Visual Working Memory and Ageing:

Do we approach cognitive tasks differently as we age?

Alicia Forsberg

PhD in Human Cognitive Neuroscience

The University of Edinburgh

2019

Declaration

I hereby declare that the research presented in this thesis was conducted by myself, and that I have composed this thesis myself, except where due acknowledgement is made in the text. I am aware of and understand the university's policy on plagiarism and I certify that this thesis is my own work, expect where indicated by referencing. The experimental work is almost entirely my own work; the collaborative contributions have been indicated clearly and acknowledged. Specifically, the research presented in Chapter 5 was produced in collaboration with Professor Matti Laine and Doctor Daniel Fellman. I confirm that the work submitted is my own, except where work which has formed part of jointly-authored journal articles has been included. My contribution and those of the other authors to this work have been explicitly indicated below. My PhD Advisors have provided input and comments on experimental design and the writing of this thesis. This work has not been submitted for any other degree or professional qualification.

Alicia Forsberg

3rd of July, 2019

Table of Contents

Abstract

Working Memory (WM) refers to cognitive functions that support the ready availability of a small amount of information temporarily, while we undertake ongoing actions and mental activities (e.g., Logie & Cowan, 2015), and is viewed as a core mechanism underpinning higher-order cognitive abilities. Moreover, the functioning of WM abilities is important for autonomy and wellbeing in older adults (Tomaszewski Farias et al., 2009). As assessed, WM suffers pronounced, linear decline during adult ageing (e.g., Borella, Carretti, & De Beni, 2008). However, merely establishing that younger adults outperform older adults on a given cognitive task (known as the 'Dull Hypothesis'; Perfect & Maylor, 2000) is of limited value given that it is uninformative regarding how and why WM declines with age. This thesis was inspired by research that has suggested that some aspects of WM decline faster than others. Indeed, verbal WM appears least susceptible, and visuospatial WM most susceptible to age-related decline (Johnson et al., 2010).

In six experiments, I moved beyond the 'Dull Hypothesis' and tested whether older adults approached WM tasks differently from younger adults, perhaps relying on relatively intact verbal abilities while performing visual memory tasks. Crucially, visual material – in everyday life as well as in memory experiments – may be remembered via verbal codes or visual traces, or both. In some WM theories, visual and verbal material is seen as maintained by separate

mechanisms. The Multiple-Component model of WM (Baddeley, 1986; 2012; Baddeley & Logie, 1999; Logie, 2011) is based on the postulate that visuospatial and verbal information is stored separately in dedicated storage buffers, which may also rely on separate rehearsal mechanisms (Baddeley, 2012; Logie, 2011). If one mechanism declines more with age, perhaps older adults strategically recruit a different mechanism. This led to our central research question: Do we approach cognitive tasks differently as we age? I investigated this in several WM paradigms, as outlined below.

The first series of experiments addressed the debate about whether older adults have a specific deficit in the ability to bind and remember conjunctions of features, by investigating the consequences of allowing verbal rehearsal in visual feature-binding tasks. In experiments 1 to 4, I studied the role of verbal labels in two different feature-binding paradigms to test whether discrepancies in the literature can be explained by verbal rehearsal of visual features, which might vary by age group. I found that overall, visual memory for difficult-to-label, non-categorical, visual information appeared especially limited for older adults, likely because it impedes engagement of other systems, such as verbal WM or long-term memory. Results regarding the potential implications for discrepancies in the feature-binding literature were mixed.

Next, I looked at the effect of instructing or preventing verbal labelling in a continuous colour memory paradigm, using a mixture model which allowed comparison of continuous (‘visual’) and categorical (corresponding to verbal labels, e.g., ’red’) memory representations. Labelling improved memory performance in both age groups, but older adults appeared to spontaneously rehearse verbal labels sub-vocally more than younger adults when simply instructed to perform the task in silence.

Finally, I investigated the role of strategic approaches in an N-back WM training paradigm. In this study, I instructed participants to use a visualisation strategy previously found to improve N-back performance in younger adults. I found that both younger and older adults benefitted from the instructed strategy and performed better than uninstructed controls, but some evidence suggested that the strategy was more difficult to implement for older adults. I also found significant associations between N-back performance and the type, and level of detail, of self-generated strategies in the uninstructed participants.

Combined, the results suggested that measures of performance and capacity partly reflect the extents to which participants apply appropriate strategies. Strategic mediation should be considered in research aiming to understand memory for visual features, continuous colour memory, and the mechanisms of WM training. Our results highlighted that strategic differences between younger and older adults may be crucial to interpret the age-related decline of memory, as measured in these paradigms, thus illustrating the importance of controlling differences in age-related strategic preferences in visual memory tasks. These differences may be informative for our understanding of age-related cognitive decline, suggesting that older adults may compensate for decline of some functions by approaching tasks differently.

Lay Summary

Working Memory can be seen as a 'mental workspace' which enables us to temporarily keep some information in mind and use it to perform some cognitive task. For instance, you might use it to calculate if you can afford another round of drinks at the pub – remembering your friends' orders, the original price, a potential 20% Happy Hour discount, and how much money you brought in the first place. Cognitive Psychologists attempt to measure Working Memory ability using various tasks, such as reading a string of digits to participants, which they then have to report back to the experimenter backwards. As measured, Working Memory appears associated with a wide range of other cognitive abilities, such as intelligence scores and academic performance in school children. However, Working Memory capacity seems to decrease as we grow older. Some researchers think that Working Memory consists of a set of separate tools. For instance, it may include a specific store for verbal things (like letter strings), and another store for visual things (such as remembering abstract shapes and their location).

While memory typically declines with age, some abilities seem to decline less than others. For instance, some research has suggested that while memory for visual things may start to decline in our early 20s, older adults perform very well on verbal memory tasks. In this thesis, I explored if older adults might use slightly different approaches than younger adults when faced with Working

Memory tasks in the laboratory. For instance, I hypothesised that they may attempt to compensate for declining visual abilities by using verbal Working Memory instead. I explored this in three tasks, used to study memory for visual feature-bindings (how we remember that a circle was green and a triangle black, rather than just two colours and two shapes), fine-grained, exact memory for colours, and a task used in 'Cognitive Training'. Cognitive training is sometimes promoted by companies as a way of training your brain to get smarter, and prevent cognitive decline – even though it is unclear whether this actually works.

I tested *how* younger and older adults approached these tasks by preventing them from using verbal strategies. I asked participants to say numbers out loud while they were doing the memory task, to make it difficult to use verbal memory – try thinking "red, blue" while also saying "two, nine" out loud to experience this. I compared memory performance while saying digits to that in silence. I found some evidence suggesting that older adults – in some cases – seemed to benefit from being able to use verbal strategies, sometimes reaching similar performance levels to younger adults. I also discuss whether some of this advantage was driven by other mechanisms, such as elaboration using Long-Term memory knowledge. It seemed like younger and older sometimes used different cognitive strategies to approach the same task. The research in this thesis adds to our understanding of *how* the cognitive system declines with age, and the circumstances which enable healthy older adults to boost their visual memory performance.

Chapter 1: Introduction

1.1 Overview

Working memory (WM) refers to cognitive functions that support the availability of temporary information while undertaking ongoing actions and mental activities (e.g., Logie & Cowan, 2015). Like other aspects of cognition, WM has been shown to be poorer in groups of healthy older adults compared to groups of younger adults (e.g., Babcock & Salthouse, 1990; Bowles & Salthouse, 2003; Craik, Luo, & Sakuta, 2010; Gazzaley, Cooney, Rissman, & D'Esposito, 2005; Johnson et al., 2010). There is also a lot of evidence suggesting that age influences different cognitive abilities and systems differently (for reviews see Logie & Morris, 2014; Perfect & Maylor, 2000; see also Johnson, Logie, & Brockmole, 2010). Specifically, memory for different materials (i.e., verbal or visual) may decline at different rates with age. For instance, several studies (e.g. Johnson et al., 2010; Park et al., 2002) and meta-analyses (e.g. Jenkins, Myerson, Joerding, & Hale, 2000) have found that visuospatial cognition declines more with aging than does verbal cognition.

