections and dwell time at the model per inspection. We expected that a higher number of model inspections would relate to shorter dwell times per inspection, and vice versa. Fewer inspections with longer dwell times (i.e., encoding) would reflect a tendency towards memorization. Correlation coefficients were reported as tau (τ) and effect sizes as z.

To characterize the natural tendencies of individuals to either rely on external sampling or internal memorizing, we classified individuals as low-loaders, medium-loaders, or high-loaders based on how often they inspected the model to place one item correctly. We extracted the number of model inspections per trial (e.g., the number of times a participant moved their cursor to reveal the model) and divided this number by six (each trial had six items to copy). The choice of cut-offs for the different categories were partly data-driven (i.e., aiming for a substantial number of participants per category), partly theory-driven, and mostly based upon task constraints. Individuals who inspected the model more than once per correctly placed item were categorized as low-loaders. These participants were considered to employ the bare minimum working memory load in the Copy Task. To illustrate, participants may have memorized the polygon shape first, and may have memorized the location upon reinspection, thus memorizing the item in a feature-by-feature manner. Individuals who inspected the model once or less than once per correctly placed item, and thus memorized one bound (shape plus location) item or more items per inspection, were categorized as medium-loaders. Individuals who correctly placed three or more items per inspection were further classified as high-loaders. Our task did not have a high enough resolution to dissociate between people loading three, four or five items, as all of them would need an additional inspection for the remaining items, yielding two inspections for trial completion. Importantly, it has been claimed that people have an estimated working memory capacity of four items (Cowan, 2001), but for more complex shapes such as polygons, this maximum capacity seems to be decreased (Alvarez & Cavanagh, 2004; Luria et al., 2010; Luria & Vogel, 2011). We therefore cannot make conclusive statements but only speculate about whether or not individuals are fully loading their capacity. Importantly, these cut-offs are arbitrary, and were solely used to characterize individual strategies used in the current task.

First, we described the strategies used in the condition in which information was freely available, as this context resembles our daily life environment the most. We did this for the participants who started with this condition (i.e., low-cost condition first), as participants were not yet influenced by the situation where information availability was manipulated (high-cost condition). Next, we investigated whether in these participants, varying the situational demands led to an adaptation in sampling behaviour (e.g., relying more on the outside world or on memorizing). Therefore, we described whether participants used different strategies in this condition, to what extent they changed their strategy from the low-cost to the high-cost condition, and whether the previously established categories were still observable.

We also explored whether and how prior experience affected strategy choice. To this end, we explored data of the group that started with the high-cost condition to investigate whether introducing higher sampling costs at the outset of the task prompted participants to adopt a strategy biased towards heavier reliance on memory, and whether they stuck to this behaviour also when there was only a low cost to sampling. We assessed whether our previously introduced categories (low-loaders, medium-loaders and high-loaders) were still observable when the order of conditions was reversed.

In addition, we investigated the influence of the order of encountered situational demands *between* participants. To this end, all trials were fed to a linear mixed-effect model (LMM; Singmann & Kellen, 2019) by using the lmer function in R (lme4 package; Bates et al., 2014). The LMM is robust against deviations from normality of the outcome variables and takes individual differences within groups into account (Schielzeth et al., 2020). Factors included were order (low-cost first, high-cost first), condition (low-cost, high-cost), the interaction of order and condition, and random slope and intercept for individuals. We ran the model to predict the influence of these factors on the *number of model inspections.* The normality of the residuals was visually examined and confirmed. Effect sizes were reported as beta-coefficients with a 95% confidence interval. Post hoc pairwise comparisons were conducted using the Tukey-Kramer method using the contrast function within the emmeans package, which accounts for multiple comparisons and controls the family-wise error rate.

2.5.2. The effect of strategy on performance. The second aim was to study how a chosen sampling strategy influenced performance. We first investigated the group that started with the low-cost condition, as we were mostly interested in initial sampling preference in a situation where information was freely available. We checked whether there was a favourable strategy category in terms of performance (reported in the Supplementary Results Chapter 4), and we used a non-parametric Kendall Rank correlation analysis to investigate whether and how initial sampling preference (number of model inspections in the low-cost condition) related to performance in the low-cost condition. Then, we evaluated whether individuals who showed larger changes in sampling frequency when moving to the high-cost condition (e.g., from low memory reliance in the low-cost condition to a higher memory load in the high-cost condition) showed more decline in performance compared to those who were more stable in strategy, or vice versa. We assessed the relation between the change factor number of model inspections and *change factor LISAS* with a Kendall Rank Correlation analysis. To more broadly assess the influence of order and condition on performance, we also fitted a linear mixed-effects model to LISAS.

3. Results

3.1. Conceptualization of memorization

We found that there was a moderate negative correlation between the number of model inspections and dwell time per inspection (τ = -.55, p < .001, z = -10.76; Supplementary Figure S4.1), indicating that fewer model inspections were related to longer inspection durations. With the current number of observations (176), for a one-tailed correlation test (α = .05) with a power of 0.8, we should be able to reliably detect effects sizes of 0.18 (Faul et al., 2009).

3.2. Individual differences in strategy and the effects of situational demands and order

Within the group of participants that started with the low-cost condition (n=43), 17 participants (39.5%) were classified as low-loaders as they inspected the model more than once to place one item correctly (Table 4.1; Figure 4.2A, black dots). Twenty-four participants (55.8%) were categorized as medium-loaders, as they made one or fewer than one inspection to place one item correctly, implying that they relied relatively more on memorization. Importantly, only two participants (4.7%) placed three or more items correctly per inspection, thereby classifying as high-loaders. These results indicate that, generally, participants relied on the external world and used working memory capacity only to a limited extent. When moving to the high-cost condition, all participants (100%) changed their strategy in the sense that they made fewer inspections as compared to in the low-cost condition. Now, only one participant (2.3%) was classified as a low-loader, against 69.8% medium-loaders and 27.9% high-loaders (Table 4.1; Figure 4.2A, red dots). This shows that when information was less readily available, participants tended to rely less on the external world and more on their internal memory capacity. However, the majority of people still did not employ a high-loading strategy.

Next, we investigated whether the presentation order of conditions affected the strategy that was used. In other words, did the context of the first condition prime the strategy that was used in the second condition? We here included the group of participants who started with the high-cost condition and completed the low-cost condition afterwards (n=45). Interestingly, none (0%) of the 45 participants were categorized as low-loader in the high-cost condition, 73.3% was categorized as medium-loader and 26.7% as high-loader (Table 4.1; Figure 4.2B, red dots). In the following low-cost condition, 3 participants (6.7%) shifted to a low-loading strategy (Table 4.1; Figure 4.2B, black dots). A medium-loading strategy was used by 35 participants (77.8%), and 7 participants (15.5%) held on to the high-loading strategy. This means that the majority of participants was consistent in their strategy, even in a context in which information in the external world became freely available. Thus, strategy is not only dependent on the current context, but also seems influenced by prior experience.

Table 4.1. *Mean number of model inspections per order (low-cost first, high-cost first), per condition (low-cost, high-cost) and per strategy category (low-loader, medium-loader, high-loader). We provide the number of individuals (n) and percentages per category based upon the number of model inspections, and the mean (M), standard deviation (SD) and the range of the number of model inspections per trial. Note that the division in categories is arbitrary and is for descriptive purposes only.*

Condition	**Strategy category**	**Order**							
		Low-cost first (n =43)				**High-cost first (n = 45)**			
		n(%)	M	SD	Range	n(%)	M	SD	Range
low-cost	low-loader	17 (39.5%)	7.69	1.43	6.12 – 10.5	3 (6.7%)	7.26	0.17	7.21 – 7.46
	medium-loader	24 (55.8%)	4.89	1.03	2.17 – 6	35 (77.8%)	4.02	1.12	2.08 – 6
	high-loader	2 (4.7%)	1.65	0.21	1.5 – 1.79	7 (15.5%)	1.60	0.38	1.04 – 2
high-cost	low-loader	1 (2.3%)	7	-	-	0 (0%)	-	-	-
	medium-loader	30 (69.8%)	2.86	0.53	2.17 – 4.58	33 (73.3%)	3.25	0.97	2.04 – 5.83
	high-loader	12 (27.9%)	1.39	0.38	1 - 2	12 (26.7%)	1.46	0.34	1.04 – 2

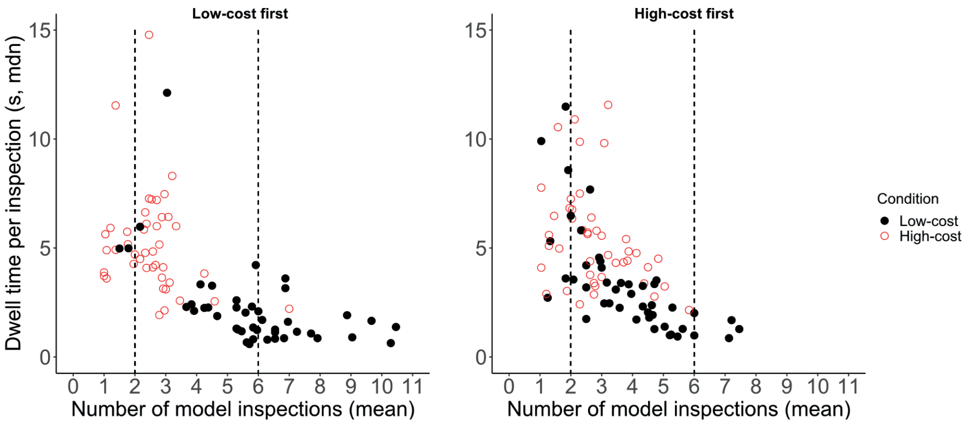


Figure 4.2. Copying behavior, presented as median model dwell time per inspection per trial and the average number of inspections per trial. **A)** Low-cost first. **B)** High-cost first. Data points represent data of the individual in the low-cost condition (filled black dots) and high-cost condition (open red dots). The vertical dashed lines represent the cut-offs used to discriminate between low-loaders (dots to the right of the right dashed line), medium-loaders (dots on and to the left of the right dashed line), and high-loaders (dots on and to the left of the left dashed line).

To further investigate the effects of situational demands and the order they are encountered in on the number of model inspections, we fitted a linear mixed-effect model (LMM) to predict the number of model inspections with these factors, while controlling for individual differences (Figure 4.3). The LMM showed a significant effect of condition (t = -14.22, p <.001, beta = −3.31 [-3.76, -2.85]), a significant effect of order (t = -5.11, p <.001, beta = −1.99 [-2.76, -1.23]), and a significant interaction effect between condition and order (t = 6.83, p <.001, beta = 2.22 [1.58, 2.86]). Importantly, this interaction effect showed that prior experience (order) differentially affected sampling behaviour across conditions. Post-hoc tests revealed that the group starting with the low-cost condition made significantly more model inspections in the low-cost condition (M = 5.85, SD =2.03) than the group that started with the high-cost condition (M = 3.86, SD = 1.62, p <.001). In the high-cost condition, the groups showed about the same number of model inspections (low-cost first: M = 2.54, SD = 1.08; high-cost first: M = 2.77, SD = 1.1, p =.77). Together, this implies that starting with the high-cost condition primed participants to keep using internal storage to a larger extent, also if the situation did not demand this per se.

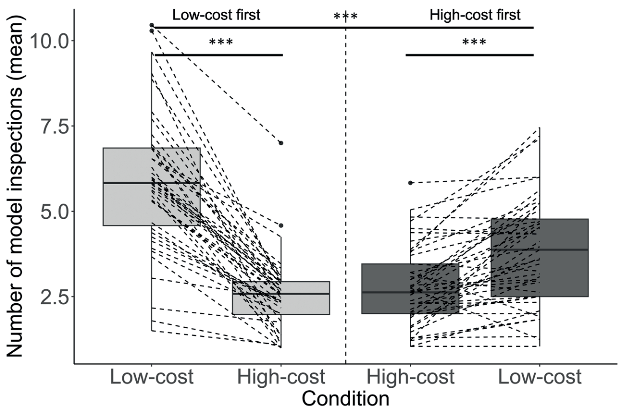


Figure 4.3. Data, presented as mean (+IQR) inspections per condition (low-cost, high-cost) for the different orders (low-cost first, high-cost first). There were more model inspections in the low-cost condition compared to the high-cost condition. Participants that started with the high-cost condition (dark grey) sampled less often than the group that started with the low-cost condition (light grey). The interaction-effect revealed that the low-cost condition differentially affected sampling behavior across groups, with the high-cost first group making significantly fewer model inspections. *p < .05 **p < .01, ***p < .001

3.3 The effect of strategy on performance

Next, we evaluated whether there was a relation between strategy and performance. For participants that started with the low-cost condition (n=43), there were no differences in performance across low-loader, medium-loader, and high-loader categories in the low-cost condition (see Supplementary Figure S4.2). When analysing continuous data from the same group, we found no significant correlation between the number of model inspections and LISAS in the low-cost condition (τ=.006, p=.96; Figure 4.4A). So, when information was freely available, there was no favourable inspection frequency yielding better performance. However, when looking at the transition from the low-cost to the high-cost condition, some people made a greater behavioural adaptation compared to others (Figure 4.4B, length of dashed lines). Therefore, we investigated whether the *magnitude of behavioural adaptation* (change factor number of model inspections) was related to performance within the individual (Figure 4.4C). For the low-cost first group, no significant correlation was found between the change in inspection frequency and the change in performance (τ=-.04, p=.69), meaning that performing the task in a non-preferred approach (hence, larger adaptations; high change factor) when moving to the high-cost condition did not affect performance more than for those who behaved more consistently across conditions.

We are aware that this analysis is underpowered (to be able to reliably detect a moderate effect size of 0.3 with a power of 0.8, we should have 64 observations instead of 43). However, due to the nature of our data we cannot treat all observations similarly. Collapsing the data across the low- and high-cost first group would bias results as their respective change factors indicate another transition (from low- to high-cost, or vice versa) and thus have different meaning. We have therefore chosen to do this analysis only for the low-cost first group as we assume this to be the most natural way of encountering information in everyday life.

However, we did want to assess the influence of the order of conditions on performance, and therefore investigated whether participants who started with the high-cost condition – and thus showed different sampling behaviour, as found in section 3.2 – performed differently than participants who started with the low-cost condition. Therefore, we decided to fit a linear mixed-effect model to LISAS to simultaneously analyse the influence of order and condition. First, a significant main effect of condition (t = 7.59, p < .001, beta = 0.37 [0.27, 0.46]) was found, indicating overall worse performance in the high-cost condition compared to the low-cost. No significant main effect of order (t = 0.4, p = .693, beta = 0.03 [-0.13, 0.19]), nor an interaction effect between order and condition (t = 0.2, p = .841, beta = 0.014 [-0.12, 0.15]) was present. Also see Supplementary Figure S4.3 and Table S4.1.

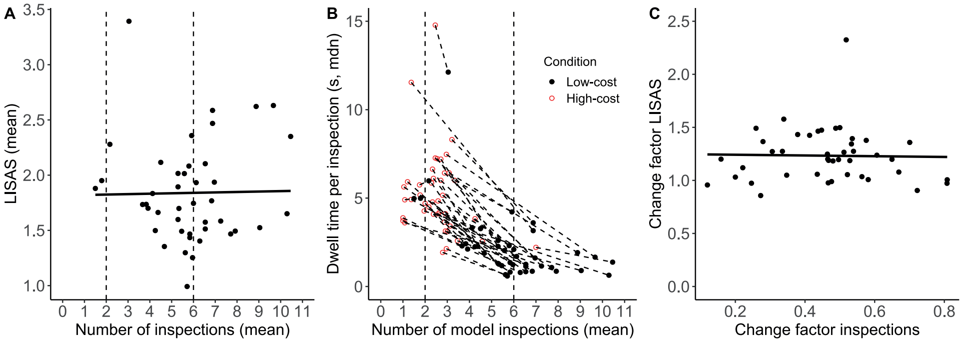


Figure 4.4. Effects of sampling behavior on performance, behavioral strategy adaptations and effects of strategy adaptations on changes in performance for the low-cost first order group only. Each point reflects an individual participant (n=43). **A)** In the low-cost condition, the mean number of inspections is not correlated to performance (mean LISAS; higher scores reflecting worse performance). The vertical dashed lines represent the cut-offs used to discriminate between low-loaders (dots to the right of the right dashed line), medium-loaders (dots on and to the left of the right dashed line), and high-loaders (dots on and to the left of the left dashed line). **B)** Representation of the behavioral shift based on mean number of model inspections per trial between the low-cost condition (closed black dots) and the high-cost condition (open red dots). The longer the line between two data points, the larger the behavioral adjustment for the individual. **C)** No significant correlation was present for change factor number of model inspections and change factor LISAS. Scores towards zero on change factor number of model inspections reflect a larger adjustment from the low-cost condition to the high-cost condition. The higher the change factor LISAS, the more performance declined in the high-cost condition compared to the low-cost condition. Change factor LISAS <1 indicates performance improvement.

To summarize the results on sampling behaviour and performance, both information availability (low-cost versus high-cost) and previous experience (low-cost first versus high-cost first) influenced sampling behaviour. There was no specific inspection frequency yielding better performance, and the magnitude of change in sampling behaviour did not affect performance within the individual. Note that including trial number in any of the reported LMM yielded qualitatively similar results, showing that when general practice effects were taken into account, the same conclusions could be drawn.

3.4. Sensitivity analyses

We detected and removed all outlier trials, and ran all analyses again. Detailed results can be found in the Supplementary Results Chapter 4 (Figures S4.4 – S4.8, Tables S4.2 – S4.3). Descriptively, some classifications changed (compare Table 4.1 and Supplementary Table S4.2). Effectively, there were two additional participants classified as high-loader (1 in low-cost condition, low-cost first; 1 in low-cost condition, high-cost first). Three participants were initially classified as low-loader, but were now classified as medium-loader (all four in low-cost condition, high-cost first). Apart from these deviations in classification numbers, outlier removal did not yield different results when rerunning the analyses, and our interpretations again held when testing for general practice effects.

4. Discussion

Although estimations of maximum capacity have proven fruitful in visual working memory research (Luck & Vogel, 2013; Ma et al., 2014), this approach falls short in grasping the usage of working memory in everyday life. Not only do capacity tasks demand full use of memory capacity, which is often not required in everyday situations, they also fail to reveal individual differences in the use of strategies that might be employed to support memory functioning in different situations. Rather than considering working memory as a system that is defined by its maximum capacity, we here addressed visual working memory as a dynamic system that is used flexibly in response to changing situational demands, *and* is subject to individual differences in how one deals with these demands. Departing from traditional paradigms that force the use of full working memory capacity to successfully complete the task, we used pre-existing data from a paradigm (Sahakian et al., 2023) that allowed individuals to store information at the preferred load or otherwise to rely on (i.e., sample from) the outside world. Their results conceptually replicate previous group-level results in that people tend to make little use of visual working memory capacity when the situation allows for rapid sampling from the external world (Ballard et al., 1995; Böing et al., 2023; Draschkow et al., 2021; Fu & Gray, 2000; Somai et al., 2020). Reliance on sampling only decreased when information became less readily available (see Sahakian et al. (2023) for this group level analysis). While the tendency to decrease sampling in response to restricted information availability has been repeatedly observed, individual differences in (sampling) strategies have largely been ignored (Logie, 2023). In the current study, we partly look at aggregated strategy data, but place emphasis on individual differences.

In line with the sparse literature on the existence of individual differences in strategy use within *capacity* tasks (Morrison et al., 2016; Morrison & Richmond, 2020), we similarly identified individual differences in the extent to which people relied on external sampling across changing task demands. When initial information was continuously available – which often is the case in everyday tasks – more than a third of people sampled more than once to correctly place one item (i.e., *resampling*). These low-loaders had a default setting to heavily rely on the outside world. Others used the external world to a lesser extent and loaded more information per iteration. Although individuals thus differed in the degree to which they relied on external sampling versus internal loading (low-loaders versus medium-loaders and high-loaders), very few participants showed infrequent sampling behaviour to a degree that they approached the limits of their working memory capacity (high-loading). Given that people *should* theoretically be able to load up three to four items in visual working memory (Cowan, 2016; Luck & Vogel, 2013), they *should* be able to place all six items with merely two inspections. Interestingly, when given the choice, people rarely do so: the majority of people inspects the model more often than strictly necessary, and few people load up three or more items at once.

When information availability was restricted, behaviour became more homogeneous. Individuals who initially *re*sampled reduced their sampling rate to such an extent that they would no longer be characterized as low-loaders. While *all* individuals reduced their sampling frequency (aligning with earlier group-level analyses), some individuals adapted more than others. Some of them even used only one or two inspections to complete a trial, meaning that they turned towards higher memory loads (three or more items). Still, even with changed task demands, the majority were reluctant to internalize information at high-load in working memory.

Interestingly, those who made greater adjustments in their sampling frequency from the low-cost to the high-cost condition did not perform worse than those who behaved more consistent across situations, suggesting that behavioural adaptations can be successfully incorporated by the individual. In other words, when information was freely available, there was no strategy that yielded better performance, and the magnitude of change in sampling frequency across conditions did not affect performance within the individual. Although we expected larger adaptations to negatively affect performance, the results imply that individuals can flexibly adjust their use of visual working memory to fit situational demands, even if this means that they approach the task in a way that does not match their initial tendency.

Notably, we found that the *order* in which one experienced changing situational demands differentially affected sampling behaviour: individuals who started with a task in which information availability was restricted, kept loading up more information during each inspection in a subsequent task in which information was freely available. This finding replicates a study of Patrick et al. (2015) that showed that starting with restricted access led to adhering to a memory-based strategy even when the restriction was lifted, whereas there was no adherence to a sampling strategy after the transition from non-restricted to restricted access. Given that some of the individuals in the high-cost first group would probably have shown low-loading behaviour if the order had been reversed (as we infer from the current study and Patrick et al., 2015), participants in the high-cost first group may have learned that they could successfully rely on their working memory. This finding is reminiscent of previous work showing that people who had a successful memory usage experience were less inclined to externalize information as compared to people who had a less (or no) successful experience (Gray & Fu, 2004; Risko & Dunn, 2015). In these cases of changing task demands, the costs associated with strategy switching may be higher than the effort associated with (the continuation of) memorization (Gilbert, 2023; Kurzban et al., 2013; Xie & Zhang, 2023). Furthermore, for this high-cost first group, the *experienced* effort of using memory at higher load may have been lower than the *anticipated* effort that the other (low-cost first) group expected to encounter when using memory at higher load (Bambrah et al., 2019). Experience and expectation therefore seem to jointly explain why order influences the choice to either offload or not. Although these factors drove behaviour in the high-cost first group towards a larger tendency to memorize, only some individuals completed trials with only one or two inspections (thus loading three or more items). The majority of participants did not load working memory capacity to such a high load, and relied on the outside world more than necessary assuming a capacity of three to four items (Cowan, 2016; Luck & Vogel, 2013). Under this assumption, we conclude that deploying visual working memory at high-load more likely is a matter of willingness rather than ability. However, we acknowledge that more complex shapes such as polygons may give rise to a different (i.e., lower) maximum capacity value (Alvarez & Cavanagh, 2004; Luria et al., 2010; Luria & Vogel, 2011). On the other hand, this capacity may be increased again when the participant could draw from semantic representations in long-term memory to verbally label the polygons (Chung et al., 2024). For example, some participants may have recognized some of the polygons as a plane or a star, thereby using a verbal strategy to support visual working memory. We therefore cannot make conclusive statements but only speculate about the extent to which individuals were making use of the low-, medium-, or high-end spectrum of their maximum capacity for the used stimuli.

Related to the previous point, we emphasize that our category conceptualization is subject to debate. One should keep in mind that our definitions of low-, medium-, and high-loader may not cover the exact underlying mechanism. We interpret our results under the assumptions that low-loaders are individuals who, in general, heavily rely on (i.e., (re)sample from) the outside world and do not load much information into working memory. However, individuals may be classified as low-loader because they inspect external information relatively often, but that does not exclude the possibility that they *may* have loaded their memory fully, but just checked themselves multiple times before they acted upon that information.

In a similar fashion, *within* trials, people can be classified differently at different points in time. Consider, for example, an action sequence in which a participant loads up three items on the first inspection, places all of them (high-loader behaviour), and then inspects and reinspects the model multiple times (low-loader behaviour) to the extent that this individual would be classified as a low-loader. The same may occur *between* trials. Further, it should be noted that some high-loaders exhibited only one inspection, implying they memorized all six items at once, while this single inspection was accompanied with a relatively short inspection time. It is certainly possible that these participants had very efficient memorization, but these data points could also reflect cheating behaviour (e.g., taking a picture of the model and using that picture to avoid inspecting the model in the online paradigm). In other instances, individuals dwelled relatively long at the model, which may reflect distractions or lapses during the task that we cannot rule out. We are aware of the ambiguities in our definition and acknowledge it's shortcomings, but are confident that our clustering approach still captures the individuals' *general* tendency. Similarly, although the relation between the number of inspections and dwell time may differ between participants, it does reflect the overall relationship.

We propose that the tendency to employ working memory at lower load is driven by effort expenditure. As information storage in visual working memory has been described as fragile (Cowan, 2001; Ricker & Cowan, 2010; Zhang & Luck, 2009) and effortful (Kardan et al., 2020; Xie & Zhang, 2023), people will likely try to circumvent the effort or uncertainty associated with maintaining multiple items. Indeed, it has been found that offloading behaviour – of which sampling can be seen as a subordinate – occurs as a result of subjective effort reduction (Risko & Dunn, 2015). Our findings are consistent with the literature describing the constant weighting of effort input and performance output at both the psychological and neurophysiological level (Kurzban et al., 2013). As there were no direct benefits (no specific rewards or importance) for the participant to fully tax memory, the decision to not memorize to high load seems a logical one. The brain is often described as a system that pursues optimal efficiency, and this leads us to speculate about optimality. Is it a matter of optimal performance or a question of optimal resource allocation? We dub a 'battle of the trade-offs': in terms of resources, there is a continuous weighting to reach an optimal equilibrium between sampling or storing expenditure, but in terms of performance, there is also an optimal balance in the speed-accuracy trade-off. These two scales operate in synchrony and are constantly pushing and pulling their respective weights. Here, reluctance to commit to loading may be an important drive to opt for low-loading (effort trade-off), especially since there does not appear to be a performance benefit for either low-, medium- or high-loading (performance trade-off).

The elegance – but also the complexity – of our paradigm is exactly that one can always freely opt for whichever strategy is most appealing. This also means that we approach strategies from a volitional point of view, and do not explicitly instruct our participants to adopt either low-, medium- or high-loading. We intentionally did not do this, as this would align more with traditional capacity tasks where a particular strategy is enforced to successfully complete the task (i.e., adhering to the instructions). While this volitional approach shows natural tendencies in engaging visual working memory and shows not to affect performance, it is possible that explicitly *instructing* a strategy might yield different performance outcomes, and that there might actually be performance benefits (or disadvantages) if we would instruct people to make more extreme changes or show a more consistent strategy.

Here, the only instruction was to complete the task as quickly and accurately as possible. Performance was therefore calculated as a linear integrated speed-accuracy trade-off score (LISAS; Vandierendonck, 2017). This measure corrects reaction time for the number of errors that were made, and therefore considers both speed and accuracy as equally favourable outcomes. Consequently, effects on either speed or accuracy of any of the strategies (one being faster but the other being more accurate) are balanced out with this measure, and do not yield a ‘better' strategy. However, in everyday settings – or due to participants' convictions about what constitutes good performance – accuracy may be favoured over speed. In such cases, it may be beneficial to *re*sample to safeguard an activated representation of the information to-be-used, even if this takes up additional time. Ensuring that the representation of memory content matches the target item by resampling then strengthens the decision to act upon the internal representation and place the item (e.g., reaching an action threshold; Sahakian et al., 2023). Potentially, low-loaders simply need more certainty to reach their action threshold as compared to medium- or high-loaders. A task-specific factor that may have influenced the perceived unequal importance of speed or accuracy, is that the paradigm incorporated 2-AFC probes to assess participants on any residual memory traces that were or were not put to use (Sahakian et al., 2023). Being aware of the possibility to be examined, participants may have placed more weight on accuracy than on speed. Unfortunately, we have no insight into participants' motivation to adhere to the task instructions, as the task was carried out online. If accuracy was deemed more important than speed, we hypothesize that frequent (re)inspecting would occur in order to check oneself, even when sampling costs were imposed. Future research should point out how changing the speed-accuracy scale (e.g., by placing more weight on either of the two outcomes) affects loading behaviour. Instructions and personal motivation should be separated in order to gain a better understanding of the variables involved.

Our study provides insight into how individuals allocate visual working memory resources to different situational demands when given the freedom of choice. Future research could investigate what exact factors underly an individual's allocation of external versus internal resources in visual working memory. Furthermore, it would be valuable to explore what additional factors might help understand this allocation question, such as stimulus familiarity and practice effects. Familiarity with specific objects could reduce effort and facilitate memorization (Blalock, 2015; Poppenk et al., 2010; Xie & Zhang, 2018), and thus could lead to reduced sampling frequency and/or improved performance *over the course of the task* (even so within blocks). Intriguingly, effects of training and experience on strategy continuation and adjustments seem to be time- and frequency dependent. For instance, Patrick et al. (2015) observed that one 'reversal' trial provoking memorization was already sufficient for people to opt for and continue with a memorization strategy for at least ten consecutive trials, even though the individuals were initially trained to use an external sampling strategy. Extending this research in an intention offloading task, Scarampi & Gilbert (2020) assessed the time-course of choice for a certain strategy after primed strategies and reversal trials. They found that a reversal trial provoked an immediate adaptation to that strategy, but that this effect wore off after a number of trials, after which people turned to the primed strategy again. To get a broad idea of behaviour and performance over the course of the current copying task, we have visualized the number of model inspections and performance as a function of trial number in Supplementary Figure S4.8 and S4.9. Similar to adaptation effects, more general task practice effects could have led to altered behaviour towards the end of the block or task. When checking for this, we indeed confirmed that practice effects influenced sampling frequency and performance, but not to the extent that the other factors were not meaningful anymore. Yet, potential detrimental effects of strategy adaptation on performance (as the individual moves from one condition to another) are not as clear-cut across conditions, and could be overshadowed by improved performance due to these learning and habituation effects.

Finally, it would be interesting to assess under what circumstances someone would switch from being a medium-loader to a high-loader, or even a full-loader, by incorporating a continuum of changing task demands rather than only two discrete conditions. Such a study could reveal a 'tipping point' for strategy adaptation. Findings regarding the linear relationship between (working) memory capacity scores and offloading behaviour reveal multifaceted contributions of memory subsystems: some studies find an effect of capacity while others do not, or only partly (in healthy subjects: Meyerhoff et al., 2021; Morrison & Richmond, 2020; Risko & Dunn, 2015; in patients: Böing et al., 2023, 2025). Differences between studies can partly be attributed to different capacity tasks used to estimate capacity. It would be interesting to see which of these findings could be replicated when the capacity scores are not derived from traditional neuropsychological tasks, but directly reflect the capacity for the stimuli used in the paradigm at hand (in the current study, those of Arnoult, 1956). Other factors, such as personality traits or neuropsychiatric tendencies (e.g. compulsive checking in obsessive-compulsive disorder) could also be of influence.

The current research only covers a small part of the strategy palette: strategically using the outside world to aid visual working memory task performance is one option, but there are many other strategies (see Gonthier (2021) for a review) that may in itself be strategically employed during our Copy Task. To illustrate, it is possible that individuals do not increase the mere *visual* representation load when the sampling cost is elevated – as implied in this study – but that the individual more actively engages in verbal recoding, thereby expanding the capacity that could be loaded (Chung et al., 2024). Although such strategy usage may have occurred within our participants, individuals still had the option to not use any of those and instead rely on the external world to circumvent (the effort associated with) internal strategies and high-loading. Thus, irrespective of whether or not participants used any of the internal strategies, we find that people tend to heavily rely on the external world.

In summary, while many factors complicate the attribution of *why* a sampling strategy occurs, having an eye for strategy use can still be insightful in seeing *how* one uses working memory. Recognizing that each individual has their own preferred approach may help future researchers and clinicians to understand the complex dynamic nature of visual working memory use outside of the lab or clinic.

5. Conclusion

Visual working memory use is clearly not solely determined by an individual's visual working memory capacity. Individuals tend to rely on the outside world more than strictly necessary given capacity, and they flexibly adapt this degree of reliance to changing situational demands. We identified low-loaders, medium-loaders, and high-loaders, and although we could distinguish these individual differences in reliance on the outside world, the majority of people is – and remains – reluctant to approach the higher end of memory capacity use. We suggest that this individual variation is the result of an ongoing weighting of resource allocation (the effort of sampling vs. storing) relative to optimizing performance (speed vs. accuracy). Prior experience, underlying personal characteristics (e.g., motivation or confidence) and the recruitment of other strategies can in turn influence these trade-offs. We conclude that visual working memory is an adaptive system that is employed based on situational demands, effort and performance expenditure, and underlying individual tendencies.

Acknowledgements: We thank Andre Sahakian and colleagues for collecting and providing their dataset on: https://osf.io/pkxdc/ (Sahakian et al., 2023).

Author Contributions: Sanne Böing: conceptualization, formal analysis, methodology, validation, visualization, writing – original draft, writing – review & editing. **Beleke de Zwart:** conceptualization, data curation, formal analysis, methodology, visualization, writing – original draft. **Antonia F. Ten Brink:** conceptualization, supervision, writing – review & editing. **Tanja C.W. Nijboer:** conceptualization, supervision. **Stefan Van der Stigchel:** conceptualization, funding acquisition, project administration, supervision, writing – review & editing

Funding: This project was supported by an ERC Consolidator Grant (grant number ERC-CoG-863732) awarded to Stefan Van der Stigchel.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the Faculty Ethics Review Board of the Faculty of Social and Behavioural Sciences at Utrecht University. The approval is filed under number 21-0297.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent was obtained to publish this paper.

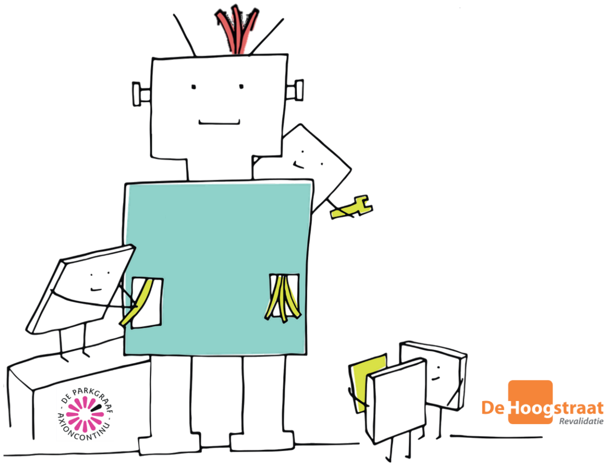
Data Availability Statement: The data presented in this study is made available in Open Science Framework at https://osf.io/pkxdc/ (Sahakian et al., 2023). Scripts used for the analyses will be made available upon publication at https://osf.io/cjgaf/.

Declaration of Interest: The authors declare no conflict of interest.

CHAPTER 5

Relying on the external world after stroke: Individual variability in compensation strategies in working memory use

Sanne Böing, Antonia F. Ten Brink, Tanja C.W. Nijboer, Stefan Van der Stigchel & Working Memory Study Group



Under review

Abstract

Capacity tasks are often used to detect abnormal working memory performance after stroke. However, in daily activities, patients may rely on the outside world by (re)inspecting information as needed (i.e., offloading), a strategy that is also advocated in memory rehabilitation. While individuals may use offloading in everyday life to support memory, capacity tasks do not allow for nor reflect this. To understand how individual stroke patients use their working memory when offloading is allowed, we recorded eye-movements of patients (n=15) and controls (n=38) as an index of offloading. Both patients and controls avoided working memory loading and relied heavily on offloading. Offloading strategies varied at the individual level, with a subset of patients showing excessive reliance on offloading. Interestingly, these patients were also those who showed abnormal capacity scores, but the reverse was not necessarily true. We conclude that low memory capacity is related to, but does not automatically lead to, offloading behaviour. Even when offloading was hampered, maintaining offloading was still more beneficial than switching to a memory-based strategy, supporting the adoption of external strategies in memory rehabilitation. The free-choice paradigm brings us a step closer to estimating working memory use in everyday life.

Keywords: offloading, memory assessment, eye movements, cerebrovascular accident, copy task

1. Introduction

Problems in working memory are frequently reported after a cerebrovascular accident (CVA, or stroke; Kimonides et al., 2018; Lugtmeijer et al., 2021). As there seems to be a central role for working memory in everyday life activities (Unsworth et al., 2009), working memory deficits can negatively impact patients' level of independence and quality of life (Kimonides et al., 2018; Nys et al., 2006; Tang et al., 2020). Next to focusing on acceptance of these cognitive changes to alleviate the long-term burden (e.g., acceptance and commitment therapy, cognitive behavioural therapy; Rauwenhoff et al., 2023; Verberne et al., 2019), memory rehabilitation aims to support defective (working) memory by optimizing efficient use of the remaining capacity (e.g., visualization techniques, chunking or grouping, or internal rehearsal; Kaschel et al., 2002; Morrison & Chein, 2011; Norris & Kalm, 2021; Tan & Ward, 2008), but also by using external compensation strategies in order to relief the internal memory load (i.e., *offloading*; Burnett & Richmond, 2023; Elliott & Parente, 2014; Gilbert, 2015a; Gilbert et al., 2023; Morrison & Chein, 2011). Memory rehabilitation thus aims to increase independence, and reintegration and participation in society, by targeting effective (working) memory use and increasing self-efficacy in activities of daily life (Cicerone et al., 2005; Saa et al., 2021). Memory rehabilitation is generally found to be effective in improving performance on memory and working memory tasks (Elliott & Parente, 2014).

Strikingly, there is a twofold discrepancy between how we approach working memory function in the clinic and how working memory is engaged in everyday activities. First, in memory rehabilitation, we aim for adequate deployment of working memory in dynamic everyday situations, and advocate using the surroundings (i.e., offloading). In memory assessment, however, patients are forced to memorize as many items as possible in a distraction-free setting, with little recruitment of other cognitive functions, and no opportunity to exploit the external world. Capacity measures thus do not reflect whether and how patients actually recruit their capacity when they have the choice not to, and thus lack specificity in testing how patients actually *use* their working memory when they are not tied to behavioural instructions. After all, instead of memorizing the entire grocery list, one can simply choose to rely on a written note and look up the required information when in the relevant aisle at the supermarket. Even with a reduced working memory capacity, individuals may not be hindered in carrying out such an activity of daily living. At the same time, individuals with a normal working memory capacity may deliberately choose not to use their maximum capacity and use offloading instead. Second, the premise is that external compensation techniques, such as trained in the clinic, will be spontaneously adopted in more complex environments. Yet, we have little insight into patients' spontaneous strategy deployment, and whether the use of the environment would occur without instruction. If we want to predict whether and how CVA patients will be affected by memory limitations in daily life, we should not only test memory capacity, but acknowledge that working memory is often used in interaction with the environment. The overarching aim of this study is therefore to understand how patients spontaneously use their working memory in interaction with the environment after a CVA.

Importantly, we know that there is individual variation in the use of strategies (Böing et al., in press; Gilbert et al., 2020). Differences in offloading may be driven by (deficient) memory capacity (Böing et al., 2023, 2025; Gilbert et al., 2020; Meyerhoff et al., 2021; Morrison & Richmond, 2020; Risko & Dunn, 2015), but also by effort minimization (Kvitelashvili & Kessler, 2024; Risko & Dunn, 2015), the desire to be accurate (Burnett & Richmond, 2023), or because some individuals may recruit other cognitive domains and can use a strategy more effectively than others. Therefore, as a primary sub-goal, we will not only compare the offloading behaviour of CVA patients and healthy controls, but also specifically address individual differences in offloading behaviour, and investigate how measures of memory capacity (as measured by traditional neuropsychological tasks) are related to offloading behaviour.

Even though the environment generally remains stable and therefore facilitates external offloading (e.g., by using calendars, whiteboards, planners and cues), information is not always readily available. For example, the grocery list may repeatedly disappear in a pocket filled with keys, gloves and cash, thereby making external information less readily available. The availability of information drastically influences the extent to which people are inclined to use the environment or to memorize; if it takes time or physical effort to retrieve external information, internal memory is relied upon more to circumvent the cost associated with retrieving external information, and vice versa (Ballard et al., 1995; Böing et al., 2023, 2025, in press; Draschkow et al., 2021; Gray et al., 2006; Hoogerbrugge, Strauch, Böing, et al., 2024; Sahakian et al., 2023; Somai et al., 2020). Given that information may be volatile in everyday life, our secondary sub aim is to test whether and how CVA patients spontaneously adapt their offloading behaviour (e.g., to a memory-based strategy) when information is not readily available, and how this affects performance.

In order to assess whether CVA patients use more offloading than controls, to address individual differences in offloading, and to test how information availability influences offloading, we designed a free-choice paradigm in which individuals were allowed to rely on the outside world as a strategy to avoid loading working memory capacity. Participants were instructed to copy a geometric jigsaw puzzle onto an empty grid as quickly and accurately as possible. Participants' eyes were tracked to measure inspection behaviour as an index of reliance on the external world, i.e. offloading. Importantly, we manipulated the availability of external information to test how this affected offloading: the example puzzle was either continuously available for inspection (low-cost condition) or became visible only after a gaze-contingent delay (high-cost condition). Crucially, there was no right or wrong strategy; participants were free to store information at their preferred working memory load and to choose the frequency with which they inspected the example puzzle. This free-choice Copy Task was specifically designed to elicit behaviour that resembles spontaneous working memory use in everyday tasks where people can rely on the external world.

2. Materials and Methods

Note that parts of the Materials and Methods section are nearly identical to those described in two previous studies (Böing et al., 2023, 2025). Parts in this section may be paraphrased or copied from the other articles. We limit the description of the measurements to those relevant for the current research.

2.1. Participants

Patients were recruited via the outpatient and inpatient rehabilitation clinic of the Centre of Excellence for Rehabilitation Medicine De Hoogstraat. One patient was recruited through the Centre for Geriatric Rehabilitation De Parkgraaf. Inclusion criteria were having suffered a CVA, aged between 18 and 85 years old, speaking Dutch, and being able to provide informed consent. Another inclusion criterion was the presence of memory deficits, which was a liberal criterion that could be based either on self-reported memory complaints, objective memory impairment based on neuropsychological assessment, or memory impairment observed by a clinician. As the inclusion pace was lower than expected, throughout the study we decided to broaden the inclusion criteria so that patients without explicit memory problems could also participate. Exclusion criteria were presence of visuospatial neglect, deficits in visual perception, moderate to severe aphasia, or when motor impairments prevented the use of a computer mouse; assessing these criteria was part of standard care and carried out by clinicians from various disciplines (e.g., (neuro)psychologists, (speech) therapists). The eligibility of patients was based on the judgment of a neuropsychologist and/or a multidisciplinary team within the rehabilitation clinic.
Healthy controls were recruited simultaneously via various public and university platforms (e.g., social media, family members, university intranet, community

centres). Post-hoc tests were performed to assure similarity between groups regarding age and education. Controls had to be aged between 18 and 85 years old, speak Dutch, and be able to provide informed consent.

All participants gave written informed consent prior to the start of the experiment. Controls were compensated for their participation with 7EU per hour paid in increments of 30 minutes, and received compensation for travel costs. Patients were not compensated. The project was approved by the Faculty Ethics Review Board of the Faculty of Social and Behavioural Sciences at Utrecht University (protocol numbers 21-0485, 22-0069 and 22-0284) and the local ethics committee of De Hoogstraat and De Parkgraaf. The protocol was conducted in accordance with the Declaration of Helsinki.

We aimed to include 25 participants in each group. This number was determined by considering previous studies that have tested inspection behaviour. The original trade-off effect on external reliance versus internal storing has been observed in a group of only 7 participants (no mention of effect size; Ballard et al., 1995), which was replicated by Somai et al. (2020) in a group of 12 participants (only unstandardized β coefficients for linear mixed-effect models were mentioned). We expected greater variability in our target group due to the heterogeneity of aetiology and a wider age range, and therefore wanted to recruit at least twice as many participants in each group. A previous study from our research group showed that this number was sufficient to detect differences in eye movement behaviour between patients with Korsakoff syndrome and healthy controls (detected standardized linear mixed-effects coefficients β were in the range of 0.05 – 0.38; Böing et al., 2023).

2.2. Measurements

2.2.1. Experimental computer task

Apparatus. We ran the experimental task on a Windows 10 Enterprise computer with an Intel Core i7-4790 CPU and 16GB RAM, and used a 27 inch LCD monitor at a resolution of 2560 x 1440 pixels at 100 Hz for experiment presentation. An EyeLink 1000 eye tracker (SR Research Ltd., Canada) was placed below the monitor to track the eyes at a sample rate of 1 kHz. Participants sat at ~67.5 cm from the monitor with their heads in a chin-rest. The lights were dimmed during administration of the experimental tasks. Eye-tracker calibration and validation were performed manually with a 9-point grid attempting to achieve a calibration error of less than 2 degrees of visual angle (dva).

Copy Task. Identical to our previous studies (Böing et al., 2023, 2025) we used a Copy Task. The task aimed to elicit behaviour that resembles spontaneous working memory use in everyday tasks where people can rely on the external world. The experiment was programmed in Python 3.7 using the PyQt5 library (Riverbank Computing Limited, 2019) for visual presentation and mouse and keyboard interaction. PyGaze (Dalmaijer et al., 2014) was used to interact with the eye tracker.

A model puzzle consisting of 6 items in a 3 x 3 example grid was shown at the left-hand side of the screen (see Figure 5.1). At the right-hand side of the screen, a 3 x 3 empty grid was presented, with a 2 x 3 resource grid presented below. The resource grid only contained items that were needed to copy the model; no distractors were present. Items were adopted from Arnoult (Arnoult, 1956; Figure 5.1A) and consisted of black geometrical shapes that could not easily be named to measure reliance on visual working memory instead of verbalisation strategies (Somai et al., 2020).

The Copy Task consisted of two experimental conditions. In the baseline or 'low-cost' condition, the example grid was visible throughout the trial (Figure 5.1B). In this way, the 'cost' to gather information from the outside world was low. In the experimental 'high-cost' condition, we raised the cost to inspect information from the external world by introducing a gaze-contingent waiting time: the example appeared after fixating the left side of the screen for a total of 2000 ms. During the waiting time an hourglass was presented (Figure 5.1C). If participants looked back to the right during the waiting interval, the delay-clock would pause, and would restart as soon as the eyes were redirected to the hourglass again, so that gaze-contingent waiting always was 2000 ms, and never more. Once the example grid became visible, it remained on screen until the participant would move their eyes towards the right side of the screen after which it would disappear.

Participants were instructed to rebuild the model puzzle as quickly and accurately as possible by dragging items from the resource grid to the empty grid using a computer mouse. No emphasis was placed on either speed or accuracy. Participants received direct feedback: if an item was placed incorrectly, the item disappeared and the cell turned red for 700ms, after which subjects could make another attempt. If the item was placed correctly, the cell turned green for 700ms and the item remained fixed. A trial ended after correct placement of six items, or when the time-limit of 42 seconds had passed. The time-limit of 42 seconds was based on the study of Somai and colleagues (2020) in which high-cost conditions with 200, 1500, and 3000 ms delays were used. The authors observed maximum completion times of 30 seconds for placing six items in either of the three variations. As we tested older adults and patients with cognitive impairments, we anticipated that participants would need more time. We therefore complemented the maximum observed completion time of Somai and colleagues (2020) by adding the gaze-contingent delay of 2000 ms for each item that had to be placed in the high-cost condition. In case someone would inspect once per item (which seems plausible from Somai et al., 2020), this would result in an additional 12 seconds. The choice to impose a time-limit was made because we wanted to have some control over the maximum task administration time, as we were bound to an extensive test protocol with limited testing time. After successful completion of a trial, positive feedback (i.e., a thumbs up symbol) was shown. If subjects failed to correctly place all items within the time-limit, they were informed that they ran out of time. By introducing the time-limit, we encouraged subjects to adopt a time-efficient strategy (Melnik et al., 2018). A drift check (max. 2 degrees visual angle) was performed before each trial, and recalibration was performed when deemed necessary.

First, three practice trials were performed in the low-cost condition to get familiar with the task. Calibration and validation of the eye-tracker were performed after the practice trials. The session started with a low-cost block of 15 trials, followed by a high-cost block of 15 trials, resulting in a total of 30 trials. This blocked design could have led to carry-over effects (Böing et al., in press; Patrick et al., 2015), but we wanted to make sure that our participants (especially older adults and/or cognitively impaired individuals) understood the basics of the task before being introduced to the more complex gaze-contingent high-cost condition. After each block, participants answered questions on their experience of commitment to and difficulty of the task (not considered in the current analysis). A session of the Copy Task took 25 to 45 minutes, dependent on the calibration time, the participants' work pace, and the number and length of breaks.

We administered one session of the Copy Task for patients, and two for controls, each session consisting of two blocks. For the controls, only data from the first session was described and analysed in the current study.

Eye movement measures. We defined and calculated several outcome measures to describe between-group inspection behaviour on the Copy Task (see Supplementary Materials: General for elaboration on those), but focus our analyses on the *number of inspections per correct placement* and inspection time per correct placement. The *number of inspections per correct placement* refers to the count of only those saccades that cross the midline from right to left, divided by the number of correct placements. This measure captures how often someone needed to inspect the model to correctly place a single item. It reflects inspection behaviour regardless of whether or not someone was able to place all items in time (hence, 'per correct placement'), as some trials were not finished in time which would bias the inspection rate. The *inspection time per correct placement* is calculated by dividing the dwell time at the model by the number of correct placements over the course of a trial. This score serves as a measure of how much viewing time (i.e., encoding time) someone needed to correctly place a single item.

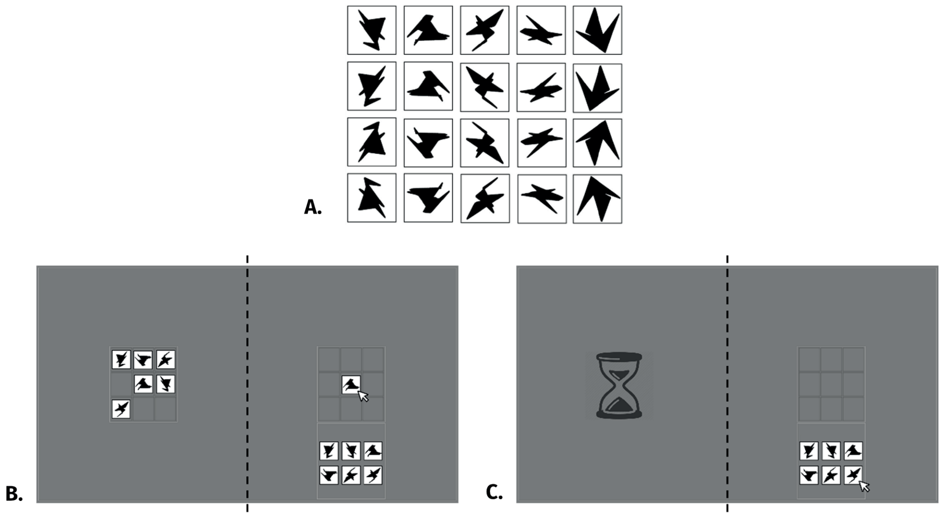


Figure 5.1. **A)** All possible stimuli in the Copy Task. Adopted from Arnoult (1956). An example trial is depicted for the **B)** low-cost condition **C)** and high-cost condition of the Copy Task. At the left-hand side of the screen, the example grid was either visible or replaced by an hourglass for 2000 ms (i.e., gaze-contingent occlusion). At the right-hand side of the screen, the empty grid to place the items (top) and the resource grid (bottom) are presented. A trial ended after 42 seconds. Note: the dotted midline is depicted for illustrative purposes and was not visible in the experiment. The Copy Task layout is adopted and adjusted from Somai and colleagues (2020), and Böing and colleagues (2023, 2025).

Performance measures. We defined and calculated several outcome measures to describe performance on the Copy Task (see Supplementary Materials: General for the way we calculated variables other than the ones highlighted here). The main outcome used was the linear integrated speed-accuracy score (LISAS; Vandierendonck, n.d., 2017, 2021). We calculated this LISAS per participant per condition (low-cost, high-cost) as:

$$LISAS = RTij + PEij \, x \frac{S_{RT}}{S_{PE}} \qquad (1)$$

where RT_{ij} (reaction time) denotes the trial *i* net copying time (completion time minus hourglass waiting time) divided by the number of correct placements for participant *j*. The reaction time data was log transformed to account for skewness associated with time measures. PE_{ij} refers to the proportion of errors on trial *i* and equals 1 minus the number of correct placements divided by the total attempts in that trial. S_{RT} denotes the participant *j*'s overall net copying time standard deviation, and S_{PE} is the participant *j*'s overall *PE* standard deviation. We calculated the standard deviations S_{RT} and S_{PE} per participant over both conditions collapsed

(Vandierendonck, n.d., 2017, 2021). The LISAS was chosen as it combines two outcomes of performance (accuracy and speed) and weighs their importance equally. Lower LISAS reflects better (i.e., more accurate and faster) performance.

Strategy and performance stability. Adapting behaviour from one situation to the other requires flexibility. Switch costs may occur in the transition from one strategy to the other, and participants may differ with regards to how easily they adjust their inspection strategy to the newly imposed conditions. Some may switch effectively and efficiently, whereas others may experience larger switch costs hampering performance. We therefore wanted to explore how spontaneous changes in inspection behaviour may have led to changes in performance. As every individual has a different starting level of performance, it is most informative to test stability/change *within* the individual. To investigate the degree to which each participant adapted their strategy from the low-cost to the high-cost condition, we divided the number of inspections per correct in the high-cost condition by the number of inspections per correct in the low-cost condition and obtained the *change factor number of inspections per correct.* A score of one indicates no change. Scores below one indicate a decrease in the number of inspections per correct, scores above one indicate an increase. The more the value deviates from one, the larger the adaptation in inspecting behaviour from the low-cost condition to the high-cost condition. Note that change factors of 0.5 and 2 indicate a similar magnitude but 0.5 indicates twice as few and 2 indicates twice as many inspections. The same rationale was followed for the performance measure (*change factor LISAS*). For visualization purposes – but not for analysis – the *change factor LISAS* was centred around zero and flipped.

2.2.2. Neuropsychological tasks (see Supplementary Materials: General for details)

We administered neuropsychological memory tasks that all had a similar task instruction: to memorize and report back as much information as possible. These tasks are all grafted on estimating a maximum capacity span. Standard stimulus set B of the modified Location Learning Task (Kessels et al., 2006, 2014) was used to assess visuospatial immediate recall, and the Rey Auditory Verbal Learning Task (15 items, Dutch version; Bouma et al., 2012; Saan & Deelman, 1986) was administered to assess verbal immediate recall. The Digit Span Forward and Backward from the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV; Wechsler, 2012) were administered to assess short-term auditory memory span and verbal working memory, respectively. A digitized version (2D) of the Corsi Block-Tapping Task was used to assess visuospatial working memory capacity (Brunetti et al., 2014; Claessen et al., 2015; Corsi, 1972; Kessels et al., 2000). Both the Forward and Backward sub-tasks were included.

2.2.3. Questionnaires (see Supplementary Materials: General for details). Several questionnaires were administered to characterize groups. Participants were asked whether they experienced memory problems (yes/no). We used the 4-statement Dutch short fatigue questionnaire ('Verkorte Vermoeidheidsvragenlijst') to assess fatigue experienced in the previous two weeks (Alberts et al., 1997; Bleijenberg et al., 2009). The Dutch Hospital Anxiety and Depression Scale was administered to screen for complaints of anxiety (7 items) and depression (7 items) (Spinhoven et al., 1997). The abridged version of the Dutch Metamemory in Adulthood questionnaire was adopted from Ponds & Jolles (1996) to characterize memory self-efficacy. To characterize coping style, the Utrecht Coping List (Gregório et al., 2014; Schreurs et al., 1984) was administered.

2.3. Procedure

2.3.1. CVA patients. For patients, we divided the test battery into two sessions over separate days (ranging from 1 to 14 days apart). Before the first session, we checked whether patients had already performed some of the neuropsychological tasks as part of rehabilitation care within six months prior to the experiment. If that was the case, they were exempt from that task; previously reported scores on those tasks were used in order to prevent unnecessary work load and possible practice effects (Bouma et al., 2012; Lezak et al., 2012). Sessions were ended after a maximum of 75 min, or when patients became too tired.
Task administration in session 1 comprised (in this order) the following: a memory complaint question (yes/no), short fatigue questionnaire ('Verkorte Vermoeidheidsvragenlijst'), Location Learning Task —direct recall, Copy Task, and Location Learning Task —delayed recall. Task administration in session 2 comprised (in this order) the following: Rey Auditory-Verbal Learning Task—direct recall, Corsi Block-Tapping Task Forward and Backward, WAIS IV Digit Span Forward and Backward, Rey Auditory-Verbal Learning Task—delayed recall. Patients were asked to fill in three questionnaires (Hospital Anxiety and Depression Scale, Metamemory in Adulthood, and Utrechtse Coping List) in between the sessions. See Supplementary Table S5.1 for an overview of the test procedure and sessions for CVA patients.

2.3.2. Controls. Participants in the control group received a link to fill out some questionnaires online at home in the period 14 to 1 day(s) before their test session, including the Verkorte Vermoeidheidsvragenlijst (fatigue), the Hospital Anxiety and Depression Scale and Utrechtse Coping List. These questionnaires were administered to characterize the group. Several other questionnaires were included (but not described here) to collect data for a parallel study (Böing et al., 2025).

At the university testing facility, the rest of the test protocol was administered in a single visit. The first and second session of the experiment were separated by a break of 10 to 20 minutes, and the total test duration was a maximum of 3 hours. Task administration in session 1 comprised (in this order): Location Learning Task – direct recall, Copy Task – first session, Location Learning Task – delayed recall, WAIS IV Digit Span Forward and Backward, and if time allowed: a Fixation and Free viewing task (not taken into account in the current study). Task administration in session 2 comprised (in this order): Rey Auditory-Verbal Learning Task – direct recall, Copy Task – second session (not taken into account in the current analysis), Rey Auditory-Verbal Learning Task – delayed recall, Corsi Block-Tapping Task Forward and Backward, and if time allowed: Change Detection Task (not taken into account in the current study).

At the end of the test protocol, the Metamemory In Adulthood questionnaire was administered. This was the case only for a subset of participants (n=15) as the questionnaire was added later to the test protocol (Böing et al., 2025). This questionnaire was added to get an extra measure on beliefs about one's memory function. See Supplementary Table S5.1 for a schematic overview of the test procedure.

2.4. Pre-processing

2.4.1. Inspection behaviour. Saccades, fixations, and timestamps were extracted using the EyeLink 1000 parser (default EyeLink saccade detection algorithm, SR Research Ltd., Canada). Data pre-processing was implemented using Python 3.10. Every first trial in each block was removed from analysis (see Results, Data Loss): this trial served to check whether the instructions had been retained (additional instructions were given when needed) and to habituate the participant to the new situation (e.g., from low-cost to high-cost). Eye-movement and performance variables were calculated as described in Measurements and Supplementary Materials: General. Data analyses were conducted using R 4.1.2 (R Core Team, 2017).

2.4.2. Individual strategy categorization. Participants were categorized based on the number of inspections per correct placement. For each individual, this outcome measure was aggregated by the mean over trials per condition (low-cost, high-cost). Those who made more than one inspection per correctly placed item were categorized as 'offloaders', i.e., those who relied relatively heavily on the external world. Participants who correctly remembered one item or more per inspection were categorized as 'loaders', i.e., those who successfully relied more on internal loading. Among the loaders, those who correctly placed three or more items per inspection could be further categorized as 'full-loaders', i.e., those who loaded up to the limits of their capacity. These category cut-offs were partly based upon the finding that people have an estimated working memory capacity of four items (Cowan, 2001), and partly based upon task constraints. The Copy Task did not have a high enough resolution to dissociate between people loading three, four, or five items, as in all of these instances, participants would need an additional inspection for the remaining items, yielding two inspections for trial completion. This resulted in a rather lenient definition of full-loading, which allowed some individual deviations but still captured the higher end of capacity use. The current categorization system was adopted from Böing and colleagues (in press) and slightly adapted to the current task characteristics.

2.4.3. Clinical classification. We used a classification to identify patients with memory impairments, defined as performing outside the normal range on neuropsychological capacity task outcomes. We defined the levels of performance in memory capacity based on a subset of the tasks administered: Location Learning Task – displacement errors, Rey Auditory-Verbal Learning Task – immediate recall score, Digit Span Forward, Digit Span Backward, Corsi Block-Tapping Task Forward, and Corsi Block-Tapping Task Backward. Each individual's scores were compared to scores of their reference group (in terms of age and education) as is common in clinical assessment. The level of performance was defined as either within or outside the normal range. An abnormal score could be any of the following: 1) a score below the 2^{nd} percentile on two or more subtasks (e.g. impaired performance), 2) a score below the 2^{nd} percentile on one subtask and/or a score between the 2^{nd} and 9^{th} percentile on two or more subtasks (e.g., below average performance). A normative score was anything outside these definitions (American Psychiatric Association, 2013; Hendriks et al., 2020).

The delayed recall scores of the Location Learning Task and Rey Auditory-Verbal Learning Task were not taken into account, as we could not assure that the delay period was equally long for all the participants and the interference tasks differed between controls and patients. We report the delayed recall scores in the Results section for completeness, but did not use them for clinical interpretation.

2.5. Data analyses

2.5.1. Group characteristics. To assure similarity between groups in terms of age and education, Mann-Whitney U tests were performed. A chi-squared test was performed to compare sex distributions between groups. Scores on neuropsychological tasks and questionnaires were reported to characterize groups, and chi-squared tests and proportion z-tests were performed to check for group differences.

5

2.5.2. Strategy conceptualization check. We ran a non-parametric Kendall Rank correlation between the *number of inspections per trial* and *inspection time per inspection* (note: not the 'per correct' placement scores). We expected that a higher number of inspections would relate to shorter inspection times per inspection, and vice versa. Fewer inspections with longer inspection times would reflect a tendency towards memorization, as more time inspecting would indicate an attempt to encode more items at once. Correlation coefficients were reported as tau (τ) and effect sizes as z.

2.5.3 Group inspection behaviour. We investigated the influence of information availability and group on inspection behaviour and performance for completeness and to provide data for a potential future meta-analysis on Copy Task behaviour. To this end, all trials were fed to a linear mixed-effect model (LMM; Singmann & Kellen, 2019) by using the lmer function in R (lme4 package; Bates et al., 2014). The LMM is robust against deviations from normality of the outcome variables, is sensitive to missing data, and takes individual differences within groups into account (Schielzeth et al., 2020). Factors included were group (CVA, control), condition (low-cost, high-cost), and the interaction of group and condition, and random slope and intercept for individuals. We ran the model to predict the influence of these factors on the *number of inspections per correct placement* and *inspection time per correct placemen*t. The normality of the residuals was visually examined and confirmed. Effect sizes were reported as beta-coefficients with a 95% confidence interval. Post hoc pairwise comparisons were conducted using the Tukey-Kramer method using the contrast function within the emmeans package (version 1.7.0) in R, which accounts for multiple comparisons and controls the family-wise error rate. We report the summarized findings in the main text, and report descriptives, statistics and graphs in the Supplementary Results Chapter 5 (Table S5.2 and Figure S5.2).

2.5.4. Individual strategy categorization, clinical classification and single case statistics. The main goal of this study was to describe individual differences in strategies. We therefore classified participants as 'offloader', 'loader', or 'full-loader', based on the *number of inspections per correct* placement (see Pre-processing). An <u>external strategy</u> translates to a high mean number of inspections per correct placement (offloading), whereas an <u>internal memorization strategy</u> translates to a low mean number of inspections per correct placement (i.e., longer encoding per iteration; (full-)loading). We provided the percentages of participants falling within each category (offloader, loader, full-loader) for each condition (low-cost, high-cost) for both groups (CVA, controls), and noted the number of inspections per correctly placed item for offloaders, loaders, and full-loaders.

Single case Bayesian Deficit Testing, with the covariate age, was used to further assess whether inspection behaviour of each individual patient statistically deviated from the performance of healthy controls. We performed a one-tailed Bayesian Deficit Test with α =0.05 and 10000 iterations on *the number of inspections per correct placement* (with the covariate age) using the single package in R (Rittmo & McIntosh, 2021). Bayesian Deficit Testing allows to assess single cases against a norm group of healthy controls: it takes a single observation and compares it to a distribution estimated by a control sample, using Bayesian methodology (Rittmo & McIntosh, 2020, 2021, 2023).

2.5.5. Strategy shifting and performance stability. We assessed the relation between the change factor for the number of inspections per correct and the change factor for LISAS performance with a Kendall Rank Correlation analysis. This provided insight into whether the change (or stability) in offloading behaviour was associated with an increase or decrease in performance.

3. Results

3.1. Group characteristics

We approached 28 patients of whom 19 agreed to participate. Of these, three were excluded due to prematurely ending the test session because the participant was not able to complete the Copy Task or because we could not track their eyes due to a failure of the eye-tracker or a medical condition. Another participant completed the protocol but was excluded a posteriori as it appeared that there was no CVA history but another medical condition. Eventually, we were able to obtain datasets of 15 patients (see Supplementary Figure S5.1 for a patient flow chart; see Table 5.1 for lesion information, demographic characteristics and test scores).

Forty-eight healthy participants were recruited as control group. Four cancelled their appointment due to personal reasons and did not wish to reschedule, four were not tested on the Copy Task due to technical problems, one completed the protocol but appeared to not meet our inclusion criteria, and for one we were unable to track the eyes. Eventually, we obtained datasets of 38 individuals (see Table 5.1 for demographic characteristics and see Supplementary Figure S5.1 for a control flow chart). Note that part of the healthy control data was reported in two previous studies (Böing et al., 2023, 2025).

With this sample size, for a one-tailed Mann-Whitney test (α = .05) with a power of .8, we would be able to reliably detect effects of Cohen's d=0.79 (Faul et al., 2009). As the sample size of the CVA patients was smaller than anticipated, we focused on individual behaviour and linear mixed-effects models (LMM; Singmann

& Kellen, 2019). LMMs weigh the number of observations and take missing data into account (Schielzeth et al., 2020).

Group characteristics, scores on neuropsychological assessment and questionnaires, and statistical comparisons between groups are displayed in Table 5.1. The level of education was characterized according to the classification of Verhage (1964, 1965), that is commonly used in Dutch clinical care, and classifies the level of education (ranging from 1 to 7) based on the number of education years. All individuals were without known visual field defects and had normal or corrected-to-normal visual acuity.

Table 5.1. *Demographic characteristics, scores on memory capacity tasks, and questionnaires, split per group (i.e., CVA patients or healthy controls), medians (IQR) or frequencies (%) are depicted.*

	CVA patients			Healthy controls			Test Statistic [a]
	n	n (%)/Mdn (IQR)	Range	*n*	n (%)/Mdn (IQR)	Range	
Demographics							
Age, years	15	61 (8.5)	49 – 84	38	60	40 – 81	U =302, p=.737, *r*=0.06
Sex, % male	15	10 (66.7%)		38	15 (39.5%)		χ^2=2.19, *p*=.139, d=-0.42
Level of educationb	15	6 (1.5)	3 – 7	38	6 (1.75)	4 – 7	U=318, *p*=.506, *r*=0.12
% low (<10 years)		4 (26.7%)			5 (13.2%)		
% medium (10 - 11 years)		3 (20%)			12 (31.6%)		
% high (>15 years)		8 (53.3%)			21 (55.3%)		
Lesion information							
Time post-stroke onset, days	15	74 (46.5)	36 – 137				
CVA history, % first	15	86.7%					
CVA type	15						
% ischemic		11 (73.3%)					
% intracerebral hemorrhage		3 (320%)					
% subarachnoid hemorrhage		1 (6.7%)					
Lesion side,	14c						
% left		11 (78.6%)					
% right		2 (14.3%)					
% bilateral		1 (7.1%)					
MoCA (0-30)	15	24 (4)	18 – 30				
SAN (Aphasia Index; 1-7)	15	6 (2)	4 – 7				
Barthel Index (0-20)	15	13 (7.5)	5 – 19				
Motricity Index Arm (0-99)d	15	76 (80)	0 – 100				
Motricity Index leg (0-99)d	15	83 (39)	0 – 100				
Mood, fatigue and coping questionnaires							
Do you experience memory problems? %yes		6 (40%)		38	9 (23.7%)		χ^2=0.72, *p*=.396, d=0.24
Fatigue, % severe fatigue	15	6 (40%)		38	6 (15.8%)		χ^2=2.35, *p*=.123, d=-0.43
Anxiety & Depression Scale (HADS)							
Anxiety (HADS)	15			38			
Not present (score 0-7)		15 (100%)			31 (81.6%)		
Potential (score 8-10)		0 (0%)			7 (18.4%)		
Likely (score ≥11)		0 (0%)			0 (0%)		

Table 5.1. *(continued) Demographic characteristics, scores on memory capacity tasks, and questionnaires, split per group (i.e., CVA patients or healthy controls), medians (IQR) or frequencies (%) are depicted.*

	CVA patients			**Healthy controls**			**Test Statistic** [a]
	n	**n (%)/Mdn (IQR)**	**Range**	***n***	**n (%)/Mdn (IQR)**	**Range**	
Depression (HADS)							
Not present (score 0-7)		11 (73.3%)			36 (94.7%)		
Potential (score 8-10)		4 (26.7%)			2 (5.3%)		
Likely (score ≥11)		0 (0%)			0 (0%)		
Total score		7 (6.5)	0 – 16		5 (7.75)	0 – 19	U =248, *p*=.470, *r*=-0.13
Coping style (UCL)	14			38			
Active		19.5 (5.74)	8 – 24		19 (7.75)	7 – 28	U =298, *p*=.522, *r*=0.12
Palliative		16 (3)	8 – 23		15.5 (4.75)	8 – 22	U =246, *p*=.693, *r*=-0.07
Avoidant		15.5 (6)	9 – 21		15 (4)	8 – 22	U =250, *p*=.748, *r*=-0.06
Social support		12 (1.75)	6 – 15		12 (4)	6 – 24	U =296, *p*=.538, *r*=0.11
Passive		9.5 (3)	8 – 14		11 (5.75)	7 – 18	U =285, *p*=.700, *r*=0.07
Emotional expression		5 (1.75)	3 – 8		6 (2)	3 – 9	U =366, *p*=**.038***, *r*=0.37
Comforting		13 (4)	6 – 15		12 (3)	5 – 19	U =189, *p*=.210, *r*=-0.24
Neuropsychological task scores							
Location learning task	15			38			
Total displacement score [e]		29 (40.5)	6 – 101		27.5 (26)	0-75	U =264, *p*=.685, *r*=-0.07
Learning index (0–1)		0.47 (0.28)	0.14 – 0.95		0.52 (0.42)	0.10 – 1	U =308, *p*=.657, *r*=0.08
Delayed recall: Placement errors		1 (7)	0 – 17	37	1 (4)	0 – 19	U =252, *p*=.587, *r*=-0.09
Rey auditory–verbal learning task	14			37			
Immediate recall: Total correct (0–75) [e]		34 (20.2)	18 – 54		42 (18)	30 – 67	U =364, *p*=**.027***, *r*=0.41
Delayed recall: Total correct (0–15)		6 (4.5)	0 – 14		8 (6)	3 – 14	U =357, *p*=**.039***, *r*=0.38
Digit span (WAIS-III/IV)	13			38			
Forward span (2–9) [e]		5 (1)	4 – 7		6 (2)	4 – 9	U =318, *p*=.112, *r*=0.29
Backward span (2–8) [e]		4 (0)	3 – 5		5 (2)	2 – 8	U =357, *p*=**.014***, *r*=0.45
Corsi block-tapping task	15						
Forward span (2–9) [e]		5 (0.5)	4 – 7	38	5 (1)	3 – 8	U =330, *p*=.345, *r*=0.16
Backward span (2–8) [e]		5 (2)	4 – 8	37	6 (1)	2 – 7	U =318, *p*=.393, *r*=0.15

	CVA patients			Healthy controls			Test Statistic [a]
	n	n (%)/Mdn (IQR)	Range	*n*	n (%)/Mdn (IQR)	Range	
Impairment within memory domain [f]							
% impaired		0 (0%)			2 (5.3%)		
% below average		6 (40%)			5 (13.2%)		
% within normal range		9 (60%)			31 (81.5%)		
Memory capacity compound, z	15	-0.17 (0.76)	-1.42 – 1.36	38	0.14 (0.91)	-1.35 – 1.43	U =378, *p*=.067, *r*=0.33
Metamemory In Adulthood	14			15			
Scale Achievement (+ = high achievement)		3.5 (0.24)	2.71 – 4.0		3.71 (0.57)	2.43 – 4.5	U =123, *p*=.442, *r*=0.17
Scale Anxiety (+ = high knowledge; lower is better) [g,h]		2.36 (0.58)	1.58 – 3.76		2.67 (1.04)	1.25 – 4.00	U =83.5, *p*=.359, *r*=-0.21
Scale Capacity (+ = high capacity) [h]		3.75 (2.86)	2.25 – 3.75		3.25 (0.58)	2.58 – 4.50	U =143, *p*=.100, *r*=0.36
Scale Change (+ = stability) [h]		3.55 (0.6)	1.60 – 4.40		3.2 (0.65)	2.11 – 4.30	U =80.5, *p*=.294, *r*=-0.23
Scale Locus (+ = internal locus)		3.22 (0.29)	1.71 – 3.86		3.14 (0.57)	2.00 – 4.14	U =82.5, *p*=.333, *r*=-0.21
Scale Task (+ = high knowledge)		3.55 (0.48)	2.80 – 4.20		3.5 (0.45)	2.70 – 4.5	U =95.5, *p*=.693, *r*=-0.09
Scale Strategies Total (+ = high use)		3.25 (0.81)	1.50 – 4.75		3.44 (0.72)	2.31 – 5.00	U =116, *p*=.662, *r*=0.10
Scale internal strategies		3 (0.69)	1.50 – 4.88		3.5 (0.69)	2.38 – 5.00	U =134, *p*=.213, *r*=0.28
Scale external strategies		3.56 (0.78)	1.50 – 4.75		3.5 (0.56)	2.25 – 5.00	U =90, *p*=.538, *r*=-0.14
Sum score							
Memory Self-Efficacy [e] (+ = higher efficacy)		3.34 (0.56)	2.58 – 4.19		3.27 (0.57)	2.26 – 4.42	U =103, *p*=.949, *r*=-0.02

CVA = cerebrovascular accident (stroke), Mdn = median, IQR = interquartile range, range (min.–max.), MoCA = Montreal Cognitive Assessment; SAN = Stichting Afasie Nederland. Sample size may differ per outcome variable. [a] Non-parametric test statistics indicating group differences and effect sizes: chi-squared, p-value, and d for binomial variables, or Mann–Whitney–Wilcoxon U, p-value, and rank-biserial correlation r for continuous data. [b] The level of education is characterized according to the classification of (Verhage, 1964, 1965), that is commonly used in Dutch clinical care, and classifies the level of education (ranging from 1 to 7) based on the number of education years. [c] One unknown. [d] Motricity Index maximum score is 99. However, if both arm and leg score 99/99, one point may be added, so a score of 100 is possible and indicates intact arm and leg function. [e] Capacity scores used in memory capacity compound z-score; Location Learning Task displacement errors are reversed so that higher scores indicate better performance on all capacity tasks. [f] Impaired: a score <2nd percentile on ≥ 2 sub taskse; Below average: a score <2nd percentile on 1 sub taske and/or a score between 2nd – 9th percentile on ≥ 2 sub taskse; Within normal range: does not fit criteria for impairment or below average. [g] Anxiety scale is reversed in calculation of the Memory Self-Efficacy sum score, so that higher scores indicate better subjective memory experience. [h] Scores used in calculation of memory self-efficacy; higher scores indicate better subjective memory experience. * $p \leq 0.05$, ** $p \leq 0.005$, *** $p \leq 0.001$

5

3.2. Data loss

Datasets were obtained for 15 participants in the CVA population. Across these 15 participants, 450 trials were planned to be collected (15 trials x 2 conditions x 15 participants). All first trials of each block were removed to assure task comprehension (30 trials). Any reason that could possibly interfere with performance (excessive movement of the participant, forgetting the task instructions, problems controlling the mouse) was logged, and the corresponding trials (19 trials) were removed from further analysis. No trials were lost due to exceeding the drift check. During data pre-processing, we checked for deviant trials that yielded invalid data due to corrupted eye-tracking logging (e.g., zero fixations per second, dwell times of zero, or missing data), of which most coincided with the logged trials. One additional trial was discarded due to a logged dwell time that exceeded the duration of the trial, which should be attributed to an eye-tracking failure. Finally, 401 trials were left for analysis.

Across 38 participants in the control group, 1140 trials were planned to be collected (15 trials x 2 conditions x 38 participants). Again, the first trials of each block were removed (76 trials). Trials that were invalid due to signal loss, excessive movement of the participant, forgetting the task instructions, or problems controlling the mouse were removed (9 trials). Despite the implementation of a drift check, some trials were started with a drift check above the 2 degrees visual angle threshold. When exceeding 5 degrees visual angle, trials were excluded. In total, 9 trials needed to be excluded because of exceeding the drift check threshold. In the control group, 1046 trials were left for analysis.

3.3. Strategy conceptualization check

There was a strong negative correlation between the number of model inspections and dwell time per inspection per trial across *all* trials (τ = -0.605, p < .001, z = -33.27), indicating that fewer model inspections were related to longer inspection durations. This finding substantiates our conceptualization that the number of inspections can be used as an index of memorization.

3.4. Group inspection behaviour

Group scores for inspection behaviour and performance across conditions (low-cost and high-cost) were calculated and reported in Supplementary Table S5.2. Statistical results of linear mixed-effects models explaining the *number of inspections per correct placement, inspection time per correct placement* and *LISAS performance* by factors group (CVA, control) and condition (low-cost, high-cost) can be found in Supplementary Figure S5.2.

For all three outcome variables, the same pattern emerged: both group and condition were significant predictors. In the high-cost condition the *number of inspections per correct placement* decreased and the inspection time per correct placement increased (e.g., memorization) as compared to the low-cost condition. Patients used more and longer inspections than controls to place one item correctly, and performed worse than controls. An interaction was only observed for the number of inspections per correct placement, where patients decreased their inspection rate more as compared to controls in the high-cost versus low-cost condition. Post hoc comparisons between conditions and groups are displayed in Supplementary Figure S5.2.

3.5. Individual strategies: strategy categorization, clinical classification and single case statistics

As a group, CVA patients used more and longer inspections to place one item correctly as compared to healthy controls, but group analyses do not necessarily reflect the individual's behaviour. To answer our main question, namely how *individual* patients use their memory, we descriptively report and visualize the distribution of used strategies across individuals, and look into the clinical meaningfulness of such a categorization. Figure 5.2 shows inspection behaviour across the low-cost and high-cost condition for individuals, separately for healthy controls and patients. We marked the individuals that classified as having abnormal memory function as measured with neuropsychological capacity tasks (i.e., extremely low or below average performance within the memory domain, see Methods). Table 5.2 displays the proportions of offloaders, loaders, and full-loaders across groups and conditions.

In the low-cost condition, almost everyone heavily relied on the external world: 100% of patients and 97.4% of controls were categorized as offloader. When imposing high-costs, the percentages dropped: 53.3% of patients and 39.5% of controls were offloading, and 46.7% of patients and 57.9% of controls displayed loading behaviour. Among the loaders, only one control was classified as full-loader. Looking at individual data points, there was quite some variability within these categories. Some offloaders inspected twice per correct item, but there were also individuals that used on average three to four inspections to place one item correctly.

Interestingly, in both groups we identified individuals that had abnormal memory capacity, indicating a deficit in memory *capacity* for these individuals. In the CVA group, the individuals with a deficit in memory load seemed to present as a cluster in their inspection behaviour (Figure 5.2B, saturated dots), showing a relatively high number of inspections per correct placement and a long inspection duration. However, in the healthy control group, those with a deficit in memory load did not show such distinct inspection behaviour (Figure 5.2A, saturated dots). To further assess whether inspection behaviour of the subgroup of CVA patients with memory impairment statistically deviated from normal, we performed a one-tailed Bayesian Deficit Test on the number of inspections per correct placement in the low-cost condition for each patient (with inclusion of the covariate age; Rittmo & McIntosh, 2023; Rittmo & McIntosh, 2021). We only compared the individual performance in the low-cost condition, as we found that groups significantly differed in this condition. Table 5.3 displays the results of the single case statistics. We found that five of the patients with an abnormal memory score also showed distinct eye-movement behaviour, while one patient with an abnormal memory score did not deviate from the norm group in terms of inspections per correct placement. Thus, whereas some individuals (in the control group) had a deficit in memory *capacity* (as measured by neuropsychological capacity tasks) but did not show deviant memory *use* (i.e., no excessive offloading in the copy task), other individuals (in the CVA population) showed both *capacity* problems <u>and</u> showed distinct memory *use* in the form of heavy reliance on offloading.

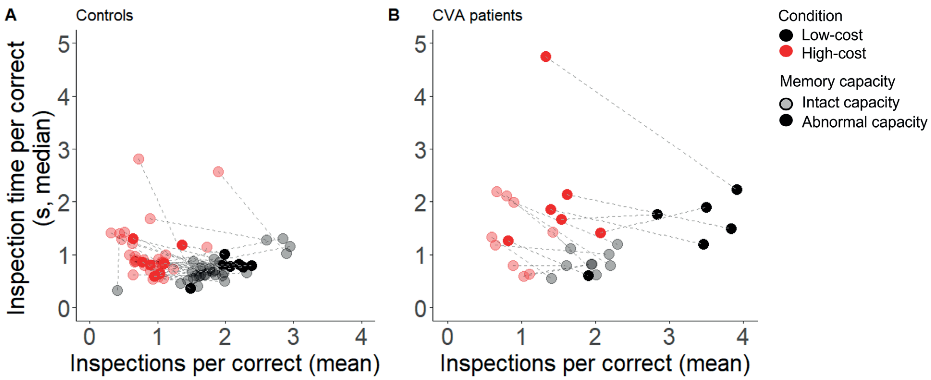


Figure 5.2. Offloading behavior, presented as median inspection time per correct placement (e.g., encoding time per item) and the average number of inspections per correct placement for the two groups: **A)** healthy controls, **B)** CVA (stroke) patients. Data points represent data of the individual aggregated over trials in the low-cost condition (black) and high-cost condition (red). Dashed lines connect the data that belongs to the same individual, and indicate the change in inspection frequency for an individual from low-cost to high-cost (i.e., change factor for inspections). Saturated dots indicate individuals that showed below average or impaired (thus, abnormal) memory performance on traditional neuropsychological memory capacity assessment.

Table 5.2. *Mean number of inspections per correctly placed item per group (CVA patients, controls), per condition (low-cost, high-cost) and per strategy category (offloader, loader, full-loader). We provide the n number of individuals (%) per category based upon the number of inspections per correct, and the mean, standard deviation (SD), and range of the number of inspections per correct.*

	Strategy category	**Group**					
		CVA patients			**Controls**		
Condition		n(%)	Mean (SD)	Range	n(%)	Mean (SD)	Range
Low-cost	Offloader	15 (100%)	2.45 (0.84)	1.41 – 3.91	37 (97.4%)	1.92 (0.41)	1.33 – 2.95
	Loader	0			1 (2.6%)	0.41 (n.a.)	n.a.
	of which full-loader	0			0		
High-cost	Offloader	8 (53.3%)	1.44 (0.32)	1.03 – 2.07	15 (39.5%)	1.19 (0.27.)	1.01 – 1.89
	Loader	7 (46.7%)	0.75 (0.12)	0.6 – 0.89	22 (57.9%)	0.72 (0.15)	0.44 – 0.95
	of which full-loader	0			1 (2.6%)	0.31 (n.a.)	n.a.