In this thesis, I attempted to move beyond testing whether younger adults' overall memory performance was better than that of older adults (i.e., 'testing the Dull Hypothesis' of cognitive ageing; Perfect & Maylor, 2000). In contrast, rejecting the 'Dull Hypothesis' involves exploring instances where older adults can perform comparatively well – thought to be a useful approach towards increased understanding of the nature of age-related cognitive decline (Logie, Horne, & Pettit, 2015; Perfect & Maylor, 2000). Therefore in this thesis I focused on exploring task conditions in which older adults may be comparatively less impaired, and aimed to systematically investigate how older adults may approach WM tasks differently. Relying on relatively intact verbal memory abilities in visual WM tasks is a potential such different approach.

However, WM in itself is far from a 'clear' concept. Different researchers have different views of what this mental workspace is (see Cowan, 2017). Views of WM as divided into different components, or sets of resources, fit with the evidence suggesting that these may decline at different rates with age. In contrast, if adhering to conceptualisations of WM as a unitary construct, one might be more inclined to presume a general cognitive decline. Therefore, I will begin by reviewing some theories of WM. Next, I will review some different theoretical perspectives related to the broader cognitive ageing literature. Then, I will present a set of four experiments investigating the potential role of different task-approaches in visual feature-binding memory tasks in ageing, followed by a similar exploration in a colour precision memory paradigm, and finally, I explore differential strategic approaches in a WM training paradigm (using the N-back task). To foreshadow, in some cases, I observed evidence suggesting that older adults were not simply like more poorly performing adults. Such differential approaches might cause confounds when comparing performance between the age groups, and be important to understand the cognitive decline associated with increasing age.

1.2 Working Memory: Definitions and Concepts

When the term ‘Working Memory’ (WM) was first introduced it referred to a temporary store for action-relevant information (Miller, Galanter, & Pribram, 1960). Since then, many theoretical models describing the properties of this temporary store have been put forward. In cognitive psychology today WM is thought to play a central role in all deliberate cognition, including language comprehension as well as reasoning abilities. Despite the vast amount of research conducted on this temporary store and its links to numerous other cognitive functions, researchers have not reached agreement on how to define this construct (see Cowan, 2017; Oberauer et al., 2018). However, most theories converge on the assumption that WM refers to a system, or a set of processes, which keep temporary mental representations available for use in thought and action (Baddeley; 2007; Barrouillet, Bernardin, & Camos, 2004; Brady, Konkle, & Alvarez, 2011; Cowan, 2017; Logie; 2011; Ma, Husain, & Bays, 2014; Morey, 2018; Oberauer et al., 2018). Note that this definition also includes what some researchers refer to as "short-term memory".

One of the earliest, most influential models depicted WM as a multiple-component system (Baddeley, 1986; 2012; Baddeley & Logie, 1999; Logie, 2011), based on the postulate that visuospatial and verbal information are stored separately in dedicated storage buffers, which may also rely on separate rehearsal mechanisms. However, since then, numerous other conceptualisations of WM have emerged (for an early review see Miyake & Shah, 1999). A key feature of debate is whether WM is best conceptualised as a distinct system (e.g., Baddeley, 2000; 2012), or as a process of controlled attention, used to maintain goal-relevant information in an active, accessible state despite conditions of interference and competition (e.g., Barrouillet et al., 2004; Barrouillet & Camos, 2015; Cowan, 2005; Engle, Kane, & Tuholski, 1999). Some suggest that differences between these two perspectives might be more apparent than real, and might derive from differences in terminology and emphasis (e.g., see Cowan & Chen, 2008, Logie, 2011, Logie & Cowan, 2015). Another point of debate is whether WM may be better conceptualised as a unitary construct (e.g., Cowan, 2005; Oberauer, 2013), rather than as consisting of different components (Baddeley, 2012; Logie, 2011). Furthermore, other theoretical approaches to WM are strongly influenced by research on individual differences in complex task performance (e.g., Engle et al., 1999; Engle & Kane, 2004; Miyake et al., 2000).

Traditionally, WM or short-term memory for verbal material was widely researched while few studies explored visual WM (for early reviews see Logie, 1995; 2003). However the last two decades have seen a substantial increase in research on the latter. For instance, Brady, Konkle, and Alvarez (2011) highlighted the value of moving beyond visual WM models that represent each memory item as distinct, to instead focus more on the role of structured representations – e.g., by considering how items may be remembered in relation to one another. Other recent re-conceptualisations view WM as a process of resource allocation (Bays, Catalao, & Husain, 2009; Ma et al., 2014; Wilken & Ma, 2004), where one resource is allocated across all visual items, and we can either remember many items with low precision, or a few items with very good precision. Some of these newer perspectives are driven by technology – such as ways to record continuous memory using a digital colour-wheel – not available previously. According to this perspective, WM is limited in capacity because we are unable to maintain representations with enough precision to enable correct recall.

An alternative model views the WM limit as best explained by a small set of discrete, fixed-resolution representations. Indeed, some suggest that there is a limit for four "slots", based on evidence that participants appear to retain about either four colours, or four orientations in visual WM at a given time (e.g. Cowan, 2001). However, when presented with four objects consisting of a colour and an orientation each, they appeared able to retain all eight features (Luck & Vogel, 1997, see also Zhang & Luck, 2008). This sparked yet another debate as to whether the visual representations comprise integrated objects or comprise individual features that are linked to retain their pairwise combinations on a temporary basis (e.g. Allen, Baddeley, & Hitch, 2006; Allen, Hitch, Mate, &

Baddeley, 2012; Hardman & Cowan, 2015; Logie, Brockmole, & Jaswal, 2011; Wheeler & Treisman; 2002; Woodman, Vogel, & Luck, 2012). This debate is ongoing (e.g. Liesefeld, Liesefeld, & Müller, 2019; Rhodes, Cowan, Hardman, & Logie, 2018; Schneegans, & Bays, 2019) and while I have used paradigms associated with both views (i.e., change-detection and precision measures), the debate is not addressed directly in the thesis.

While the emphasis on verbal rehearsal and the possible recruitment of separate resources by participants in this thesis was inspired by research conducted within the Multiple-Component model framework, the present work was not designed to distinguish between models of WM, nor to investigate whether verbal information and visual information are stored and processed within a single system or in multiple, domain-specific systems. However, to set the context for the work, below, I review the key features of the multiple component model that are relevant for the reported experiments. One particularly important feature of this model and for my experiments is the use of sub-vocal rehearsal to support memory for verbal information. While most of the other frameworks reviewed do not emphasise the role of sub-vocal rehearsal of materials, there is broad recognition of its importance (e.g. Cowan, Saults, & Blume, 2014; Camos, Lagner, & Barrouillet, 2009; although see Lewandowsky & Oberauer, 2015 for an exception), and no framework explicitly rejects the notion that sub-vocal rehearsal of verbal material is a separate mechanism that can support memory. Similarly, it is generally agreed that perceptual input results in different types of mental codes, which may interfere with one another in complex ways (e.g., Morey & Cowan, 2004). Below, I review key features of several influential WM models, and the research on which they are based.