Table 5.3. *Single case statistics for the mean number of inspections per correctly placed item in the low-cost condition for each individual in the patient group. We provide patients' mean scores, Bayesian Deficit Testing standardized effect sizes (Z-CCC) with 95% confidence intervals (CI) of task difference between the case and controls, p-values, and an estimation of the proportion of controls that would exhibit a more extreme conditioned score than the patient case. Those who showed an abnormal score on neuropsychological capacity tasks are labelled.*

ID	***M* inspections per correct in the low-cost condition**	**Z-CCC [95% CI]**	***p*-value**		**Proportion controls scoring higher (% [95% CI])**	**Abnormal memory capacity**
3001	3.84	3.92 [2.66, 5.03]	<.001	***	0.01 [0.00, 0.08]	Yes
3002	3.91	4.19 [3.04, 5.22]	<.001	***	0.02 [0.00, 0.12]	Yes
3003	1.66	-0.51 [-0.84, -0.15]	.687		68.71 [55.74, 80.1]	
3006	2.2	0.78 [0.27, 1.25]	.230		22.97 [10.55, 39.21]	
3007	1.61	-0.45 [-0.99, 0.10]	.664		66.39 [46.03, 83.78]	
3008	2.18	0.63 [0.27, 0.96]	.273		27.32 [16.81, 39.42]	
3009	1.95	0.19 [-0.16, 0.53]	.429		42.85 [29.86, 56,45]	
3010	1.9	0.11 [-0.27, 0.47]	.459		45.87 [31.79, 60.65]	Yes
3011	1.4	-0.96 [-1.38, -0.52]	.820		81.97 [69.74, 91.65]	
3012	3.5	3.45 [2.55, 4.24]	<.001	***	0.09 [0.00, 0.54]	Yes
3013	2.01	0.16 [-0.33, 0.63]	.441		44.12 [26.31, 62.85]	
3014	2.84	2.08 [1.46, 2.63]	.026	*	2.57 [0.42, 7.23]	Yes
3016	2.3	0.93 [0.50, 1.33]	.186		18.62 [9.25, 31.02]	
3018	3.46	3.35 [2.48, 4.11]	.001	***	0.12 [0.00, 0.65]	Yes
3019	1.94	0.09 [-0.25, 0.43]	.466		46.64 [33.38, 59.86]	

* *p* ≤ 0.05, ** *p* ≤ 0.005, *** *p* ≤ 0.001

To summarize, the vast majority of people relies on the external world when information is readily available. When high-costs are imposed, some individuals switch to loading, but seldom to full-loading, and many individuals are inclined to stick to offloading. Individuals that show deficits in their maximum *load* can, but do not necessarily, show distinct memory *use* as indexed by inspection behaviour. This indicates that inspection behaviour has the potential to reveal individual signatures in memory usage that go beyond the mere measure of memory capacity.

3.6. Strategy shifting and performance stability

We analysed whether a larger change in inspection frequency from the low- to high-cost condition was related to a change in performance. To this end, each individual's *change factor number of inspections per correct and change factor LISAS* were calculated. Change factors for inspection behaviour ranged from 0.19 to 1.1 (M = 0.48, SD = 0.17), where a value of 1 indicates a stable inspection frequency, values below 1 a decrease in inspections, and values above 1 an increase in inspections. Change factors for LISAS ranged from 0.86 to 2.52 (M = 1.39, SD = 0.36). A change factor of 1 indicates stable performance. Note that higher values of LISAS indicate worse performance, so that change scores above 1 indicate decreased performance, and scores below 1 indicate improvement.

Kendall's rank correlation analysis showed a significant negative relation between the *change factor number of inspections per correct* and the *change factor LISAS* (τ = -0.24, p = .01, z = -2.51), indicating that a decrease in inspection frequency related to an increase in LISAS and thus a decrease in performance. To make interpretation of this relation more intuitive, the *change factor LISAS* was centred around zero (so that change factor zero indicates stable performance) and flipped (so that negative numbers indicate a decrease in performance from the low-cost to the high-cost condition). A change factor above zero indicates improvement. Figure 5.3A illustrates the negative relation between the *change factor number of inspections per correct* and the *change factor LISAS*.

Memory rehabilitation after stroke emphasizes using external strategies. We questioned whether adherence to the use of external strategies -even when the external world is not readily available anymore- would show an advantage over using a relatively more memory-based strategy for CVA patients specifically. We therefore visualized the relation between the *change factor number of inspections per correct* and the *change factor LISAS* for the patient group in Figure 5.3B, where the same relation became apparent. This relation suggests that it is indeed beneficial for this group to adhere to behaviour that was applied previously.

In summary, the patients who stuck to their initial inspection frequency managed to maintain a more stable performance. Those who started to rely relatively less on the outside world and more on internal memory showed decreased performance. This suggests that it may not be wise to start using memory to a higher degree, even though the environment provokes memorization. Especially with fallible memory, as is the case in those with abnormal memory capacity scores, external strategies would be the most beneficial option.

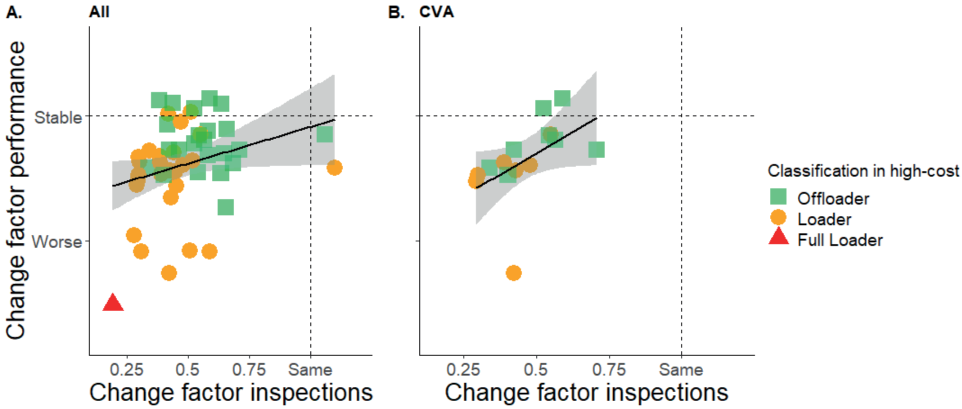


Figure 5.3. The negative relation between the change in the number of inspections per correct and change factor LISAS performance for **A)** all participants and **B)** the CVA (stroke) population specifically. Dashed vertical and horizontal lines indicate a stable score. Values below 1 for the change factor inspections indicate a decrease in the number of inspections from the low-cost to the high-cost condition. A value of 0.5 would indicate half the amount of inspections from the low-cost to the high-cost condition. For the change factor performance, values below the dashed line indicate a decrease in performance. Values above this line indicate an improvement. Green squares indicate those who kept using an offloading strategy in the high-cost condition, orange circles indicate those who used or started using loading, and the red triangle displays the one person that used full-loading.

4. Discussion

Working memory problems are common after stroke (Kimonides et al., 2018; Lugtmeijer et al., 2021). Memory rehabilitation aims to support defective working memory and to relief the internal memory load by advocating the use of external compensation strategies (i.e., *offloading*; (Burnett & Richmond, 2023; Elliott & Parente, 2014; Gilbert, 2015a; Gilbert et al., 2023; Morrison & Chein, 2011). While a growing body of literature shows that engaging working memory naturally co-occurs with exploiting the external world (Ballard et al., 1995; Böing et al., 2023, 2025, in press; Draschkow et al., 2021; Droll & Hayhoe, 2008; Gray et al., 2006; Gray & Fu, 2004; Grinschgl, Papenmeier, et al., 2021; Hoogerbrugge, Strauch, Böing, et al., 2024; Kvitelashvili & Kessler, 2024; Melnik et al., 2018; Meyerhoff et al., 2021; Risko & Gilbert, 2016; Sahakian et al., 2023; Somai et al., 2020; Waldron et al., 2007), memory assessment generally does not allow for nor reflect the use of such external strategies but requires full memory capacity for successful task completion. Measures of capacity (e.g., how much one *can* remember) therefore lack specificity in testing how one *uses* their working memory when given the opportunity to use the external world as memory buffer. We have little objective insight into patients' spontaneous offloading behaviour when engaging working memory in interaction with the environment, while it is exactly this behaviour that is targeted with memory rehabilitation. With the overarching aim to objectively approximate individuals' working memory use after stroke, we tracked the eyes of inpatient and outpatient survivors of a cerebrovascular accident (CVA) and healthy controls while they performed a copying task. Participants could choose their preferred working memory load, or could exploit the outside world as a strategy to avoid loading working memory (thus, offloading by (re)inspecting external information). Importantly, external information was either immediately available to inspect (low-cost) or after a delay (high-cost) to investigate whether and how individuals would adjust their offloading behaviour in response to this environmental change. As sub aims, we 1) compared CVA patients and healthy controls in offloading behaviour, 2) explored offloading at the individual level to distinguish different and find predominant strategies across individuals, 3) interpreted offloading behaviour from a clinical viewpoint, and 4) explored whether and how strategy was adjusted when information was less readily available, and how this influenced performance for the individual.

We observed distinct inspection behaviour for the CVA population as compared to healthy controls. The majority of patients relied heavily on offloading, where it was common to inspect multiple times to place one item correctly. Critically, a subset of patients showed excessive offloading when information was readily available: they showed up to four inspections with long encoding times to place a single item correctly. Interestingly, this subset of patients comprised those who had decreased memory capacity as measured by traditional neuropsychological tasks. Similar to controls and the remainder of the CVA population, these patients decreased their inspection frequency in response to a situation where information was less readily accessible. However, they still relied heavily on external information, making up to two inspections for a single placement. Intriguingly, there was also a number of individuals (one patient and multiple controls) that showed abnormal memory capacity scores but did *not* show distinct inspection behaviour.

Eye-movement behaviour in our Copy Task thus reveals that one can have a *capacity* problem while not being excessively reliant on the outside world, whereas others' *capacity* problems actually result in heightened offloading levels.

Our findings suggest that while all patients who showed deviations in offloading also showed reduced memory capacity, capacity deficits do not automatically result in distinct offloading behaviour. This difference in directionality may be explained by the extent to which the individual can use other cognitive functions that support working memory function in dynamic situations, such as decision making, monitoring/updating, and planning. Although we did not collect data on other cognitive functions to test this hypothesis, we speculate that some patients exhibit (subtle) deficiencies in these executive functions that could influence the way they have dealt with the more complex and interactive Copy Task. For example, systematicity – systematically copying items from left to right, and top to bottom – may be a monitoring and planning strategy that supports memory functioning on our task. Patients with difficulties in monitoring/updating and planning would not have the ability to draw from this source and show more disorganized, and hence, worse, performance instead (Sahakian et al., in prep.). By lack of structure, a higher number of eye-movements towards the example puzzle may be needed to (re)localize remaining items. Further, as was also observed in our patient population, stroke often has motoric consequences (Hendricks et al., 2002; Langhorne et al., 2009). Slowed motor responses (here, longer mouse movements) increase the delay over which information has to be retained in working memory. Longer delays necessitate longer encoding times (Sahakian et al., 2024), and may explain decay of information for which an additional inspection is then executed. In addition, hemiparesis of the executive hand leads to decreased motor skills (Hatem et al., 2016), meaning that -even though one may be able to dress, eat, and operate the mouse- the patient may have had to put in more effort (physically, but also mentally) to initiate and act out goal-directed mouse movements. Logically, motoric deficits can be present for the contralesional hand, but even the ipsilesional hand may show decreased motor functioning (Johnson & Westlake, 2021; Smith et al., 2023; Winstein & Pohl, 1995). The simultaneous use of mental resources for a motor task, specifically for precision grip movements, and for internal memory storage, may result in reduced working memory performance (Xie & Zhang, 2023), driving a greater tendency to rely on the external world.

Apart from influencing dexterity and upper extremity functioning (Mani et al., 2013), lesion side is also likely to have influenced the type of memory problems that are encountered. Individuals with damage to the right hemisphere are described to show impaired immediate visual memory while immediate verbal memory abilities are intact, whereas left-sided lesions more often result in impairments in verbal memory (Logie, 2011). Regarding copying behaviour and visuospatial memory, dissociations in functionality between both hemispheres can be very subtle: van Asselen et al. (2008, 2009) used a task that assessed categorical versus coordinate spatial object-location bindings. Only a small difference in task layout, such as adding grid cells to a screen instead of presenting an empty screen, already elicited differential processing between the hemispheres with a dominance for the left versus right hemisphere in processing categorical and coordinate spatial representations, respectively. However, as soon as spatial information had to be integrated with object information and recalled (as is needed in our Copy Task), this lateralization effect disappeared. Unfortunately, right sided lesions were underrepresented in our sample, and our sample was too small in general to draw conclusions on the effect of laterality of the CVA on inspection behaviour and the type of memory subprocesses involved. Further research should elucidate potential lateralization effects.

Irrespective of these considerations, using inspection behaviour as an index of offloading advances our knowledge of how memory impairment drives working memory deployment in dynamic visual tasks. In a previous and similar study, we found that patients with severe amnesia (Korsakoff's syndrome) relied on the external world disproportionately as compared to healthy controls (Böing et al., 2023), and another study further confirmed that memory capacity was related to inspection behaviour (Böing et al., 2025). Those with low capacity relied more on the external world than those with high capacity, and those failing neuropsychological capacity tasks were most inclined to use a high number of inspections, which aligns with the current findings. Importantly, however, mere capacity could not fully account for inspection behaviour (Böing et al., 2025). That is, an increment in memory capacity does not directly translate to a similarly large increment in actual memory deployment. Theoretically, visual working memory capacity is about four items (Cowan, 2001). Practically, the lowest verbal and visual working memory capacity spans in the current study were three and four items, respectively, but we barely observed anyone loading three or more items per iteration and thus using their full potential. Thus, people generally avoid using their full memory capacity.

It has been suggested that effort minimization is at the root of offloading behaviour (Burnett & Richmond, 2023; Gilbert, 2015a; Meyerhoff et al., 2021; Risko & Dunn, 2015; Van der Stigchel, 2020). Although we did not specifically test this, one may argue that those who experience memory deficits have to put in relatively more effort (e.g., longer encoding, more conscious processing of information) to arrive at the same memory performance as compared to individuals that do not experience such difficulties. This would explain the high degree of spontaneous reliance on the external world for those who show abnormal capacity scores (also in Böing et al., 2023, 2025), even if inspecting information was hampered by adding a delay. Although we cannot state that individuals who are recovering from stroke do really *need* these frequent inspections per se, we do argue that frequent inspecting is the behaviour of choice when given the opportunity. Further, regardless of baseline inspection frequency, it seems beneficial to adhere to one's inspection frequency even when the environment may encourage a higher degree of memorization (e.g., due to the delay). Those who remain relatively constant in their inspection frequency perform more stable as compared to those who show a larger decrease in inspection frequency. This indicates that changing one's strategy to a more memory-based strategy comes at a cost of performance, which would support the use of external strategies in memory rehabilitation. However, this finding does not align with results of an earlier study that used a similar web-based copy task (Böing et al., in press) where we found that individuals who made larger changes in inspection frequency did not show worsened performance. We attribute this difference to the assumed homogeneity and characteristics of the population tested in the web-based study (mostly adults under 40, without neurological and psychiatric conditions), and task-specific differences with regards to inspecting external information (using a cursor on the web versus eye-movements in the lab). Although making saccades towards external information comes at a cost (Koevoet et al., 2024), other physical processes do as well (Mehta, 2016; Morel et al., 2017; Xie & Zhang, 2023). Directing the cursor to uncover external information (e.g., hand-eye coordination) may thus require more resources than making just a saccade. This would make frequent inspecting in a cursor-based version less attractive due to heightened costs, and would lower the threshold for memorizing to be more cost-efficient. The differences between studies again emphasize the highly interactive and fluid nature of a trade-off between inspecting information externally versus internalizing information in working memory.

Altogether, our study highlights that a vast majority of people – both patients and controls – avoids memory loading and heavily relies on the environment. Offloading strategies vary between individuals, with a subset of patients showing excessive reliance on offloading. We found that reduced memory capacity does relate to, but does not automatically result in, offloading behaviour, meaning that capacity tasks disguise nuances in everyday memory use. People start to memorize more if external information is less readily available, but adhering to offloading seems more adaptive than starting to use a more memory-based strategy, supporting the use of external strategies in memory rehabilitation. Although strategies employed on our Copy Task are, of course, not identical to how one engages working memory in interaction with a home or city environment, they do bring us a step closer to estimating memory functioning in activities of daily living. The assessment and rehabilitation of memory should acknowledge these (nuances in) the use of strategies in general and during stroke recovery.

Supplementary Materials: Additional descriptions, tables, figures and analyses that were reported in the manuscript can be found in the Supplementary Materials.

Author Contributions: Sanne Böing: conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, visualization, writing – original draft, writing – review and editing. **Antonia F. Ten Brink:** conceptualization, methodology, resources, software, supervision, writing – review and editing. **Tanja C.W. Nijboer:** conceptualization, methodology, supervision, writing – review and editing. **Stefan Van der Stigchel:** conceptualization, funding acquisition, methodology, supervision, writing – review and editing. All authors have read and agreed to the published version of the manuscript.

Collaborators: Working Memory Study Group: Marike Jansen, Mirjam Kouwenhoven, Cindy van de Moosdijk, Patricia Roggeveen

Acknowledgements: The Working Memory Study Group from De Hoogstraat Rehabilitation consisted of two psychologists and an occupational therapist. We explicitly thank M. Jansen, C. Van de Moosdijk, P. Roggeveen, and D. Leendertse who were involved during recruitment and/or helped collecting data that was part of standard clinical care. K. Wouda and M. Kouwenhoven assisted with embedding the logistics of the study in the care facility. We thank A. Visser-Meilly for supporting the enrollment of the study and her perspective on the topic. M. Wilms assisted us in embedding the study and patient recruitment in the Center for Geriatric Rehabilitation De Parkgraaf. The authors further thank T. Hamers and B. de Zwart for helping with data collection of healthy controls. We thank A.J. Hoogerbrugge for programming the copy task, and supporting with data pre-processing.

Funding: This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement 863732).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the Faculty Ethics Review Board of the Faculty of Social and Behavioural Sciences at Utrecht University (protocol numbers protocol numbers 21-0485, 22-0069 and 22-0284). The local ethics committees of De Hoogstraat and De Parkgraaf also approved the study.

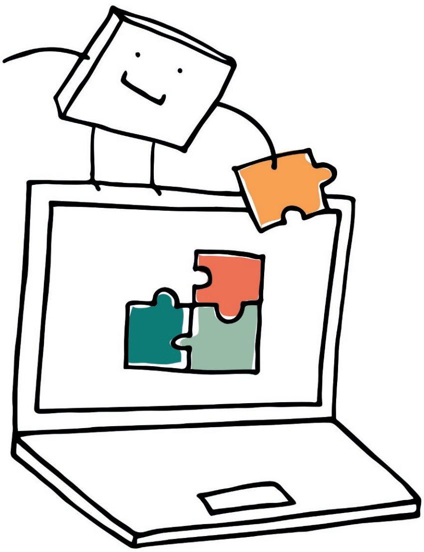
Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent was obtained from the patient(s) to use their data in this paper.

Data Availability Statement: The data presented in this study will be openly available at Open Science Framework upon paper publication at https://osf.io/95zx7/.

Declaration of Interest: The authors declare no conflict of interest.

CHAPTER 6

Summary and General Discussion



The general objective of this dissertation was to investigate how variability in the *internal storage capacity* influences the trade-off in working memory, and how memory *potential* translates to memory *usage*. In **Part I**, *I* summarize the main finding of each chapter, and provide an integrative analysis to further substantiate my claims. In **Part II**, *I* discuss theoretical and clinical considerations and mark suggestions for future research and clinical practice.

Part I: summary and additional analysis

Summary of main findings

The trade-off between external sampling (i.e., offloading) and internal storing (memorizing) has been often investigated by manipulating the cost of sampling. For example, experimental paradigms implemented delays to access external information (e.g., Sahakian et al., 2023; Somai et al., 2020), a physical distance to reach external information (e.g., Draschkow et al., 2021), or varied the reliability of external information (e.g., Hoogerbrugge et al., 2024). While higher costs to accessing external information evoke a greater use of internal memory storage, these studies also showed that healthy individuals are reluctant to use full memory capacity if not directly necessary. These findings hint at working memory as an interactive and adaptive system whereby internal storage is tuned to environmental demands. In this thesis, *I* approached this subject from a different perspective, and tested how interactive working memory would be influenced with changing costs of internal storage due to brain injuries that may result in decreased memory functioning. *I* hypothesized that patients would rely more on offloading (i.e., external sampling) to alleviate memory burden than those with normal memory function, even when it was relatively costly to access external information.

In **Chapter 2** we showed that patients with severe amnesia (Korsakoff's syndrome; n=24) indeed had a strong tendency to rely on the outside world (expressed in inspection frequency and inspection time), more so than healthy controls. Although patients did attempt to adjust their behaviour to changing information availability by decreasing reliance on the outside world and increasing use of internal memory, they failed to do so effectively as shown by a disproportionate decrease in accuracy and an increase in speed as compared to controls. This indicates that this memory-impaired group shows a baseline reliance on the external world that is higher and remains higher even when the external world is less readily available.

For individuals referred to an outpatient memory clinic (n=29, **Chapter 3**) a similar result was expected. Yet, these individuals inspected external information as often as controls, but used longer inspections and showed worse speed-accuracy performance. The absence of a difference in inspection frequency can be explained by the fact that memory capacity was more similar between controls and referred individuals than between controls and patients with Korsakoff's syndrome (**Chapter 2**). When assessing the effect of memory capacity rather than the effect of referral, we found that higher memory capacity was related to both fewer and shorter inspections. Memory potential thus translated to memory usage. Interestingly, there was no effect of memory self-efficacy, that is, negative beliefs about one's memory function did not yield a heavier reliance on the outside world.

In **Chapter 5** we found that relying on the external world was also common among CVA patients (n=15). Some patients demonstrated an unusually strong dependence on the external world. Notably, these individuals also scored abnormally on conventional memory capacity tests, but the inverse was not always true. As controls with abnormal capacity scores did not show such excessive reliance on the external world, we cannot attribute the excessive reliance on the external world of the CVA subgroup to their decreased memory function solely. *I* suggest that there are additional factors driving reliance on the external world in the CVA subgroup.

The extent to which one relies on the external world varied at the individual level. In **Chapter 4** and **Chapter 5** low-loaders, medium-loaders and high-loaders were discriminated to describe their working memory strategy. Low-loaders were those who inspected external information more than once per item. This translates to a memory load of fewer than one item. Medium-loaders loaded one or more items per inspection. High-loaders (or 'full-loaders') were defined as those who memorized three items or more per inspection. Crucially, low- and medium-loading strategies were common, but individuals seldom exhibited a high-loading strategy that approached working memory capacity limits. Despite individual differences in sampling, healthy participants in an online study (**Chapter 4**) flexibly adapted sampling frequency to changing task demands: neither baseline sampling frequency nor adjustments in sampling frequency across conditions impacted speed-accuracy performance. Contrarily, in the lab-based paradigm, healthy adults and CVA patients (**Chapter 5**) who made larger adaptations in inspection frequency showed a larger drop in speed-accuracy performance. For these groups it seemed more beneficial to adhere to baseline inspection frequency – even when conditions changed – and those who continued to offload seemed to be least affected on performance.

The translational value of traditional memory capacity tasks across clinical populations

A sub aim of this thesis was not solely to detect differences between patients and controls, but also to address the influence of (impaired) memory capacity on inspection behaviour and the trade-off between external sampling and internal storing per se. This perspective helps to narrow the gap between rigid memory usage in clinical capacity tasks and more dynamic memory usage in everyday life. The relation between memory capacity scores and reliance on the outside world was investigated in **Chapter 3**, using data from healthy controls and individuals referred to an outpatient memory clinic. In **Box 1** and **Box 2**, *I* extend this analysis by adding the remainder of the patient data collected for this dissertation (patients with Korsakoff's syndrome and those recovering from a CVA).

On this larger scale, lower memory capacity is again associated with a heavier reliance on the outside world. The difference between those with low and high memory capacity is more pronounced when external information is continuously available to inspect than when external information is less readily available to inspect. This is contrary to the expectation that group differences would express itself mostly in the high-cost condition; we expected that those with memory problems would take the increased sampling costs for granted to alleviate the memory burden, and thus sample relatively more often than controls. It is thus surprising that inspection behaviour in the low-cost condition is more sensitive to differences in capacity than the high-cost condition.

Across all groups, high-loading – remembering three items per inspection or more – occurred only in rare cases. The majority inspected the example almost twice to place one item correctly when information was freely available, meaning that they remembered only half an item per inspection. When information was less readily available, the majority inspected the model slightly more than once per correctly placed item. This translates to memorization of only one item per iteration. Generally speaking, only one item or less was placed, which is well below the maximum memory capacity.

Taken together, inspection behaviour in the low-cost condition shows a more pronounced difference between groups. In the high-cost condition, however, behaviour appeared to become more condensed, indicating that a freedom-of-choice task in which information is readily available best captures individual variation related to memory capacity. Nevertheless, in either of the two conditions, the absolute number of inspections per correct placement indicates that there is a clear gap between memory potential and memory usage in a freedom-of-choice task. Capacity is thus related to the degree to which one uses external sampling, but cannot fully account for the heavy reliance on the external world in patients and controls.

Box 1. Memory capacity translates to inspection behaviour (1/2)
Note on methods
Differences across populations in protocol and design are equalized by only including the first 15 trials of each condition for each group. Data used are therefore slightly different from those used in **Chapter 2** and **3**. Different sample sizes are present due to continued inclusion of healthy controls after completion of the study reported in **Chapter 2**. Memory capacity compound scores were calculated for each individual by averaging the individual's z-scores on subtasks Rey Auditory Verbal Learning Task – direct recall (over five trials), Location Learning Task – displacement errors (over five trials), Digit Span forward span, Digit Span backward span, Corsi Block Tapping Task forward span, and Corsi Block Tapping Task backward span. The analysis largely follows the methods that were described in **Chapter 3** (Böing et al., 2025). The **Supplementary Materials: General** provides a group overview for demographics, scores on memory capacity tasks, and group inspection behaviour and individual strategy behaviour on the Copy Task.

Results
Fewer inspections per correct were made in the high-cost condition (t = -17.34, p <.001, beta = -1.13 [-1.25, -1.00]) and higher memory capacity compound scores further led to fewer inspections (z-scores; t = -4.15, p <.001, beta = -0.43 [-0.63, -0.23]), while accounting for effects of age (t = 4.46, p <.001, beta = 0.01 [0.01, 0.02]) and education (t = -2.95, p = <.005, beta = - 0.07 [-0.12, -0.02]). Those with a higher capacity thus memorized more information per iteration than those with a lower capacity, see Figure 6.1A. A marginal interaction showed that when information was freely available, a higher capacity resulted in a larger decrease in the number of inspections (beta = -0.54) than when information was less readily available (beta = -0.25; t = 2.05, p = .043 , beta = 0.18 [0.01, 0.36]).

Similar effects are found for the relation between the memory capacity compound and inspection time per correct placement and performance (see Figure 6.1): those with a higher capacity dwelled shorter to place one item correctly (t = -5.07, p <.001 , beta = -0.47 [-0.65, -0.29]), and their performance was better (lower LISAS indicates better performance) than performance of those with lower capacity (t = -5.59, p <.001 , beta = -0.37 [-0.50, -0.24]), while controlling for effects of age and education (Age: t = -1.48, p =.141 , beta = 0.01 [-0.003, 0.02]; t = 2.64, p < .01, beta = 0.01 [0.00, 0.02], respectively; Education: t = 0.41, p = .685, beta = 0.02 [-0.08, 0.13]; t = -1.61, p = .11, beta = -0.06 [-0.14, 0.01], respectively). No interaction effects were present. Covariates are not taken into account in Figure 6.1, so actual slopes may differ from those depicted.

Box 1. Memory capacity translates to inspection behaviour (2/2)

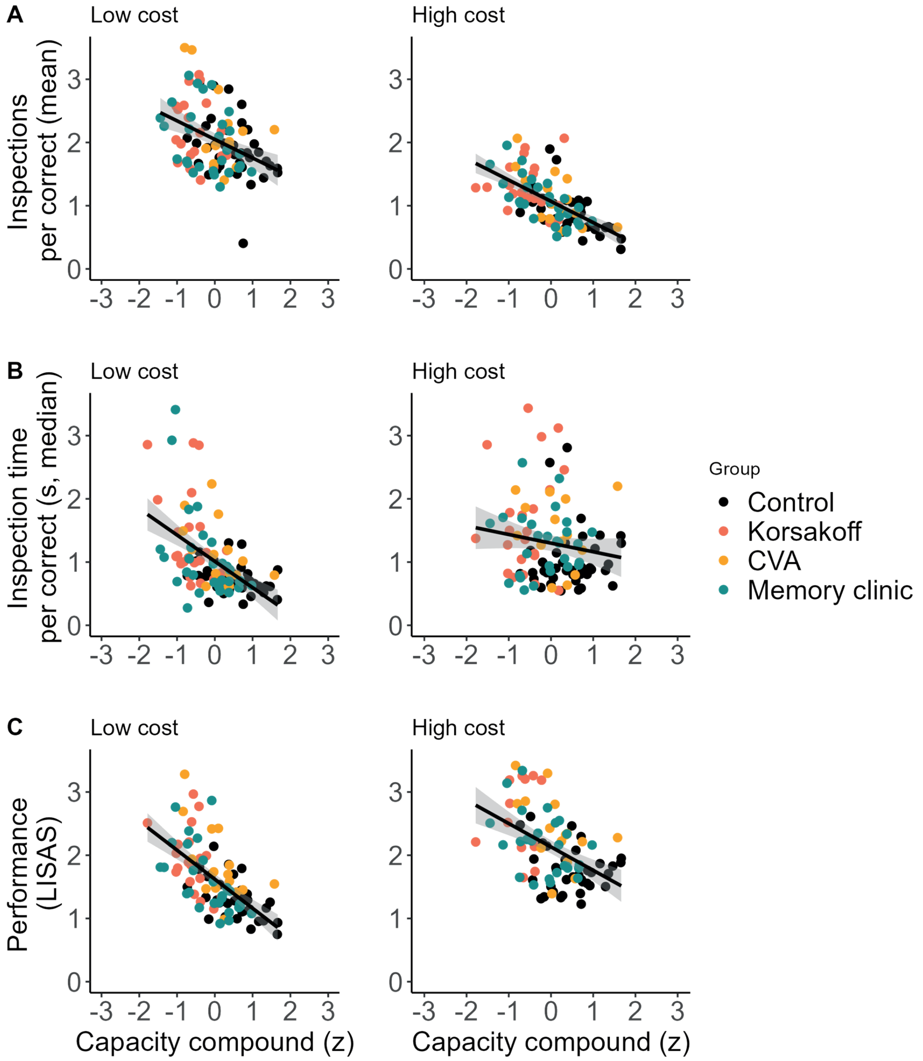


Figure 6.1. **A)** The number of inspections per correct placement, **B)** inspection time per correct, and **C)** performance LISAS as function of memory capacity compound z-score, split on condition, where each datapoint represents an individual from one of the four groups. Black = Controls, red = patients with Korsakoff's syndrome, orange = patients recovering from a cerebrovascular accident (CVA), cyan = referred to an outpatient memory clinic. A smoothed linear correlation coefficient is added in black with confidence intervals in grey. Note that covariates are not taken into account in this figure, and that analyses are performed using linear mixed-effect models. Factor estimates may therefore differ from the correlation depicted in the figure.

Box 2. Inspection behaviour as function of memory classification.

The capacity compound includes absolute capacity scores (e.g., verbal working memory span of four, ten displacement errors etc.) but does not give a clinical interpretation. Therefore, we assessed memory function for each individual (regardless of patient group) and indicated whether their memory function was intact, below average or extremely low (following classification labels of Hendriks et al., 2020). A factorial ANOVA showed that inspection frequency was higher for those with an extremely low memory capacity than for those with a below average or intact performance (see Figure 6.2. The latter two groups did not differ. The effects between groups were driven by behaviour in the low-cost condition: when comparing the number of inspections for the three classes (with Tukey-corrected post-hoc comparisons) across the low-cost condition the exact same effects were found, while no differences were present in the high-cost condition. Surprisingly, thus, inspection behaviour in the low-cost condition is more sensitive to differences in capacity than the high-cost condition, which is also subserved by the earlier linear mixed-effect model analyses.

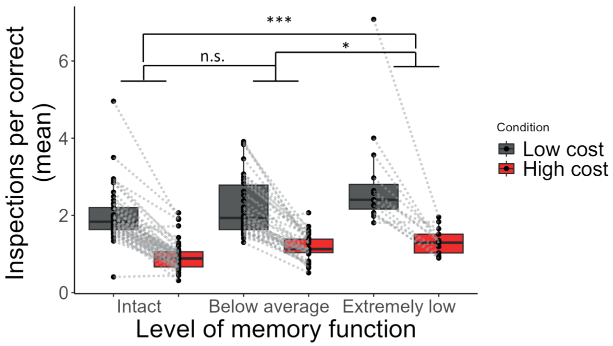
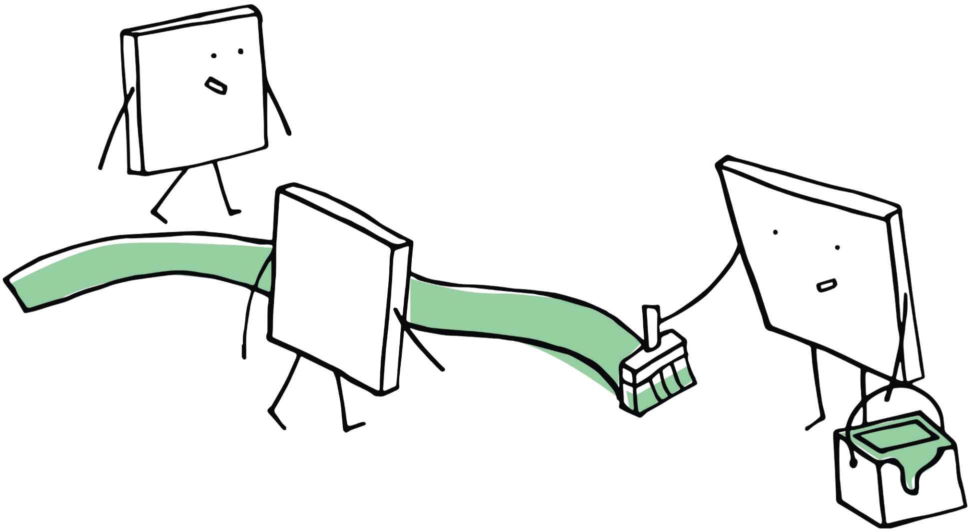


Figure 6.2. Inspections per correct as a function of memory performance classification. Groups were merged, and performance was categorized as intact, below average or extremely low according to clinical consensus; Raw capacity scores were compared with the appropriate norm groups (for age and education). If performance was below the 2nd percentile on two or more subtasks, memory capacity was categorized as extremely low. If performance was below the 2nd percentile on one subtask *(and/)or* below the 9th percentile on two or more subtasks, memory performance was categorized as below average. If these criteria were not met, memory performance was deemed within normal range. Both classification and condition significantly influenced inspection frequency ($F(2)=11.90$, $p<.001$, $F(1)=145.07$, $p<.001$, respectively). An interaction effect was absent ($F(2)=1.67$, $p = 0.19$). Tukey corrected post-hoc comparisons revealed that the extremely low group (n=14) differed significantly from the intact (n=53) and below average group (n=38), but that the intact and below average group did not differ from each other. This was true for the low-cost condition, but no differences were present for the high-cost condition.

6

Part II: general discussion

The aim of the general discussion is to paint the picture from theory to practice: *I* discuss a non-exhaustive list of theoretical and clinical considerations that should be taken into account when interpreting inspections as an index of memorization. *I* mark suggestions for future research and clinical practice, and *I* end with some closing statements meant to spark debate.



Inspections as index of internal memory storage
One of our main claims is that memory capacity is not fully utilized, which we infer from the number of inspections needed to place all information correctly. Is it valid to use inspection behaviour as an approximation of memory storage? In this section *I* discuss various factors that one should be alert to when making claims about using inspections as index of internal memory storage.

It is by virtue of the visual system that we are able to internalize, and thus memorize, external visual information. The eyes must be directed towards an object within a two-degree visual angle for the visual system to reliably encode it (Nelson & Loftus, 1980), with longer viewing times allowing for better encoding (Koevoet et al., 2023; Sahakian et al., 2024; Vogel et al., 2006). It is therefore appealing to infer memorization from gaze location and duration, hence inspection behaviour. Conclusions drawn in this thesis indeed rest upon the assumption that inspections carry information about memory usage. Many other studies have also relied on inspection metrics to make inferences about working memory utilization across conditions in freedom-of-choice paradigms, where the *number* of model inspections seems most specific as an indicator of shifts between external reliance and internal memory use (Qing et al., 2024). The duration of inspections was found to be somewhat less specific to strategy changes, and even more unspecific were outcomes such as completion time and incorrect placements (Qing et al., 2024). The latter do not directly capture the *shift* per se, but rather convey the *effects* of a shift. For example, patients with Korsakoff's syndrome (**Chapter 2**) tried to shift their strategy (as indexed by a decrease in the number of inspections and an increase in inspection duration to encode information), but their attempt was unsuccessful: their performance in terms of both speed and accuracy dropped disproportionately as compared to that of healthy controls. Similarly, we showed that those who adhered to a more stable inspection frequency across conditions were less hampered in performance than those who tried to adapt their inspection frequency to a larger extent. Comparing performance behaviour from one condition to the other is thus more suited to specify switch costs within the individual (**Chapter 5**) than to specify a strategy change per se (Qing et al., 2024).

Changes in the number of inspections are thus used to index *shifts* in internal memory usage, but what does the standalone number of inspections indicate? By dividing the number of correctly placed items per trial by the number of inspections per trial we infer the average memory load per inspection within a trial. However, fluctuations in memory use within a trial are discarded by this measure. To illustrate, an individual may load up three items on the first inspection, places all of them (memorizing three items, i.e., high-loading behaviour), and then inspects and reinspects the model for a total of five times for the remaining three items (memorizing fewer than one item, i.e., low-loading behaviour). With a total of six inspections to correctly place six items, this trial would be marked as a trial in which the individual would memorize one item per inspection (i.e., medium-loading). The measure thus only approximates the general memorization tendency of the individual. The same holds when aggregating over trials. Inspections are thus a rather crude measure on the level of the task as well as on the level of the condition or trial.

Even in its basic operationalization, a single inspection includes more than one gaze: inspections are defined as saccades that were directed from the right to the left side of the screen (i.e., a *crossing*) with a model viewing as result. Saccades and fixations that occurred within the area of the example model after its appearance are thus compressed into a single inspection.

A further note is warranted when using the number of inspections *per correct*. This measure only allows statements about a *successful* memorization sequence. In order to place an item correctly, successful encoding, maintenance, recognition and guidance of behaviour are required. Only if all these phases are completed successfully would the individual be able to place the item correctly. If one of these processes fails, an erroneous placement would occur, or the individual would refrain from placing an item. Consider an example where an individual may fixate and encode three items with one inspection, and places two of these items. According to our definition, the individual would have memorized two items. By means of the number of inspections per correctly placed item, we cannot possibly know about the attempt to encode the third item. Due to the lack of item-specific encoding information, and the absence of information on consecutive item-specific placements, in our data we cannot dissociate encoding, maintenance and retrieval phases within the memorization process.

More fine-grained analyses on precise fixation locations and consecutive item placements would allow to better extract what exact external information is processed and used at what point in time. For example, in a healthy sample who completed an online task, the placement sequence after an inspection showed to be sensitive to the strength of the memory representation created during encoding: the item that is placed first is most probably the item that has the strongest internal memory representation, while later item placements indicate 'squeezing out' more uncertain memory content with a higher risk of erroneous placements (Sahakian et al., 2023). If such placement sequences are not only linked to inspections per se, but linked to fixation sequences within the inspection, this may serve as a more direct measure of memorization in patients, also allowing to capture, for example, primacy and recency effects (McAteer et al., 2023) or item prioritisation (Allen et al., 2024). Nevertheless, fixations do not necessarily indicate attention to the fixated object (Schad et al., 2012), nor does reinspecting a previously fixated object necessarily reflect that there was no (remaining) internal memory for that specific object (Sahakian et al., 2023).

Inspections are driven by more than mere information uptake. Even when participants were on their way to reinspect information, they performed above chance on a 2-AFC when their memory content was assessed (Sahakian et al., 2023). This indicates that reinspecting occurred despite a leftover memory trace, implying that reinspecting not only serves to encode novel information, but also serves to reduce uncertainty before making the decision to act upon the available information. Similarly, Hoogerbrugge and colleagues (2024) suggest that individuals use reinspecting to boost metacognitive confidence regarding visual working memory representations: persistent reinspecting occurred even after many repetitions (and thus, learning) of the same material.

Reinspecting can have a twofold effect on boosting memory representations and memory confidence. We know that memory representations become less precise with longer delays between encoding and report/action (Sahakian et al., 2024). Each increment of inspection time can guard against this decay (Sahakian et al., 2024). Therefore, reinspecting may occur not only to strengthen confidence and to accumulate memory representations (i.e., encode), but also to reset the timer of memory decay. Together, a higher number of inspections and longer inspection durations contribute to stronger or more precise memory representations, facilitate a shorter memorization interval, and safeguard a feeling of certainty about the memory representation.

The conclusion that memory capacity is not fully utilized (**Chapter, 2, 3, 5,** and many other studies) does not only rest upon the assumption that inspections index memory usage. It also relies on the assumption that working memory capacity is three to four items (Cowan, 2001). Within our sample of 106 individuals, only five people had a visuospatial working memory span (Corsi Backward) of less than three items. However, for more complex shapes such as the polygons used in our paradigm, this maximum capacity may be decreased (Alvarez & Cavanagh, 2004; Luria et al., 2010; Luria & Vogel, 2011). We therefore cannot make conclusive statements but only speculate about whether or not individuals are fully loading their capacity for the items in our task (Arnoult, 1956). For a one-on-one translation from capacity potential to usage, it would be elegant to implement a memory capacity task with the same stimuli as used in the freedom-of-choice task.