1.2.1. The Multiple-Component model

Initially, Baddeley (1983; Baddeley & Hitch, 1974) presented the Multiple-Component model as a development of Atkinson and Shiffrin's (1968) Modal Model. This model was derived from earlier debates regarding the unitary versus separate nature of memory (Keppel & Underwood, 1962), and inspired by Broadbent's (1958) proposals that short-term memory may be conceptualised as an attentional filter and temporary store, which is separate from long-term memory. In Atkinson and Shiffrin's modal model, new information was thought to pass through visual, auditory, or haptic sensory registers, and then through a short-term memory store where information could be rehearsed and finally passed on to a permanent long-term memory store. Some data from individuals with focal brain damage suggested that an impairment of short-term memory did not prevent information from accessing long-term memory. For example, patient KF (Shallice & Warrington, 1970) exhibited intact long-term despite impaired short-term memory performance (for a recent review of 20 such cases see Shallice & Papagno, 2019). In contrast, the amnesic patient HM demonstrated poor performance on long-term memory tasks, despite good performance on short-term retention tasks (e.g., Milner, Corkin, & Teuber, 1968). This apparent double dissociation between impairment of short-term memory but intact long-term memory and vice versa across different patients did not fit with Atkinson and Shiffrin's proposal that would require an intact short-term memory in order to access long-term memory.

Baddeley and Hitch (1974) suggested a definition of WM as a control system, with limits both on its storage and processing capabilities. They separated a special verbal short-term memory system from general control and attentional processes, and suggested that this WM system had access to 'phonemically' coded information – possibly by controlling a rehearsal buffer. This proposal was supported by investigation of the influence of remembering a sequence of numbers while carrying out comprehension and reasoning tasks. When sequences of three digits had to be remembered, there was no impact on performance of the other tasks. However, the time to respond to reasoning and comprehension tasks was slowed by a requirement concurrently to remember 6 digits. This suggested that there was a limited capacity verbal memory system, perhaps supported by sub-vocal rehearsal that could operate independently of more complex cognition. But when the capacity of that memory system was exceeded, there was an overlap with the more demanding tasks of reasoning and comprehension. Such findings supported their case for the theoretical concept that was later developed into the Multiple-Component model (Baddeley et al., 1984, 1975; Baddeley, 1986; Baddeley & Logie, 1999). While the original 1974 model did not elaborate on the proposed system for visuospatial information, Baddeley (1983; 1986) later proposed two domain-specific memory stores concerned with storing visual or verbal information, respectively, and a third a-modal central executive, as an attentional resource pool. The conceptualisation of the Multiple-Component model proposed by Baddeley (1986) included a *central executive*, thought to be responsible for attentional control both within and beyond working memory. This was supported by two specialised limited-capacity stores, the *phonological loop,* and the *visuospatial sketchpad*; all three systems are thought to be limited in capacity. Baddeley (2000) introduced a fourth component, the *episodic buffer*, thought to store cross-modality representations on a temporary basis. The *phonological loop* was thought to be a temporary store for acoustic and verbal information, and the contents of that stored can be maintained using an articulatory rehearsal system. The *visuospatial sketchpad* was thought to maintain visual and spatial information, and was assumed to support mental imagery. Logie (1995) proposed that a temporary visual store (‘the visual cache’) can be distinguished from a temporary store for movement sequences (‘the inner scribe’). Crucially, the inner scribe was proposed to serve as a rehearsal function for the visual cache. The separation of verbal versus visual stores was partly based on neuropsychological cases of verbal and visuospatial short-term memory deficits (e.g. Hanley & Young, 2019; Hanley, Young, & Pearson, 1991; for reviews see Logie, 1995; Logie & Della Sala, 2005). Similarly, some known neuropsychological disorders affect visual cognition substantially more than non-visual cognition (e.g., Lissauer & Jackson, 1988; Parra, Della Sala, Logie, & Abrahams, 2009; Zeman, Dewar, & Della Sala, 2015). The notion that visual and verbal information are stored by different systems is also supported by findings that participants may perform concurrent visual and verbal tasks with small decrement in performance, while performing two tasks from the same domain results in a greater decline in task performance (Cocchini, Logie, Della Sala, MacPherson, & Baddeley, 2002; Fougnie et al., 2015, Logie, Zucco, & Baddeley, 1990; Logie et al., 1990; Thalmann & Oberauer, 2017).

The proposition that processes supporting temporary memory (i.e., the two domain-specific memory systems) are different from executive control processes distinguishes the Multiple Component Model from other models. Also, one version of the upgraded Multiple-Component Model contains an *episodic buffer*, which is a limited capacity store that forms an interface between WM and long-term memory, and is capable of combining representations from different domains (Baddeley, 2000). See Figure 1.1 for a widely cited version of Baddeley’s model, and note differences with Logie’s model in Figure 1.2. These models are both widely used in research. I now describe these different components in more detail.

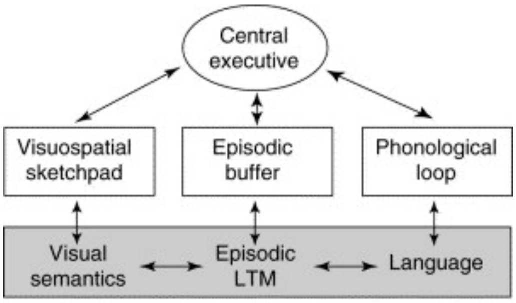


Figure 1.1. Baddeley's model. Adapted from "The episodic buffer: a new component of working memory?", by Baddeley, A., 2000, *Trends in cognitive sciences*, *4*(11), p. 421.

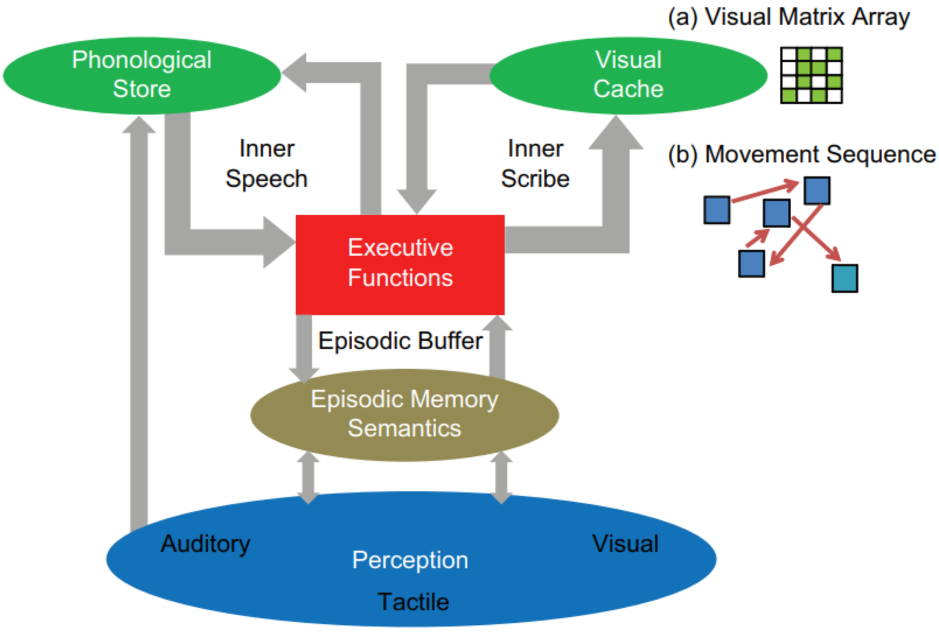


Figure 1.2. Logie's model. Adapted from "The functional organization and capacity limits of working memory", by Logie, R. H., 2011, *Current directions in Psychological science*, *20*(4), p. 241.

1.2.1.1 The phonological loop

The phonological loop is thought to comprise a passive phonological store and an active sub-vocal articulatory rehearsal loop (Vallar & Baddeley, 1984, Salamé & Baddeley, 1982). The temporary phonological store is presumed to hold auditory memory for a few seconds, before it decays – if it is not revived by articulatory rehearsal: i.e. active sub-vocal speech: repeating the information to oneself (e.g., Baddeley, 1986, 2003). This construct appears to play a role in language acquisition (Gathercole & Baddeley, 1989) and has been proposed to have evolved to facilitate language acquisition (Baddeley, Gathercole, & Papagno, 1998).

Early studies on the verbal short-term memory indicated that sequences of phonologically similar materials (such as B, D, T, V) are harder to remember in serial order than are sequences of dissimilar material (such as F, L, Q, R). This led to the notion of a verbal memory store linked to the speech system (e.g., Baddeley, 1966; Conrad, 1964; Conrad & Hull, 1964; Salamé & Baddeley, 1986). The phonological similarity effect has also been found for words, while semantic similarity (e.g., Big, Huge, Large; Baddeley, 1966) does not impair memory substantially for immediate recall but does disrupt delayed recall. This implies that letters and words are retained on a temporary basis using an acoustic (phonological) code. Similarly, participants have been shown to remember sequences of short, monosyllabic words (e.g., Book, Chair, Spoon, Bus) significantly better than the same number of longer, polysyllabic words (e.g., University, Refrigerator, Hippopotamus, Mississippi; Baddeley et al., 1975). This effect of word length was replicated using shorter words and longer words for country names, ensuring that all words belonged to the same semantic category, and using both visual and auditory presentation. The effect of word length on memory indicated that the time allowed to rehearse was a crucial feature, and led to the conclusion that the articulatory loop’s capacity was limited by temporal duration. It was estimated that participants could accurately remember the number of words they were able to articulate in around two seconds (Baddeley, 1986). Crucially, the word length effect in Baddeley et al. (1975) was eliminated when articulatory rehearsal was blocked using articulatory suppression (i.e., instructing participants to repeat task-irrelevant words such as random numbers while they perform the task). There was no benefit for short compared to long words when verbal rehearsal was prevented by articulatory suppression. Remarkably, articulatory suppression also abolished the detrimental effect of phonologically similar verbal items on memory, when presented visually (e.g., Estes, 1973; Peterson & Johnson, 1971), but not when items were presented auditorily (Murray, 1968). This indicated that auditory information automatically gains access to the phonological store, whereas visually presented verbal information needs to be actively sub-vocalised to enter the loop (Baddeley et al., 1984).

1.2.1.2 The visuospatial sketchpad

The visuospatial sketchpad component (Baddeley, 1983) is proposed to serve the function of integrating spatial, visual and perhaps even kinaesthetic information into a unified representation, for temporary storage and manipulation (Baddeley, 2003). In the original Baddeley and Hitch (1974) paper, a temporary visual store was only mentioned in passing. However, such a store was the focus of two subsequent book chapters (Baddeley, Grant, Wight, & Thomson, 1975; Baddeley & Lieberman, 1980). Indeed, much earlier research focused on the verbal component, and the visuospatial component remained comparatively unexplored (with some exceptions; see Hitch, Halliday, Schaafstal, & Schraagen, 1988; Logie, 1986). Originally, this visuospatial component's function was proposed to be similar to the phonological loop, but for visual memory (e.g., Baddeley et al., 1975; Baddeley & Lieberman, 1980; Logie, 1986; 1995; Quinn & Ralston, 1986). One of the very first studies operationalising this component compared a condition which required participants to remember instructions which could be stored using visual imagery, with one where verbal encoding could be used. Participants performed the memory task with or without a concurrent spatial tracking task, which required use of a tool to track a moving light-dot. This spatial tracking task impaired performance in the imagery condition, but not in the verbal-only task, indicating that spatial abilities were not used in the verbal task (Baddeley et al., 1973; Baddeley & Lieberman, 1980).

Further research suggested that visual information was stored separately from verbal information, as for instance, visual matrix task performance was more impaired when paired with a concurrent task that also contained visuospatial material but not in a non-visual spatial mental arithmetic task, while this pattern was reversed for a verbal memory task (Logie et al., 1990).

Next, it was suggested that the visuospatial sketchpad contained a ‘visual cache’, which retains temporary static visual images (such as of a recently presented array of coloured shapes), and an ‘inner scribe’, which enables encoding and maintenance of spatial and movement information (Logie, 1995). Essentially Logie proposed a division of the single visuospatial sketchpad into a system with separate storage and rehearsal components. The ‘visual cache’ was thought to be limited by the visual complexity of the representation, and to maintain the representations for a few seconds after seeing it. However, the ‘inner scribe’ was proposed to be able to rehearse the contents of the visual cache, and thus maintain the representation for more than a few seconds (Logie, 1995).

Additional research investigated the potential fractionation of this visual memory system into visual and spatial memory components. Resulting evidence suggested that visual and spatial memory had different developmental trajectories (e.g., Logie & Pearson, 1997), such that the difference in performance between visual and spatial memory tasks grew larger between the ages of five to

12 years. Furthermore, evidence comparing a visual task (i.e., the visual patterns task, which tests memory of which squares on a grid were coloured as opposed to blank) and a spatial task (i.e., the Corsi blocks tasks: requiring memory for which blocks the experimenter tapped, in the correct order) found that a visual interference task had greater detrimental impact on a visual memory task, and a spatial interference task had a greater negative effect on a spatial memory task (Della Sala, Gray, Allamano, Baddeley, & Wilson, 1999). Indeed, various other studies support the claim for a dissociation between visual and spatial working memory, as visual memory tasks are more strongly disrupted by visual interference than spatial, and vice versa (Klauer & Zhao, 2004). For instance, playing tones from different locations appears to disrupt spatial memory, but not impair the vividness of mental imagery (Smyth & Scholey, 1994).

However, the workings of the proposed rehearsal mechanism for visuospatial representations have proven more controversial than sub-vocal rehearsal of verbal representations. Some research suggested that rehearsal of visuospatial representations is linked to covert attention, without eye-movements (Awh & Jonides, 2001; Godijn & Theeuwes, 2012). Others suggest that the eye-movements are an important part of this process (Baddeley, 1986; Belopolsky, & Theeuwes, 2009; Postle, 2006), such that spatial locations are rehearsed by planned eye-movements to those locations. For instance, eye-movements performed between encoding and recall appeared to disrupt memory for locations (Pearson & Sahraie, 2003). Research preventing participants from performing eye-movement during all stages of visual and spatial WM tasks found that eye-movements performed during encoding or maintenance impaired spatial memory. In contrast, preventing eye-movements only during the retrieval phase did not have a negative impact on spatial memory (Pearson, Ball, & Smith, 2014), suggesting that eye movements do play a role in the maintenance of spatial information in WM.

Some claim that a specialised visual system is not necessary. A meta-analysis of dual-task costs in visual and spatial short-term memory tasks found robust dual-task costs regardless of the domain tapped by the secondary task (Morey, 2018). This suggested that visual memory was vulnerable to interference from a variety of sources. Morey interpreted this as supporting the notion of one, unified WM system. However, previous dual task studies in this area have acknowledge cross-domain dual task costs, but emphasise that the within-domain dual task costs are very much larger than the cross-domain costs (e.g. Cocchini et al., 2002; Logie et al., 1990; Salway & Logie, 1995).

1.2.1.3 The central executive

In the original Multi-Component model, the third and final component was thought to coordinate the two other components – making the whole system work together. In the initial model, this system was thought to both store and manipulate information. For instance, the central executive was thought to make processing decisions and to move information between the two memory stores, amongst other possibilities. However, the idea that the central executive had a storage capacity was later rejected because having both storage and processing abilities effectively made this component a too-powerful homunculus (a little person inside the brain who in some mysterious, empirically un-testable way coordinates cognition), which hindered scientific testing of the model predictions (see Baddeley & Logie, 1999; Logie, 2016).