In summary, inspections are suited to reflect a shift in reliance between internal memory and the external world. The absolute number of inspections, however, does clearly carry more implicit information than mere memory storage.

Inspections in the context of memory confidence

In the previous section, *I* discussed how inspections can serve to boost memory confidence. External sampling is often suggested as a way to avoid uncertainty or safeguard a feeling of security in visual working memory tasks (Hoogerbrugge et al., 2024; Sahakian et al., 2023), and this finding extends across various tasks and types of offloading (Boldt & Gilbert, 2019; Burnett & Richmond, 2023; Risko & Dunn, 2015; Risko & Gilbert, 2016). This perspective provides a useful framework for interpreting frequent reinspecting, particularly in patients with reduced memory confidence: we hypothesized that (re)inspecting behaviour on the Copy Task could serve as a proxy for an individual's *belief* about their own memory functioning, e.g. metamemory or memory self-efficacy. We expected reinspecting to occur as an instance of checking behaviour in patients (or controls) with general concerns about their memory. In the following section, *I* discuss our findings on inspection behaviour in the context of memory confidence.

To our surprise, we found no relation between perceived memory self-efficacy and reliance on the external world in controls and patients referred to the memory clinic (**Chapter 3**). This conclusion held when adding the CVA population to the analysis: lower memory self-efficacy did not relate to a higher reliance on the outside world. The discrepancy between the previously mentioned studies and ours probably arises because of differences in the used construct of memory confidence; metacognitive beliefs, metacognitive experiences, uncertainty and memory self-efficacy are related but not synonyms, and they can vary across different situations. For example, one may not be confident in terms of their memory use in daily life but confident in relying on their internal representation when copying geometrical items on a computer task (this dissertation), or one may be confident on their performance on one type of memory task without this influencing behaviour on another task (Grinschgl, Meyerhoff, et al., 2021). These cases illustrate the distinction between metacognitive *beliefs,* i.e., a generic gist about one's memory function, and metacognitive *experiences*, reflecting confidence about memory performance on a specific task (or trial) that had just been completed (Hertzog and colleagues 1987, as cited in Beaudoin & Desrichard, 2011; Grinschgl, Meyerhoff, et al., 2021). Our study investigated metacognitive beliefs rather than metacognitive experiences: the questionnaire that we used (Metamemory in Adulthood; Ponds & Jolles, 1996a, 1996b) inquired about general memory self-efficacy in everyday life, but did not capture metacognitive experiences across the test procedure (e.g., after an (un)successful copy task trial). A relation between metacognitive experiences and inspecting behaviour could emerge if confidence in working memory performance was assessed just before or after each trial in the Copy Task. The timing of such a confidence assessment is also important. Anticipated performance may differ from experienced performance, and therefore differentially affect the tendency to either rely on external information or not. Those who had a successful memory experience would be less inclined to externalize information as compared to people who had a less (or no) successful experience (Gray & Fu, 2004; Risko & Dunn, 2015). As such, one could learn over the course of trials what the most optimal strategy would be given memory ability, and therefore metacognitive experience can be fluid over time even within a task.

Metacognition is a complex construct in itself (Fleming & Lau, 2014; Terneusen et al., 2022). Similar to the discrepancy between subjective memory concern and objective memory capacity (Beaudoin & Desrichard, 2011; Mattos et al., 2003; Ponds et al., 1997; Ponds & Jolles, 1996a), metacognitive insight may or may not be tuned to actual memory performance across a variety of tasks. This makes a specifically interesting case for those who lack illness insight, which is often found in individuals with Korsakoff's syndrome (Arts et al., 2017; Walvoort et al., 2016) or a progressed neurodegenerative disease (Howorth & Saper, 2003). Metacognition requires an overarching sense of task requirements, task complexity, cognitive ability and assessing the distance between what is required and what can be reached for the individual. Therefore, it is reminiscent of the capacity to self-monitor and links to executive functioning (Rhodes & Kelley, 2005; Roebers, 2017). Executive functioning may therefore modulate metamemory more than memory functioning itself. Similarly, since working memory is required for complex cognition, decision-making, and problem-solving (Baddeley et al., 2021; Logie et al., 2020; Shelton et al., 2010), it is also often seen as part of executive functions. If we consider memory span within the memory domain, but flexible working memory use and metamemory within the executive domain, the ability to adaptively employ the working memory trade-off may depend more on executive functioning than on memory capacity. Being able to assess, monitor and flexibly adopt the most optimal strategy *given* memory capacity would then depend on intact executive functioning. With impaired executive functioning, one may not have the insight to adaptively employ the trade-off using their (knowledge on) memory capacity. In patients with Korsakoff's syndrome, executive deficits are often present next to memory deficits (Arts et al., 2017; Brand, 2007), and this may explain why these patients were not able to adaptively adjust their trade-off *given* their memory deficits; they may be overconfident and not attuned to their ability.

Individual characteristics can also play a role in the mismatch between objective memory ability and metamemory. Individuals referred to the memory clinic that presented with concerns but did not show objective impairments were often highly educated. Adults with a higher level of education may experience a steeper decline of subjective memory performance (Hülür et al., 2015), thereby underestimating their actual memory function. These individuals are used to setting a high bar, and may be extra aware of small changes in cognition. It could also be that these individuals score high on trait neuroticism, which is found to result in a higher degree of memory complaints (Ponds & Jolles, 1996a). To elucidate to what extent personal characteristics may play a role in deciding for a strategy in the trade-off, it would be valuable to test people with specific tendencies as observed in, for example, individuals with obsessive-compulsive disorder (Karadag et al., 2005; Tolin et al., 2001), or high levels of performance failure anxiety. Observationally, one participant in our sample experienced an extreme level of failure anxiety but had no memory deficits. This participant would be expected to rely more heavily on the outside world (here as expression of checking) than other controls due to this failure anxiety. However, this hypothesis was not supported by single case statistics (Bayesian Test of Deficit with Covariate age, p = .07). Future research should elucidate whether and how these characteristics are influencing offloading in general, and external sampling specifically.

What does all this imply for assessment of memory confidence using eye-tracking? To most important take-home message is that inspection behaviour cannot serve as an objective marker of subjective beliefs about memory function in everyday life activities, nor can it detect those with subtle deviations in reliance on the external world due to these potential memory insecurities. However, when metacognition is tied to the task at hand, inspection behaviour *may* allow to detect subtle differences between those with positive and negative memory appraisal.

Inspections in the context of effort

It is suggested that effort minimization is at the core of choosing for either internal storage or external sampling (Van der Stigchel, 2020). The definition of effort varies cross-disciplinary while it has gotten considerable attention in psychology, neuroscience and behavioural economics (for a review of the concept of psychological effort, see Thomson & Oppenheimer, 2022). *I* discuss various factors that influence the effort expenditure in healthy controls and patients in the trade-off between external sampling and internal storing.

Increased cognitive effort at higher memory loads

Higher working memory loads are associated with higher cognitive effort, which becomes apparent in subjective reports (Crawford et al., 2023; Kurzban et al., 2013) as well as brain activation (Engstrom et al., 2013; Kardan et al., 2020) and comparisons with physical exertion (Xie & Zhang, 2023). Although higher loads proportionally increase the associated effort, increased cognitive effort may even appear beyond working memory capacity (Kardan et al., 2020). Our results fit well with the idea of effort minimization in the context of effortful memorization: even though individuals may be able to successfully store information at higher working memory loads, the effort associated with memorizing at higher loads may be avoided by all. The trade-off thus slightly tilts towards memorization when sampling becomes more effortful, but both patients and controls continue to structurally underutilize memory capacity (**Chapter 2, 3, 5**; Cowan, 2001; Luck & Vogel, 2013). To quantify this interpretation, a subjective and/or objective measure of effort exertion could be taken into account in future studies using the trade-off framework.

Anticipated versus experienced cognitive effort

Anticipated effort can be different from experienced effort (Bambrah et al., 2019; Kurzban et al., 2013). Specific to working memory tasks, anticipated effort seems to be higher than real-time experienced effort (Bambrah et al., 2019), implying that people overestimate how much effort they have to exert during a working memory task. This misconception could drive a higher degree of offloading in anticipation of the idea that working memory needs to be exerted. If real-time experience appears not to be as effortful as expected, the trade-off outcome could somewhat tilt towards internal memorization. This idea fits with our findings in **Chapter 4:** individuals that started with the high-cost condition were, by means of the manipulation, primed to start using memorization right away. As such, this group showed a relatively stable use of internal storage, even when transitioning to the low-cost condition. These individuals may have *experienced* that the effort of using memory at higher load was lower than the *anticipated* effort of memorization for those who started with the low-cost condition. This may explain why the groups differed in the internal memory load in the low-cost condition: the presentation order determined whether the individual had experienced a positive effort evaluation for using internal memory (high-cost first) or was left with a misconception of anticipated effort, leading to memory avoidance right away (low-cost first).

Motivation to engage in high effort tasks

Although people generally avert cognitive effort when they have the opportunity (Thomson & Oppenheimer, 2022), higher incentives increase the likelihood of engaging in a higher effort task (Inzlicht et al., 2018). In our paradigm there was no direct incentive to engage in effortful memorization, so the incentive to use working memory at higher load did not weigh out the increased effort cost. In other words, there was no motivation to engage in a high effort option. If we were to couple a reward to memorization at higher loads, the trade-off would tilt towards a higher degree of memorization (Inzlicht et al., 2018; Kurzban et al., 2013; Shenhav et al., 2013). Capacity tasks in neuropsychological assessment can be compared to a high-incentive task. Outcomes on such a task are used to guide diagnoses and further care, and patients should therefore be inclined to use their maximum capacity even though they have to exert higher effort.

Sampling effort

Accessing external information requires a physical action, such as the execution of a saccade, head movement, or walking to a different room. Physical acts are in themselves associated with an effort cost (e.g., Chiu & Gilbert, 2024; Morel et al., 2017; Xie & Zhang, 2023) and have to be taken into account in the trade-off to arrive at the most optimal cost-benefit equilibrium. Even at the level of saccades, effort minimization drives target location selection (Koevoet et al., 2024). By using a gaze-contingent waiting time in the Copy Task, sampling costs were not only added by a delay, but were also tied to the effort associated with the execution of the saccade per se. Our paradigm recruited the ‘cheapest’ type of saccades: the example model and the copying space were presented on the horizontal plane, and horizontal saccades are found to be less costly as compared to vertical and oblique ones (Koevoet et al., 2024). As the task layout remained stable across conditions, differences found across conditions can be attributed to the increased delay and not by saccade costs per se. However, to sample from the example model in the high-cost condition, gaze has to be sustained for two seconds. There may be individual differences with regards to the ability to sustain fixation and inhibit visually driven saccades towards the right side of the screen where the stimuli remained visible (Jarvstad & Gilchrist, 2019; Krauzlis et al., 2017; Maron et al., 2021; Unsworth et al., 2009). In this case, exerting control to keep fixation adds another cost to sampling effort in the high-cost condition.

The paradigm in **Chapter 4** used a mouse-contingent manipulation rather than a gaze-contingent manipulation (**Chapter 2, 3, 5**). Mouse-contingent, head-contingent and gaze-contingent paradigms show qualitatively similar results (Draschkow et al., 2021; Sahakian et al., 2023; Somai et al., 2020), but can yield slight quantitative differences regarding inferred memory load. While mouse-contingent paradigms likely engage the oculomotor system to guide the mouse movement (Anwyl-Irvine et al., 2022; Gray & Fu, 2004), gaze-contingent paradigms do not require the execution of a mouse movement during sampling. Task constraints thus differ (Gray et al., 2006), with the mouse-contingent paradigm carrying a higher physical sampling cost. The trade-off may therefore be somewhat tilted towards memorization in the online paradigm.

Increased effort in patients

On top of the factors discussed above, several factors may influence the effort expenditure for patients specifically. Below, *I* discus increased cognitive effort and increased effort to execute arm, hand, or eye movements as factors that may influence the trade-off between external sampling and internal storage in patients.

1. Cognitive effort.

We hypothesized that individuals with memory complaints or objectified deficits may anticipate or experience a higher degree of effort to use their memory in order to arrive at the same memory performance as those without problems. A hint in favor of this idea lies in findings on brain activation patterns of patients with Kleine-Levin Syndrome, a specific form of periodic hypersomnia which is accompanied by problems in working memory. In order to perform a four item working memory task, patients showed a larger BOLD response than controls throughout the anterior cingulate cortex (ACC) and the anterior insular cortex (AIC), which the authors interpreted as an effort-related activation (Engstrom et al., 2013). While such a difference in effort would go unnoticed in neuropsychological capacity assessment, it could be a factor influencing the tipping point of the trade-off in our patients: those who have to exert higher effort for internal storage may be inclined to take higher sampling costs for granted, and thus rather rely on the outside world to circumvent the increased effort associated with memorization.

Interestingly, though, Aschenbrenner and colleagues (2023) found that lower working memory capacity was not related to higher subjective cognitive effort. In contrast, increased subjective cognitive effort was rather explained by older age in a domain-general manner (Aschenbrenner et al., 2023; Crawford et al., 2022). Clinical populations and healthy controls in our studies (**Chapter 2, 3,** and **5**) were all over 40 years old, as opposed to the sample that partook in the online study (**Chapter 4**). Comparing sampling frequencies and inferred memory loads, we descriptively observe that the younger sample in the online study showed fewer inspections and thus higher memory loads than the older sample in the lab-based studies. Further, we found an effect of covariate age on the number of inspections per correct (see **Box 1**), where higher age led to more inspections and thus lower memory load. These findings may partly be explained by a difference in age-related effort minimization, although it should be noted that the paradigms differed in other influential ways (e.g., coloured items, mouse-contingent vs. gaze-contingent paradigm).

2. Hand and arm movements.

Effort associated with executing hand or arm movements may influence the trade-off between external sampling and internal storage in patients. Motoric consequences are common after a CVA (Hendricks et al., 2002; Langhorne et al., 2009). Hemiparesis of the dominant hand results in reduced motor skills (Hatem et al., 2016). Motoric deficits can affect the contralesional hand, but the ipsilesional hand may also exhibit reduced motor functioning (Johnson & Westlake, 2021; Smith et al., 2023; Winstein & Pohl, 1995). It is plausible that patients therefore have to put in more effort (physically, but also mentally) to initiate and act out goal-directed and coordinated computer mouse movement, regardless of the hand that is used. Similarly, patients with Korsakoff's syndrome may experience impaired fine motor skills (Welch et al., 1997), adding a cost to the execution of a drag-and-drop mouse movement. Using mental resources for a motor task, particularly precision grip movements, and for internal memory storage simultaneously could lead to reduced working memory performance (Xie & Zhang, 2023), driving a greater need to rely on the external world. Further, slowed motor responses (here, longer mouse movements) extend the period over which information has to be retained in working memory. Increased delays require longer encoding periods (Sahakian et al., 2024), and may explain decay of information and the need for an additional inspection.

3. Eye movements.

Eye movement difficulties may further influence the trade-off in patients. Patients with Korsakoff syndrome can show residual nystagmus, involuntary jerk-like movements of the eyes during intended fixation that is not straight ahead (Isen & Kline, 2020; Kattah, 2017; Kopelman et al., 2009). Nystagmus may hamper visual perception and sustained fixation. Although we have not objectively tested patients with Korsakoff's syndrome for this condition, observationally, some patients did indeed show oculomotor deviations (although they did pass eye-tracker calibration and validation). These patients may have experienced difficulty with perceiving the stimuli in the first place, adding a perceptual processing cost. Further, nystagmus may have interfered with keeping gaze at the hourglass in the high-cost condition; if gaze was not at the left side of the screen, the hourglass timer would pause, thereby adding an extra delay to access and encode the example model.

Not mere memory: executive functioning

The Copy Task is not a mere working memory task. *I* already posited that working memory can be considered an integral part of executive functioning as well. Executive functioning was mentioned in this discussion in relation to metacognitive assessment, but the task itself also recruits subdomains of executive functioning (attention, planning, monitoring). Impairments in executive functioning are often observed in Korsakoff's syndrome (Arts et al., 2017; Brand, 2007; Janssen et al., 2023; Kopelman et al., 2009; Maharasingam et al., 2013; Moya et al., 2021). Impairments in executive functioning result in disinhibition, impulsive behaviour and a decreased ability to (self-)monitor, plan and pay attention (Suchy, 2009). A diminished ability to concentrate may result in attentional lapses, which are described as occurrences where acting (here, making an eye movement) is not based on sensory evidence, but rather is an expression of a momentary lapse in attention or memory (Ashwood et al., 2022). Patients with Korsakoff's syndrome further show more impulsive behaviour which may express in both a heightened frequency of inspections, but also in a larger error rate in this population (less conservative criterion to place items, Sahakian et al., 2023). We speculate that some patients recovering from a CVA (Leśniak et al., 2008) and some of those referred to the outpatient memory clinic (e.g., in early Parkinson's disease or FTD; Lees & Smith, 1983; Stopford et al., 2012) also exhibit (subtle) deficiencies in these executive functions that could have influenced the way in which patients have dealt with the more complex and interactive Copy Task. For example, systematicity — systematically copying items from left to right, and top to bottom — may serve as a monitoring and planning strategy that supports memory functioning during our task. Patients with difficulties in monitoring, updating, or planning may lack the ability to employ this strategy, leading to more disorganized and, consequently, poorer performance (Sahakian et al., in prep.). Without a structure to adhere to, these patients may require a higher number of eye movements toward the example puzzle to (re)locate remaining items. As we have not tested executive functioning in our patients, we cannot exclude the effect of (impaired) executive functioning on the Copy Task.

Another aspect of executive functioning is the ability to switch between tasks. Although the same instructions hold for both the low-cost and the high-cost condition (i.e., to complete the trial as fast and accurately as possible), one may argue that transitioning from the low-cost to the high-cost condition involves such a task switch; the individual has to evaluate and monitor whether the strategy that was applied in the first situation is still a good fit for the second situation (Altmann & Gray, 2008). Patients with executive deficits may have difficulty with appraising this novel condition, and may show perseverative behaviour (Oscar-Berman et al., 2004). Moreover, even if one is able to evaluate the fit of their strategy, costs associated with switching may be higher than effort costs associated with the (continuation of) memorization (Gilbert, 2023; Kurzban et al., 2013; Luwel et al., 2009; Xie & Zhang, 2023).

Not mere memory: visual search

When a target item from the example model is encoded in working memory, it has to be searched for in the resource grid before it can be correctly reported (i.e., dragged to the correct cell). As such, items in the resource grid have to be inspected and compared to the internal item representation to select the appropriate target for further use. Thus, while change detection paradigms (e.g., Luck & Vogel, 2013) and neuropsychological capacity assessment (e.g., Corsi, 1972; Kessels et al., 2014; Saan & Deelman, 1986; Wechsler, 2012) require encoding, maintenance and report, the Copy Task additionally requires visual search.

Even in single target search, participants are inclined to make additional inspections to ascertain accurate target selection whenever this is possible (e.g., Hoogerbrugge, Strauch, Nijboer, et al., 2024). When the number of search items is increased (i.e., when copying six items in the Copy Task) multitarget search arises. Searching for two items is more difficult than single item search, and difficulty increases even further for four items (Hoogerbrugge et al., in prep.). Consequently, when target items are not available for reinspection and thus have to be encoded in memory (e.g., reminiscent of the high-cost condition) an increase in search difficulty arises. Most individuals adopt a sequential search pattern when dealing with multiple search targets: they inspect and search for the first item, after which they inspect and search for the second item and so on (Hoogerbrugge et al., in prep.). Nevertheless, some individuals show concurrent search. Opting for either of the two search strategies may again require a metacognitive assessment of which search strategy is most suited/least effortful in a given situation, and switching between strategies may again require executive control.

Additionally, correctly placed items do not disappear from the resources, meaning that five distractors are present when searching for a single item. These distractors may interfere with the memory representation of the search target (Kumle et al., 2024) driving extra inspections (Hoogerbrugge et al., in prep.; Hoogerbrugge, Strauch, Nijboer, et al., 2024; and see 'inspections as index of internal memory storage'). To monitor and suppress items from the resource grid that were already placed, one needs to exert cognitive control and thus use additional cognitive resources. Especially for those with diminished executive functioning, this may be quite a costly challenge. The costs associated with multitarget visual search may be an additional reason for participants to keep their memory load low, leading to an underutilization of memory capacity.

External strategies in memory rehabilitation

Spontaneously relying on information from the external world is reminiscent of external compensation strategies used in memory rehabilitation, where patients are often trained to rely on the outside world to support memory function. Findings from this dissertation suggest that many patients (**Chapter 2, 3, 5**) use the external world even without explicit instruction. In **Chapter 4**, we found similar spontaneous behaviour. When assessing whether any advantage exists for either a low-loading, medium-loading or high-loading there were no differences in performance between categories (in the low-cost condition) and the number of inspections did not correlate with performance. However, for patients and healthy controls (collapsed data from **Chapter 2, 3,** and **5**), a higher number of inspections (thus, low-loading) yields worse performance ($\tau = 0.267$, $p < .001$, $z = 13.821$), indicating that offloading comes at a cost of performance. This is contrary to what would be expected from practices in memory rehabilitation. An important remark here is that it could be that those using a high number of inspections are those who perform worse in general, and that the effect on performance is not necessarily driven by the tendency to offload, but rather by decreased global cognitive functioning. Baseline performance differences should be accounted for. Therefore, we assessed whether changes in inspection frequency (hence, low-loaders become high-loaders, or low-loaders continue to low-load) result in changed performance. Adhering to a similar sampling frequency in response to a changed environment was found to be more beneficial than making a large adjustment to sampling frequency (**Chapter 5**, and also when adding other populations: $\tau = -0.159$, $p = .02$, $z = -2.29$).

An important consideration is that performance is defined as an integrated speed-accuracy score (LISAS; Vandierendonck, 2017, 2021). This integrated score treats both speed and accuracy equally important, while accuracy may be favoured over speed in clinical care. Future research using freedom-of-choice tasks could help to objectify the spontaneous use of external compensation strategies and the effectiveness of instructed behaviour on an individual level.

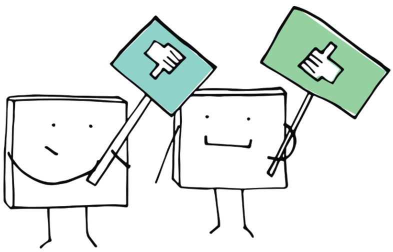
Now what? Considerations for clinical practice

Scores on memory capacity tasks showed to translate to memory *usage* in a freedom-of-choice task. This finding is encouraging for neuropsychological practice. Although traditional neuropsychological tasks are often criticized for their lack of ecological validity, and efforts have been made to incorporate innovative technologies to bridge the gap between clinical assessment and real-world applications (Chaytor & Schmitter-Edgecombe, 2003; Parsey & Schmitter-Edgecombe, 2013), the present study demonstrates that traditional pencil-and-paper tasks do actually translate to behaviour on a more dynamic task. However, they do not tell the entire story as we repeatedly observed that individuals with and without memory impairments mostly underutilize their maximum capacity. Generally speaking, people tend to only memorize one item or fewer, even when information is not readily accessible, and this behaviour seems to be exaggerated in patients with acquired brain injury. It is not just the ability to internally store information that has an influence on how the trade-off between external sampling and internal storage is employed, but many more factors play a role. How should we deal with this knowledge in neuropsychological assessment? In the **Debate Boxes,** *I* provide an experience-based opinion on several questions regarding the administration of capacity tasks and freedom-of-choice tasks in clinical care.

Debate Box 1: Should capacity tasks in neuropsychological assessment be discarded?
Working memory capacity is structurally underutilized, a tendency that is not captured by traditional capacity tasks. Might we thus as well get rid of memory capacity assessment? I argue not. Capacity tasks do not allow for compensation with external strategies, but they do give valuable insight in performance in the most ‘optimal’ circumstances of a distraction-free clinical setting with little recruitment of other cognitive domains. As such, they serve a different purpose than freedom-of-choice tasks: they capture the limits of memory if memory is recruited in isolation. Forced capacity tasks allow to assess this performance against a threshold, and are at this point effective at classifying individual’s performance as extremely low, below average or intact (Hendriks et al., 2020). Therefore, capacity tasks provide a nearly pure and clean assessment of memory subprocesses within different modalities. Freedom-of-choice tasks provide a more dynamic and ‘dirty’ assessment of memory use, confounded with the involvement of other cognitive domains. Capacity tasks should not be discarded as they serve as a solid starting point to test the limits, are well documented, and norm data are widely available.

Debate Box 2: Should we add freedom-of-choice tasks to neuropsychological assessment?
Capacity tasks do not allow for external strategies that individuals may use for whatever reason to alleviate memory burden. Conversely, freedom-of-choice tasks provide information on patients' interaction with the external world while using memory. In an ideal setting, test batteries are attuned to the specific individual and the question at hand. As such, if the goal is not to classify but to explore the flexibility and spontaneous use of workarounds, freedom-of-choice tasks may be particularly insightful. Such insights can help to dissociate those who can maintain performance using workarounds given a potential capacity deficit from those who cannot. Although there are (to date) no threshold or norm values for inspection behaviour, insights into inspection behaviour can offer a more complete picture of how one may (adaptively) use memory in interaction with the environment. Taken together, I advocate using freedom-of-choice paradigms for a more comprehensive cognitive assessment, but only when they are relevant to the clinical question.

Debate Box 3: Is it feasible to administer the Copy Task during clinical memory assessment?
Not in its current form. Several issues should be resolved before clinical application can be considered. First, the task consists of multiple trials per condition (in our design at least fifteen), and the eye-tracker needs to be calibrated before testing. Together, this easily adds up to half an hour of testing time. As patients often have reduced cognitive workload, time is scarce. The length and ease of administration should be adjusted in order for a version of the Copy Task to be suited for clinical assessment. Second, the Copy Task is a computer task. Although digitized neuropsychological testing is considered feasible for patients (Spreij et al., 2020), digitized testing is not yet widely adopted (Parsons & Duffield, 2020). Software and devices are rapidly changing, and so the clinician has to keep up with all these developments to be able to work with and reliably interpret data from these sources. Moreover, the clinician must possess adequate technological proficiency to set up the task and read out the resulting data effectively. Eye-tracking data is not straightforwardly interpreted and requires pre-processing and calculations before a readily interpretable outcome will emerge. Future freedom-of-choice paradigms designed for clinical application should ideally provide easily interpretable, unambiguous values that can be generated with the push of a single button.



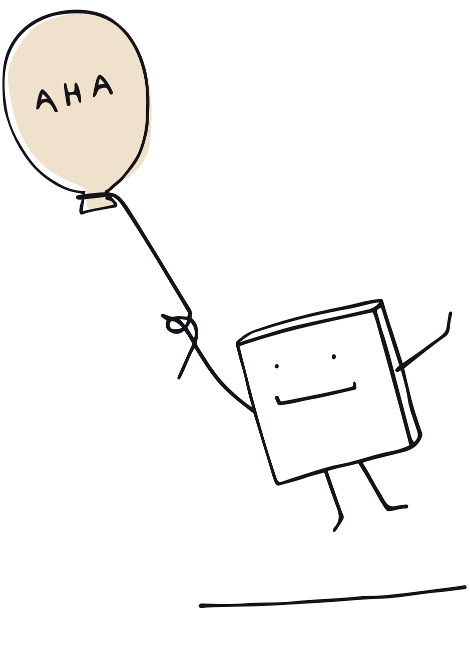
Debate Box 4: Is eye-tracking feasible in patient populations?
Although eye-tracking is a promising technique and can provide a wide range of metrics with regards to cognition, obtaining valid eye-tracking data can be quite a challenge. Whether it is feasible to collect eye-tracking data from patients depends on the paradigm. For example, collecting valid eye-tracking data may be feasible when the patient only has to keep fixating at a dot, but when dynamic gaze patterns should be tracked (e.g., in the gaze-contingent paradigm) this feasibility can drastically decrease. Patients may have droopy eyelids, nystagmus, other oculomotor deficits, or may move excessively, thereby disrupting the eye-tracker signal necessitating repeated calibration of the signal and disturbing gaze-contingent manipulations. Data loss is common. In order for the Copy Task (or other freedom-of-choice task) to be suited for all patients, eye-tracking should be avoided. Here, finger/hand/head movements – or an option to choose for either depending on the patient – may offer a solution.

Debate Box 5: Is relying on the external world beneficial?
Many studies find a benefit from using external strategies in everyday activities on a task level (for a review, see Cicerone et al., 2019), which is why reliance on the external world is advocated in memory rehabilitation. However, on our Copy Task, higher inspection frequencies (thus, higher reliance on the outside world) yielded a worse speed-accuracy score performance (lab-based) or yielded no difference in performance as compared to lower inspection frequencies (web-based). A heavier reliance on the external world thus was not beneficial on our task. However, when assessing performance across conditions within the individual, adhering to a higher inspection frequency was the more beneficial strategy in terms of the speed-accuracy score. Taken together, external sampling is not inherently superior, but it seems that those who persist in heavily relying on the external world across situational changes benefit from it.

Whether or not it is beneficial to rely on the outside world is mostly dependent on the outcome of interest. External offloading may lengthen task duration, but can also improve task accuracy. At the same time, external offloading may lead to worse internal memory: executing saccades during offloading may hamper memory precision (e.g., Schut et al., 2017), and the opportunity to fall back onto external aids can impair internal memory for externally offloaded information. Given these effects on internal memory accuracy, it is important to consider the goal of the patient. Is task accuracy emphasized? Rely on the outside world even if it comes at a time cost. Is memory accuracy targeted? Use internal strategies or longer inspection times for deeper encoding and stronger memory traces.

CHAPTER 7

Summary in Dutch
(Nederlandse samenvatting)



Het geheugen speelt een sleutelrol in ons dagelijks leven. Problemen die ervaren worden in de geheugenfunctie kunnen dan ook veel zorgen opleveren, voor zowel het individu zelf als de naasten. Hoewel geheugenproblemen deels passen bij ouder worden, kunnen ze ook een symptoom zijn van onderliggende pathologie van de hersenen, zoals een neurodegeneratieve ziekte of een beroerte. Het is van belang om de geheugenfunctie goed te kunnen testen, zowel in het kader van diagnostiek als behandeling. Een complicerende factor is dat er niet zo iets bestaat als 'het' geheugen; het geheugen bestaat uit verscheidene subsystemen. Eén van die subsystemen betreft het werkgeheugen. Het werkgeheugen stelt ons in staat om informatie kortdurend vast te houden en te bewerken voor verder gebruik. Daarmee dient het als doorgeefluik tussen waarneming, langetermijngeheugen en actie. Het werkgeheugen wordt in verband gebracht met verschillende hogere orde denkfuncties (cognitieve functies), en kan – naast als onderdeel van het geheugen – worden beschouwd als onderdeel van het executief functioneren, waar ook processen zoals aandacht, planning en mentale flexibiliteit onder vallen. Om de werkgeheugenfunctie goed in kaart te kunnen brengen, is gedegen werkgeheugenonderzoek nodig.

Traditioneel berust werkgeheugenonderzoek in zowel de neuropsychologische praktijk als de experimentele psychologie op het schatten van *hoeveel* informatie iemand kan onthouden (maximale werkgeheugenspanne). Dankzij deze benadering kunnen clinici onderscheiden welke individuen tekorten of stoornissen vertonen in de werkgeheugenopslag- en bewerkingscapaciteit. Capaciteitstaken leren ons veel over werkgeheugenlimieten, bijvoorbeeld voor verbale of visuele informatie, maar zeggen weinig over hoe patiënten het werkgeheugen gebruiken in de dynamiek van het dagelijks leven. Capaciteitstaken sturen aan op inzet van de maximale werkgeheugenspanne, terwijl mensen in het dagelijks leven meestal kunnen terugvallen op externe hulpmiddelen om het werkgeheugen te ontlasten, zoals boodschappenlijstjes of instructietekeningen. Het is dan niet noodzakelijk om het werkgeheugen volledig te gebruiken, want men kan simpelweg terugkijken naar de benodigde informatie. Zulke interacties met de buitenwereld worden niet in kaart gebracht met capaciteitstaken. Capaciteitstaken testen dus het werkgeheugen-*potentieel* maar niet *wanneer* en *hoe* iemand ervoor kiest de capaciteit van het werkgeheugen in te zetten.

Dit proefschrift had als doel om te testen *hoe* patiënten met en zonder geheugenproblemen hun werkgeheugen inzetten als ze konden terugvallen op de buitenwereld, en welke factoren hierin een rol speelden. Dit testten we door oogbewegingen te meten tijdens een kopieertaak op de computer. Deelnemers moesten zo snel en goed mogelijk een voorbeeldfiguur nabouwen door de puzzelstukjes met de computermuis naar de goede plek te verslepen. De taak was ontworpen om gedrag uit te lokken dat wat dichter in de buurt komt van gedrag bij taken in het dagelijks leven, zoals het in elkaar zetten van een meubel met behulp van een instructietekening. Soms was het voorbeeldfiguur de hele tijd zichtbaar, maar soms moesten deelnemers wachten voordat het voorbeeldfiguur zichtbaar was. Vergelijk het met een situatie waarin de instructietekening steeds dichtvalt en het tijd kost terug te bladeren naar de juiste pagina. We verwachtten dat patiënten vaker dan gezonde volwassenen zouden terugkijken naar het voorbeeldfiguur – zelfs als ze daarvoor langer moesten wachten – omdat ze simpelweg niet alle informatie in één keer zouden kunnen onthouden, of omdat ze door terug te kijken minder moeite zouden hoeven doen om hun werkgeheugen te gebruiken. Vaak terugkijken zou dan een efficiëntere oplossing zijn dan informatie onthouden.

In **hoofdstuk 2** vergeleken we de oogbewegingen van gezonde controles met die van patiënten met het syndroom van Korsakov. Dat is een neurocognitieve stoornis die wordt veroorzaakt door een vitamine B-tekort, waardoor hersengebieden die betrokken zijn bij de geheugenfunctie worden aangedaan. Het syndroom van Korsakov kenmerkt zich door ernstige (werk)geheugenproblemen, maar ook door problemen in executieve functies zoals aandacht, inhibitie en planning. We vonden dat deze patiënten meer gebruikmaakten van de buitenwereld dan controles: ze keken vaker en langer terug naar het voorbeeldfiguur. Hun prestatie – waarbij snelheid en accuratesse werden meegenomen – was ook slechter dan die van gezonde controles. Het lukte patiënten niet goed om meer informatie te gaan onthouden, ook al probeerden ze dat wel toen het tijd kostte om het voorbeeldfiguur te inspecteren.

Ook testten we in **hoofdstuk 3** het kijkgedrag van mensen die vanwege geheugenproblemen naar een geheugenpoli werden verwezen. Een verwijzing betekent niet direct dat er een geheugenstoornis is, maar het kán wel. Soms heeft een persoon geheugenklachten, en daarmee een verminderd vertrouwen in het geheugen, maar wordt er op de neuropsychologische capaciteitstaken geen afwijkende geheugenfunctie gevonden. Dan kan het zijn dat er andere factoren voor zorgen dat iemand geheugenklachten ervaart, zoals problemen in de aandacht, stress of somberheid, of een hoge verwachting van het eigen functioneren. De patiënten met een verwijzing naar de geheugenpoli vormen dus een heterogene groep zonder of met objectiveerbare geheugenproblemen in variërende ernst. De mate waarin deze patiënten en de gezonde controles terugkeken naar het voorbeeldfiguur was gelijk. Dat kan deels verklaard worden doordat een deel van de verwezen personen geen objectiveerbare problemen hadden, terwijl sommige controles juist wél objectiveerbare geheugenproblemen vertoonden. Wat hebben capaciteitslimieten en overtuigingen nou voor effect op het wel of niet (her) inspecteren van externe informatie? We vonden tegen onze verwachting in dat een lager vertrouwen in je eigen geheugen niet leidde tot vaker terugvallen op de buitenwereld. Maar hoe meer je *kunt* opslaan, hoe meer je ook *zult* opslaan. Er bleef echter flink wat speelruimte over: zelfs met een hogere geheugenscore keken veel mensen vaak terug naar het voorbeeld. Het gros van de mensen maakte geen gebruik van hun volledige capaciteit als dat niet noodzakelijk was.

Het is belangrijk om individuele verschillen te erkennen. Elk individu is uniek, niet enkel in termen van geheugencapaciteit; ook andere cognitieve functies en persoonlijke eigenschappen kunnen beïnvloeden hoe iemand ervoor kiest met externe informatie te interacteren. Met andere woorden, wat voor de één een goede werkwijze is, is voor de ander suboptimaal. Dat kan leiden tot verschillende voorkeursstrategieën. Om een beeld te krijgen van individuele verschillen in de keuze voor terugkijken of onthouden classificeerden we in **hoofdstuk 4** en **5** deelnemers als iemand die het minimale onthield (minder dan één item) en vaak terugkeek, als iemand die iets meer onthield (één tot drie items) en minder maar nog steeds vrij vaak terugkeek, of als iemand die veel onthield (drie of meer items) en weinig terugkeek. Verreweg de meeste mensen onthielden weinig en keken vaak terug als het voorbeeldfiguur continu zichtbaar was. Bijna niemand onthield drie items of meer. Als er een wachttijd was voor het bekijken van het voorbeeldfiguur, gingen mensen wat meer onthouden om de wachttijd te omzeilen. Echter, er waren nog steeds niet veel mensen die drie of meer items onthielden. Interessant is dat deelnemers die meededen aan een online variant van de kopieertaak (**hoofdstuk 4**) hun kijk- en onthoudgedrag aan konden passen naar de wachttijd zonder dat dit nadelig was voor hun prestatie. In **hoofdstuk 5** zagen we echter dat degenen die een grotere aanpassing maakten de taak ook minder goed gingen uitvoeren. Bij die groep was het beter om evenveel te blijven kijken als dat ze deden zonder wachttijd: het lukte deze groep niet goed genoeg om zich aan te passen. Het verschil tussen de deelnemers in **hoofdstuk 4** en **hoofdstuk 5** kan verklaard worden door verschillen in de taak (online of in het lab), maar ook door de demografische karakteristieken van de deelnemers (leeftijd) en/of onderliggende pathologie (cerebrovasculair accident of neurologisch gezond). Patiënten die een cerebrovasculair accident (beroerte) hebben doorgemaakt en hiervan revalideren, kunnen problemen ervaren in de (werk)geheugenfunctie. In **hoofdstuk 5** vonden we dat een subset van deze patiënten excessief vaak terugkeek naar het voorbeeldfiguur; sommigen keken maar liefst vier keer om één puzzelstukje goed neer te leggen. Het is interessant dat dit precies de patiënten waren die ook een verminderde geheugenfunctie hadden. In tegenstelling tot deze patiënten, lieten gezonde controles met verminderde geheugenfunctie geen dusdanig hoge terugkijkfrequentie zien. Het lijkt er dus op dat geheugencapaciteit niet per definitie bepaalt hoe vaak iemand terugkijkt, maar dat er bij de subset patiënten nog iets anders meespeelde. Dit betreft bijvoorbeeld het vermogen om taakvoortgang te monitoren of een verminderde motorische functie die muisbewegingen kan belemmeren. Daarmee kan het gebruik van de computermuis bij patiënten extra mentale denkkracht opeisen die dan niet meer beschikbaar is voor het onthouden van informatie.

In totaal verzamelde ik gegevens over het kijkgedrag van 106 deelnemers uit drie patiëntgroepen (patiënten met het syndroom van Korsakov, verwezen naar een geheugenpoli, of herstellend van een beroerte) en een controlegroep. Ook onderzocht ik inspectiegedrag van 88 online deelnemers. Ik vond dat een grotere geheugencapaciteit resulteerde in minder vaak terugkijken naar de buitenwereld, en dus een verhoogd gebruik van de interne opslag. Dit betekent dat geheugencapaciteit zoals gemeten op traditionele capaciteitstaken voorspellend is voor kijkgedrag, en dus voor *hoe* iemand het geheugen gebruikt wanneer daartoe de vrije keuze is.

Dit is bemoedigend nieuws voor de neuropsychologische praktijk. Er is vaak kritiek op de traditionele neuropsychologische taken met betrekking tot de vertaalslag naar gedrag in het dagelijks leven. Daarnaast klinkt er een aanhoudende roep om innovatieve technologische ontwikkelingen te implementeren om een gat te dichten tussen kliniek en praktijk. Dit proefschrift toont echter dat de oude vertrouwde pen-en-papier taken het gedrag op een dynamischere taak wel degelijk voorspellen.

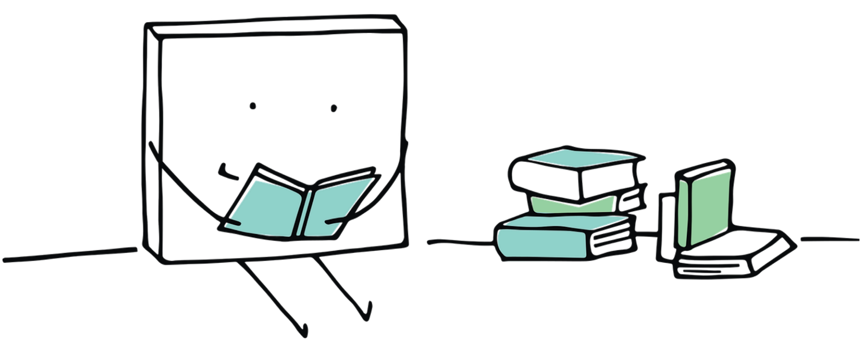
Het is desalniettemin belangrijk om alert te zijn op individuele verschillen en om te erkennen dat mensen zelden hun volledige werkgeheugencapaciteit gebruiken als dat niet direct noodzakelijk is. Over het algemeen werd maximaal één item per keer onthouden, wat impliceert dat werkgeheugengebruik wordt vermeden wanneer dit mogelijk is. Zelfs wanneer informatie niet direct toegankelijk is — waarbij minder en langere inspecties gebruikt worden en men dus meer op het werkgeheugen begint te vertrouwen — blijven mensen nog steeds in grote mate terugvallen op informatie in de buitenwereld. Dit geldt voor gezonde adolescenten (**hoofdstuk 4**) en gezonde volwassenen (**hoofdstuk 2, 3,** en **5**), maar des te meer bij patiënten met een neurologische aandoening (**hoofdstuk 2** en **5**) of een verminderde geheugencapaciteit (**hoofdstuk 3**, ondersteund door een analyse van alle deelnemers aan het labonderzoek).