Research on this component has been developing slowly, despite its great importance in general cognition. The initial plan was to research the more simplistic sub-systems first, and subsequently approach this more complex coordinator role. However, Baddeley has since referred to this neglect as an 'embarrassment' to the model (p. 6, 1996), only starting to be addressed in the late 80s (e.g., Baddeley, 1986). The upgraded approach was inspired by the Supervisory Attention System (Norman & Shallice, 1986). This dual attention system is partly automatic, and enables behaviour via habits (or schemas), while the other part is able to switch or divide attention between two cognitive tasks. The latter system was thought to be able to override habits in situations where such responses were inadequate (e.g., remembering to not bring lunch to work if you have been invited to a lunch meeting at a restaurant). The two systems were distinguished following observations of patients with an impaired 'supervisory attentional system' following frontal lobe damage, associated with failures in carrying out several fairly simple but open-ended tasks over a 15 – 30 minute period (apparently stemming from an inability to coordinate when to move on from one task to the next; Shallice & Burgess, 1991).

Research testing the role of executive functions in the Multi-Component model used the same idea as that employed to separate the visual and verbal sub-components, namely selective interference. Typically, some primary memory task was completed at the same time as a secondary task, one thought to require the central executive, or with a control task (hypothesised not to require the central executive). Performance could then be compared to that under a secondary control task which should not require the central executive (such as articulatory suppression: i.e., repeating digits out loud). In contrast, having to count backwards in increments of seven, or generate random numbers, was assumed to require more executive resources. While both tasks – on the surface level – involve saying digits out loud, articulatory suppression only requires simple repeating of digits, while the backwards counting task requires a combination of monitoring the previous number and performing subtractions, (see Baddeley, 1966; 1986). This type of paradigm can thus be used to test whether the central executive is recruited in a given memory task. For instance, spatio-sequential tasks were found to involve central executive processes more than a static visual task (Rudkin, Pearson, & Logie, 2007), as spatio-sequential memory performance was impaired more by random number generation than articulatory suppression, while there was less of a difference between these two manipulations for static visual memory performance.

Baddeley (1996) subsequently outlined four sub-functions of the central executive: (1) Coordinating performance on two separate tasks (i.e., ability to focus attention while doing two tasks at once), (2) Dividing attention between two important targets or stimulus streams, (3) Selective Attention – the capacity to attend selectively to one stimulus and inhibit the disrupting effect of others, and (4) interface with long-term memory: the ability to use long-term memory knowledge in WM tasks. Other, non-WM specific attempts to fractionate executive functions using factor analysis on performance scores of various tasks has separated it into three separate functions: shifting between different tasks, inhibition of pre-potent responses, and the updating of working memory representations (Miyake et al., 2000). While these three abilities may partly rely on a general executive capacity, they were also distinguishable between as well as within people. Miyake et al. (2000), also identified the ability to perform two tasks concurrently, or dual-tasking as fourth executive function, but did not investigate this systematically.

Within the Multi-Component model, the central executive may be seen as a tool to explore the attentional characteristics of WM, and its links to the two assistant systems outlined above. However, Baddeley (1996) argued that the central executive as a component may perhaps be better conceptualised as a way of labelling the problem of how executive functions are implemented, not as a proposed solution. In fact, when the central executive was robbed of its storage capacity in the updated Multiple-Component model (see Baddeley & Logie, 1999), this created a new problem – it was unclear how information processed by different subsystems could be combined and remembered, if there was no general storage function. This led Baddeley (2000) to a proposed addition to the Multiple-Component model: the episodic buffer, which I will now discuss.

1.2.1.4 The episodic buffer

If the central executive was an attentional system without storage capacity (e.g., Baddeley & Logie, 1999), and the sub-systems held verbal, visual and spatial information respectively, how could such different types of information be integrated? For instance, how could such a system handle seeing a person's face and hearing them say their name, and then allow us greet them by their name later on? And also, why are words forming prose or sentences much easier to remember than random words presented at the same rate? In 2000, Baddeley added the episodic buffer, equipping the model with a mechanism to deal with these problems (see Figure 1). This new system was a temporary multidimensional store, forming an interface between the WM components reviewed above, as well as long-term memory. It was thought to allow the verbal and visuospatial systems to interact, and allow the binding of different types of information into ‘chunks’, such as remembering that a face and a name belong together. The episodic buffer was presupposed to be limited by a specific number of chunks (perhaps four; Baddeley; 2012; in line with Cowan, Chen, & Rouder, 2004; Luck & Vogel, 1997). This type of multi-feature binding mechanism was suggested to be a key function of consciousness (Baars, 2002). In the context of the original 1974 model, the buffer can be regarded as dividing the original central executive into a pure attentional system (the central executive) and a separate storage system (i.e., the episodic buffer; Baddeley, 2012). Furthermore, the buffer was thought to enable information from long-term memory to supplement the information held in the WM sub-stores (Baddeley & Wilson, 2002), and thus helping explain why we can remember more words on a word list in our own language than from an unfamiliar one. While the episodic buffer enables the model to explain a variety of phenomena, it could be criticised for having homunculus-like properties, and perhaps limiting its usefulness as an addition to the model.

The Multiple-Component model was constructed by identifying sub-systems by testing the extent to which different tasks interfere – or do not interfere – with one another, allowing overlap to be detected. This fractionation of cognition has been supported by a large body of research illustrating distinctions between visual and verbal stores (Logie et al., 1990), as well as between visual and spatial stores (Della Sala et al., 1999; Logie & Pearson, 1997), and differences in the extent of executive recruitment needed to successfully complete different cognitive tasks (Rudkin, Pearson, & Logie, 2007). This fractionation differs from the next models I will review. Generally, in the construction of other models, there has typically been less concern regarding the identification of sub-systems at play in producing a certain level of performance. Instead, more focus has been placed on the attentional resources or capacity available to individuals as they approach tasks.

1.2.2 Embedded Processes

This model was developed by Cowan (1988, 1995, 1999, 2005, 2008), and stems from the notion that the part of WM that we can consciously access must be unitary in nature, under the assumption that consciousness cannot have parts that do not know about each other (e.g., see Logie & Cowan, 2015). The model contains two main features. Firstly, when information is presented, representations of that information enter WM as subsets of momentarily activated long-term memory information. For instance, presenting digits to participants will activate long-term memory knowledge of what digits look like, and represent (Cowan, 2005, p. 41). Crucially, this activation is subject to decay over time. Secondly, a subset of this activated information is held in the ‘focus of attention’, which represents an individual's WM capacity. This is estimated to be limited to three to five representational units (Cowan, et al., 2004) which may contain more than one feature each (e.g., a visual object such as a coloured shape could be one item, Cowan, 2005). However, some evidence suggests that the WM capacity limit may be limited by interference among temporary bindings of items, rather than a specific number of items or chunks (see Oberauer, 2013). The Embedded Process model also includes a store for brief sensory after-images, which may hold iconic information for around 250 ms. Re-entering representations into the focus of attention is seen as a way to keep them fresh in memory, and thus shield them from decay. This process is sometimes referred to as *attentional refreshing* (e.g., Camos et al., 2009; Camos et al., 2018). Cowan (1999) referred to a *central attentional controller* which supervises such processes to maintain information in memory for longer. This central attentional controller is proposed to be domain-general (Cowan, 1999). Therefore, the model posits that there is a general limit to WM, regardless of representation domain (Cowan, 2001; Cowan & Morey, 2007, Cowan et al., 2014; Saults & Cowan, 2007; Vergauwe, Barrouillet, & Camos, 2010). This is a contrast to the Multiple-component model described above, where verbal and visual materials are thought to have different stores. However, Cowan has referred to ‘central’ and ‘peripheral’ components of WM (Saults & Cowan, 2007; Cowan et al., 2014). *Central* WM may be allocated to different kinds of stimuli, while *peripheral* components are used only to store material from one specific modality (e.g., verbal), which cannot be traded for more storage in the other modality (e.g., visual), thus seemingly non-attentional stores. Evidence for peripheral storage abilities indicated that the notion of the focus of attention as a central holding area was too simplistic. Cowan et al. (2014) acknowledged that adding peripheral components makes the Embedded Processes framework more similar to the Multiple-component model.

1.2.3 TBRS (Time-Based Resource-Sharing)

A different approach was provided by Barrouillet and colleagues (2004), who conceptualised WM function in terms of the allocation of a shared attentional resource, across both memory and processing tasks. This model is similar to Cowan's Embedded Processes framework in that the functional limit of WM depend on a shared attentional resource. However, the TBRS model also focuses on the *time* available during processing. A recent version of the TBRS model is based on four key tenets. First, both processing and maintenance of memory items require attention, which is a limited resource. Therefore, sharing of attention between these is needed. Second, when attention is withdrawn from memory representations, they suffer from time-related decay. Items can be refreshed by attentional re-focusing. Third, any process which captures attention disrupts the maintenance of the memory traces, because this disrupts the refreshing of items via attentional re-focusing. Finally, given this central bottleneck, which only allows one (central) process at a time, sharing attention is time-based. Attention is shared by rapid switching between processing and memory maintenance (Barrouillet & Camos, 2007). This notion of time-based resource sharing is supported by findings that memory span decreases as the proportion of time spent on any one processing task increases (also termed cognitive load; e.g., Barrouillet et al., 2004, 2007). However, while attentional resource sharing is central to this model, it is not a 'pure' resource sharing model, because in addition to attentional WM maintenance (such as attentional refreshment), the model also allows non-attentional mechanisms to contribute, such as sub-vocal rehearsal and activated long-term memory (e.g., see Camos et al., 2009; for a review see Barrouillet & Camos, 2015).

1.2.4 Primary and Secondary Memory – Individual Differences

Unsworth and Engle (2006) adopted another approach. They distinguished between 'Primary' and 'Secondary' memory, drawing on the dissociation originally proposed by James, (1890). Unsworth and Engle noted that findings that individual differences in WM capacity were associated with various other aspects of cognitive performance. For instance, WM capacity is associated with multitasking ability (Hambrick, Oswald, Darowski, Rench, & Brou, 2010) and susceptibility to hindsight bias (Calvillo, 2012). Also, measures of WM capacity appeared to share at least half their statistical variance with measures of general fluid intelligence (thought to measure the ability to reason with unfamiliar information; Kane, Hambrick, & Conway, 2005).

In this framework, limitations in WM are seen as arising from both Primary and Secondary memory components (Unsworth & Engle, 2007). Primary memory is posited to maintain a distinct number of separate representations in an active state for ongoing processing via continuous allocation of attention (Unsworth & Engle, 2007), and is thought to hold approximately four items, similar to other accounts of limited storage capacity (e.g., Cowan, 2001; Luck & Vogel, 1997). If the Primary system is overtaxed with more items than it can hold, items are displaced from Primary memory and must be recalled from Secondary memory, via cue-dependent search processes. Also, if attention is removed from Primary memory (for instance, by requesting participants to perform an attentionally demanding concurrent task), all items leave Primary Memory. If items have been displaced from Primary memory, they must be retrieved from Secondary memory for successful memory performance (Unsworth & Engle, 2007). This process is highly competitive.

The framework also has a component of attention control (see Shipstead, Lindsey, Marshall, & Engle, 2014), which accounts for the ability to ignore distracting, irrelevant information. Engle (2002) viewed WM capacity as being equivalent to the ability to deploy attention to select relevant information, as well as access memories beyond current conscious awareness (see Kane, Conway, Hambrick, & Engle, 2007). This account emphasises that high WM capacity is produced by better ability to ignore irrelevant, distracting information, and focus on what is relevant. This notion is supported by findings that higher WM capacity is correlated with better performance on various attention capture tasks, such as testing the ability to ignore attention-grabbing peripheral events (e.g., Fukuda & Vogel, 2011; Unsworth & Spillers, 2010).

Finally, proponents of this framework do not deny that domain specific representations exist – and conflict to a greater degree within than between domains – but instead focus mainly on the hypothesised process which drives the predictive utility of WM capacity on various complex cognitive tasks. Thus, this framework differs from the MCM framework in terms of which aspect of cognition they are attempting to describe.

1.2.5 Resource Models: Slots vs. Precision?

The idea of set numbers of items that our cognitive systems can process has been pervasive in Psychology for over half a century (e.g., Miller's classic ‘magic number' seven, plus or minus two, 1956). WM has been thought to hold a specific, small number of items, such or Cowan's four (2001). Similarly, Luck and Vogel ’s discrete slot model (1997) proposed that visual WM can hold objects in three or four independent object 'slots', one for each stored item. Recently, (visual) WM has been re-conceptualised as being limited in terms of the quality of representations of remembered items (Ma et al., 2014), rather than the quantity of objects. Ma et al. (2014) proposed that a limited resource can be distributed flexibly among all items that are maintained in WM, challenging the idea that an object is either allocated a ‘slot’ and is remembered, or otherwise forgotten.

Resource models of working memory (e.g., Bays & Husain, 2008; Palmer, 1990; Wilken & Ma, 2004) are based on two premises: (1) Internal representations of sensory stimuli (and/or the measurement of such representations) are noisy, and (2) The noise levels of these representations increase as the number of to-be-remembered items increases, because a limited representational medium is allocated among all items in memory. Thus, for more items, this resource is spread thin, which reduces the precision of recall of items (Ma et al., 2014). This theoretical framework was developed in conjunction with continuous delayed-estimation paradigms of WM, which allowed continuous manipulation of the signal-to-noise ratio (Wilken & Ma, 2004). Using this technique, both the to-be-remembered items and the response space are continuous, rather than discrete as assumed by a ‘slot’ model. Precision measures of visual WM capacity may require reproduction of specific shades of colour from a continuous colour-wheel (e.g., Bae, Olkkonen, Allred, Wilson, & Flombaum, 2014) or orientations of arrows (e.g., Fallon, Mattiesing, Muhammed, Manohar, & Husain, 2017). Such precision measures may produce stronger experimental paradigms by avoiding misleading cut-off points (remembered vs.

forgotten) in performance, and distinguishing participants who were just slightly incorrect from those who appear to guess randomly.

However, some evidence appears to contradict the notion of WM as a general resource spread across any number of items in a visual array. Zhang and Luck (2008) found that when presented with more than a few simple objects, participants appeared to store a high-resolution representation of a subset of those items and forget the other items. Therefore, they argued that visual WM is better conceptualised as holding a small set of discrete representations with fixed resolution, rather than distributed resources from a general pool (Zhang & Luck, 2008).

A debate has emerged regarding how to distinguish ultra-low precision representations from guessing (Nosofsky & Donkin, 2016; Sewell, Lilburn, & Smith, 2014). For instance, Adam, Vogel, and Awh (2017) found that variable precision models which do not include a measure of guessing identify memory representations that are indistinguishable from guesses. The frequency of those representations was very similar to the estimated rate of guessing in models which acknowledge item limits. Indeed, Awh (2018) argued that if one scales the relative precision of an item beyond the hypothesised slots to be equal to the thickness of a piece of paper, then the comparative precision for items in the slots would equal the height of the Burj Khalifa. Arguably, if the resource spread to items outside of the slots is that minuscule it might be equal to guessing for all practical intents and purposes. Perhaps the distinction between a very unprecise representation and a ‘guess' is a moot point.