Tot slot

Dit brengt ons terug bij de vraag: hoe testen we het werkgeheugen op een zinvolle manier? De heterogeniteit in taken, populaties, cognitieve profielen en persoonlijke kenmerken maakt het ingewikkeld om factoren die het werkgeheugengebruik beïnvloeden van elkaar los te trekken. Dit weerspiegelt de complexiteit waarmee neuropsychologen worden geconfronteerd bij het beoordelen van cognitief functioneren in de klinische praktijk. Ik onderstreep het belang van een veelzijdige beoordeling van de geheugenfunctie. Dit proefschrift is dan ook een oproep aan diagnostici: test de capaciteitslimieten, maar erken en houd rekening met individuele kenmerken en compensatiestrategieën die werkgeheugengebruik in de dynamiek van het dagelijks leven beïnvloeden.

APPENDICES

References



Abdulrab, K., & Heun, R. (2008). Subjective Memory Impairment. A review of its definitions indicates the need for a comprehensive set of standardised and validated criteria. *European Psychiatry*, *23*(5), 321–330. https://doi.org/10.1016/j.eurpsy.2008.02.004

Alberts, M., Smets, E. M. A., Vercoulen, J. H. M. M., Garssen, B., & Bleijenberg, G. (1997). 'Verkorte vermoeidheidsvragenlijst': Een praktisch hulpmiddel bij het scoren van vermoeidheid. *Nederlands Tijdschrift voor Geneeskunde*, *141*(32), 1526–1530.

Allen, R. J., Atkinson, A. L., & Hitch, G. J. (2024). Getting value out of working memory through strategic prioritisation: Implications for storage and control. *Quarterly Journal of Experimental Psychology*, 17470218241258102. https://doi.org/10.1177/17470218241258102

Altmann, E. M., & Gray, W. D. (2008). An integrated model of cognitive control in task switching. *Psychological Review*, *115*(3), 602–639. https://doi.org/10.1037/0033-295X.115.3.602

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

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.).* https://doi.org/10.1176/appi.books.9780890425596

Anwyl-Irvine, A. L., Armstrong, T., & Dalmaijer, E. S. (2022). MouseView.js: Reliable and valid attention tracking in web-based experiments using a cursor-directed aperture. *Behavior Research Methods*, *54*(4), 1663–1687. https://doi.org/10.3758/s13428-021-01703-5

Arnoult, M. (1956). Familiarity and recognition of nonsense shapes. *Journal of Experimental Psychology*, *51*, 269–276. https://doi.org/10.1037/h0047772

Arts, N., Walvoort, S., & Kessels, R. (2017). Korsakoff's syndrome: A critical review. *Neuropsychiatric Disease and Treatment*, *Volume 13*, 2875–2890. https://doi.org/10.2147/NDT.S130078

Aschenbrenner, A. J., Crawford, J. L., Peelle, J. E., Fagan, A. M., Benzinger, T. L. S., Morris, J. C., Hassenstab, J., & Braver, T. S. (2023). Increased cognitive effort costs in healthy aging and preclinical Alzheimer's disease. *Psychology and Aging*, *38*(5), 428–442. https://doi.org/10.1037/pag0000742

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), Article 10. https://doi.org/10.1038/nrn1201

Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2021). A multicomponent model of working memory. In *Working Memory: The state of the science* (pp. 10–43). Oxford University Press.

Ballard, D., Hayhoe, M., & Pelz, J. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, *7*, 66–80. https://doi.org/10.1162/jocn.1995.7.1.66

Bambrah, V., Hsu, C.-F., Toplak, M. E., & Eastwood, J. D. (2019). Anticipated, experienced, and remembered subjective effort and discomfort on sustained attention versus working memory tasks. *Consciousness and Cognition*, *75*, 102812. https://doi.org/10.1016/j.concog.2019.102812

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting Linear Mixed-Effects Models using* *lme4* (arXiv:1406.5823). arXiv. http://arxiv.org/abs/1406.5823

Beaudoin, M., & Desrichard, O. (2011). Are memory self-efficacy and memory performance related? A meta-analysis. *Psychological Bulletin*, *137*(2), 211–241. https://doi.org/10.1037/a0022106

Berg, E. van den, Kant, N., & Postma, A. (2012). Remember to Buy Milk on the Way Home! A Meta-analytic Review of Prospective Memory in Mild Cognitive Impairment and Dementia. *Journal of the International Neuropsychological Society*, *18*(4), 706–716. https://doi.org/10.1017/S1355617712000331

Bilgel, M., An, Y., Lang, A., Prince, J., Ferrucci, L., Jedynak, B., & Resnick, S. M. (2014). Trajectories of Alzheimer disease-related cognitive measures in a longitudinal sample. *Alzheimer's & Dementia*, *10*(6), 735-742.e4. https://doi.org/10.1016/j.jalz.2014.04.520

Blalock, L. D. (2015). Stimulus familiarity improves consolidation of visual working memory representations. *Attention, Perception, & Psychophysics*, *77*(4), 1143–1158. https://doi.org/10.3758/s13414-014-0823-z

Bleijenberg, G., Knoop, H., & Gielissen, M. (2009). De Verkorte VermoeidheidsVragenlijst voor het vaststellen van de ernst van chronische vermoeidheid. *Bijblijven*, *25*(1), 19–21. https://doi.org/10.1007/BF03087615

Böing, S., Brink, A. F. T., Ruis, C., Schielen, Z. A., van den Berg, E., Biesbroek, J. M., Nijboer, T. C. W., & Van der Stigchel, S. (2025). Inspecting the external world: Memory capacity, but not memory self-efficacy, predicts offloading in working memory. *Journal of Clinical and Experimental Neuropsychology*. https://www.tandfonline.com/doi/abs/10.1080/13803395.2024.2447263

Böing, S., de Zwart, B., Ten Brink, A. F., Nijboer, T. C. W., & Van Der Stigchel, S. (in press). Relying on the external world: Individuals variably use low- and medium-loading, but rarely high-loading, strategies when engaging visual working memory. Visual Cognition. https://doi.org/10.1080/13506285.2025.2484523

Böing, S., Ten Brink, A. F., Hoogerbrugge, A. J., Oudman, E., Postma, A., Nijboer, T. C. W., & Van der Stigchel, S. (2023). Eye movements as proxy for visual working memory usage: Increased reliance on the external world in Korsakoff Syndrome. *Journal of Clinical Medicine*, *12*(11), Article 11. https://doi.org/10.3390/jcm12113630

Boldt, A., & Gilbert, S. J. (2019). Confidence guides spontaneous cognitive offloading. *Cognitive Research: Principles and Implications*, 4(1), 45. https://doi.org/10.1186/s41235-019-0195-y

Bouma, A., Mulder, J., Lindeboom, J., Schmand, B., & Brein en Cognitie (Psychologie, FMG). (2012). *Handboek neuropsychologische diagnostiek*. Pearson. https://dare.uva.nl/personal/pure/en/publications/handboek-neuropsychologische-diagnostiek(633b0632-36fd-4ab5-92cc-3d868ff18266).html

Brand, M. (2007). Cognitive profile of patients with alcoholic Korsakoff's syndrome. *International Journal on Disability and Human Development*, *6*(2), 161–170. https://doi.org/10.1515/IJDHD.2007.6.2.161

R

Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology*, *21*(1), 1–16. https://doi.org/10.1111/j.2044-8260.1982.tb01421.x

Brockmole, J., & Logie, R. (2013). Age-related change in visual working memory: A Study of 55,753 participants aged 8–75. *Frontiers in Psychology*, 4. https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00012

Brunetti, R., Del Gatto, C., & Delogu, F. (2014). eCorsi: Implementation and testing of the Corsi block-tapping task for digital tablets. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00939

Burnett, L. K., & Richmond, L. L. (2023). Just write it down: Similarity in the benefit from cognitive offloading in young and older adults. *Memory & Cognition*. https://doi.org/10.3758/s13421-023-01413-7

Calamia, M., Markon, K., & Tranel, D. (2012). Scoring Higher the Second Time Around: Meta-Analyses of Practice Effects in Neuropsychological Assessment. *The Clinical Neuropsychologist*, *26*(4), 543–570. https://doi.org/10.1080/13854046.2012.680913

Chiu, G., & Gilbert, S. J. (2024). Influence of the physical effort of reminder-setting on strategic offloading of delayed intentions. *Quarterly Journal of Experimental Psychology*, *77*(6), 1295–1311. https://doi.org/10.1177/17470218231199977

Chung, Y. H., Brady, T. F., & Störmer, V. S. (2024). Meaningfulness and Familiarity Expand Visual Working Memory Capacity. *Current Directions in Psychological Science*, *33*(5), 275–282. https://doi.org/10.1177/09637214241262334

Cicerone, K. D., Dahlberg, C., Malec, J. F., Langenbahn, D. M., Felicetti, T., Kneipp, S., Ellmo, W., Kalmar, K., Giacino, J. T., Harley, J. P., Laatsch, L., Morse, P. A., & Catanese, J. (2005). Evidence-Based Cognitive Rehabilitation: Updated Review of the Literature From 1998 Through 2002. *Archives of Physical Medicine and Rehabilitation*, *86*(8), 1681–1692. https://doi.org/10.1016/j.apmr.2005.03.024

Claessen, M. H. G., van der Ham, I. J. M., & van Zandvoort, M. J. E. (2015). Computerization of the Standard Corsi Block-Tapping Task Affects Its Underlying Cognitive Concepts: A Pilot Study. *Applied Neuropsychology: Adult*, *22*(3), 180–188. https://doi.org/10.1080/23279095.2014.892488

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. https://doi.org/10.3758/BF03196772

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain [PhD thesis]. *Montreal: McGill University, Montreal.*, 85.

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

Cowan, N. (2016). *Working Memory Capacity: Classic Edition*. Psychology Press.

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

Crawford, J. L., Eisenstein, S. A., Peelle, J. E., & Braver, T. S. (2022). Domain-general cognitive motivation: Evidence from economic decision-making – Final Registered Report. *Cognitive Research: Principles and Implications*, *7*(1), 23. https://doi.org/10.1186/s41235-022-00363-z

Crawford, J. L., English, T., & Braver, T. S. (2023). Cognitive Effort-Based Decision-Making Across Experimental and Daily Life Indices in Younger and Older Adults. *The Journals of Gerontology: Series B*, *78*(1), 40–50. https://doi.org/10.1093/geronb/gbac167

Dalmaijer, E. S., Mathôt, S., & Van der Stigchel, S. (2014). PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods*, *46*(4), 913–921. https://doi.org/10.3758/s13428-013-0422-2

de Joode, E. A., van Boxtel, M. P. J., Hartjes, P., Verhey, F. R. J., & van Heugten, C. M. (2013). Use of an electronic cognitive aid by a person with Korsakoff syndrome. *Scandinavian Journal of Occupational Therapy*, *20*(6), 446–453. https://doi.org/10.3109/11038128.2013.821161

Doherty, J. M., Belletier, C., Rhodes, S., Jaroslawska, A., Barrouillet, P., Camos, V., Cowan, N., Naveh-Benjamin, M., & Logie, R. H. (2019). Dual-task costs in working memory: An adversarial collaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(9), 1529–1551. https://doi.org/10.1037/xlm0000668

Drag, L. L., & Bieliauskas, L. A. (2010). Contemporary Review 2009: Cognitive Aging. *Journal of Geriatric Psychiatry and Neurology*, *23*(2), 75–93. https://doi.org/10.1177/0891988709358590

Draschkow, D., Kallmayer, M., & Nobre, A. C. (2021). When Natural Behavior Engages Working Memory. *Current Biology*. https://doi.org/10.1016/j.cub.2020.11.013

Droll, J., & Hayhoe, M. (2008). Trade-offs Between Gaze and Working Memory Use. *Journal of Experimental Psychology. Human Perception and Performance*, *33*, 1352–1365. https://doi.org/10.1037/0096-1523.33.6.1352

Elliott, M., & Parente, F. (2014). Efficacy of memory rehabilitation therapy: A meta-analysis of TBI and stroke cognitive rehabilitation literature. *Brain Injury*, *28*(12), 1610–1616. https://doi.org/10.3109/02699052.2014.934921

Engstrom, M., Landtblom, A.-M., & Karlsson, T. (2013). Brain and effort: Brain activation and effort-related working memory in healthy participants and patients with working memory deficits. *Frontiers in Human Neuroscience*, *7*. https://www.frontiersin.org/articles/10.3389/fnhum.2013.00140

Faul, F., Erdfelder, E., Lang, L.-G., & Buchner, A. (2009). *G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences.* https://link.springer.com/article/10.3758/BF03193146

Fegen, D., Buchsbaum, B. R., & *D'*Esposito, M. (2015). The Effect of Rehearsal Rate and Memory Load on Verbal Working Memory. *NeuroImage*, *105*, 120–131. https://doi.org/10.1016/j.neuroimage.2014.10.034

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision*. Oxford University Press.

R

Fissler, P., Küster, O. C., Laptinskaya, D., Loy, L. S., von Arnim, C. A. F., & Kolassa, I.-T. (2018). Jigsaw Puzzling Taps Multiple Cognitive Abilities and Is a Potential Protective Factor for Cognitive Aging. *Frontiers in Aging Neuroscience*, *10*. https://www.frontiersin.org/articles/10.3389/fnagi.2018.00299

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*, 443. https://doi.org/10.3389/fnhum.2014.00443

Fu, W.-T., & Gray, W. D. (2000). Memory versus Perceptual-Motor Tradeoffs in a Blocks World Task. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *22*. Retrieved from https://escholarship.org/uc/item/5mn4s2zt

Gerridzen, I. J., Hertogh, C. M. P. M., Depla, M. F., Veenhuizen, R. B., Verschuur, E. M. L., & Joling, K. J. (2018). Neuropsychiatric Symptoms in People With Korsakoff Syndrome and Other Alcohol-Related Cognitive Disorders Living in Specialized Long-Term Care Facilities: Prevalence, Severity, and Associated Caregiver Distress. *Journal of the American Medical Directors Association*, *19*(3), 240–247. https://doi.org/10.1016/j.jamda.2017.09.013

Gilbert, S. J. (2015a). Strategic offloading of delayed intentions into the external environment. *Quarterly Journal of Experimental Psychology*, *68*(5), 971–992. https://doi.org/10.1080/17470218.2014.972963

Gilbert, S. J. (2015b). Strategic use of reminders: Influence of both domain-general and task-specific metacognitive confidence, independent of objective memory ability. *Consciousness and Cognition*, *33*, 245–260. https://doi.org/10.1016/j.concog.2015.01.006

Gilbert, S. J. (2023). *Cognitive offloading is value-based decision making: Modelling cognitive effort and the expected value of memory*.

Gilbert, S. J., Bird, A., Carpenter, J. M., Fleming, S. M., Sachdeva, C., & Tsai, P.-C. (2020). Optimal use of reminders: Metacognition, effort, and cognitive offloading. *Journal of Experimental Psychology: General*, *149*(3), 501–517. https://doi.org/10.1037/xge0000652

Gilbert, S. J., Boldt, A., Sachdeva, C., Scarampi, C., & Tsai, P.-C. (2023). Outsourcing Memory to External Tools: A Review of 'Intention Offloading'. *Psychonomic Bulletin & Review*, *30*(1), 60–76. https://doi.org/10.3758/s13423-022-02139-4

Gonthier, C. (2021). Charting the Diversity of Strategic Processes in Visuospatial Short-Term Memory. *Perspectives on Psychological Science*, *16*(2), 294–318. https://doi.org/10.1177/1745691620950697

Gray, W. D., & Fu, W.-T. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science*, *28*(3), 359–382. https://doi.org/10.1207/s15516709cog2803_3

Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review*, *113*(3), 461–482. https://doi.org/10.1037/0033-295X.113.3.461

Gregório, G. W., Brands, I., Stapert, S., Verhey, F. R., & van Heugten, C. M. (2014). Assessments of Coping After Acquired Brain Injury: A Systematic Review of Instrument Conceptualization, Feasibility, and Psychometric Properties. *The Journal of Head Trauma Rehabilitation*, *29*(3), E30. https://doi.org/10.1097/HTR.0b013e31828f93db

Grinschgl, S., Meyerhoff, H. S., Schwan, S., & Papenmeier, F. (2021). From metacognitive beliefs to strategy selection: Does fake performance feedback influence cognitive offloading? *Psychological Research*, *85*(7), 2654–2666. https://doi.org/10.1007/s00426-020-01435-9

Grinschgl, S., Papenmeier, F., & Meyerhoff, H. S. (2021). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology*, *74*(9), 1477–1496. https://doi.org/10.1177/17470218211008060

Harper, C. (1983). The incidence of Wernicke's encephalopathy in Australia—A neuropathological study of 131 cases. *Journal of Neurology, Neurosurgery & Psychiatry*, *46*(7), 593–598. https://doi.org/10.1136/jnnp.46.7.593

Hatem, S. M., Saussez, G., della Faille, M., Prist, V., Zhang, X., Dispa, D., & Bleyenheuft, Y. (2016). Rehabilitation of Motor Function after Stroke: A Multiple Systematic Review Focused on Techniques to Stimulate Upper Extremity Recovery. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00442

Hendricks, H. T., van Limbeek, J., Geurts, A. C., & Zwarts, M. J. (2002). Motor recovery after stroke: A systematic review of the literature. *Archives of Physical Medicine and Rehabilitation*, *83*(11), 1629–1637. https://doi.org/10.1053/apmr.2002.35473

Hendriks, M. P. H., Kessels, R. P. C., Gorissen, M., Schmand, B., & Duits, A. (2014). *Neuropsychologische Diagnostiek. De klinische praktijk*. Uigeverij Boom.

Hendriks, M. P. H., Mol, B. A. W., & Kessles, R. P. C. (2020). Opinie: Uniformiteit in de kwalitatieve beschrijving van scores op prestatietaken. *Tijdschrift voor Neuropsychologie*, *15*(3). https://www.tvnp.nl/inhoud/tijdschrift_artikel

Hitch, G. J., Allen, R. J., & Baddeley, A. D. (2024). The multicomponent model of working memory fifty years on. *Quarterly Journal of Experimental Psychology*, 17470218241290908. https://doi.org/10.1177/17470218241290909

Hoogerbrugge, A. J., Strauch, C., Böing, S., Nijboer, T. C. W., & Stigchel, S. V. der. (2024). Just-in-Time Encoding Into Visual Working Memory Is Contingent Upon Constant Availability of External Information. *Journal of Cognition*, *7*(1), Article 1. https://doi.org/10.5334/joc.364

Hoogerbrugge, A. J., Strauch, C., Hoevers, N., Olivers, C. N. L., Nijboer, T. C. W., & Van der Stigchel, S. (in prep.). *Multi-target visual search dynamically switches between concurrent and sequential modes*.

Hoogerbrugge, A. J., Strauch, C., Nijboer, T. C. W., & Van der Stigchel, S. (2023). Don't hide the instruction manual: A dynamic trade-off between using internal and external templates during visual search. *Journal of Vision*, *23*(7), 14. https://doi.org/10.1167/jov.23.7.14

Hoogerbrugge, A. J., Strauch, C., Nijboer, T. C. W., & Van Der Stigchel, S. (2024). Persistent resampling of external information despite 25 repetitions of the same visual search templates. *Attention, Perception, & Psychophysics*, *86*(7), 2301–2314. https://doi.

R

org/10.3758/s13414-024-02953-z
Howorth, P., & Saper, J. (2003). The dimensions of insight in people with dementia. *Aging & Mental Health*. https://doi.org/10.1080/1360786031000072286
Hülür, G., Hertzog, C., Pearman, A. M., & Gerstorf, D. (2015). Correlates and Moderators of Change in Subjective Memory and Memory Performance: Findings from the Health and Retirement Study. *Gerontology*, *61*(3), 232–240. https://doi.org/10.1159/000369010
Huntley, J. D., & Howard, R. J. (2010). Working memory in early Alzheimer's disease: A neuropsychological review. *International Journal of Geriatric Psychiatry*, *25*(2), 121–132. https://doi.org/10.1002/gps.2314
Hurt, C. S., Burns, A., Brown, R. G., & Barrowclough, C. (2012). Why don't older adults with subjective memory complaints seek help? *International Journal of Geriatric Psychiatry*, *27*(4), 394–400. https://doi.org/10.1002/gps.2731
Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The Effort Paradox: Effort Is Both Costly and Valued. *Trends in Cognitive Sciences*, *22*(4), 337–349. https://doi.org/10.1016/j.tics.2018.01.007
Isen, D. R., & Kline, L. B. (2020). Neuro-ophthalmic Manifestations of Wernicke Encephalopathy. *Eye and Brain*, *Volume 12*, 49–60. https://doi.org/10.2147/EB.S234078
Janssen, C. P., & Gray, W. D. (2012). When, What, and How Much to Reward in Reinforcement Learning-Based Models of Cognition. *Cognitive Science*, *36*(2), 333–358. https://doi.org/10.1111/j.1551-6709.2011.01222.x
Janssen, G. T. L., Egger, J. I. M., & Kessels, R. P. C. (2023). Impaired Executive Functioning Associated with Alcohol-Related Neurocognitive Disorder including Korsakoff's Syndrome. *Journal of Clinical Medicine*, *12*(20), Article 20. https://doi.org/10.3390/jcm12206477
Jarvstad, A., & Gilchrist, I. D. (2019). Cognitive control of saccadic selection and inhibition from within the core cortical saccadic network. *The Journal of Neuroscience*, 1419–18. https://doi.org/10.1523/JNEUROSCI.1419-18.2018
Jessen, F., Wiese, B., Bachmann, C., Eifflaender-Gorfer, S., Haller, F., Kölsch, H., Luck, T., Mösch, E., van den Bussche, H., Wagner, M., Wollny, A., Zimmermann, T., Pentzek, M., Riedel-Heller, S. G., Romberg, H.-P., Weyerer, S., Kaduszkiewicz, H., Maier, W., Bickel, H., & German Study on Aging, C. and D. in P. C. P. S. G. (2010). Prediction of Dementia by Subjective Memory Impairment: Effects of Severity and Temporal Association With Cognitive Impairment. *Archives of General Psychiatry*, *67*(4), 414–422. https://doi.org/10.1001/archgenpsychiatry.2010.30
Johansson, B., Allen-Burge, R., & Zarit, S. H. (1997). Self-Reports on Memory Functioning in a Longitudinal Study of the Oldest Old: Relation to Current, Prospective, and Retrospective Performance. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *52B*(3), P139–P146. https://doi.org/10.1093/geronb/52B.3.P139
Johnson, B. P., & Westlake, K. P. (2021). Chronic Poststroke Deficits in Gross and Fine Motor Control of the Ipsilesional Upper Limb. *American Journal of Physical Medicine &*

Rehabilitation, *100*(4), 345. https://doi.org/10.1097/PHM.0000000000001569

Jungen, M., van Engelen, E., & Bokhorst, M. (2019). *Uitgebreide toelichting van het meetinstrument: HADS*. https://meetinstrumentenzorg.nl/wp-content/uploads/instrumenten/HADS-form.pdf

Karadag, F., Oguzhanoglu, N., Ozdel, O., & Atesci, F. C. (2005). Memory Function in Patients with Obsessive Compulsive Disorder and the Problem of Confidence in Their Memories: A Clinical Study. *Croat Med J*.

Kardan, O., Adam, K. C. S., Mance, I., Churchill, N. W., Vogel, E. K., & Berman, M. G. (2020). Distinguishing cognitive effort and working memory load using scale-invariance and alpha suppression in EEG. *NeuroImage*, *211*, 116622. https://doi.org/10.1016/j.neuroimage.2020.116622

Kaschel, R., Sala, S., Cantagallo, A., Fahlboeck, A., Laaksonen, R., & Kazén, M. (2002). Imagery mnemonics for the rehabilitation of memory: A randomised group controlled trial. *Neuropsychological Rehabilitation: An International Journal*, *12*, 127–153. https://doi.org/10.1080/09602010143000211

Kasse, E., Oudman, E., Olivier, M., Wijnia, J. W., & Postma, A. (2019). Subtle object location perception deficits in Korsakoff's syndrome. *Journal of Clinical and Experimental Neuropsychology*, *41*(9), 881–887. https://doi.org/10.1080/13803395.2019.1640864

Kattah, J. C. (2017). The Spectrum of Vestibular and Ocular Motor Abnormalities in Thiamine Deficiency. *Current Neurology and Neuroscience Reports*, *17*(5), 40. https://doi.org/10.1007/s11910-017-0747-9

Kessels, R. P. C., Bucks, R. S., Willison, J. R., & Byrne, L. M. T. (2014). *Location Learning Test (3th ed.).* Hogrefe Uitgevers.

Kessels, R. P. C., & Kopelman, M. D. (2012). Context Memory in Korsakoff's Syndrome. *Neuropsychology Review*, *22*(2), 117–131. https://doi.org/10.1007/s11065-012-9202-5

Kessels, R. P. C., Nys, G. M. S., Brands, A. M. A., Berg, E. van den, & Zandvoort, V. M. J. E. (2006). The modified Location Learning Test: Norms for the assessment of spatial memory function in neuropsychological patients. *Archives of Clinical Neuropsychology*, *21*(8), 841–846. https://doi.org/10.1016/j.acn.2006.06.015

Kessels, R. P. C., van Zandvoort, M. J. E., Postma, A., Kappelle, L. J., & de Haan, E. H. F. (2000). The Corsi Block-Tapping Task: Standardization and Normative Data. *Applied Neuropsychology*, *7*(4), 252–258. https://doi.org/10.1207/S15324826AN0704_8

Kimonides, S., Cavuoto, M. G., Silva, L. D., & Kinsella, G. J. (2018). The role of subjective cognitive complaints and depressive symptoms in social re- integration following stroke: A mediation explanation in a cross-sectional sample. *Topics in Stroke Rehabilitation*. https://doi.org/10.1080/10749357.2018.1489570

Koevoet, D., Naber, M., Strauch, C., Somai, R. S., & Van der Stigchel, S. (2023). Differential aspects of attention predict the depth of visual working memory encoding: Evidence from pupillometry. *Journal of Vision*, *23*(6), 9. https://doi.org/10.1167/jov.23.6.9

Koevoet, D., Zantwijk, L. V., & Naber, M. (2024). *Effort Drives Saccade Selection*.

Kopelman, M. D., Thomson, A. D., Guerrini, I., & Marshall, E. J. (2009). The Korsakoff Syndrome: Clinical Aspects, Psychology and Treatment. *Alcohol and Alcoholism*, *44*(2), 148–154. https://doi.org/10.1093/alcalc/agn118

Krauzlis, R. J., Goffart, L., & Hafed, Z. M. (2017). Neuronal control of fixation and fixational eye movements. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1718), 20160205. https://doi.org/10.1098/rstb.2016.0205

Kumle, L., Võ, M. L.-H., Nobre, A. C., & Draschkow, D. (2024). Multifaceted consequences of visual distraction during natural behaviour. *Communications Psychology*, *2*(1), 1–13. https://doi.org/10.1038/s44271-024-00099-0

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, *36*(6), 661–679. https://doi.org/10.1017/S0140525X12003196

Kvitelashvili, S., & Kessler, Y. (2024). *Memory at Will: Investigating Voluntary Utilization of Visual Working Memory Capacity*. https://doi.org/10.21203/rs.3.rs-3834000/v1

Langhorne, P., Coupar, F., & Pollock, A. (2009). Motor recovery after stroke: A systematic review. *The Lancet Neurology*, *8*(8), 741–754. https://doi.org/10.1016/S1474-4422(09)70150-4

Lee, D. G., Daunizeau, J., & Pezzulo, G. (2023). Evidence or Confidence: What Is Really Monitored during a Decision? *Psychonomic Bulletin & Review*, *30*(4), 1360–1379. https://doi.org/10.3758/s13423-023-02255-9

Lees, A. J., & Smith, E. (1983). Cognitive deficits in the early stages of Parkins's disease. *Brain*, *106*(2), 257–270. https://doi.org/10.1093/brain/106.2.257

Lenth, R. V. (2024). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (Version 1.10.4).

Leśniak, M., Bak, T., Czepiel, W., Seniów, J., & Członkowska, A. (2008). Frequency and Prognostic Value of Cognitive Disorders in Stroke Patients. *Dementia and Geriatric Cognitive Disorders*, *26*(4), 356–363. https://doi.org/10.1159/000162262

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological Assessment* (fifth). Oxford University Press.

Logie, R., Camos, V., & Cowan, N. (2020). *Working Memory: The state of the science*. Oxford University Press.

Logie, R. H. (2011). The Functional Organization and Capacity Limits of Working Memory. *Current Directions in Psychological Science*, *20*(4), 240–245. https://doi.org/10.1177/0963721411415340

Logie, R. H. (2023). Strategies, debates, and adversarial collaboration in working memory: The 51st Bartlett Lecture. *Quarterly Journal of Experimental Psychology*, *76*(11), 2431–2460. https://doi.org/10.1177/17470218231194037

Lu, W., Richards, M., Werring, D., & Bobak, M. (2022). Memory Trajectories Before and After First and Recurrent Strokes. *Neurology*, *98*(6), e589–e600. https://doi.org/10.1212/

WNL.0000000000013171

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), Article 6657. https://doi.org/10.1038/36846

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

Lugtmeijer, S., Lammers, N. A., de Haan, E. H. F., de Leeuw, F.-E., & Kessels, R. P. C. (2021). Post-Stroke Working Memory Dysfunction: A Meta-Analysis and Systematic Review. *Neuropsychology Review*, *31*(1), 202–219. https://doi.org/10.1007/s11065-020-09462-4

Luria, R., Sessa, P., Gotler, A., Jolicœur, P., & Dell'Acqua, R. (2010). Visual Short-term Memory Capacity for Simple and Complex Objects. *Journal of Cognitive Neuroscience*, *22*(3), 496–512. https://doi.org/10.1162/jocn.2009.21214

Luria, R., & Vogel, E. K. (2011). Shape and color conjunction stimuli are represented as bound objects in visual working memory. *Neuropsychologia*, *49*(6), 1632–1639. https://doi.org/10.1016/j.neuropsychologia.2010.11.031

Luwel, K., Schillemans, V., Onghena, P., & Verschaffel, L. (2009). Does switching between strategies within the same task involve a cost? *British Journal of Psychology*, *100*(4), 753–771. https://doi.org/10.1348/000712609X402801

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

Magen, H., Emmanouil, T.-A., McMains, S. A., Kastner, S., & Treisman, A. (2009). Attentional demands predict short-term memory load response in posterior parietal cortex. *Neuropsychologia*, *47*(8), 1790–1798. https://doi.org/10.1016/j.neuropsychologia.2009.02.015

Maharasingam, M., Macniven, J. A. B., & Mason, O. J. (2013). Executive functioning in chronic alcoholism and Korsakoff syndrome. *Journal of Clinical and Experimental Neuropsychology*, *35*(5), 501–508. https://doi.org/10.1080/13803395.2013.795527

Mani, S., Mutha, P. K., Przybyla, A., Haaland, K. Y., Good, D. C., & Sainburg, R. L. (2013). Contralesional motor deficits after unilateral stroke reflect hemisphere-specific control mechanisms. *Brain*, *136*(4), 1288–1303. https://doi.org/10.1093/brain/aws283

Maron, D. N., Bowe, S. J., Spencer-Smith, M., Mellahn, O. J., Perrykkad, K., Bellgrove, M. A., & Johnson, B. P. (2021). Oculomotor deficits in attention deficit hyperactivity disorder (ADHD): A systematic review and comprehensive meta-analysis. *Neuroscience & Biobehavioral Reviews*, *131*, 1198–1213. https://doi.org/10.1016/j.neubiorev.2021.10.012

Martin, R. C. (2021). The Critical Role of Semantic Working Memory in Language Comprehension and Production. *Current Directions in Psychological Science*, *30*(4), 283–291. https://doi.org/10.1177/0963721421995178

Mattos, P., Lino, V., Rizo, L., Alfano, Â., Araújo, C., & Raggio, R. (2003). Memory complaints and test performance in healthy elderly persons. *Arquivos de Neuro-Psiquiatria*, *61*(4), 920–924. https://doi.org/10.1590/S0004-282X2003000600006

R

McAteer, S. M., Ablott, E., McGregor, A., & Smith, D. T. (2023). Dynamic resource allocation in spatial working memory during full and partial report tasks. *Journal of Vision*, *23*(2), 10. https://doi.org/10.1167/jov.23.2.10

Mehta, R. K. (2016). Integrating Physical and Cognitive Ergonomics. *IIE Transactions on Occupational Ergonomics and Human Factors*, 4(2–3), 83–87. https://doi.org/10.1080/21577323.2016.1207475

Melnik, A., Schüler, F., Rothkopf, C. A., & König, P. (2018). The World as an External Memory: The Price of Saccades in a Sensorimotor Task. *Frontiers in Behavioral Neuroscience*, *12*, 253. https://doi.org/10.3389/fnbeh.2018.00253

Meyerhoff, H. S., Grinschgl, S., Papenmeier, F., & Gilbert, S. J. (2021). Individual differences in cognitive offloading: A comparison of intention offloading, pattern copy, and short-term memory capacity. *Cognitive Research: Principles and Implications*, *6*(1), 34. https://doi.org/10.1186/s41235-021-00298-x

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

Morel, P., Ulbrich, P., & Gail, A. (2017). What makes a reach movement effortful? Physical effort discounting supports common minimization principles in decision making and motor control. *PLOS Biology*, *15*(6), e2001323. https://doi.org/10.1371/journal.pbio.2001323

Morgan, P., Patrick, J., Waldron, S., King, S., & Patrick, T. (2009). Improving Memory after Interruption: Exploiting Soft Constraints and Manipulating Information Access Cost. *Journal of Experimental Psychology. Applied*, *15*, 291–306. https://doi.org/10.1037/a0018008

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 & Review*, *18*(1), 46–60. https://doi.org/10.3758/s13423-010-0034-0

Morrison, A. B., & Richmond, L. L. (2020). Offloading items from memory: Individual differences in cognitive offloading in a short-term memory task. *Cognitive Research: Principles and Implications*, *5*(1), 1. https://doi.org/10.1186/s41235-019-0201-4

Morrison, A. B., Rosenbaum, G. M., Fair, D., & Chein, J. M. (2016). Variation in strategy use across measures of verbal working memory. *Memory & Cognition*, *44*(6), 922–936. https://doi.org/10.3758/s13421-016-0608-9

Moya, M., López-Valencia, L., García-Bueno, B., & Orio, L. (2021). *Disinhibited-Like Behavior Correlates With Frontal Cortex Damage in Alcohol-Induced Wernicke-Korsakoff Syndrome*. Preprints. https://doi.org/10.20944/preprints202111.0484.v1

Nelson, W. W., & Loftus, G. R. (1980). The Functional Visual Field During Picture Viewing. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(4), 391–399. https://doi.org/10.1037/0278-7393.6.4.391

Norris, D., & Kalm, K. (2021). Chunking and data compression in verbal short-term memory. *Cognition*, *208*, 104534. https://doi.org/10.1016/j.cognition.2020.104534

Nys, G. M. S., van Zandvoort, M. J. E., van der Worp, H. B., de Haan, E. H. F., de Kort, P. L. M., Jansen, B. P. W., & Kappelle, L. J. (2006). Early cognitive impairment predicts long-term depressive symptoms and quality of life after stroke. *Journal of the Neurological Sciences*, *247*(2), 149–156. https://doi.org/10.1016/j.jns.2006.04.005

Oscar-Berman, M., Kirkley, S. M., Gansler, D. A., & Couture, A. (2004). Comparisons of Korsakoff and Non-Korsakoff Alcoholics on Neuropsychological Tests of Prefrontal Brain Functioning. *Alcohol: Clinical and Experimental Research*, *28*(4), 667–675. https://doi.org/10.1097/01.ALC.0000122761.09179.B9

Oudman, E., Nijboer, T. C. W., Postma, A., Wijnia, J. W., Kerklaan, S., Lindsen, K., & Van der Stigchel, S. (2013). Acquisition of an instrumental activity of daily living in patients with Korsakoff's syndrome: A comparison of trial and error and errorless learning. *Neuropsychological Rehabilitation*, *23*(6), 888–913. https://doi.org/10.1080/09602011.2013.835738

Oudman, E., Nijboer, T. C. W., Postma, A., Wijnia, J. W., & Van der Stigchel, S. (2015). Procedural Learning and Memory Rehabilitation in Korsakoff's Syndrome—A Review of the Literature. *Neuropsychology Review*, *25*(2), 134–148. https://doi.org/10.1007/s11065-015-9288-7

Oudman, E., Schut, M. J., Ten Brink, A. F., Postma, A., & Van der Stigchel, S. (2020). Visual working memory capacity in Korsakoff's amnesia. *Journal of Clinical and Experimental Neuropsychology*, *42*(4), 363–370. https://doi.org/10.1080/13803395.2020.1722800

Oudman, E., Wijnia, J. W., Oey, M. J., van Dam, M., & Postma, A. (2021). Wernicke-Korsakoff syndrome despite no alcohol abuse: A summary of systematic reports. *Journal of the Neurological Sciences*, *426*, 117482. https://doi.org/10.1016/j.jns.2021.117482

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

Patrick, J., Morgan, P. L., Smy, V., Tiley, L., Seeby, H., Patrick, T., & Evans, J. (2015). The influence of training and experience on memory strategy. *Memory & Cognition*, *43*(5), 775–787. https://doi.org/10.3758/s13421-014-0501-3

Piolino, P., Desgranges, B., & Eustache, F. (2009). Episodic autobiographical memories over the course of time: Cognitive, neuropsychological and neuroimaging findings. *Neuropsychologia*, *47*(11), 2314–2329. https://doi.org/10.1016/j.neuropsychologia.2009.01.020

Pitel, A. L., Beaunieux, H., Witkowski, T., Vabret, F., De La Sayette, V., Viader, F., Desgranges, B., & Eustache, F. (2008). Episodic and Working Memory Deficits in Alcoholic Korsakoff Patients: The Continuity Theory Revisited. *Alcoholism: Clinical and Experimental Research*, *32*(7), 1229–1241. https://doi.org/10.1111/j.1530-0277.2008.00677.x

Pizzonia, K. L., & Suhr, J. A. (2022). Systematic Review of Correlates of Internal and External Memory Strategy Use in Older Adults. *Journal of Applied Gerontology*, *41*(5), 1491–1499. https://doi.org/10.1177/07334648211065427

Ponds, R. W. H. M., Boxtel, M., & Jolles, J. (2006). De 'Cognitive Failure Questionnaire' als maat

R

voor subjectief cognitief functioneren. *Tijdschrift Voor Neuropsychologie*, *1*, 37–45.
Ponds, R. W. H. M., Commissaris, K., & Jolles, J. (1997). Prevalence and Covariates of Subjective Forgetfulness in a Normal Population in the Netherlands. *The International Journal of Aging and Human Development*, *45*(3), 207–221. https://doi.org/10.2190/MVQ1-WB58-875H-Y4X0
Ponds, R. W. H. M., & Jolles, J. (1996a). Memory Complaints in Elderly People: The Role of Memory Abilities, Metamemory, Depression, and Personality. *Educational Gerontology*, *22*(4), 341–357. https://doi.org/10.1080/0360127960220404
Ponds, R. W. H. M., & Jolles, J. (1996b). The abridged Dutch Metamemory in Adulthood (MIA) Questionnaire: Structure and effects of age, sex, and education. *Psychology and Aging*, *11*(2), 324–332. https://doi.org/10.1037/0882-7974.11.2.324
Poppenk, J., Köhler, S., & Moscovitch, M. (2010). Revisiting the novelty effect: When familiarity, not novelty, enhances memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1321–1330. https://doi.org/10.1037/a0019900
Qing, T., Strauch, C., Van Maanen, L., & Van der Stigchel, S. (2024). Shifting reliance between the internal and external world: A meta-analysis on visual-working memory use. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-024-02623-z
R Core Team. (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www. R-project. org
R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/.
Rauwenhoff, J. C. C., Bol, Y., Peeters, F., Van Den Hout, A. J. H. C., Geusgens, C. A. V., & Van Heugten, C. M. (2023). Acceptance and commitment therapy for individuals with depressive and anxiety symptoms following acquired brain injury: A non-concurrent multiple baseline design across four cases. *Neuropsychological Rehabilitation*, *33*(6), 1018–1048. https://doi.org/10.1080/09602011.2022.2053169
Rhodes, M. G., & Kelley, C. M. (2005). Executive processes, memory accuracy, and memory monitoring: An aging and individual difference analysis. *Journal of Memory and Language*, *52*(4), 578–594. https://doi.org/10.1016/j.jml.2005.01.014
Ricker, T. J., & Cowan, N. (2010). Loss of Visual Working Memory within Seconds: The Combined Use of Refreshable and Non-Refreshable Features. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *36*(6), 1355–1368. https://doi.org/10.1037/a0020356
Risko, E. F., & Dunn, T. L. (2015). Storing information in-the-world: Metacognition and cognitive offloading in a short-term memory task. *Consciousness and Cognition*, *36*, 61–74. https://doi.org/10.1016/j.concog.2015.05.014
Risko, E. F., & Gilbert, S. J. (2016). Cognitive Offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002
Rittmo, J., & McIntosh, R. (2021). singcar: Comparing single cases to small samples in R. *Journal of Open Source Software*, *6*(68), 3887. https://doi.org/10.21105/joss.03887

Rittmo, J., & McIntosh, R. (2023). *singcar: Comparing Single Cases to Small Samples* (Version 0.1.5). https://CRAN.R-project.org/package=singcar

Rittmo, J., & McIntosh, R. D. (2020). *The R package singcar: Comparing Single Cases to Small Samples* (Version R Package version 0.1.3.). https://cran.r-project.org/web/packages/singcar/vignettes/singcar_vignette.html

Riverbank Computing Limited. (2019). *PyQt5 library*.

Rizopoulos, D. (2006). ltm: Latent Trait Models under IRT. *Journal of Statistical Software*, *17*(5). https://doi.org/10.18637/jss.v017.i05

Roebers, C. M. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. *Developmental Review*, *45*, 31–51. https://doi.org/10.1016/j.dr.2017.04.001

Rogers, T., Hodges, J., Patterson, K., & Lambon Ralph, M. (2003). Object recognition under semantic impairment: The effects of conceptual regularities on perceptual decisions. *Language and Cognitive Processes*, *18*(5–6), 625–662. https://doi.org/10.1080/01690960344000053

Saa, J. P., Tse, T., Baum, C. M., Cumming, T., Josman, N., Rose, M., O'Keefe, S., Sewell, K., Nguyen, V., & Carey, L. M. (2021). Cognitive Recovery After Stroke: A Meta-analysis and Metaregression of Intervention and Cohort Studies. *Neurorehabilitation and Neural Repair*, *35*(7), 585–600. https://doi.org/10.1177/15459683211017501

Saan, R. J., & Deelman, B. G. (1986). *De 15-Woordentests A en B. Een voorlopige handleiding (Intern rapport).* Groningen: AZG, afd. Neuropsychologie.

Sahakian, A., Böing, S., Gayet, S., Paffen, C. L. E., & Van der Stigchel, S. (in prep.). *Task-specific strategies in visual working memory use*.

Sahakian, A., Gayet, S., Paffen, C. L. E., & Van der Stigchel, S. (2023). Mountains of memory in a sea of uncertainty: Sampling the external world despite useful information in visual working memory. *Cognition*, *234*, 105381. https://doi.org/10.1016/j.cognition.2023.105381

Sahakian, A., Gayet, S., Paffen, C., & Van der Stigchel, S. (2024). *The rise and fall of memories: Temporal dynamics of visual working memory*. https://osf.io/sgycr/download

Saykin, A. J., Wishart, H. A., Rabin, L. A., Santulli, R. B., Flashman, L. A., West, J. D., McHugh, T. L., & Mamourian, A. C. (2006). Older adults with cognitive complaints show brain atrophy similar to that of amnestic MCI. *Neurology*, *67*(5), 834–842. https://doi.org/10.1212/01.wnl.0000234032.77541.a2

Scarampi, C., & Gilbert, S. J. (2020). The effect of recent reminder setting on subsequent strategy and performance in a prospective memory task. *Memory*, *28*(5), 677–691. https://doi.org/10.1080/09658211.2020.1764974

Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, *11*(9), 1141–1152. https://doi.org/10.1111/2041-210X.13434

Schmand, B., Jonker, C., Geerlings, M. I., & Lindeboom, J. (1997). Subjective memory complaints in the elderly: Depressive symptoms and future dementia. *British Journal of Psychiatry*, *171*(4), 373–376. https://doi.org/10.1192/bjp.171.4.373

Schmand, B., Jonker, C., Hooijer, C., & Lindeboom, J. (1996). Subjective memory complaints may announce dementia. *Neurology*, *46*(1), 121–125. https://doi.org/10.1212/WNL.46.1.121

Schreurs, P. J., Tellegen, B., & Willige, G. V. (1984). Gezondheid, stress en coping: De ontwikkeling van de Utrechtse coping-lijst. [Health, stress and coping: The development of the Utrechtse Coping Scale.]. *Gedrag: Tijdschrift Voor Psychologie*, *12*(1–2), 101–117.

Shelton, J. T., Elliott, E. M., Matthews, R. A., Hill, B. D., & Gouvier, Wm. D. (2010). The relationships of working memory, secondary memory, and general fluid intelligence: Working memory is special. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 813–820. https://doi.org/10.1037/a0019046

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function. *Neuron*, *79*(2), 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Singmann, H., & Kellen, D. (2019). An Introduction to Mixed Models for Experimental Psychology. In *New Methods in Cognitive Psychology*. Routledge.

Smith, D. B., Scott, S. H., Semrau, J. A., & Dukelow, S. P. (2023). Impairments of the ipsilesional upper-extremity in the first 6-months post-stroke. *Journal of NeuroEngineering and Rehabilitation*, *20*(1), 106. https://doi.org/10.1186/s12984-023-01230-8

Smits, S., Oudman, E., Altgassen, M., & Postma, A. (2022). Smartwatch reminders are as effective as verbal reminders in patients with Korsakoff's syndrome: Three case studies. *Neurocase*, *28*(1), 48–62. https://doi.org/10.1080/13554794.2021.2024237

Sohlberg, M. M., Kennedy, M., Avery, J., Coelho, C., Turkstra, L., Ylvisaker, M., & Yorkston, K. (2007). *Evidence-based practice for the use of external aids as a memory compensation technique*. *15*(1).

Somai, R. S., Schut, M. J., & Van der Stigchel, S. (2020). Evidence for the world as an external memory: A trade-off between internal and external visual memory storage. *Cortex*, *122*, 108–114. https://doi.org/10.1016/j.cortex.2018.12.017

Spinhoven, Ph., Ormel, J., Sloekers, P. P. A., Kempen, G. I. J. M., Speckens, A. E. M., & Hemert, A. M. V. (1997). A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. *Psychological Medicine*, *27*(2), 363–370. https://doi.org/10.1017/S0033291796004382

Steinberg, S. I., Negash, S., Sammel, M. D., Bogner, H., Harel, B. T., Livney, M. G., McCoubrey, H., Wolk, D. A., Kling, M. A., & Arnold, S. E. (2013). Subjective Memory Complaints, Cognitive Performance, and Psychological Factors in Healthy Older Adults. *American Journal of Alzheimer's Disease & Other Dementias®*, *28*(8), 776–783. https://doi.org/10.1177/1533317513504817

Stopford, C. L., Thompson, J. C., Neary, D., Richardson, A. M. T., & Snowden, J. S. (2012). Working memory, attention, and executive function in Alzheimer's disease and frontotemporal dementia. *Cortex*, *48*(4), 429–446. https://doi.org/10.1016/j.cortex.2010.12.002

Storm, B. C., Stone, S. M., & Benjamin, A. S. (2017). Using the Internet to access information inflates future use of the Internet to access other information. *Memory*, *25*(6), 717–723. https://doi.org/10.1080/09658211.2016.1210171

Suchy, Y. (2009). Executive Functioning: Overview, Assessment, and Research Issues for Non-Neuropsychologists. *Annals of Behavioral Medicine*, *37*(2), 106–116. https://doi.org/10.1007/s12160-009-9097-4

Takahashi, J., Hirano, Y., Miura, K., Morita, K., Fujimoto, M., Yamamori, H., Yasuda, Y., Kudo, N., Shishido, E., Okazaki, K., Shiino, T., Nakao, T., Kasai, K., Hashimoto, R., & Onitsuka, T. (2021). Eye Movement Abnormalities in Major Depressive Disorder. *Frontiers in Psychiatry*, *12*. https://www.frontiersin.org/articles/10.3389/fpsyt.2021.673443

Tan, L., & Ward, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin & Review*, *15*(3), 535–542. https://doi.org/10.3758/PBR.15.3.535

Tang, E. Y. H., Price, C., Stephan, B. C. M., Robinson, L., & Exley, C. (2020). Impact of Memory Problems Post-stroke on Patients and Their Family Carers: A Qualitative Study. *Frontiers in Medicine*, *7*. https://doi.org/10.3389/fmed.2020.00267

Terneusen, A., Quaedflieg, C. W. E. M., Heugten, C. van, Ponds, R., & Winkens, I. (2022). *The Many Facets of Metacognition: Comparing Multiple Measures of Metacognition in Healthy Individuals*. PsyArXiv. https://doi.org/10.31234/osf.io/hwa65

Thomson, K. S., & Oppenheimer, D. M. (2022). The "Effort Elephant" in the Room: What Is Effort, Anyway? *Perspectives on Psychological Science*, *17*(6), 1633–1652. https://doi.org/10.1177/17456916211064896

Tolin, D. F., Abramowitz, J. S., Brigidi, B. D., Amir, N., Street, G. P., & Foa, E. B. (2001). Memory and memory confidence in obsessive–compulsive disorder. *Behaviour Research and Therapy*, *39*(8), 913–927. https://doi.org/10.1016/S0005-7967(00)00064-4

Tulving, E. (2002). Episodic Memory: From Mind to Brain. *Annual Review of Psychology*, *53*(1), 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114

Turvey, C. L., Schultz, S., Arndt, S., Wallace, R. B., & Herzog, R. (2000). Memory Complaint in a Community Sample Aged 70 and Older. *Journal of the American Geriatrics Society*, *48*(11), 1435–1441. https://doi.org/10.1111/j.1532-5415.2000.tb02634.x

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*(6), 635–654. https://doi.org/10.1080/09658210902998047

van Asselen, M., Kessels, R. P. C., Frijns, C. J. M., Jaap Kappelle, L., Neggers, S. F. W., & Postma, A. (2009). Object-location memory: A lesion-behavior mapping study in stroke patients. *Brain and Cognition*, *71*(3), 287–294. https://doi.org/10.1016/j.bandc.2009.07.012

van Asselen, M., Kessels, R. P. C., Kappelle, L. J., & Postma, A. (2008). Categorical and coordinate

R

spatial representations within object-location memory. *Cortex*, *44*(3), 249–256. https://doi.org/10.1016/j.cortex.2006.05.005

Van Asselen, M., Kessels, R. P. C., Wester, A. J., & Postma, A. (2005). Spatial Working Memory and Contextual Cueing in Patients with Korsakoff Amnesia. *Journal of Clinical and Experimental Neuropsychology*, *27*(6), 645–655. https://doi.org/10.1081/13803390490919281

Van der Stigchel, S. (2020). An embodied account of visual working memory. *Visual Cognition*, *28*(5–8), 414–419. https://doi.org/10.1080/13506285.2020.1742827

Vandierendonck, A. (n.d.). *Note on the usage of the linear integrated speed-accuracy score (LISAS)*. https://www.researchgate.net/profile/Andre-Vandierendonck/publication/330483132_lisas-usage-notepdf/data/5c41ffba458515a4c72f8a5a/lisas-usage-note.pdf?origin=publication_list

Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior Research Methods*, *49*(2), 653–673. https://doi.org/10.3758/s13428-016-0721-5

Vandierendonck, A. (2021). On the Utility of Integrated Speed-Accuracy Measures when Speed-Accuracy Trade-off is Present. *Journal of Cognition*, *4*(1), 22. https://doi.org/10.5334/joc.154

Verberne, D. P. J., Spauwen, P. J. J., & Van Heugten, C. M. (2019). Psychological interventions for treating neuropsychiatric consequences of acquired brain injury: A systematic review. *Neuropsychological Rehabilitation*, *29*(10), 1509–1542. https://doi.org/10.1080/09602011.2018.1433049

Verhage, F. (1964). *Intelligentie en leeftijd: Onderzoek bij Nederlanders van twaalf tot zevenenzeventig jaar.*

Verhage, F. (1965). Intelligence and Age in a Dutch Sample. *Human Development*, *8*(4), 238–245.

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

Waldron, S. M., Patrick, J., Morgan, P. L., & King, S. (2007). Influencing Cognitive Strategy by Manipulating Information Access. *The Computer Journal*, *50*(6), 694–702. https://doi.org/10.1093/comjnl/bxm064

Walvoort, S. J. W., van der Heijden, P. T., Wester, A. J., Kessels, R. P. C., & Egger, J. I. M. (2016). Self-awareness of cognitive dysfunction: Self-reported complaints and cognitive performance in patients with alcohol-induced mild or major neurocognitive disorder. *Psychiatry Research*, *245*, 291–296. https://doi.org/10.1016/j.psychres.2016.08.007

Wechsler, D. (2012). *WAIS-IV-NL: Wechsler adult intelligence scale – Nederlandstalige bewerking. Technische Handleiding.* (4th ed.). Pearson Assessment and Information BV.

Welch, L. W., Cunningham, A. T., Eckardt, M. J., & Martin, P. R. (1997). Fine Motor Speed Deficits in Alcoholic Korsakoff's Syndrome. *Alcohol: Clinical and Experimental Research*, *21*(1), 134–138. https://doi.org/10.1111/j.1530-0277.1997.tb03741.x

Williams, J. R., Robinson, M. M., Schurgin, M., Wixted, J., & Brady, T. (2022). *You can't "count" how many items people remember in working memory: The importance of signal detection-based measures for understanding change detection performance*. PsyArXiv. https://doi.org/10.31234/osf.io/jnt7r

Winstein, C. J., & Pohl, P. S. (1995). Effects of unilateral brain damage on the control of goal-directed hand movements. *Experimental Brain Research*, *105*(1), 163–174. https://doi.org/10.1007/BF00242191

Winters, B. D., Saksida, L. M., & Bussey, T. J. (2008). Object recognition memory: Neurobiological mechanisms of encoding, consolidation and retrieval. *Neuroscience & Biobehavioral Reviews*, *32*(5), 1055–1070. https://doi.org/10.1016/j.neubiorev.2008.04.004

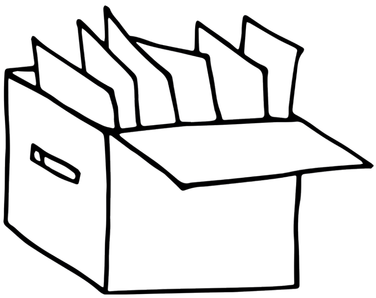
Xie, W., & Zhang, W. (2018). Familiarity Speeds Up Visual Short-term Memory Consolidation: Electrophysiological Evidence from Contralateral Delay Activities. *Journal of Cognitive Neuroscience*, *30*(1), 1–13. https://doi.org/10.1162/jocn_a_01188

Xie, W., & Zhang, W. (2023). *Effortfulness of Visual Working Memory: Gauged by Physical Exertion*. https://oce-ovid-com.proxy.library.uu.nl/article/00004785-202307000-00015/HTML

Zhang, W., & Luck, S. J. (2009). Sudden Death and Gradual Decay in Visual Working Memory. *Psychological Science*, *20*(4), 423–428. https://doi.org/10.1111/j.1467-9280.2009.02322.x

R

Supplementary Materials



Supplementary Materials: General contains supplementary materials for Chapter 2, 3 and 5 that overlap (e.g., material descriptions), and displays group demographics, memory capacity scores and inspection behaviour for an overview.

Supplementary Materials: Chapter-specific lists chapter-specific supplemental results.

Supplementary Materials: General

Contains supplementary materials for Chapter 2, 3 and 5 that overlap (e.g., detailed material descriptions), and displays demographics and memory capacity scores and inspection behaviour values per group for an overview as supplement to the General Discussion.

Contents

Detailed task and outcome descriptions

Copy Task

The Copy Task allows to extract many different variables. The variables of main interest are described in the chapters (sometimes with a slightly different variable name). Here, other outcome measures are listed for completeness.

Performance measures

- Completion time (s, median): median time (in seconds) someone took to finish the trial, or to reach the time limit of 42 seconds.
- Net copying time (s, median): the median time (in seconds) someone was actively copying, i.e., completion time minus the gaze-contingent waiting time
- Number of correct placements: count of correct item placements within a trial. The maximum is six item placements per trial.
- Number of incorrect placements: count of erroneous item placements within a trial. A higher count indicates worse performance.
- Total attempts: the sum of the number of correct and incorrect placements.
- Wrong per correct ratio: mean ratio of how many mistakes were made per correct item placement.
- *Success rate*: ratio between the number of correct placements and the total attempts. High scores indicate accurate performance.
- *Speed score*: net copy time divided by the number of correct placements, that is, the net copy time per correctly placed item. High scores indicate worse speed performance.
- *Linear integrated speed-accuracy score (LISAS),* following

$$LISAS = RTij + \ PEij\ x \frac{S_{RT}}{S_{PE}}$$

where RT_{ij} (reaction time) denotes the trial *i* net copying time divided by the number of correct placements for individual j. The reaction time data was log transformed to account for skewness associated with time measures. PE_{ij} refers to the proportion of errors on trial *i* and equals 1 minus the number of correct placements divided by the total attempts in that trial. S_{RT} denotes the individual *j*'s overall net copying time standard deviation, and S_{PE} is the individual *j*'s overall *PE* standard deviation. Standard deviations were calculated for individual j by collapsing all trials without split on condition. The *LISAS* was chosen as it combines two outcomes of performance (accuracy and speed) and weighs their importance equally. Lower LISAS reflects better (i.e., more accurate and faster) performance.

Eye-movement measures

Note: 'inspection' and 'crossing' can be used interchangeably, but indicate the same construct. Similarly, 'dwell time', 'inspection time' and 'encoding time' are used interchangeably, but indicate the same construct.

- *N*umber of inspections (or: crossings): An inspection is defined as a saccade (eye-movement) that crosses the midline from right (workspace) to left (example model), i.e., crossing. This measure reflects the mean count of how frequently someone inspected the example model per trial.
- Inspection (or: dwell, encoding) time at the model (s, median): median total time per trial that someone inspected the example model in seconds.
- *Inspection* (or: dwell, encoding) *time per inspection*: the time someone took to inspect the example model per inspection. Equals the total inspection time divided by the number of inspections over the course of the trial.
- *Number of inspections* per correct: mean ratio of *only* those saccades that cross the midline from right to left (see *number of inspections)*, divided by the number of correct placements.
- *Inspection* (or: dwell, encoding) time per correct (s, median): median inspection time at the model divided by the number of correct placements over the course of a trial. This score serves as a measure of how much viewing time (i.e., encoding time) someone needed to correctly place a single item.

Conceptually, an external sampling strategy would translate to a relatively high number of inspections. A memorization strategy would translate to a relatively low number of inspections. Memorization, then, would also encompass longer inspection times per inspection in order to encode more items.

Change Detection Task

In lab settings, working memory capacity is commonly assessed using Change Detection Tasks (Luck & Vogel, 2013). In the current protocol, a simplified version of the Change Detection Task from Luck and Vogel was used (Luck & Vogel, 1997; Oudman et al., 2020; see Figure S1.1). With a varying set size of 2, 3, 4, or 6 items, white bars in different orientations (0°, 30°, 60°, 90°, 120°, 150°) were presented on a black screen for 1000 ms, followed by a gaussian random visual white noise mask for 300 ms. After the white noise screen, the bars were presented again, one of which was cued by a surrounding red square. The orientation of the non-target bars did not change, but the orientation of the cued bar changed in 50% of trials.

Participants verbally reported whether or not they detected a change in the orientation of the cued bar.

Five practice trials with feedback were completed prior to the experiment. After practice, 4 blocks of each 20 trials without feedback were completed, adding up to an approximate duration of 10 minutes. Every set size was presented 20 times in random order. *d'* (dprime) was calculated as capacity outcome measure. *d'* is stated to yield a robust outcome for visual working memory performance that is less prone to biases in response tendency than, for example, Kmax (Williams et al., 2022).

$$d' = z[p(hits)] - z[p(falsealarms)]$$

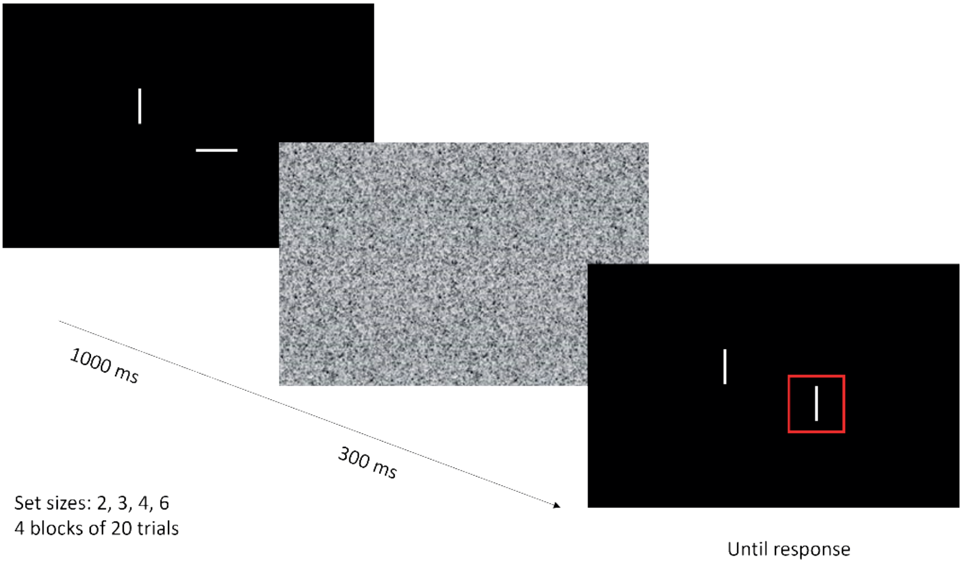


Figure S1.1. Trial overview of the Change Detection Task. Set sizes vary from 2, 3, 4, to 6 white bars in orientations 0°, 30°, 60°, 90°, 120°, 150°.

Neuropsychological memory capacity tasks

Location Learning Task (LLT). Standard stimulus set B of the modified Location Learning Task (Kessels et al., 2006, 2014) was used to assess visuospatial immediate and long-term recall. Subjects were given the instruction to closely inspect a board with a 5 x 5 matrix containing 10 line drawings of objects for 15 seconds, and to memorize the locations of the objects as accurately as possible. This procedure was repeated for five times and after each presentation patients were instructed to place the items on the correct position (the correct cell) in an empty matrix. The ten object cards were given one by one in random order. Before the start, one practice trial (2 x 2 matrix containing two items) was performed to ensure task comprehension. After a delay phase (ideally 20-30 minutes, but due to various reasons in our sample ranging from 25 to 50 minutes), patients were unexpectedly asked to locate the objects again without seeing the stimulus board.

Primary outcomes measures are the learning index (amount of learning over five trials), displacement errors (sum of errors over five trials), and the delayed recall score (subtraction of delayed recall placement error minus placement error of fifth trial). Higher displacement error scores indicate worse performance, a higher learning index indicates better performance (Kessels et al., 2014). A negative delayed recall score indicates loss of information during retention phase, whereas a positive score indicates a better memory after the retention phase (Kessels et al., 2014). The displacement error score is reversed in pre-processing of the data to ensure that higher numbers reflect better performance.

Rey Auditory Verbal Learning Task (RAVLT). The Rey Auditory Verbal Learning Task (RAVLT; 15 items, Dutch version; Bouma et al., 2012; Saan & Deelman, 1986) was administered to assess verbal immediate and long-term recall. Participants were instructed to memorize a long list of words, without time or order restrictions. Fifteen unrelated, but easy to visualize words (subtest A) were read out loud (1 word every 2 seconds). The procedure was repeated five times. After each repetition, participants needed to recall all the words they memorized, also the ones that they mentioned in a previous trial. After a delay phase (ideally 20-30 minutes, but due to various reasons in our sample ranging from 25 to 50 minutes), participants were unexpectedly asked to recall the words again without hearing them again. Outcome measure used are: total number of correct words (range: 0-75) and number of correct words during the delayed recall (range: 0-15). Higher scores indicates better memory capacity.

Digit Span Test (WAIS-IV). We used the Digit Span subtest Forward and Backward from the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV; Wechsler, 2012) to assess verbal working memory. The test administrator reads a sequence of digits out loud. Each part consists of eight items of each two series, that increase in length up to a maximum of 8 (backward) or 9 (forward) digits. During the DSTF, short-term auditory memory is measured, and the participant has to repeat the sequence in the same order. During the DSTB, the participant has to repeat the items backward to measure verbal working memory. The longest sequence that was correctly repeated was used as an outcome measure for maximum capacity (range 2–8 or 2-9). As such, higher scores indicate better performance.

Corsi Block Tapping Task. A digitized version (2D) of the Corsi Block-Tapping Task was used to assess visuospatial working memory (Brunetti et al., 2014; Claessen et al., 2015; Corsi, 1972; Kessels et al., 2000) where nine blue squares (30x30mm) were presented on a tablet (255x205 mm, see Figure S1.2). A sequence of squares, that increases in length up to a maximum of 8 (backward) or 9 (forward), lights up in yellow (500ms flashing time, 1000ms interval, Brunetti et al., 2014; Claessen et al., 2015). Participants were instructed to tap the squares in the same sequence or to tap them backwards. The forward subtest assesses short-term visuospatial memory; the backward subtests assesses visuospatial working memory. The longest sequence that was correctly repeated was used as an outcome measure for maximum capacity (forward range 2–9, backward range 2-8), and total scores were calculated by multiplying this capacity score with the series-score (e.g., to gain insight in whether people only had one or two sequences correct for that span). Higher scores indicate better performance.

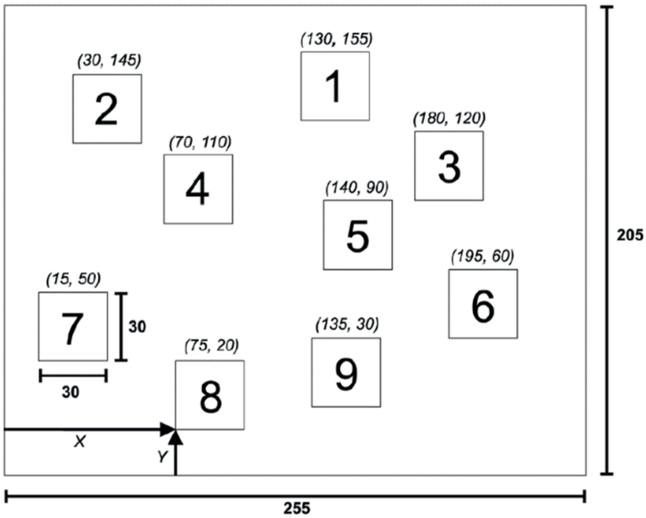


Figure S1.2. Layout of the tablet version of the Corsi Block Tapping Task.
Adapted from Kessels et al., 2000.

S

Questionnaires

Memory complaints. Participants were asked whether they experienced memory problems (yes/no). The answer was used to categorize participants as having or not having subjective memory problems.

Verkorte Vermoeidheidsvragenlijst (fatigue). We used the 4-statement Dutch short fatigue questionnaire (*Verkorte Vermoeidheidsvragenlijst*) to assess fatigue experienced in the previous two weeks (Alberts et al., 1997; Bleijenberg et al., 2009). One of the statements is: “I feel tired”. On a 7-point scale, participants were asked to indicate to what extent the statement was true, with higher scores indicating more fatigue. One statement was rephrased ("I feel fit"), so that lower scores indicated more fatigue, and needed to be reversed in scoring. Total scores range from 4 to 28, and a score ≥ 18 indicates severe fatigue. These were reported as a group descriptive.

Hospital Anxiety and Depression Scale. The Dutch Hospital Anxiety and Depression Scale is a 14-item self-report questionnaire that is often administered in clinical care as a screener to assess complaints of anxiety (7 items) and depression (7 items), without focusing on physical complaints (Spinhoven et al., 1997). Scores can be interpreted per subdomain. Scores within the range of 0–7 indicate no anxiety or depression, 8–10 indicate possible anxiety or depression, and scores of 11–21 indicate probable anxiety or depression (Jungen et al., 2019). Note that these results alone cannot be used to make a clinical diagnosis, but rather serve as an indicator of the presence of distress (Spinhoven et al., 1997).

Metamemory in Adulthood. The abridged version of the Dutch Metamemory in Adulthood questionnaire was adapted from Ponds & Jolles (1996). It consists of 58 items that inquire about memory and attention, and an additional 16 items that ask about strategies people apply to support memory in daily life. Participants indicated the extent to which they agree with the statement on a 5-point scale. Several scale scores can be computed: Task, Capacity, Change, Anxiety, Achievement, Locus, External Strategies, and Internal Strategies. A memory self-efficacy score can be derived from the Capacity, Change, and Anxiety subscale together on the same 5-point scale.

Utrechtse Coping List. This questionnaire asks about cognitive and behavioural strategies that individuals may use when confronted with problematic situations (e.g., coping; Gregório et al., 2014; Schreurs et al., 1984). The questionnaire consists of 47 4-point scale items divided into 7 subscales that capture emotional, behavioural and/or cognitive response tendencies. The 7 subscales are Active confrontation (7 items), Palliative reaction (8 items), Avoiding (8 items), Seeking social support (6 items), Passive reaction (7 items), Emotional expression (3 items), and Comforting thoughts/Optimism (5 items).

Group overview: Korsakoff syndrome, outpatient memory clinic, CVA, controls

Group demographics

Table S1.1. *Demographics per group (Korsakoff syndrome, memory clinic, cerebrovascular accident (CVA), controls). Median (IQR) and frequencies (%) are depicted.*

	Korsakoff syndrome (chapter 2)			**Memory clinic** (chapter 3)			**CVA** (chapter 5)			**Controls** (chapter 2, 3, 5)		
	n	**n (%)/ Mdn (IQR)**	**Range**	***n***	**n (%)/ Mdn (IQR)**	**Range**	***n***	**n (%)/ Mdn (IQR)**	**Range**	***n***	**n (%)/ Mdn (IQR)**	**Range**
Demographics												
Age, years	24	64 (8.5)	47 – 74	29	67 (10)	37-80	15	61 (8.5)	49 – 84	38	60	40 – 81
Sex, % male	24	16 (66.7%)		29	17 (58.6%)		15	10 (66.7%)		38	15 (39.5%)	
Level of education[a]	24	4.5 (1.25)	3 – 7	29	6 (1)	4-7	15	6 (1.5)	3 – 7	38	6 (1.75)	4 – 7
% low (<10 years)					5 (17.2%)			4 (26.7%)			5 (13.2%)	
% medium (10 - 11 years)					8 (27.6%)			3 (20%)			12 (31.6%)	
% high (>15 years)					16 (66.7%)			8 (53.3%)			21 (55.3%)	
Aetiology information												
Time post-admission, year	24	3.1 (7.4)										
Suspected neurological aetiology	24	24 (100%)		29			15	15 (100%)				
% yes					17 (58.6%)							
% no					11 (37.9%)							
% ambiguous diagnosis					1 (3.4%)							
Time post-stroke onset, days							15	74 (46.5)	36 – 137			
Stroke history, % first							15	86.7%				
Stroke type							15					
% ischemic								11 (73.3%)				
% intracerebral hemorrhage								3 (320%)				
% subarachnoid hemorrhage								1 (6.7%)				
Lesion side,							14[b]					
% left								11 (78.6%)				
% right								2 (14.3%)				
% bilateral								1 (7.1%)				
MoCA (0-30)							15	24 (4)	18 – 30			
SAN (Aphasia Index; 1-7)							15	6 (2)	4 – 7			
Barthel Index (0-20)							15	13 (7.5)	5 – 19			
Motricity Index Arm (0-99)[c]							15	76 (80)	0 – 100			
Motricity Index leg (0-99)[c]							15	83 (39)	0 – 100			

n = sample size, Mdn = median, IQR = interquartile range, range (min.–max.). Sample size may differ per outcome variable. a The level of education is characterized according to the classification of (Verhage, 1964, 1965), that is commonly used in Dutch clinical care, and classifies the level of education (ranging from 1 to 7) based on the number of education years. b One unknown. c Motricity Index maximum score is 99. However, if both arm and leg score 99/99, one point may be added, so a score of 100 is possible and indicates intact arm and leg function.

Memory capacity task scores

Table S1.2. *Scores on memory capacity tasks, and questionnaires, split per group (Korsakoff syndrome (Chapter 2), referred to memory clinic (Chapter 3), CVA (Chapter 5), and controls (Chapter 2,3 and 5), medians (IQR) or frequencies, and score ranges are depicted.*

	Korsakoff syndrome (chapter 2)			**Memory clinic** (chapter 3)			**CVA** (chapter 5)			**Controls** (chapter 2, 3, 5)		
	n	**n (%)/ Mdn (IQR)**	**Range**	***n***	**n (%)/ Mdn (IQR)**	**Range**	***n***	**n (%)/ Mdn (IQR)**	**Range**	***n***	**n (%)/ Mdn (IQR)**	**Range**
Neuropsychological task scores												
Location learning task	23			29			15			38		
Total displacement score [a]		85 (50.5)	45 – 129		49 (35)	5 – 150		29 (40.5)	6 – 101		27.5 (26)	0-75
Learning index (0–1)		0.11 (0.08)	0.03 – 0.3		0.29 (0.21)	0.05 – 1		0.47 (0.28)	0.14 – 0.95		0.52 (0.42)	0.10 – 1
Delayed recall: Placement errors				28	3.5 (10.5)	0 – 39		1 (7)	0 – 17	37	1 (4)	0 – 19
Rey auditory–verbal learning task	22			28			14			37		
Immediate recall: Total correct (0–75) [a]		25 (7.5)	14 – 36		36.5 (15.2)	13 – 51		34 (20.2)	18 – 54		42 (18)	30 – 67
Delayed recall: Total correct (0–15)		1 (2)	0 – 4		4.5 (5)	0 – 13		6 (4.5)	0 – 14		8 (6)	3 – 14
Digit span (WAIS-III/IV)	24			29			13			38		
Forward span (2–9) [a]		5 (1)	4 – 8		5 (1)	4 – 7		5 (1)	4 – 7		6 (2)	4 – 9
Backward span (2–8) [a]		4 (2)	2 – 7		4 (1)	2 – 7		4 (0)	3 – 5		5 (2)	2 – 8
Corsi block-tapping task	23			29			15					
Forward span (2–9) [a]		5 (0)	1 – 8		5 (2)	2 – 7		5 (0.5)	4 – 7	38	5 (1)	3 – 8
Backward span (2–8) [a]		5 (1)	2 – 7		5 (2)	2 – 7		5 (2)	4 – 8	37	6 (1)	2 – 7
Impairment within memory domain [b]	23[c]			29			15			38		
% impaired		8 (34.8%)			5 (17.25%)			0 (0%)			2 (5.3%)	
% below average		13 (56.5%)			15 (51.7%)			6 (40%)			5 (13.2%)	
% within normal range		2 (8.7%)			9 (31.05%)			9 (60%)			31 (81.5%)	
Memory capacity compound, *z*	24	-0.63 (0.53)	-1.78 – 0.31	29	-0.08 (1.01)	-1.44 – 0.97	15	0.10 (0.54)	-0.84 – 1.58	38	0.51 (0.85)	-0.73 – 1.66

n = sample size, Mdn = median, IQR = interquartile range, range (min.–max.). Sample size may differ per outcome variable. [a] Capacity scores used in memory capacity compound z-score; Location Learning Task displacement errors are reversed so that higher scores indicate better performance on all capacity tasks. [b] Impaired: a score <2nd percentile on ≥ 2 sub tasks[a]; Below average: a score <2nd percentile on 1 sub task[a] and/or a score between 2nd – 9th percentile on ≥ 2 sub tasks[a]; Within normal range: does not fit criteria for impairment or below average. [c] If scores ≥ 2 (sub)tasks, not taken into consideration.

Inspection behaviour

Table S1.3. *Group scores per group (Korsakoff syndrome (Chapter 2), referred to memory clinic (Chapter 3), CVA (Chapter 5), and controls (Chapter 2,3 and 5), medians (IQR), and score ranges are depicted for outcomes of performance and inspection behaviour across conditions (low-cost and high-cost).*

Copy Task Scores	Korsakoff syndrome (chapter 2)			Memory clinic (chapter 3)			CVA patients (chapter 5)			Healthy controls (chapter 2, 3, 5)		
	n	Mdn (IQR)	Range	*n*	Mdn (IQR)	Range	*n*	Mdn (IQR)	Range	*n*	Mdn (IQR)	Range
Completion time, s	24			29			15			38		
Low-cost		31.85 (13.97)	17.46 – 42		21.5 (11)	13.3 – 42		28.5 (12.8)	15.9-42.0		19.8 (6.38)	12.6-33.4
High-cost		42 (0.001)	33.68 – 42		38.2 (6.42)	30.8 – 42		42.0 (2.14)	35.1-42.0		34.6 (11.2)	26.0-42.0
Net copying time, s	24			29			15			38		
Low-cost		31.85 (13.97)	17.46 – 42		21.5 (11)	13.3 – 42		28.5 (12.8)	15.9-42.0		19.8 (6.38)	12.6-33.4
High-cost		32.34 (4.96)	21.65 – 37.77		28 (5.45)	22.4 – 32.33		31.3 (4.09)	22.8-36.0		24.6 (4.77)	18.1-36.0
Correct placements (0-6)	24			29			15			38		
Low-cost		5.84 (0.45)	3.22 – 6		6 (0.154)	4.39 – 6		5.93 (0.64)	2.57-6		6 (0)	5.57 -6
High-cost		3.54 (1.33)	1.78 – 6		5.46 (1.52)	3.07 – 6		4.43 (1.49)	2.36-6.71		5.79 (0.70)	2.23-6
Success rate (0-1)	20			27			15			36		
Low-cost		0.97 (0.08)	0.86 – 1		0.973 (0.031)	0.779 – 1		0.96 (0.04)	0.77-1		0.97 (0.06)	0.844-1
High-cost		0.82 (0.22)	0.37 – 0.99		0.874 (0.058)	0.738 – 0.981		0.85 (0.14)	0.55-0.98		0.91 (0.13)	0.497-0.99
Speed score, s	24			29			15			38		
Low-cost		5.622 (2.4)	3 – 14.17		3.7 (2.18)	2.35 – 11.3		5.31 (3.08)	2.72-18.3		3.46 (1.37)	2.13-6.73
High-cost		11.95 (6.02)	3.68 – 20.77		5.24 (3.16)	3.75 – 11.3		7.84 (4.81)	3.8-18.7		4.43 (1.21)	3.33-15.2
LISAS	20						15			36		
Low-cost		1.89 (0.58)	1.15 – 2.97	26	1.38 (0.63)	0.94 – 2.52		1.7 (0.7)	0.99-3.28		1.28 (0.46)	0.75-2.14
High-cost		3.19 (1.33)	1.65 – 5.12	27	2.03 (0.54)	1.49 – 2.75		2.32 (0.65)	1.39-3.49		1.72 (0.42)	1.23-3.71
Number of inspections	24			29			15			38		
Low-cost		11.7 (4.68)	8.07 – 18.78		9.89 (3.29)	5.96 – 18.6		11.8 (3.04)	8.21-17.1		10.9 (2.25)	2.43-17.4
High-cost		3.82 (1.9)	2.04 – 6.3		4.73 (1.16)	2.93 – 6.18		4.14 (1.95)	2.57-6.07		4.56 (1.83)	1.86-7.77
Inspection time per crossing, s	24			29			15			38		
Low-cost		0.49 (0.12)	3.56 – 1.06		0.465 (0.137)	0.27 – 1.058		0.44 (0.19)	0.36-0.71		0.4 (0.12)	0.26-0.71
High-cost		1.22 (1.03)	0.54 – 4.22		1.172 (0.661)	0.557 – 3.729		1.56 (1.30)	0.59-4.32		1.01 (0.71)	0.53-5.19
Number of inspections per correct placement	24			29			15			38		
Low-cost		2.14 (1.1)	1.35 – 4.65		1.88 (0.673)	0.994 – 5.44		2.18 (1.23)	1.4-3.91		1.84 (0.44)	0.41-2.95
High-cost		1.2 (0.31)	0.75 – 2.29		1.05 (0.546)	0.488 – 1.66		1.03 (0.61)	0.60-2.06		0.89 (0.39)	0.31-1.90
Inspection time per correct placement, s	24			29			15			38		
Low-cost		1.01 (0.69)	0.56 – 3.23		0.742 (0.376)	0.444 – 2.61		1.02 (0.55)	0.56-2.24		0.70 (0.20)	0.33-1.30
High-cost		1.49 (1.03)	0.54 – 3.05		1.14 (0.638)	0.558 – 2.8		1.43 (0.83)	0.59-4.75		0.87 (0.37)	0.55-2.81

Valid datasets *n*, median Mdn, interquartile range (IQR), and range (min.–max.). Completion time and net copying time in the low-cost condition are the same. In the high-cost condition, the net copying time is the completion time minus the hourglass waiting time.

Individual strategies

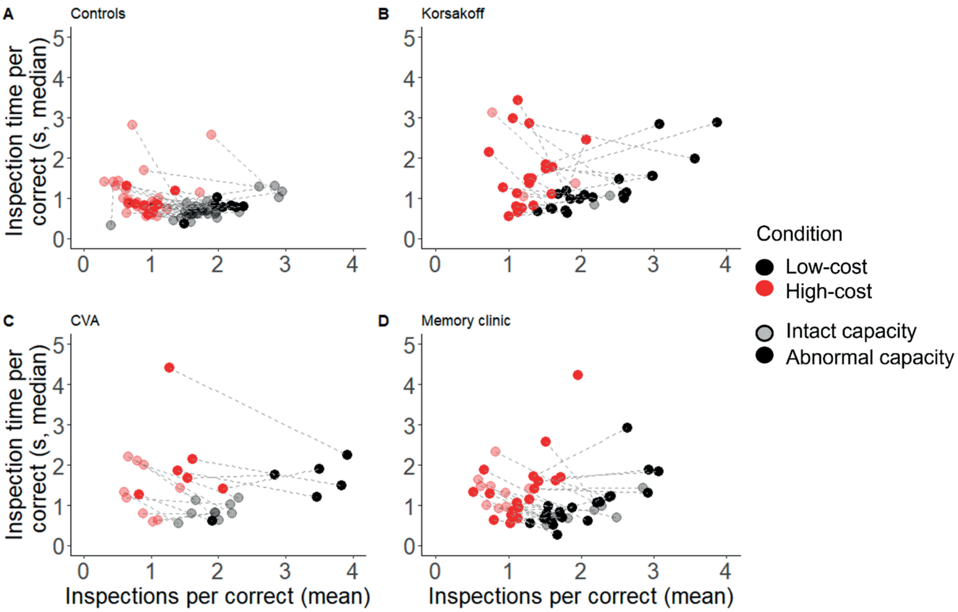


Figure S1.3. Reliance on the external world, presented as median inspection time per correct placement (e.g., encoding time per item) and the number of inspections per correct placement for the four groups: ***A)*** healthy controls, ***B)*** Korsakoff syndrome, ***C)*** CVA (stroke) patients, ***D)*** patients referred to memory clinic. Data points represent data of the individual aggregated over trials in the low-cost condition (black) and high-cost condition (red). Dashed lines connect the data that belong to the same individual, and indicate the change in inspection frequency for an individual from low-cost to high-cost (i.e., change factor for inspections). Saturated dots indicate individuals that showed below average or extremely low (thus, abnormal) memory performance on traditional neuropsychological memory capacity assessment. Inspection per correct values higher than 1 indicate low-loading behaviour. Values equal to or less than 1 but greater than 0.33 indicate medium-loading, and values equal to or less than 0.33 indicate high-loading.