Some evidence also suggests that even continuous visual tasks may be approached verbally (Hardman et al., 2017; Souza & Skora, 2017). Verbal WM follows different rules. For instance, words can be rehearsed using a relatively non-attentional process of sub-vocal rehearsal (repeating the words to oneself). Also, a verbal colour label such as ‘red' could be applied to a set of different reddish shades, and would therefore not be helpful to identify a specific shade of that colour. Verbal memory representations may follow a ‘sudden death' pattern of decay (either you remember the label or not) while visual representations appear to decay gradually (see Donkin, Nosofsky, Gold, & Shiffrin, 2015). Application of verbal strategies in visual precision tasks may thus also contribute to discrepant findings, and one possible resolution would be to suggest that verbal representations are retained in slots, whereas visual representations are limited by precision and quality of those representations. I discuss precision measures, and the role of verbal labels in such paradigms, in more detail in Chapter 3.

1.2.6. Ensemble Representations

Brady and Alvarez (2011) questioned the tendency of many other models of visual WM to assume that each item is stored as an independent unit, and not account for how item representations may interact with one another (Alvarez &

Cavanagh, 2004; Bays, Catalao, & Husain, 2009; Luck & Vogel, 1997; Rouder et al., 2008; Wilken & Ma, 2004; Zhang & Luck, 2008; although see Lin & Luck, 2008, and Johnson, Spencer, Luck, & Schöner, 2009). In real-world scenes, we typically process and remember numerous items, as well as their relation to one another. Thus, Brady and Alvarez (2011) studied how the structure of the items together may provide higher-order constraints on memory for individual items. Even in a quick display of simple stimuli (e.g., a set of circles) participants may encode and use a global statistic of all items, such as the mean circle size (Ariely, 2001; Chong & Treisman, 2003), rather than remembering each circle entirely independently. Indeed, participants' responses appeared biased towards the mean circle size of the sets of items (Brady & Alvarez, 2011). Similar higher-order statistics could influence memory in a variety of tasks. For instance, a display of coloured dots might be easier to remember if all items on the top are 'warm' colours, and the bottom items are 'cold', compared to a set of memory items without such patterns (see Brady & Alvarez, 2015), even though displays are typically randomised to minimise patterns. This research suggests that to really understand the (visual) WM system, and how it operates outside of the laboratory, maybe such global factors are essential. I discuss the potential influence of such global factors in relation to our results in Chapters 2 and 3.

1.2.7 WM Models: Summary

This review does not include all models of WM. I aimed to provide a general overview of the theoretical background and current debates that influenced the research in this thesis. I will discuss each model as relevant to the development of research questions and the interpretation of results throughout the thesis.

As illustrated, WM is conceptualised differently by different researchers. The notion of a separate verbal store and rehearsal mechanism drove my hypotheses in Experiments 1 to 5. Tension between domain-general and domain-specific models is illustrated by contrasting findings of substantial modality- and domain-specific phenomena; doing two tasks either both involving verbal or visual-spatial representations results in poorer performance than performing two tasks that rely on different types of representations (e.g., Fougnie et al., 2015; Thalmann & Oberauer, 2017), contrasting with evidence suggesting that there is a general limit to WM, regardless of representation domain (e.g., Cowan & Morey, 2007; Vergauwe et al., 2010). However, proponents of each have suggested that the differences may lie more in the type of theoretical approach, rather than in fundamental model disagreement (Cowan, 2005; Logie, 2011). The various models of WM are not necessarily mutually exclusive, and may differ primarily in emphasis placed on specific aspects of the constituent cognitive processes (i.e. goal directed cognitive control versus task specific representations for general capacity and multicomponent models, respectively). Recent collaboration between researchers associated with the MCM, TBRS and EP frameworks has highlighted similarities between the models, in that some results may be compatible with several models, often regarded as opposed (see Doherty et al., 2018). Similarly, recent reference to peripheral components of WM (Cowan et al., 2014) may make the Multiple-component and Embedded Processes models more compatible, by allowing a specialised verbal store in the latter. Indeed, while domain-specific stores are less emphasised in unitary conceptions of WM, they are not explicitly rejected (Cowan, 2005; Cowan et al., 2014; Oberauer, 2013), and researchers associated with models which do not emphasise verbal/visual separation generally agree that sub-vocal rehearsal of verbal material is a separate mechanism that can support memory (e.g., see Camos, Mora, & Oberauer, 2011). I discuss debates regarding how resources are allocated to visual scenes (e.g., slots vs. precision) in relation to Exp. 5, where I used a precision paradigm to explore verbal approaches to visual WM. Finally, the domain-general or specific debate influenced Exp. 6; in relation to the notion that WM can be trained. In the next part of this literature review, I discuss theories of the decline of WM with aging.

1.3 Working Memory and Cognitive Ageing

The past century has seen rapid growth in the number of people who live to an older age. Early empirical evidence of the decline of cognitive abilities with age was presented in the 1930s, in an evaluation of perceptual, motor, and cognitive abilities in 1600 people across the age range (6 to 95 years) which found that all these abilities started to decline after age 30 (Miles, 1933). Since then, a large body of research has suggested that cognitive ability generally seems to decline with age (Cattell, 1943; Deary et al. 2007; Salthouse, 1996). Similarly, poorer WM performance in older than in groups of younger healthy adults is also a well-established finding (e.g., Babcock & Salthouse, 1990; Bowles & Salthouse, 2003; Bromley, 1958; Craik, Luo, & Sakuta, 2010; Gazzaley, Cooney, Rissman, & D'Esposito, 2005; Johnson et al., 2010; Park, Lautenschlager, Hedden, Davidson, & Smith, 2002). It is generally agreed that this decline with age is important to understand, since WM is believed to underpin effective operation of other cognitive functions, such as perception and problem-solving (e.g., Ma, Husain, & Bays, 2014), and to be related to general intelligence (e.g., Unsworth, Fukuda, Awh, & Vogel, 2014) and reasoning ability (e.g., Kyllonen & Christal, 1990; Conway, Kane, & Engle, 2003). Also, WM abilities are important for autonomy and wellbeing in older adults (Tomaszewski Farias et al., 2009). However, in addition to ongoing debate regarding the characteristics of WM (as illustrated in the previous section), there are also several different theories about why and how cognitive ability, and specifically memory, declines with age. Below, I review some of these theories, before discussing evidence suggesting that different cognitive abilities decline at different rates as we age.

1.3.1 The Processing Speed Theory of Adult Age Differences

Birren (1965) found that processing time increased with age across various cognitive tasks. He hypothesised that reduced processing speed was a central mechanism behind various age-related deficits, including reduced memory performance (also supported by Salthouse, 1985; 1996). Salthouse proposed that deficiencies in two mechanisms account for age-related differences in memory (Salthouse, 1996). Firstly, a *limited time mechanism,* time available for later cognitive operations, is restricted when earlier operations occupy much of the available time. Secondly, a *simultaneity mechanism;* the products of early processing may be lost before later processing is completed. For instance, if older adults perform a task more slowly, they would need to retain the crucial information for longer, which increases memory decay. Thus, age-related slowing in processing was thought to be a major factor in poorer overall memory performance. Salthouse suggested that this general slowing might impact performance on different memory tasks differently. He also proposed that older adults might approach tasks differently, using some strategy to attempt to 'compensate' for declining processing speed (Salthouse, 1985).

1.3.2 The Processing Resource Model of Memory Deficits in Cognitive Aging

Another theory stemmed from Craik and Lockhart's (1972) level of processing theory of memory. They found that participants who were guided to perform deep semantic processing performed equally well or better than participants who studied materials simply with the intention to remember (see also Hyde & Jenkins, 1969). As understanding about the importance of both the type and level of processing during encoding developed, it was proposed that older adults' memory deficits might stem from a failure to engage in deep processing during encoding spontaneously. The gap between younger and older adults appeared extra-large when instructed to intentionally learn the material, suggesting that older adults were less able to employ suitable higher-level encoding strategies (*production deficit hypothesis*; Kausler, 1970). Encouraging participants to process meaning and engage in elaborative encoding facilitated memorisation of items, particularly for older adults (e.g., Craik, 1986). Furthermore, older adults performed more similarly to younger adults in familiar situations, while the performance gap appeared wider in novel situations, which require more self-initiated processing (e.g., see Craik, 1994; Park & Gutchess, 2000, but see also Park & Shaw, 1992). Findings of how environmental cues could benefit older adults led to the hypothesis that older adults have less *processing resources* (also referred to as 'mental energy'), which may reduce both the quality and quantity of their memory operations (Craik & Byrd, 1982).