Supplementary Materials: Chapter-specific

Lists the chapter-specific supplemental results as referred to in the separate chapters (with exclusion of task descriptions, see Supplementary Materials: General)

Contents

Supplementary Results: Chapter 2

Patient flow chart

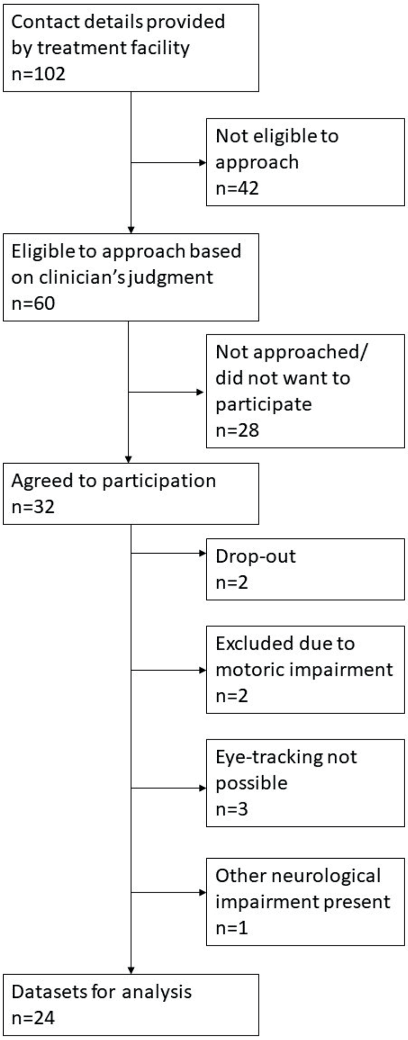


Figure S2.1. Patient flow chart of the recruitment and inclusion process.

Procedure

Table S2.1. *Test procedure for healthy controls and patients with KS.*

	Day 1			Day 2
	Session 1	break	Session 2	Session 2
Healthy controls	LLT Copy Task session 1 LLT – delayed Digit Span WAIS IV If time allowed: Fixation and Free viewing		RAVLT Copy Task session 2 RAVLT – delayed Corsi Block Tapping Task If time allowed: Change Detection Task	n.a.
Patients with KS	LLT Copy Task session 1 LLT – delayed Digit Span WAIS IV If time allowed: Fixation and Free viewing	n.a.		RAVLT Copy Task session 2 RAVLT – delayed Corsi Block Tapping Task If time allowed: Change Detection Task

Correlations with Copy Task outcome measures

Table S2.2. *Spearman's correlations (ρ, raw p-value) for correlations of educational level and age with the outcome measures on the Copy Task that are used in our prediction models for patients with Korsakoff's syndrome (KS) and healthy controls in both conditions of the Copy Task (baseline, high cost). Significant correlations were found for age and both performance and eye-movement measures on the Copy Task in both healthy controls and patients, but these effects are not bothersome in further group comparisons, as groups were age-matched.*

	Baseline				**High cost**			
	Patients with KS		**Healthy controls**		**Patients with KS**		**Healthy controls**	
	ρ	*p*	*ρ*	*p*	*ρ*	*p*	*ρ*	*p*
Correlated variable: Educational level								
Success rate	0.073	.756	0.270	.191	0.14	.557	0.089	.673
Speed score	0.106	.623	-0.34	.083	0.059	.786	-0.191	.341
Number of crossings	-0.187	.384	-0.094	.642	0.17	.428	-0.240	.227
Dwell time per crossing	0.384	.064	-0.115	.569	-0.108	.612	0.192	.337
Number of crossings per correct	-0.015	.945	-0.96	.632	0.205	.338	-0.323	.1
Encoding time per crossing per correct	0.136	.526	-0.105	.601	0.1	.64	0.117	.56
Correlated variable: Age								
Success rate	-0.084	.726	-0.3	.146	0.541	.014*	-0.072	.733
Speed score	0.362	.082	0.533	.004**	-0.153	.476	0.197	.325
Number of crossings	0.112	.591	0.125	.536	0.43	.036*	0.554	.003**
Dwell time per crossing	0.543	.006**	0.453	.018*	-0.399	.054	-0.5	.008**
Number of crossings per correct	0.216	.312	0.148	.461	0.462	.023*	0.573	.002**
Encoding time per crossing per correct	0.403	.051	0.313	.111	-0.26	.223	-0.212	.289

*$p \leq .05$, **$p \leq .01$, ***$p \leq .001$

S

Dynamic VWM strategy

Data loss – drift check descriptives.

A drift check was performed before the start of each trial, by computing the root mean squared error (RMS) of the gaze prediction on a central fixation cross which was shown for two seconds. When the RMS was greater than > 2º (degree visual angle), a warning message was displayed. Large RMSs could occur due to the participant not paying attention, not fixating stably, intermittent head movements during the trial due to which participants' position was changed, or drift of the eye tracker. Whenever a warning message was displayed, the experimenter could opt for either a second try, recalibration, or to move forward with the measurement error regardless. All RMS values were logged. Regardless of drift check implementation, some trials were initiated with large RMS. *Crossings* are quite crude of an outcome measure, but we decided to remove trials that were initiated with a mean measurement error of >5º nonetheless. 44 trials exceeded the threshold, and they were all from healthy controls. For patients with KS, the mean drift check value was 0.9 (range 0.11 – 4.34), and mean SD was 0.45 (range 0.03 – 5.18). For controls, the mean drift check value was 0.9 (range 0.09 – 4.97), and mean SD was 0.53 (range 0.03 – 6.0).

Outlier removal

Table S2.3. *Linear mixed-effects coefficient estimate and test-statistics (t, p) for outcome measures on the Copy Task predicted by factors Group, Condition, and Group * Condition with and without inclusion of outliers (N observations for controls (HC) and patients (KS).*

		Outliers included			Outliers excluded		
	N outlier (HC, KS)	Estimate	t	*p*	Estimate	t	*p*
Outcome variable							
Success rate	2 (2,0)						
- Group		-0.02	-1.53	.133	-0.03	-2.16	.037*
- Condition		-0.09	-3.37	.002**	-0.07	-2.81	.008**
- Group*Condition		-0.10	-2.77	.009**	-0.12	-3.03	.004**
Speed score	5 (3,2)						
- Group		3.19	5.57	<.001***	2.66	6.05	<.001***
- Condition		1.63	2.97	<.005**	1.20	2.12	.04*
- Group*Condition		3.73	4.65	<.001***	4.21	5.1	<.001***
Number of crossings	0						
- Group		2.17	3.47	.001**			
- Condition		-5.64	-12.1	<.001***			
- Group*Condition		-2.34	-3.44	.001**			
Dwell time per crossing	7 (3,4)						
- Group		136.93	2.92	.004**	104.81	2.25	.025
- Condition		1296.85	4.97	<.001***	1026.71	6.96	<.001
- Group*Condition		45.13	0.12	.91		NA[A]	NA[A]
Number of crossings per correct	2 (0,2)						
- Group		0.71	4.08	<.001***	0.56	3.85	<.001***
- Condition		-0.88	-8.65	<.001***	-0.88	-9.64	<.001***
- Group*Condition		-0.24	-1.59	.12	-0.15	-1.11	.27
Dwell time per correct	6 (3,3)						
- Group		0.71	4.61	<.001***	0.45	4.99	<.001***
- Condition		0.41	2.16	.039*	0.29	1.7	.096
- Group*Condition		0.27	0.97	.34	0.62	2.48	.017*

*p≤05, **p≤.01, ***p≤.001

A After outlier exclusion, the linear mixed-effects model failed to converge, suspectedly because the removal of seven participants led to insufficient data to make predictions. We simplified the model by removing the (earlier non-significant) interaction effect, after which the main effects were again observed.

S

Discussion - Forward Span analysis

Table S2.4. *Linear mixed-effects coefficient estimates and raw p-values for additionally analysed factors (fixed covariates are level of education and age, and forward span score of interest) within the linear mixed-effects regression models to predict sampling behaviour (crossings and dwell time per correct placement) for the patients with Korsakoff's Syndrome on the Copy Task split on condition (baseline, high cost).*

	No. of crossings per correct placement				**Dwell time per correct placement**			
	Baseline		**High cost**		**Baseline**		**High cost**	
Digit Span – FW span	**Est.**	**Raw *p***	**Est.**	**Raw *p***	**Est.**	**Raw *p***	**Est.**	**Raw *p***
N=24								
Education	-0.088	.622	0.075	.280	0.080	.586	0.138	.424
Age	0.038	.184	0.011	.303	0.040	.096.	-0.033	.229
FW Span	-0.102	.622	-0.068	.396	-0.160	.352	-0.076	.703
Corsi – FW span								
N=23								
Education	-0.258	.146	0.052	.489	-0.062	.683	0.086	.637
Age	0.017	.568	0.016	.220	0.035	.187	-0.029	.356
FW Span	-0.401	.062	0.005	.953	-0.225	.219	-0.018	.933

Note. Digit Span – FW span = forward span on the WAIS IV Digit Span, Corsi – FW span = forward span on the Corsi Block Tapping Test. *$p \leq .05$.

Discussion - Ineffective crossings

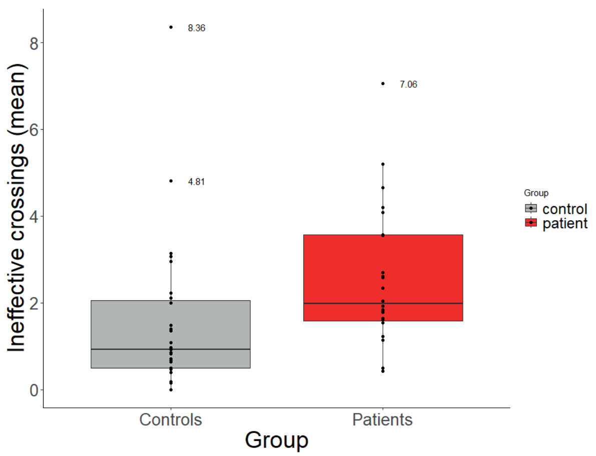


Figure S2.2. The aggregated number of ineffective crossings in the high cost condition (in baseline, the example was not occluded, and therefore, all crossings could be used for sampling) for controls (grey) and patients with Korsakoff's syndrome (red). The number of ineffective crossings differed significantly across groups (Mann-Whitney-Wilcoxon U=140099, p<.001, rank-biserial correlation r=-0.31). Black dots represent outcomes of individual participants. Outlier values (1.5*interquartile range) are indicated.

Supplementary Results: Chapter 3

Patient flowchart

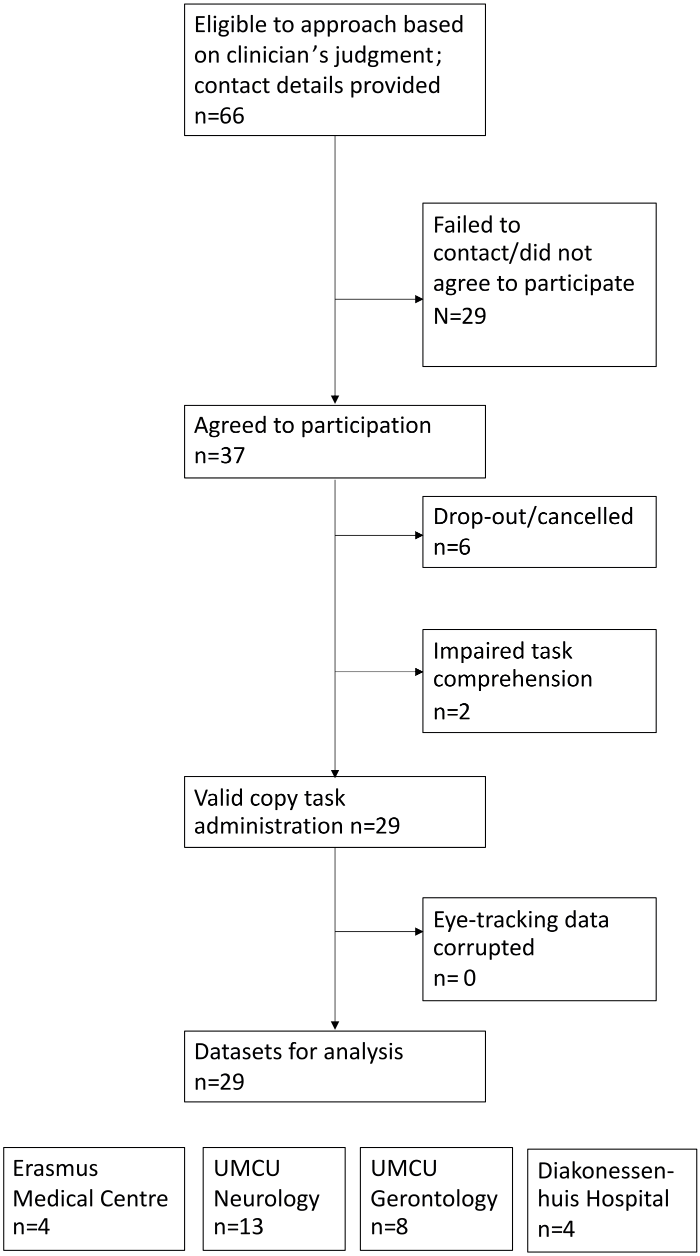


Figure S3.1. Patient flowchart: division of recruitment and flow outcome per outpatient clinic

Procedure

Table S3.1. *Test procedure*

At home	At testing facility		
14 – 1 day(s) before testing	Session 1		Session 2
1. Memory complaints 2. CFQ 3. VVV 4. HADS* 5. COCO-P	LLT* Copy Task session 1 LLT – delayed* Digit Span WAIS IV* If time allowed: Fixation and Free viewing	break	RAVLT* Copy Task session 2 RAVLT – delayed* Corsi Block Tapping Task If time allowed: Change Detection Task 1. MIA

*Task/questionnaire administration only when not yet administered as part of standard care and within six months prior to lab visit

Suspected neurological aetiology within referred sample

Non-neurological / neurological (17)

- Mild Cognitive Impairment (8)
- Dementia (5)
  - Alzheimer's Disease (2)
  - Frontotemporal dementia (1)
  - Mixed (2)
- Epilepsy (1)
- Parkinson spectrum (2)
- Unclear, but suspected neurological (1)
- Unclear whether neurological or psychological
- Unclear, but suspected psychological
- No impairment

neurodegenerative

other neurological

Figure S3.2. Overview of suspected neurological aetiology within referred sample. Of 29 referred individuals, 17 had a suspected neurological syndrome. 12 individuals did not have objectifiable impairments, or their cognitive complaints were explained by psychological factors.

Non-referred controls flowchart

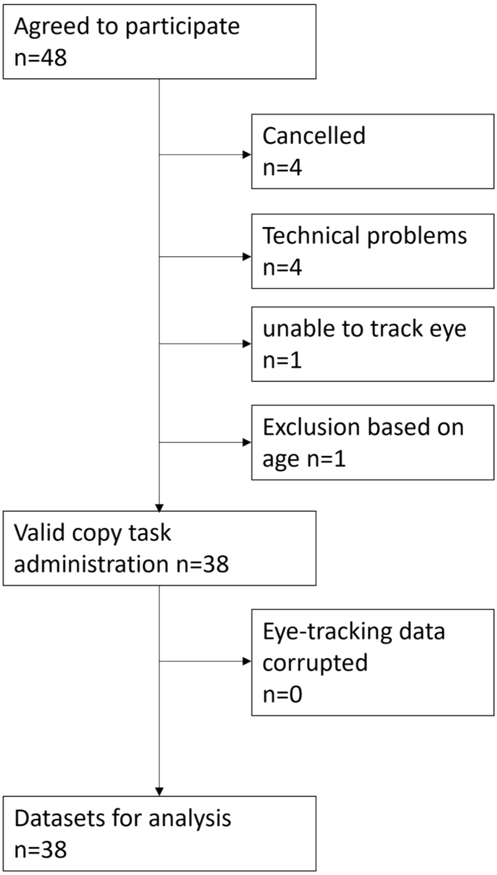


Figure S3.3. Flowchart of age- and education matched controls.

Subjective and objective memory performance overlap across groups

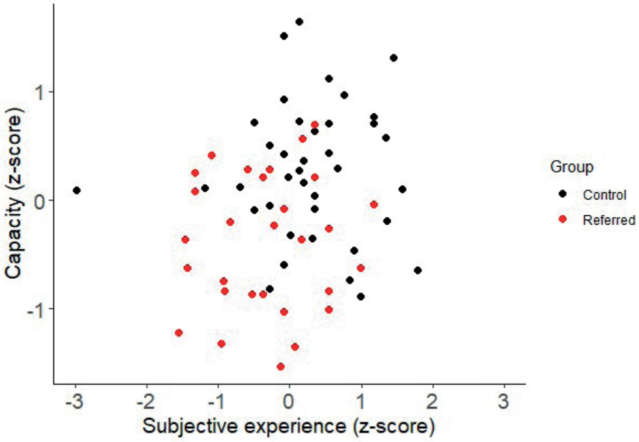


Figure S3.4. Visualization of overlap in groups (red = referred, black = control) for memory self-efficacy compound (subjective experience z-score) and objective capacity compound (z-score).

Inspection behaviour based on memory capacity and CFQ.

As part of the memory self-efficacy component, we administered both the Cognitive Failure Questionnaire (CFQ) and the Metamemory in Adulthood (MIA) questionnaire. The CFQ was included in the battery from the start. However, when piloting the tasks and questionnaires we realized that the CFQ gives a limited view of self-perceived memory self-efficacy as it is focused on broader cognition, while the MIA specifically focuses on this. Unfortunately, we did not have immediate access to the Dutch version of the MIA questionnaire, which is why we could only administer it after the study had already started. Therefore, not all participants completed the (n=18 for referred individuals, n=15 for non-referred controls).

To subserve this hiatus in the number of data sets, we have checked the correlation of the subscale Absentmindedness of the CFQ that was used in the memory self-efficacy composite and the MSE sum score that is derived from the MIA. They correlated well (p=.009, r=0.44), so we are confident that the CFQ is apt to describe memory self-efficacy for those that did not complete the MIA. Further, all scores were transformed to z-scores, so the composite score weighs the two questionnaires equally.

However, to make sure that the missing data does not bias interpretation, we performed the analyses for memory self-efficacy again but without inclusion of the MIA scores. The interpretation of results did not change (all p >0.25 for any effect in the LMM for CFQ Absentmindedness on *number of inspections, dwell time, and LISAS*). In conclusion, it is not problematic that the MIA scores are not available for the entire sample.

Predictive value of memory capacity subtasks and level of memory functioning.

Each of the capacity scores was included in a separate regression model to predict behaviour in the copy task. We ran the models for both conditions separately, as we hypothesised that memory capacity would influence behaviour mostly in a situation where it is beneficial to tax working memory (high-cost condition) and not necessarily when information is freely available.

S

Table S3.2. *Unstandardized coefficient estimates, raw p-values and Holm-Bonferroni adjusted p-values for factors (fixed covariates level of education and age, and capacity score of interest) within the regression models to predict sampling behaviour for both referred and non-referred individuals (taken together) on the Copy Task (crossings per correct placement, dwell time per correct) split on condition (baseline, high cost).*

	No. of crossings per correct placement						**Dwell time per correct placement**					
	Baseline			**High cost**			**Baseline**			**High cost**		
Digit Span – FW span	**Est.**	**Raw *p***	**Holm**	**Est.**	**Raw *p***	**Holm**	**Est.**	**Raw *p***	**Holm**	**Est.**	**Raw *p***	**Holm**
N=67												
Education	0.073	0.314		-0.057	0.067		0.023	0.606		-0.041	0.536	
Age	0.009	0.226		0.016	**.000**		0.01	**.026**		-0.006	0.395	
Digit Span Forward	-0.2	**.002**	**.030**	-0.1	**.000**	**.005**	-0.156	**.000**	**.003**	-0.091	0.108	1
Digit Span – BW span												
N=67												
Education	0.065	0.378		-0.056	0.07		0.008	0.869		-0.049	0.459	
Age	0.012	0.109		0.017	**.000**		0.012	**.009**		-0.004	0.526	
Digit Span Backward	-0.16	**.007**	.091	-0.096	**.000**	**.003**	-0.103	**.008**	.096	-0.062	0.243	1
Corsi – FW span												
N= 67												
Education	0.039	0.6		-0.065	**.034**		-0.009	0.843		-0.058	0.376	
Age	0.009	0.24		0.015	**.000**		0.011	**.029**		-0.005	0.433	
Corsi Span Forward	-0.138	**.047**	.376	-0.107	**.000**	**.004**	-0.084	.058	.376	-0.057	0.345	1
Corsi – BW span												
N=66												
Education	0.022	0.751		-0.08	**.004**		-0.018	0.684		-0.067	0.301	
Age	0.006	0.383		0.014	**.000**		0.009	0.052		-0.005	0.495	
Corsi Span Backward	-0.217	**.0006**	.084	-0.12	**.000**	**.0001**	-0.144	**.000**	**.005**	-0.045	0.424	1
D’												
N=59												
Education	0.014	0.759		-0.061	**.043**		-0.005	0.814		-0.061	0.323	
Age	0.003	0.492		0.012	**.000**		0.004	0.055		-0.012	0.072	
D’	-0.145	**.025**	.225	-0.145	**.000**	**.008**	-0.105	**.000**	**.014**	-0.038	0.644	1
LLT - Placement Errors												
N=67												
Education	0.008	0.92		-0.089	**.006**		-0.027	0.561		-0.073	0.255	
Age	0.011	0.149		0.017	**.000**		0.011	**.022**		-0.003	0.656	
Placement errors	0.002	0.43	1	0.002	**.04**	.480	0.002	0.158	.632	-0.001	0.641	1

LLT - Learning index												
N=67												
Education	0.022	0.767		-0.077	**.013**		-0.0167	0.72		-0.069	0.288	
Age	0.011	0.144		0.017	**.000**		0.012	**.015**		-0.004	0.545	
Learning index	-0.429	0.115	.575	-0.369	**.001**	**.013**	-0.34	**.048**	.376	-0.076	0.748	1
LLT – delayed recall												
N=65												
Education	-0.001	0.991		-0.086	**.009**		-0.037	0.451		-0.074	0.265	
Age	0.012	0.142		0.018	**.000**		0.012	**.019**		-0.004	0.588	
Delayed recall score	0.003	0.74	1	0.004	0.3	1	0.004	0.512	1	-0.002	0.755	1
RAVLT – total score												
N=65												
Education	0.138	0.065		-0.012	0.648		0.044	0.356		0.017	0.794	
Age	0.001	0.865		0.012	**.000**		0.006	0.231		-0.01	0.137	
Total score	-0.024	**.000**	**.008**	-0.013	**.000**	**.000**	-0.016	**.000**	**.006**	-0.012	**.045**	.495
RAVLT – delayed recall												
N=65												
Education	0.081	0.282		-0.038	0.17		0.006	0.908		-0.011	0.858	
Age	0.003	0.675		0.012	**.000**		0.007	0.166		-0.009	0.206	
Delayed recall score	-0.055	**.018**	.198	-0.036	**.000**	**.001**	-0.035	**.019**	0.198	-0.026	0.178	1

Note. Digit Span – FW span = forward span on the WAIS III or IV Digit Span, Digit Span – BW span = backward span on the WAIS III or IV Digit Span, Corsi – FW span = forward span on the Corsi Block Tapping Test, Corsi – BW span = backward span on the Corsi Block Tapping Test, *D'* = dprime on the Change Detection Task, LLT – Placement Errors = total number of placement errors on the Location Learning Task, LLT – Learning Index = learning ratio over the course of five trials, LLT – Delayed recall = number of items retained after delay, RAVLT – total score = number of correctly reproduced words over five trials from the Rey Auditory Verbal Learning Task, RAVLT – delayed recall score = number of correctly reproduced words on the Rey Auditory Verbal Learning Task after a delay period. *p<.05.

Number of inspections per correct as function of memory functioning category

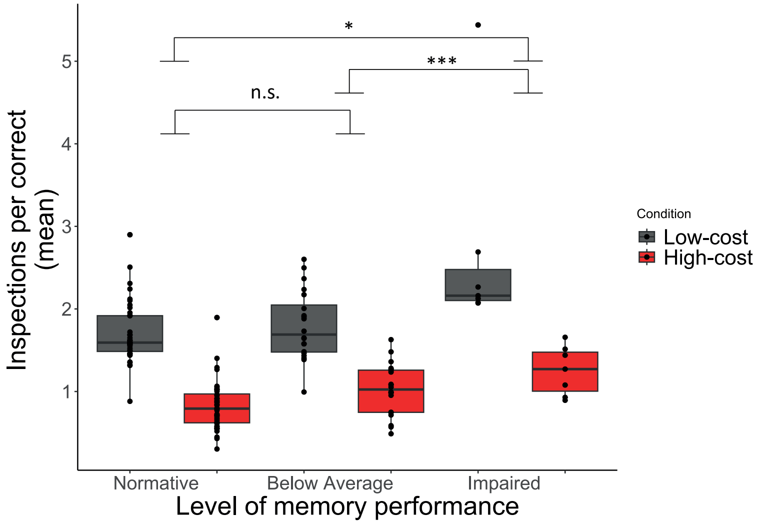


Figure S3.5. Inspections per correct as a function of memory performance category. Groups were merged, and performance was categorized as intact (normative), below average or impaired according to clinical consensus; Raw scores were compared with the appropriate norm groups (for age and education). If performance was below the 2nd percentile on two or more subtasks, memory performance was categorized as impaired. If performance was below the 2nd percentile on one subtask (and/)or below the 9th percentile on two or more subtasks, memory performance was categorized as below average. If these criteria were not met, memory performance was deemed within normal range. Both category and condition significantly influence inspection frequency ($F(2)=16.571$, $p<.001$, $F(1)=118.333$, $p<.001$, respectively). Tukey corrected post-hoc comparisons reveal that the impaired group (n=7) differs significantly from the normative (n=40) and below average group (n=20), but that the normative and below average group do not differ from each other. When removing outliers on the number of inspections per correct, the effects of category and condition held ($F(2)=14.370$, $p<.001$, $F(1)=155.445$, $p<.001$, respectively). Post-hoc comparisons now showed that only the difference between the normative and impaired group was significantly different.

Supplementary Results: Chapter 4

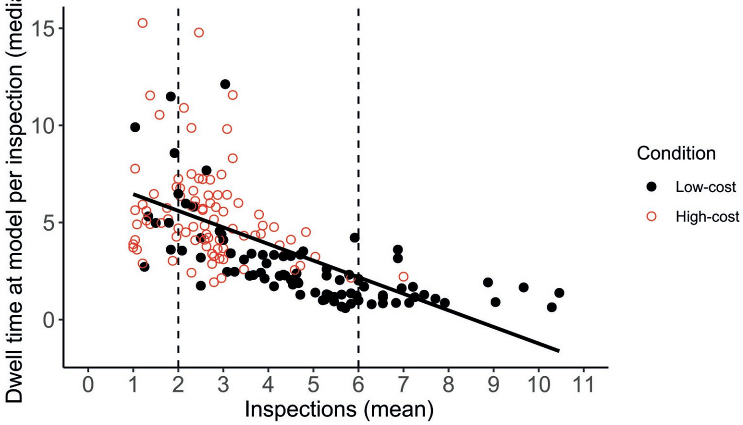


Figure S4.1. Correlational analysis between mean number of inspections per trial and median dwell time at the Model per inspection across conditions for all 88 participants. Dots represent individual datapoints. The vertical dashed lines represent the cut-offs used to discriminate between low-loaders (dots to the right of the right dashed line), medium-loaders (dots on and to the left of the right dashed line), and high-loaders (dots on and to the left of the left dashed line). Colors indicate conditions (black, filled = low-cost; red, open = high-cost). $\tau = -.55$, $p<.001$, $z = -10.76$

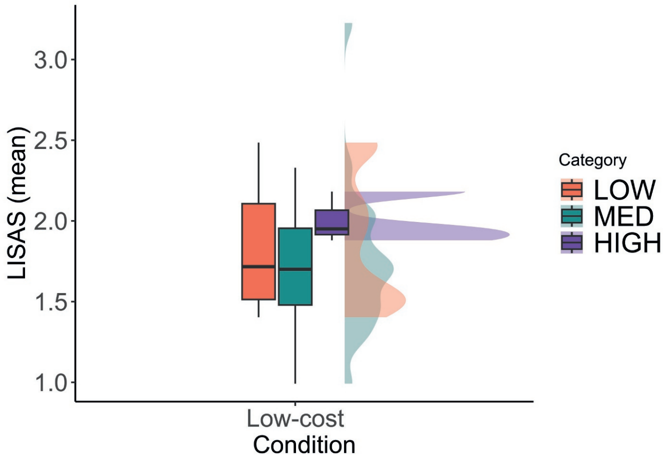


Figure S4.2. LISAS performance across initial sampling preference (low-loader, medium-loader, high-loader) in the low-cost condition for participants that started with the low-cost condition. No differences between categories were present (Kruskal-Wallis test showed no significant difference in LISAS across categories, $\chi^2(2, N= 43) = 1.447$, $p = .485$).

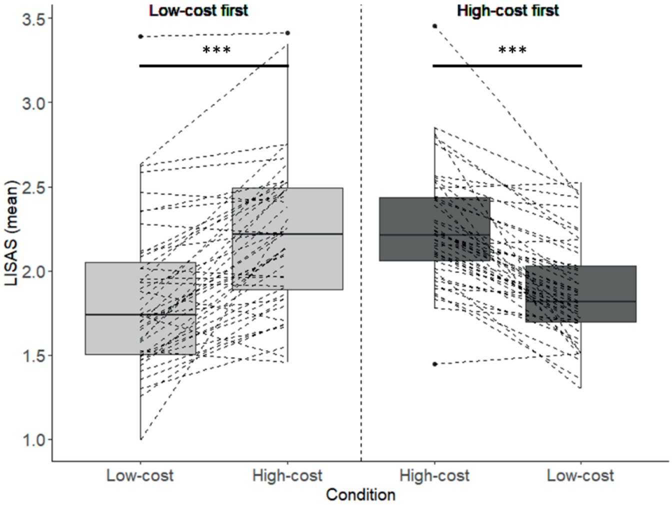


Figure S4.3. Data, presented as mean LISAS (+IQR) per condition (low-cost vs. high-cost) for the different starting conditions (low-cost first (light grey) vs. high-cost first (dark grey)). Higher LISAS values indicate worse performance. LISAS was significantly worse in the high-cost condition (t = 7.59, p < .001, beta = 0.37 [0.27, 0.46]), with no difference between low-cost first vs. high-cost first (t = 0.4, p = .693, beta = 0.03 [-0.13, 0.19]), and no interaction effects (t = 0.20, p = .841, beta =0.01 [-0.12, 0.15]).

Table S4.1. *Overview of the accuracy scores, speed scores and LISAS per order (low-cost first, high-cost first) and per condition (low-cost, high-cost). We provide the mean, median, standard deviation and the range per outcome measure.*

Outcome Measure	**Condition**	**Order**							
		Low-cost first (n =43)				**High-cost first (n = 45)**			
		Mean	**Median**	**Sd**	**Range**	**Mean**	**Median**	**Sd**	**Range**
Accuracy score	low-cost	0.87	0.87	0.05	0.75 – 0.97	0.86	0.87	0.06	0.72 – 0.97
	high-cost	0.78	0.79	0.10	0.56 – 0.95	0.80	0.82	0.10	0.52 – 0.95
Speed score	low-cost	5.56	4.66	2.89	2.43 – 18.72	5.37	5.05	1.62	2.98 – 9.88
	high-cost	6.94	5.81	3.23	3.21 – 20.85	7.27	6.51	2.45	3.88 – 18.06
LISAS	low-cost	1.84	1.74	0.46	0.99 – 3.39	1.87	1.82	0.304	1.3 – 2.52
	high-cost	2.21	2.22	0.44	1.46 – 3.41	2.25	2.21	0.35	1.44 – 3.46

Outlier extraction: reanalysis without outliers

Initially, datasets of all participants were analysed without outlier removal (n = 4219 trials). To make sure that findings were not driven by outliers, we ran the analyses again after removing those trials where scores were ≥1.5 times the interquartile range apart from the group median for that specific outcome measure (number of model inspections, LISAS) in that specific condition (low-cost, high-cost) per order of condition (low-cost first, high-cost first). Outlier extraction resulted in removal of 247 trials (5.8%). 188 (76.1%) of all trials were removed based on the number of model inspections (low-cost first + low-cos*t* = -35; low-cost first + high-cos*t* = -0; high-cost first + low-cos*t* = -125; high-cost first + high-cos*t* = -28). The other 59 (23.8%) were excluded based on LISAS (low-cost first + low-cos*t* = -10; low-cost first + high-cos*t* = -20; high-cost first + low-cos*t* = -3; high-cost first + high-cos*t* = -26). We ran the analysis with a total of 3972 trials (1997 low-cost first; 1975 high-cost first).

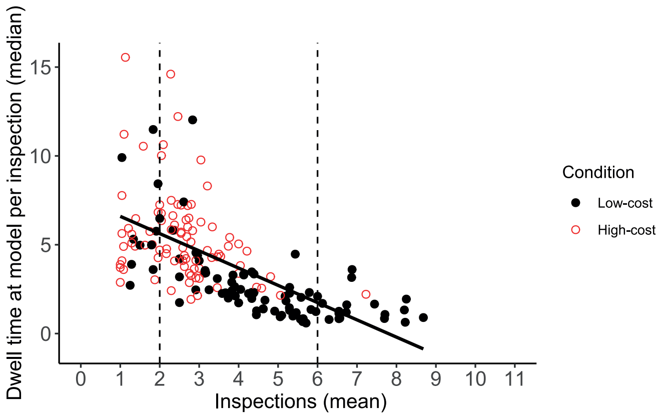


Figure S4.4. Correlational analysis between mean number of inspections per trial and median dwell time at the Model per inspection across conditions for all 88 participants without outlier trials, τ = -.55, p < .001, z = -10.83. Dots represent individual datapoints. Dashed lines indicate cut-offs for strategy categorizations. The vertical dashed lines represent the cut-offs used to discriminate between low-loaders (dots to the right of the right dashed line), medium-loaders (dots on and to the left of the right dashed line), and high-loaders (dots on and to the left of the left dashed line). Colors indicate conditions (black, filled = low-cost; red, open = high-cost).

Table S4.2. *Mean number of model inspections per order (low-cost first, high-cost first), per condition (low-cost, high-cost) and per strategy category (low-loader, high-loader, full-loader) without outlier trials. We provide the number of individuals per category (n) and percentages per category based upon the number of model inspections, and the mean (M), standard deviation (SD) and the range of the number of model inspections per trial. Note that the division in categories is arbitrary and is for descriptive purposes only.*

Condition	**Strategy category**	**Order**							
		Low-cost first (n =43)				**High-cost first (n = 45)**			
		n(%)	**M**	**SD**	**Range**	**n(%)**	**M**	**SD**	**Range**
low-cost	low-loader	17 (39.5%)	7.18	0.80	6.12 – 8.68	0 (0%)	-	-	-
	medium-loader	23 (53.5%)	4.94	0.88	2.83 - 6	37 (82.2%)	3.89	0.94	2.33 – 5.64
	high-loader	3 (7%)	1.74	0.21	1.5 – 1.91	8 (17.8%)	1.57	0.38	1.04 – 2
high-cost	low-loader	1 (2.3%)	7.23	-	-	0 (0%)	-	-	-
	medium-loader	30 (69.8%)	2.83	0.53	2.17 – 4.58	33 (73.3%)	3.08	0.82	2.04 – 5.06
	high-loader	12 (27.9%)	1.37	0.40	1 - 2	12 (26.7%)	1.44	0.37	1 – 2

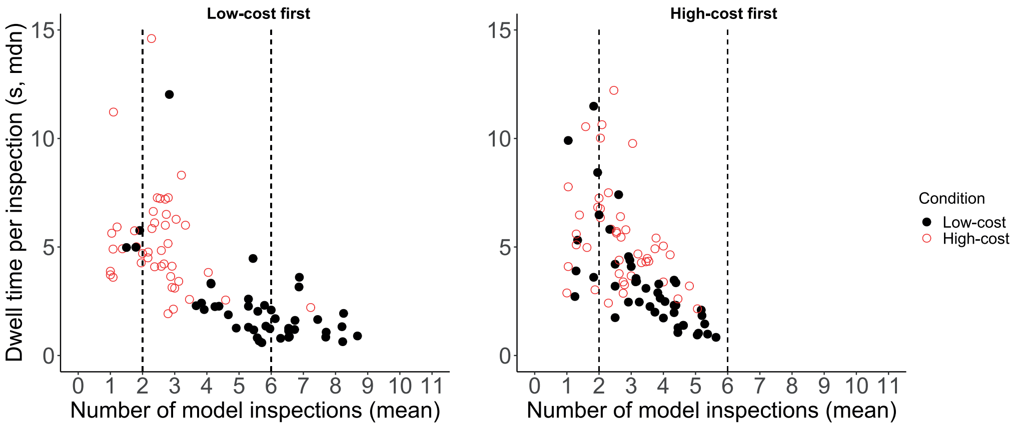


Figure S4.5. Inspection behavior without outlier trials, presented as medial model dwell time per inspection per trial and average number of model inspections per trial. Order: **left)** Low-cost first, **right)** High-cost first. Each datapoint reflects an individual, with the closed black dots representing the low-cost condition and the open red dots the high-cost condition. The vertical dashed lines represent the cut-offs used to discriminate between low-loaders (dots to the right of the right dashed line), medium-loaders (dots on and to the left of the right dashed line), and high-loaders (dots on and to the left of the left dashed line).

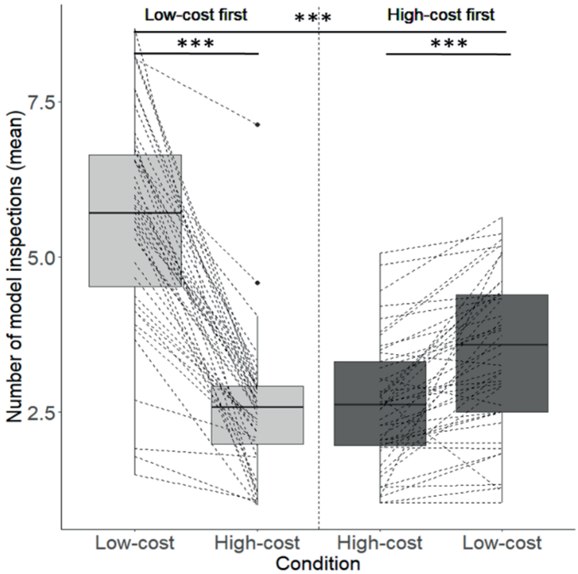


Figure S4.6. Data without outlier trials, presented as mean (+IQR) number of model inspections per condition (low-cost condition, high-cost condition) for the different starting conditions (low-cost first, high-cost first). There are more model inspections in the low-cost condition compared to the high-cost condition (t = -15.54, p < .001, beta =-3.08 [-3.46, -2.69]). The group who started with the high-cost condition (dark-grey) sampled significantly less compared to the participants starting with the low-cost condition(light grey) (t = -6.64, p < .001 beta = -2.12 [-2.75, -1.50]). The interaction-effect revealed that the low-cost condition differentially affected sampling behavior across groups (t = 8.07, p = <.001, beta = 2.24 [1.69, 2.78]), with the high-cost first group making significantly fewer inspections in the low-cost condition.

S

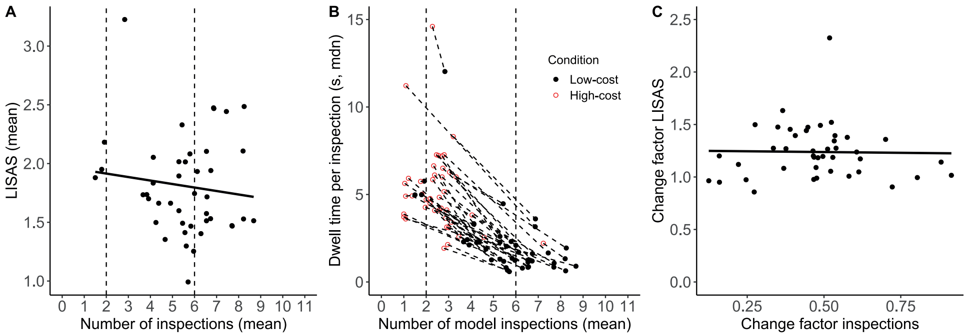


Figure S4.7. Effects of sampling behavior on performance, behavioral strategy adaptations and effects of strategy adaptations on changes in performance for the low-cost first order group without outlier trials. Each point reflects an individual participant (n=43). **A)** In the low-cost condition, the mean number of inspections is not correlated to performance (mean LISAS; higher scores reflecting worse performance; τ = -.03, *p* = .75). The vertical dashed lines represent the cut-offs used to discriminate between low-loaders (dots to the right of the right dashed line), medium-loaders (dots on and to the left of the right dashed line), and high-loaders (dots on and to the left of the left dashed line). **B)** Representation of the behavioral shift based on mean inspections per trial between the low-cost condition (closed black dots) and the high-cost condition (open red dots). The longer the line between two data points, the larger the behavioral adjustment (hence, change) for the individual. **C)** No significant correlation between change factor inspections and change factor LISAS (τ = -.05, *p* = .62). Scores towards zero on change factor number of model inspections reflect a larger adjustment from the low-cost condition to the high-cost condition. The higher the change factor LISAS, the more performance declined in the high-cost condition compared to the low-cost condition. Change factor LISAS <1 indicates performance improvement.

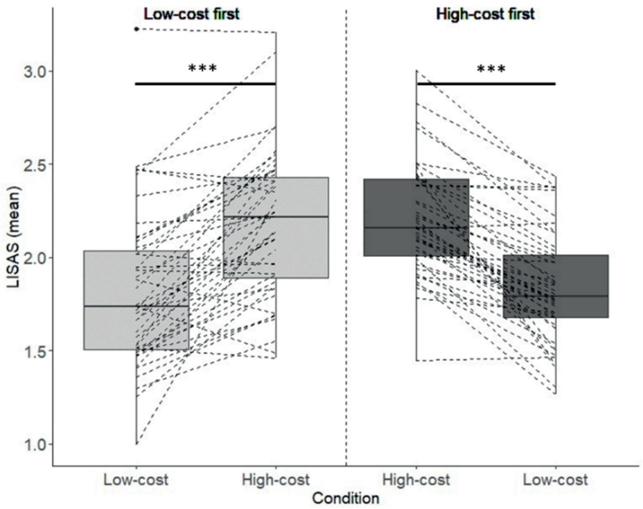


Figure S4.8. Data without outlier trials, presented as mean LISAS (+IQR) per condition (low-cost vs. high-cost) for the different starting conditions (low-cost first; light grey, high-cost first; dark grey). High LISAS values indicate worse performance. LISAS was significantly worse in the high-cost condition (t = 8.07, p < .001, beta =-0.37 [0.28, 0.46]), with no difference between the starting condition (low-cost first vs. high-cost first; t = 0.38, p = .706, beta = 0.03 [-0.12, 0.18]), and no interaction effect (t = -0.14, p = .89, beta = -0.01 [-0.14, 0.12]).

Table S4.3. *Overview of the accuracy scores, speed scores and LISAS per order (low-cost first, high-cost first) and per condition (low-cost, high-cost) without outlier trials. We provide the mean, median, standard deviation and the range per outcome measure.*

Outcome Measure	Condition	Order							
		Low-cost first (n =43)				High-cost first (n = 45)			
		Mean	Median	Sd	Range	Mean	Median	Sd	Range
Accuracy score	low-cost	0.87	0.88	0.05	0.78 – 0.97	0.87	0.88	0.06	0.73 – 0.97
	high-cost	0.79	0.80	0.10	0.56 – 0.95	0.81	0.82	0.10	0.54 – 0.95
Speed score	low-cost	5.38	4.66	2.59	2.43 – 17.2	5.22	4.93	1.55	2.98 – 9.39
	high-cost	6.64	5.80	2.70	3.21 – 19.0	6.83	6.36	2	3.88 – 15.3
LISAS	low-cost	1.81	1.73	0.42	0.99 – 3.23	1.83	1.79	0.3	1.27 – 2.43
	high-cost	2.18	2.22	0.4	1.46 – 3.21	2.2	2.16	0.3	1.44 – 3.0

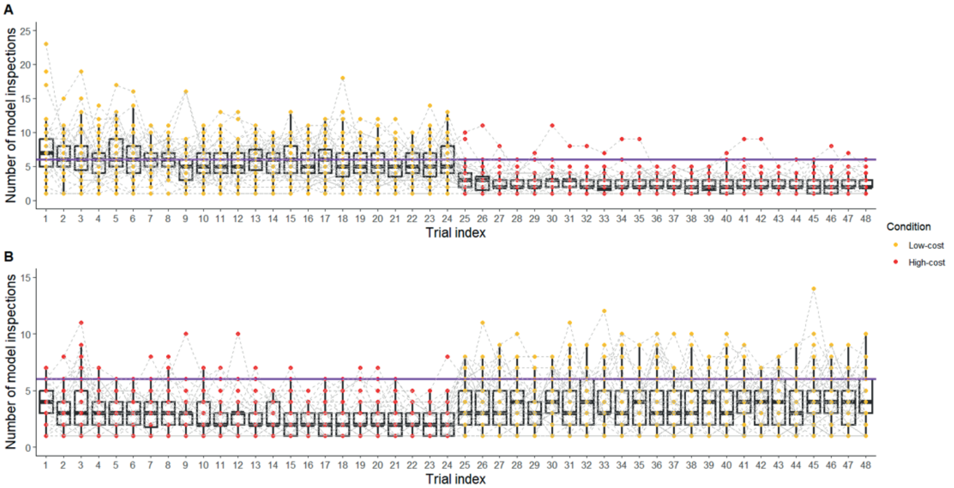


Figure S4.9. Visualization of inspection behavior per individual (grey lines) over the course of the task (as a function of trial number) across the low-cost (yellow) and high-cost (red) condition, separated on the order of condition presentation (**A:** low-cost first, **B:** high-cost first). The purple line indicates our cut-off for low-loading (≥ 6 inspections) versus high-loading (<6 inspections) categorization. NB. Outlier trials are not excluded in this figure.

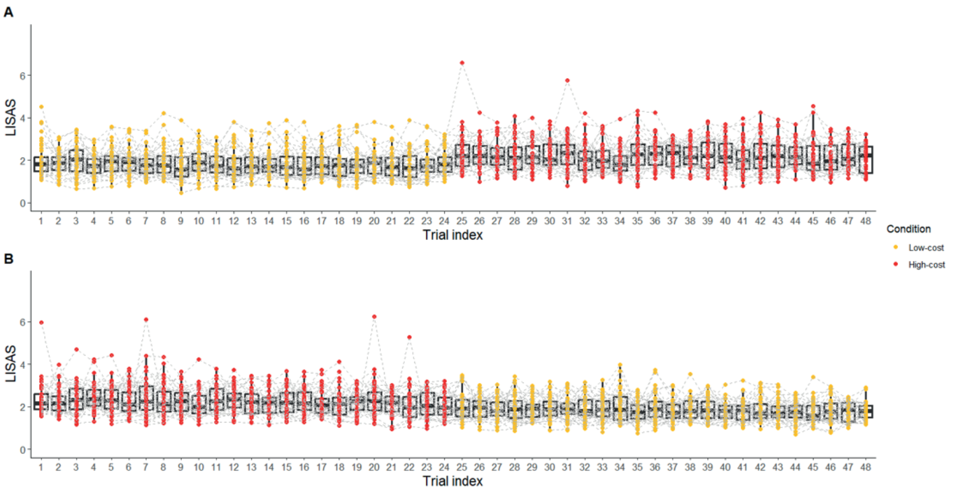


Figure S4.10. Visualization of LISAS performance per individual (grey lines) over the course of the task (as a function of trial number) across the low-cost (yellow) and high-cost (red) condition, separated on the order of condition presentation (**A:** low-cost first, **B:** high-cost first). Higher LISAS indicates worse performance. NB. Outlier trials are not excluded in this figure.

Supplementary Results: Chapter 5

Procedures

Table S5.1. *Test procedures for* ***A)*** *healthy controls and* ***B)*** *CVA patients*

A.

At home	At university testing facility		
14 – 1 day(s) before testing	Session 1		Session 2
1. Memory complaints? 2. CFQ* 3. VVV 4. HADS 5. COCO-P*	LLT Copy Task session 1 LLT – delayed Digit Span WAIS IV If time allowed: Fixation and Free viewing*	break	RAVLT Copy Task session 2* RAVLT – delayed Corsi Block Tapping Task If time allowed: Change Detection Task* 1. MIA

B.

At rehabilitation centre testing facility	At home/room	At rehabilitation centre testing facility
Session 1	1. HADS 2. MIA 3. UCL	Session 2
Memory complaints? VVV LLT Copy Task session 1 LLT – delayed		RAVLT Corsi Block Tapping Task Digit Span WAIS IV RAVLT – delayed

*Data from these questionnaires and tasks were collected for another study, and are discarded in the current study. Abbreviations: CVA = cerebrovascular accident, CFQ = Cognitive Failure Questionnaire, VVV = Verkorte Vermoeidheidsvragenlijst (fatigue), HADS = Hospital Anxiety and Depression Scale, COCO-P = Cognitive Complaints – Participation, LLT = Location Learning Task, RAVLT = Rey Auditory Verbal Learning Task, MIA = Metamemory in Adulthood.

Results

Patient and healthy control flowcharts

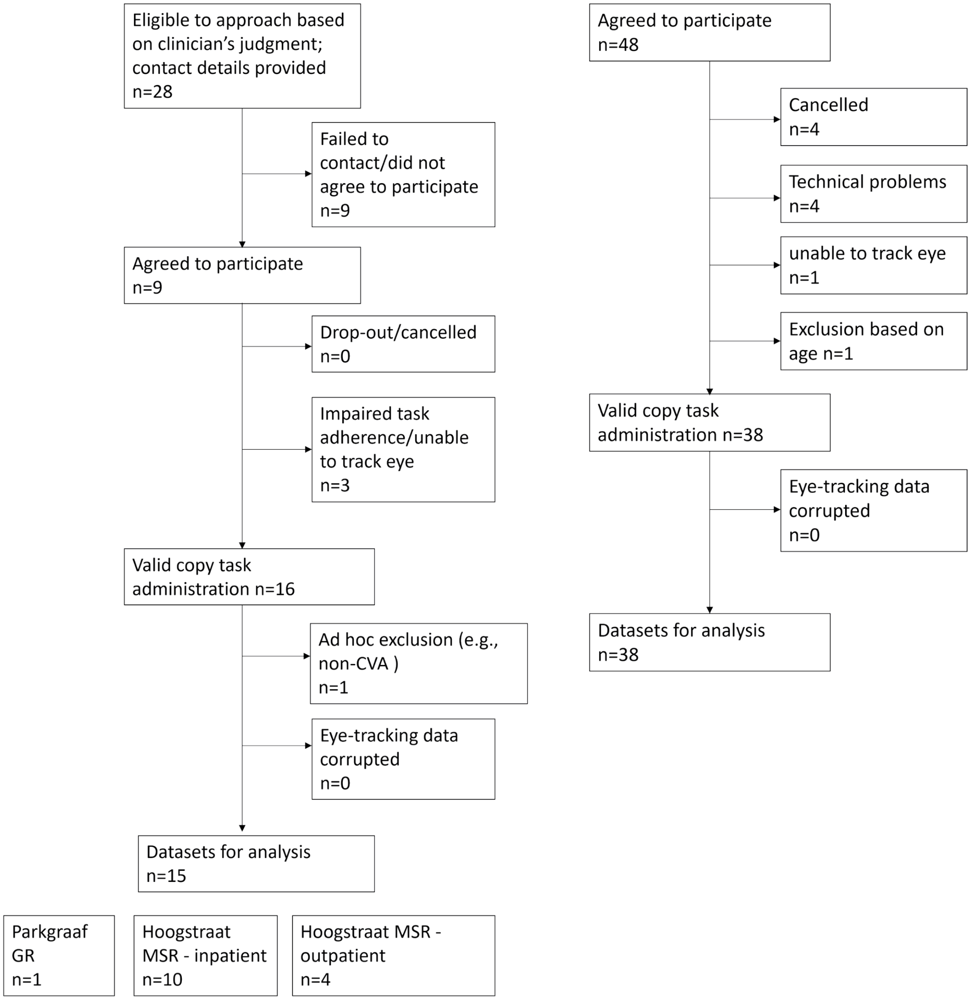


Figure S5.1. Patient and control flow chart

Group inspection behaviour

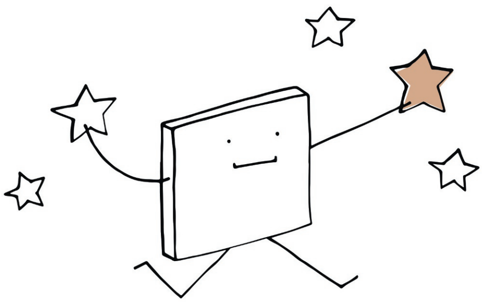
Table S5.2. *Descriptive group scores (CVA patients, healthy controls) for outcomes of performance and inspection behaviour across Copy Task conditions (low-cost and high-cost). Variables in bold are used as index for performance and strategy.*

Copy Task Scores	CVA patients			Healthy controls		
	n	Mdn (IQR)	Range	*n*	Mdn (IQR)	Range
Completion time, s	15			38		
Low-cost		28.5 (12.8)	15.9-42.0		19.8 (6.38)	12.6-33.4
High-cost		42.0 (2.14)	35.1-42.0		34.6 (11.2)	26.0-42.0
Net copying time, s	15			38		
Low-cost		28.5 (12.8)	15.9-42.0		19.8 (6.38)	12.6-33.4
High-cost		31.3 (4.09)	22.8-36.0		24.6 (4.77)	18.1-36.0
Correct placements (0-6)	15			38		
Low-cost		5.93 (0.64)	2.57-6		6 (0)	5.57 -6
High-cost		4.43 (1.49)	2.36-6.71		5.79 (0.70)	2.23-6
Success rate (0-1)	15			36		
Low-cost		0.96 (0.04)	0.77-1		0.97 (0.06)	0.844-1
High-cost		0.85 (0.14)	0.55-0.98		0.91 (0.13)	0.497-0.99
Speed score, s	15			38		
Low-cost		5.31 (3.08)	2.72-18.3		3.46 (1.37)	2.13-6.73
High-cost		7.84 (4.81)	3.8-18.7		4.43 (1.21)	3.33-15.2
LISAS	15			36		
Low-cost		1.7 (0.7)	0.99-3.28		1.28 (0.46)	0.75-2.14
High-cost		2.32 (0.65)	1.39-3.49		1.72 (0.42)	1.23-3.71
Number of crossings	15			38		
Low-cost		11.8 (3.04)	8.21-17.1		10.9 (2.25)	2.43-17.4
High-cost		4.14 (1.95)	2.57-6.07		4.56 (1.83)	1.86-7.77
Inspection time per crossing, s	15			38		
Low-cost		0.44 (0.19)	0.36-0.71		0.4 (0.12)	0.26-0.71
High-cost		1.56 (1.30)	0.59-4.32		1.01 (0.71)	0.53-5.19
Number of inspections per correct placement	15			38		
Low-cost		2.18 (1.23)	1.4-3.91		1.84 (0.44)	0.41-2.95
High-cost		1.03 (0.61)	0.60-2.06		0.89 (0.39)	0.31-1.90
Inspection time per correct placement, s	15			38		
Low-cost		1.02 (0.55)	0.56-2.24		0.70 (0.20)	0.33-1.30
High-cost		1.43 (0.83)	0.59-4.75		0.87 (0.37)	0.55-2.81

CVA = cerebrovascular accident (stroke), Mdn = median, IQR = interquartile range, range (min.–max.). Variables depicted in bold are used in subsequent interpretation of performance (LISAS) and strategy behaviour (number of inspections per correct/dwell time per correct)

S

Acknowledgements (dankwoord)



Mijn onderzoek draaide om *external sampling*: nauwgezet de omgeving inspecteren opdat je het voorbeeld zo goed mogelijk kunt nabootsen. Voeg de termen samen en je eindigt met het woord *example*. Ik heb het geluk gehad dat ik de afgelopen jaren genoeg *examples* om mij heen had waarop ik kon terugvallen, naar wiens voorbeeld ik door *external sampling* vaardigheden heb kunnen afkijken, of die simpelweg mijn buitenwereld waren om mijn binnenwereld aan toe te vertrouwen.

Stefan, hoewel je wist dat ik niet op zoek was naar een promotieplek, haalde je me binnen. Je schonk me niet alleen het vertrouwen, maar ook je relativeringsvermogen, een relaxte werkhouding, een ontvankelijke leiderschapsstijl, en een razend enthousiasme. Ook buiten de muren van de universiteit klikte het: we voerden gesprekken over retraites en zweethutten, bezochten grimmige kunst in Amelisweerd, deelden als haringen in een ton mijn Fred Again awakening op Lowlands, voltooiden een ijzige BorcuLoop en doften ons op voor de Avond van Wetenschap en Maatschappij. Gedegen onderzoek is belangrijk, maar sfeer zeker zoveel. Jij zorgde ervoor dat de ERC niet alleen wetenschappelijk, maar ook sociaal een succes kon worden. Je bent een fantastische promotor. Doe je snel je Boing-shirt weer aan?

Van jou, **Tanja**, leerde ik ‘aaibaarder’ schrijven. Met jouw expertise van en netwerk in het klinische speelveld leerde ik balanceren tussen wetenschappelijk grondig en klinisch uitvoerbaar. Ondanks het scherm dat zich meestal tussen ons in bevond, leidde deze virtuele barrière er niet toe dat je minder scherp op de inhoud was. Onze interactie heeft me veel geleerd over mijn communicatie en mijn (en ons) verwachtingsmanagement verbeterd. Dank dat je me betrok bij Betweter en Lowlands Science; die hoogtepunten had ik voor geen goud willen missen! Ik ben dankbaar voor het wederzijds begrip en de persoonlijk ruimte die we elkaar gunden op momenten dat dat belangrijker was dan werk.

Dan mijn co-pro en dagelijkse vraagbaak chatGP**Teuni**. Als er iemand op dagelijkse basis een *example* was, dan was jij het. Ik bewonder hoeveel kennis en kunde jij in huis hebt, hoe je een laagdrempelige maar constructief kritische voedingsbodem weet te bieden aan eenieder met wie je werkt, en hoe je staat voor wat je doet. Het is bizar (en soms een tikkeltje confronterend) hoeveel wij gemeen hebben. Hoewel de grenzen tussen professioneel en privé soms vervaagden (denk: GinTeunics, GeenTeeni, je deelname aan mijn sportles of openhartigheid in het park) ben ik trots op hoe we ons daar doorheen hebben weten te navigeren. Ik ben blij dat onze samenwerking een vervolg krijgt!

Veel dank aan de scherpe geest en het inclusietalent van **Albert**, **Carla**, **Cindy**, **Erik**, **Esther**, **Matthijs, Patricia**, en **Zoë**. Zonder jullie wil om patiënten op

mijn onderzoek te attenderen was er weinig van de grond gekomen. Dank ook aan het **team van de Hoogstraat en het KCRU,** in het bijzonder **Cindy, Dorrith, Eefje, Laurine, Nienke, Patricia** en **Willemijn,** dat jullie mij wilden adopteren op de woensdagen, bij de kerstlunch en de vakdemo's, en me hielpen met de patiëntdossiers.

Dank aan de **leescommissie** voor hun tijd, aandacht en opponeren bij de verdediging: Prof. dr. C.M. van Heugten, Prof. dr. L.J. Kappelle, Prof. dr. C.N.L. Olivers, Dr. I.M. Wiegand en Prof. dr. M.J.E. van Zandvoort.

Ook een bedankje aan een onbekende; de **schepper van Plankie**. Ere wie ere toekomt: thanks, Manfred Steger.

Lieve **kantoorbuddies**, wat zou H0.15 zijn zonder jullie. Van Teams koffiepraat naar spoffen naar goffen; zoals de koffie beter werd, werd onze band dat ook. **Alex**, zonder jou zou ik compleet verloren zijn geweest in het woud der oogbewegingsdata en die klotecode. Los van alle academische steun die ik aan je gehad heb, hield je me op persoonlijk vlak ook gevraagd en ongevraagd een spiegel voor en ging je kwetsbare communicatie niet uit de weg. Ik ben dankbaar voor je eerlijkheid, zelfs als je van kantoor wegvluchtte vanwege mijn wervelwindenergie. Ik zal het missen om te strijden om het voetenkrukje, om samen pinguïntruien te dragen met kerst, om je koffie-reviews te horen, en, gewoon, je aanwezigheid. **Andre,** An, Annie (sorry not sorry voor de afko), jouw '*hmm*'s', '*goh*'s', en andere uitingen van genialiteit vulden H0.15 met een aangename ruis van aanwezigheid. Ondanks je trage loop- en spreektempo (zucht) heb je een triatletisch snel begrip en barst je van de freerunning ideeën. Blijf vooral doorgaan met BFs flexen en hou vast aan die onverstoorbare relaxedheid die jou zo typeert. Oja, en voor de lezers thuis een raadsel: Andre en ik zijn op respectievelijk de 2^{e} en de 1^{e} van dezelfde maand in hetzelfde jaar geboren. Toch is Andre ouder. Rara, hoe kan dat? **Damian**, Deem, de ultieme LMM-master, is het al siggi? Van dat jaartje later starten is niets te merken; je piekt als wetenschappelijk paperkanon, en hebt je plek in ons midden verankerd. Zonder jou raak ik alles kwijt, maar mét jou evengoed (lees: headbanger). Ik zal je bijdehandheid, je eigen taaltje, en je liefde voor plankie missen. Gelukkig houden King Siggi en King Stroekoe een oogje in het zeil. **Kantoorbuddies**, fliep flap lekker verder en wats de drip. Laat me weten wat er in dat lege pand komt.

Je kunt niet altijd zes gooien. Maar met drie keer vijf kun je ook winnen. **Zoë,** ZuZu, collega, vriendin, co-auteur, coach, fitchecker, factchecker, fan en zoveel meer. Het was slaytastic dat jij de afgelopen vier jaar om de hoek zat en het kantoorbestaan verrijkte met galgjewoorden (als je ze tenminste goed spelde). Of het nou is op goodie two shoes, op boevenpad, of zoëfend op de fiets, je bent geweldig in alle hoedanigheden. Never loose your glitter. **Jeanita,** a.k.a. JayJay, Sjaan, Jea, de ‘geëvolueerde versie van Janet’. Wat begon als valentijnsfling ontwikkelde zich tot een diepgaande vriendschap met wederzijdse biechtmomenten en whiteboardtherapiesessies. Ik leer elke keer weer meer van (en over) je. Ik had het niet zonder jouw openheid, vrolijkheid, energie, aanstekelijke lach, verstopte eendjes en wandelrondjes willen doen, parel.

Van bachelorscriptiebegeleider (goed galgjewoord) naar chococcino-mikado partner: **Chris P.**, van wie moet je verliezen nu ik wegga? Je ouderdom en ongeëvenaarde eetgewoonten – waarom zou je een bord gebruiken?! – zijn voer voor grappen, maar op het gebied van pingpongen en mangotangoën doe je niet onder voor de jongelui. Dank voor je openhartigheid, gezelligheid en humor. Laten we af en toe nog een 2223 doen!

Dan de Ass Profs. **Christoph,** Jigsaw, I came, I conquered. I will remember you by your laughs, the questionable sweets you brought, your German jokes, the animal emoji’s you sent on Teams, the numerous counts your phone fell out of your pocket, the sneakiness with which you tried to go cycling a ‘little bit’ further, the mountain gossip, and your dried tomato quiche. **Sam**, our shared office presence developed from in-phase to anti-phase (sadly). I will never be able to order whipped cream without hearing you call me Sahne. Thanks for you stoic humor and your sincere interest in whatever happened in my life. **Surya**, van geknakte ruggengraat tot gebroken scheenbeen; hopelijk heeft versie 2.0 een niet-rennen-voor-het-OV modus. Dankjewel voor je eeuwige geduld met mijn LISAS-frustraties, je hulp bij mijn LISAS-frustraties, en je gelach om mijn LISAS-frustraties. **Sjoerd**, met een bak décaf (ik) of slootwaterkoffie (jij) in de hand bleek: zo intimiderend ben je helemaal niet. Door jou heb ik geleerd over maskers, enzo. Thanks. Ofzo. Weet ik veel.

Kabir, always suitably wearing your bluegrey-ish (?!) vest as our cookiemonster. As the forever baby newbee you had quite a distance to bridge. Luckily you own the morphs so hard that you blew everyone away, quickly stole our jealous hearts and filled our belly’s with your Snackxpert products. I couldn’t have been more successful to ‘catch’ you as my council successor. Same holds for **Dominique**! Dank ook voor het overnemen van de geweldige organisatie van de introcourse samen met **Martijn**, en de korte maar warme updates in de gang of de pantry! **Larissa**, dank dat jij je opwierp als mede PhD Rep. Jouw down-to-earth manier van doen inspireert me; jij bent de belichaming van ‘waar een wil is, is een planning’. **Marinke** en **Renee**, ik heb genoten van de vete’s, de eendjeslol, de heart to heart conversations en de ik-loop-even-kort-naar-binnen-maar-blijf-dan-toch-hangen momentjes in H0.19. Dat geldt eigenlijk voor de hele bezetting van H0.19: ook bij **Famke, Lara** en **Liselotte** stond de deur altijd open (letterlijk en figuurlijk). Van raamzwaaien tot koffiemomentjes en blèren in de Duke, dank voor jullie haha0.19 enthousiasme en het uitwisselen van ervaringen en verhalen. **Evi,** je bubbly persoonlijkheid en je vruchtenhagelliefde hebben me meer dan eens aan het grinniken gebracht. Paddenstoelenhoedje af voor jou! **Beleke** en **Tessie**, mede dankzij jullie heb ik zo'n berg NPO-data kunnen verzamelen. Ik had me geen betere onderzoekassistenten kunnen wensen! Het was ontzettend leuk om jullie te begeleiden en evenzo cool om te zien dat jullie je eigen PhD traject gestart zijn. Dankjewel, **Martin,** voor je Go-Pro tutorials, je onstuitbare energie en het opbrengen van onderwerpen waar ik waarschijnlijk nog nooit een seconde eerder over nagedacht had. **Sterre** en **Jasmijn**, van voetbal kijken tot ijsjes eten en kletsen; Porto was zoveel leuker dankzij jullie. Obrigada! **Neville**, naast fietsen en hardlopen kon ik ook je praattempo maar moeilijk bijbenen. Dankjewel voor je energie! **Jinyu** and **Luzi**, you have such a kind heart. You always engaged in a conversation with me, and never stopped being curious. **Dan, Daan, Lasse**, **Liangyou, Rick, Tianying, Yichen, Yuqing**, and all **other (PhD-)students**: I really enjoyed our conversations during lunches, conferences or department outings, and getting to know you all. Ook SMARTies **Eileen**, **Charlotte** en **Judy**, dank voor de kletsmomentjes. **Elissa**, op Lowlands Science vonden we elkaar al snel; na de intense diensten is er heus nog wat energie om te dansen, zot! Merci!

Dank aan de **neurostaf**, **attentionlabmembers** en alle andere **collega's van PF,** die mijn interesse in ons vakgebied wisten te wekken én wisten vast te houden over al die jaren. **Carla**, **Irene**, en **Martine**, als stagebegeleiders wisten jullie mij al enorm te inspireren, en dat is nooit gestopt. **Irene**, onwijs bedankt voor de tijd die je ook de latere jaren nam om te kletsen over je loopbaan. Tot gauw bij BASE! **Chris D.**, laten we vaker speculeren over jouw persoonlijke achtergrondvragen bij mijn presentaties. Dank voor je interesse en vertrouwen, en het aanschuiven bij mijn verdediging! **Anouk**, ik vind het superleuk dat je mij op de radar houdt voor alumnibijeenkomsten en baankansen. Onze interactie voelt altijd relaxed. Thanks! **Nathan** en **Jasper**, het zal een flinke kluif geweest zijn mij als programmeer-groentje op code-heavy projecten te begeleiden, maar met geduld, zweet en tranen hebben we er toch artikelen uit weten te pompen. Racen en survivallen hebben voor altijd een dubbele betekenis. Over survivallen gesproken, **Jim**, life is a bumpy road. Dankjewel dat je me meenam op je reis; via de kinderopvang naar de Uithof, of gewoon boven de damp van een warme drank. Ik ken weinig mensen die zo oordeelloos kunnen luisteren als jij. **Michael,** je was nooit te beroerd me te helpen met laptopcrashes, connectiviteitsissues, disfunctionele dongels en wat niet meer, of het nu via Teams was of wanneer ik weer aan de 'bar' kwam hangen. Dankjewel voor het lachen om mijn slechte woordgrappen. Ook **Mark** en **Halim**, dank voor het oplossen van mijn tech-issues! **Janet, Eline, Eric met een C,** en **Karin**, jullie waren vertrouwde gezichten tussen de wetenschappers. Dank voor jullie hulp bij véél meer dan etiketten printen, mappen bestellen, het uitlenen van de PF-borrelpas, en bij personeelsvragen die ik drie keer moest stellen voor ik het antwoord snapte. **Murat**, bedankt voor de gezellige praatjes over vakantie en je elektrische fiets! Een speciale shout-out (dit zou eigenlijk door de intercom moeten) aan **Niyazi**, die binnenkomst voor mij al sinds 2013 vrolijk en vertrouwd maakt.

Alba, Tambó! I don't know whether to write this in English or Dutch, aangezien we elkaar in het Engels ontmoetten bij de master maar jij inmiddels full-on Dutchie bent. I thank you from the bottom of my heart for your wisdom, craziness, wild dances, hospitality, and your support when going through large transitions. You show tough love with a softness surrounding it. Give **Jesse** a hug for me. **Coen**, ik ben zo blij dat het voorzittersjaar slechts het begin van onze vriendschap bleek te zijn. Ik koester onze (te) sporadische koffiemomentjes en je goedlachsheid. Dank aan de **Lunsj**genootjes, met een speciale juich voor **Emily**! Ik bewonder je kracht in het omarmen van de kwetsbaardere momenten in het leven. Ook **Laura** verdient een dikke mention; opdat er nog vele ontboezemwandelingen mogen volgen.

Lief **26e**, lieve **Maatje**, **Essie**, **Lous**, **Nikie**, **Lot** en **Midg** (uiteraard volgens constitutie), we zien en spreken elkaar niet meer zoveel als vroeger, maar ik voel jullie aanwezigheid over land en zee heen en weet dat we altijd op elkaar kunnen terugvallen. Jullie zijn de upper bij mijn downers. Een propje en een sticker voor onze psychologenvriendenpraat. **Janneke**, ik bewonder je zelfstandigheid, onverschrokken(broek)heid, communicatie en warmte. Dankjewel voor de wederzijdse bezoekjes, telefoontjes en aanmoediging.

Forever roomies van de Kerkstraat, **Suze, Nynke, Jade** en **Judith**, ook als we niet meer samenwonen voelt het als thuiskomen bij jullie. Ik had me geen fijnere huisgenootjes kunnen wensen om mee thee te leuten, chocola te verorberen, mee te lachen en te huilen, en in slonzig huispak bij elkaar op de bank te vallen. **Alessa, Janine** en **Jolien**, dank jullie wel voor de praatjes in de keuken, de check-ins vanuit de deurpost, de deur opendoen als ik daar geen zin in had, en de kneuterige kaarsverf- en friettafelmomentjes.

Miek, Bieb, Cel, Char, Es, Jip, Juul, Lee, Lot, Mies, Niño, Ries en **Spits**, dank jullie wel voor jullie (ongetwijfeld soms geveinsde) interesse in en het aanhoren van mijn werkperikelen, voor de nodige afleiding tijdens clubweekenden, etentjes en borrels, en de armen om mij heen als het wat minder ging. **Char** en **Lee**, jullie waren initieel misschien wat sceptisch over mijn beslissing de PhD aan te nemen, maar coachten me met alle geduld van de wereld door die keuze en alles wat daarna kwam heen. Dank!

Zonder **koffie** was dit proefschrift niet tot stand gekomen. **Michel** en alle **30ML**lers, ik genoot (en geniet) enorm van jullie tweede-huiskamer-vibes, de grote cappuccino's, en de ins en outs over het komen en gaan in de zaak. **Ray** en **Nick**, een flat white of een dirty chai latte met havermelk, graag. Dank voor de kletspraat en schunnige latte art.

Zonder **Koffi** was dit proefschrift niet tot stand gekomen: **Serge Koffi**, jouw afrolessen waren hét moment om mijn hoofdzaken bijzaak te maken en stoom af te blazen. Dank **Annemieke**, **Esther**, **Giorgio**, **Ragna, Stéphanie** en alle andere **dansers** die me al dan niet via de spiegel opzweepten om een tandje bij te zetten. **Marleen**, dank dat je in mij zag wat ik zelf nog niet had gezien. De barre gaf me houvast als ik wiebelde.

Allerliefste vriendinnetjes **Lon** en **Em**, wij gaan inmiddels al zo'n twintig jaar als besties door het leven, en jullie hebben veel versies van mij meegemaakt; de leuke en de minder leuke, van brugklasnerd tot PhD nerd, van tiener tot dertiger, van Haarlemmer tot Utrechter. Ik kan me simpelweg geen leven zonder jullie voorstellen. Dank dat jullie alle dalen met me doorgaan en alle mijlpalen met me vieren. I hope your boobs sag first.

Lieve **mama**, ondanks dat ik altijd met mijn ogen rol als iemand zegt hoeveel ik op jou lijk, ben ik er enorm trots op. Je bent de meest zachtaardige persoon die ik ken en staat altijd voor een ander klaar. Bij niemand kan ik zo zacht in een knuffel zinken als bij jou. Dankjewel voor het luisteren, vragen en aanhoren. Ik put enorm veel steun uit de onverwachte attenties, en het gegeven dat jij er altijd en overal bent. Lieve **pap**, padre, sinds mijn studententijd zijn we veel hechter geworden en ik ben daar zo ontzettend dankbaar voor. Je bent een voorbeeld als het gaat om assertiviteit, opgewektheid en sociale openheid. Dankjewel dat je me het gevoel geeft een trotse vader te zijn; ik ben een trotse dochter. Ik kijk uit naar nog meer wandel-, fiets- en hardlooptochten, biertjes drinken en voor de honderdste keer LOTR kijken. **Pap** en **mam**, ik hou van jullie.

Lieve **(bonus-/schoon)familie**, bedankt voor de support langs de zijlijn, voor het inchecken, bevragen en toejuichen, voor het meedenken en begrijpen, voor jullie gastvrijheid en verwelkoming, voor jullie zorgen en ontzorgen, voor het samen delen en samen doormaken. Ik ben dankbaar voor jullie als familie en voor de banden die daar ondanks het niet-biologische toch van in de buurt komen.

Lieve **Merijn,** Njirem, niemand heeft de afgelopen jaren een grotere rol in mijn leven gespeeld dan jij. Nog meer dan een academische leerschool was mijn PhD voor ons een weg van persoonlijke uitdagingen. Ik ben zo zo zo dankbaar dat jij voor, naast en achter me stond om me te gidsen, vergezellen of een schop onder mijn kont te geven. Je liet me los maar hield me vast, bent blauw maar mengt met groen, bent de Sam to my Frodo, bent BOSS maar ook kachel, hebt een helder doel voor ogen (NOAD) en bent een ware lyrical genius (maar wat dan wel?). Hierbij de grappen-credits die je toekomen. Dankjewel voor jou.

About the author

Profiel

Sanne werd geboren in december 1993 in Haarlem. Een maatschappelijke profiel met biologie als add-on aan het Sancta Maria Lyceum creëerde de basis voor een studie Psychologie aan de Universiteit Utrecht. Na een tussenjaar om haar Frans bij te spijkeren in Parijs startte Sanne in 2013 aan de opleiding, waarin zij zich toespitste op de neuropsychologie en cognitieve neurobiologische psychologie. Haar bachelor werd tijdelijk onderbroken voor een jaar voorzitterschap bij de Utrechtse Faculteitsvereniging Sociale Wetenschappen Alcmaeon, waarna zij in 2017 haar bachelor behaalde en een dubbele master startte. In 2019 voltooide zij de master Neuropsychologie en in 2020 de onderzoeksmaster Neuroscience & Cognition, beiden cum laude aan de Universiteit Utrecht.

Haar ervaring met fundamenteel cognitief onderzoek en haar klinische onderlegdheid leidden ertoe dat zij werd aangenomen op het promotietraject dat zij in 2021 startte bij Prof. Dr. Stefan van der Stigchel, Dr. Tanja Nijboer en Dr. Teuni ten Brink. Gedurende haar promotietraject maakte ze deel uit van de facultaire PhD-raad en nam ze deel aan verschillende public engagement activiteiten. Ze was tevens bestuurslid bij Stichting DARE2BYOU, ze ondersteunde als neuropsycholoog in het Universitair Medisch Centrum Utrecht, en ze coachte scholieren bij hun studiekeuze. Ook ontdekte ze haar liefde voor dans vanuit een nieuwe hoek en startte met het geven van dans- en bewegingslessen. Vanaf april 2025 is zij als projectmanager/postdoctoraal onderzoeker aan de slag bij de Universiteit Leiden, waar zij implementatieonderzoek uitvoert in samenwerking met Prof. Dr. Ineke van der Ham en de afdeling Revalidatiegeneeskunde van het UMCU, revalidatiecentrum De Hoogstraat en revalidatiecentrum Basalt.

List of publications

Published manuscripts

Böing, S., Ten Brink, A. F., de Zwart, B., Nijboer, T. C. W., & Van der Stigchel, S. (2025). Relying on the external world: individuals variably use low- and medium-loading, but rarely high-loading, strategies when engaging visual working memory. *Visual Cognition.* https://doi.org/10.1080/13506285.2025.2484523

Böing, S., Brink, A. F. T., Ruis, C., Schielen, Z. A., van den Berg, E., Biesbroek, J. M., Nijboer, T. C. W., & Van der Stigchel, S. (2025). Inspecting the external world: Memory capacity, but not memory self-efficacy, predicts offloading in working memory. *Journal of Clinical and Experimental Neuropsychology.*

Hoogerbrugge, A. J., Strauch, C., **Böing, S**., Nijboer, T. C. W., & Stigchel, S. V. der. (2024). Just-in-Time Encoding Into Visual Working Memory Is Contingent Upon Constant Availability of External Information. *Journal of Cognition,* 7(1), Article 1.

Böing, S., Van der Stigchel, S., & Van der Stoep, N. (2024). The impact of acute asymmetric hearing loss on multisensory integration. *European Journal of Neuroscience,* 59(9), 2373–2390.

Böing, S., Ten Brink, A. F., Hoogerbrugge, A. J., Oudman, E., Postma, A., Nijboer, T. C., & Van der Stigchel, S. (2023). Eye Movements as Proxy for Visual Working Memory Usage: Increased Reliance on the External World in Korsakoff Syndrome. *Journal of Clinical Medicine,* 12(11), 3630.

Böing, S., Fabius, J. H., Hakkenberg, M., Nijboer, T. C., & Van der Stigchel, S. (2023). More (corrective) consecutive saccades after a lesion to the posterior parietal cortex. *European Journal of Neuroscience*, 58(7), 3650-3670.

Manuscripts in press

Böing, S., Ten Brink, A. F., Nijboer, T. C., & Van der Stigchel, S. (in press). De externe wereld als geheugensteun: capaciteitstaken vertalen naar kijkgedrag. Tijdschrift voor Neuropsychologie

Manuscripts under review

Böing, S., Ten Brink, A. F., Nijboer, T. C., & Van der Stigchel, S. (under review). Relying on the External World: Individual Variability in Compensation Strategies in Working Memory Usage After Stroke.

Manuscripts in preparation

Sahakian, A., **Böing, S**., Gayet, S., Paffen, C. L. E., & Van der Stigchel, S. (in prep.). Strategies in visual working memory use in healthy and clinical populations.

Conference proceedings

2024	**Perception Day**, Utrecht (NL). *Relying on the external world: inspection behavior reveals offloading strategies in visual working memory after brain injury.*	Talk
2024	**Annual Dutch Conference on Rehabilitation Medicine (DCRM)**, Utrecht (NL). *Relying on the external world: Uncovering individual strategies in working memory usage after stroke.*	Poster
2024	**Global Neuropsychology Conference of the International Neuropsychological Society (INS)**, Porto (PO). *Relying on the external world: Memory capacity, but not memory self-efficacy, predicts sampling behavior.*	Poster
2024	**21st Conference of the Neuropsychological Rehabilitation Special Interest Group of the World Federation for Neurorehabilitation (NR-SIG-WFNR).** Coimbra (PO). *Relying on the external world: Uncovering individual strategies in working memory usage after stroke.*	Poster
2024	**4th International Conference on Working Memory,** Leeds (UK). *Relying on the external world: Low- and high-loading individuals in dynamic visual working memory usage.*	Poster
2024	**Voorjaarscongres Nederlandse Vereniging voor Neuropsychologie (NVN),** Nijmegen (NL). *Relying on the external world: Memory capacity, but not memory self-efficacy, predicts inspection behavior.*	Poster
2023	**19th NVP Winter Conference on Brain & Cognition,** Egmond aan Zee (NL). *Relying on the external world: Uncovering individual strategies in dynamic visual working memory usage*	Poster
2023	**8th Scientific Meeting of the Federation of the European Societies of Neuropsychology (FESN),** Thessaloniki (GR). *Eye movements as proxy for memory deficits in daily situations: innovative neuropsychological assessment of visual working memory.*	Poster
2023	**Annual Dutch Congress on Rehabilitation Medicine,** Den Bosch (NL). *Mini-symposium - innovation in cognitive rehabilitation: strengths and limitations.*	Talk
2023	**Perception Day,** Helmholtz Institute, Utrecht (NL). *Eye movements as proxy for visual working memory usage: increased reliance on the external world in Korsakoff's syndrome.*	Poster
2022	**Najaarscongres Nederlandse Vereniging voor Neuropsychologie (NVN),** Maastricht (NL). *Eye-tracking: innovative neuropsychological assessment of visual working memory.*	Poster

2022	**19th Conference of the Neuropsychological Rehabilitation Special Interest Group of the World Federation for Neurorehabilitation (NR-SIG-WFNR),** Maastricht (NL). *Eye-tracking – innovative neuropsychological assessment of visual working memory.*	Talk
2022	**21st European Conference on Eye Movements (ECEM)**, Leicester (UK). *Eye-tracking in innovative neuropsychological assessment of visual working memory.*	Poster
2022	**Vision Sciences Society (VSS),** St. Pete's (FL, US). *The impact of simulated asymmetrical hearing loss on multisensory integration of auditory and visual spatial information.*	Poster
2022	**18th NVP Conference on Brain & Cognition,** Egmond aan Zee (NL). *More secondary saccades after a lesion to the posterior parietal cortex.*	Poster
2021	**Najaarscongres Nederlandse Vereniging voor Neuropsychologie (NVN**; cancelled). *Novel neuropsychological assessment of visual working memory.*	Poster
2021	**Annual Meeting of the Organisation for Psychological Research into Stroke (OPSYRIS),** Norwich (UK; hybrid). *Novel neuropsychological assessment of visual working memory.*	Poster pitch
2021	**Vision Sciences Society (VSS),** St. Pete's, FL (US; virtual). *More secondary saccades after a lesion to the posterior parietal cortex.*	Poster

A