1.3.3 The Inhibitory Theory of Memory Deficits with Age

In a seminal study, Rabbitt (1965) asked younger and older participants to sort cards into different piles quickly. Each card contained either the letter 'A' or 'B', and the sorting rule was to put cards with 'A' in one pile and those with 'B' in another. However, each card also contained either one, four, eight, or zero irrelevant letters. The experimenter split the cards into different decks by the amount of irrelevant information and used a stopwatch to record the sorting of each deck. Slower sorting times indicated that older adults struggled more than younger adults to ignore irrelevant information. Rabbitt suggested that this might be due to reduced perceptual grouping efficiency.

Hasher and Zacks (1988) suggested that this specific deficit in ignoring task-irrelevant information is a crucial factor in the cognitive decline – e.g., in WM (Hasher, Lustig, & Zacks, 2007; Hasher & Zacks, 1988; but see Kramer, Humphrey, Larish, & Logan, 1994). Evidence for an inhibitory deficit in older adults has also been found in attentional flanker tasks (Zeef, Sonke, Kok, Buiten, & Kenemans, 1996), as well as in Stroop tasks (Milham et al., 2002; West & Alain, 2000), and in the negative priming literature (Kane, Hasher, Stoltzfus, Zacks, & Connelly, 1994).

Older adults' inability to inhibit irrelevant information is thought to result in such information 'occupying space' in WM, hence reducing their storage and processing abilities compared to younger adults (Hasher & Zacks, 1988). Some behavioural evidence supports this notion that irrelevant information affects older adults WM performance differently (e.g., McNab et al., 2015), but as Lustig and Jantz (2015) pointed out in their review, the evidence is far from unequivocal. However, even when (differential) behavioural impairments of irrelevant distractors are not found, neural measures often indicate impaired processing by older adults. For instance, based on a meta-analysis, Turner and Spreng (2012) suggested that the pattern associated with inhibition in older adults was like a "young plus" pattern, i.e., that older adults activated similar inhibitory control regions as younger adults (particularly the right inferior frontal gyrus), but to a greater degree. Similarly, evidence for 'unnecessary' activity in older adults in response to distractors, has been found by comparing the amount of maintenance-related activity when distractors are presented with activity without distractors (e.g., Gazzaley et al., 2005; Gazzaley et al., 2008; Jost, Bryck, Vogel, & Mayr, 2011). Neither of these studies found a clear behavioural difference in effect of distracting information between the age groups. Turner and Spreng (2012) suggested that such findings might indicate that inhibitory control is comparatively more demanding for older adults, even when their behavioural performance is not (differentially) impaired by distracters, perhaps as a result of increased cognitive effort to compensate for the age-related decline (see Baltes & Baltes, 1990). Other work suggested that inhibition may not be a unitary construct, which might also explain why age-related deficits are only sometimes observed (Kramer & Madden, 2008). This thesis did not directly test the effects of irrelevant information in ageing, but I discuss potential confusion of irrelevant items from the original memory array with the probed target item in Chapter 2.

1.3.4 The Scaffolding Theory of Age and Cognition

The advent of structural and functional neuroimaging transformed the study of cognitive ageing and inspired brain-based theories of age-related memory decline. Brain atrophy with age is now a well-established finding; while neuronal growth may continue throughout the lifespan, it is not sufficient to compensate for age-related loss (e.g., Sailor, Schinder, & Liedo, 2017). As we age, the brain undergoes a global decline including thinning of the cerebral cortex (Salat et al., 2004), global reduction in grey matter (Good et al., 2001) and sulcal depth (Rettman et al., 2006). Brain atrophy is also linked to impaired cognition. For instance, older adults with comparatively smaller brain volume (i.e., hippocampal, parahippocampal) were found to have impaired explicit memory (e.g., Raz, Gunning-Dixon, Head, Dupuis, & Acker, 1998). Functional neuroimaging allowed comparison between activity levels in brains of younger and older participants as they encoded and retrieved information. For instance, older adults were found to exhibit higher levels of neural activity than younger adults in certain task conditions (e.g., Cabeza et al., 1997; Reuter-Lorenz et al., 2000). This increased activation was thought to reflect compensatory processes (e.g., Cabeza, 2002; Reuter-Lorenz & Cappell, 2008).

The Scaffolding Theory of Age and Cognition (Park & Reuter-Lorenz, 2009) explained such findings of increased frontal activation in older adults with the proposition that as we age, more is needed to do less. Specifically, they proposed that the age-related decrease in volume of some brain areas (e.g., the Caudate, Hippocampus, and Cerebellum) was compensated for by increased bilateral prefrontal cortex activation (Park & Reuter-Lorenz, 2009). Such 'scaffolding' is seen as a neurocognitive response to a challenge, which can be either internal (i.e., an intrinsic feature of an ageing brain, compensating for lower processing speed or storage capacity) or external (i.e., a response to a challenging task). Interestingly, some evidence suggested that younger adults exhibited similar increased bilateralisation as a result of increased task difficulty (Banich, 1998). The Scaffolding Theory is based on neuroimaging studies suggesting that ageing participants approach cognitive tasks differently. In this thesis, I tested whether this can also be observed in behavioural performance patterns.

1.3.5 Preventing Cognitive Decline

Loss of cognitive abilities with age – for instance due to dementia – is considered by many as worse than death (Patrick, Starks, Cain, Uhlmann, & Pearlman, 1994). Perhaps reflecting this fear, many researchers have explored ways to prevent such decline. The evidence for the efficiency of various ‘interventions’, including physical activity, hormone therapy, vitamins, ‘antidementia’ drugs, and cognitive training, in preventing age-related cognitive decline has tended to yield mixed results (see Kane et al., 2017 for a review).

Here, I focus on cognitive training; an increasingly popular non-pharmacological intervention. Cognitive training typically requires participants to complete some computerised task several times a week for several weeks. Training thus appears more low-risk than pharmacological intervention and seems to be preferred by older adults (Rodakowski, Saghafi, Butters, & Skidmore, 2015). Early promising results suggested that such training improved cognition in healthy young participants (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008) and children with ADHD (Klingberg, Forssberg, & Westerberg, 2002). However, further research with more appropriate experimental controls suggested that while WM training typically improved performance on the trained task itself, transfer effects to other measures of cognitive ability such as reasoning or fluid intelligence appeared at most small (for comprehensive meta-analyses see; Melby-Lervåg & Hulme, 2013; Melby-Lervåg, Redick, & Hulme, 2016; Schwaighofer, Fischer, & Buhner, 2015; Weicker, Villringer, & Thöne-Otto, 2016).

Evidence regarding the effects of cognitive training in older adults has also been mixed. Broadly speaking, the evidence that cognitive training can prevent age-related cognitive decline is not very strong (see Butler et al., 2018 for a review). Younger adults have been found to improve more than older adults (Burki et al., 2014; Dahlin et al., 2008; Heinzel et al., 2014; Zinke et al., 2013). However, others have observed gains of similar magnitude on trained tasks in younger and older adults (e.g., Bürki et al., 2014; Li et al., 2008; Richmond et al., 2011; von Bastian et al., 2013; Zając-Lamparska & Trempała, 2016). Larger training gains in younger adults might appear consistent with animal models suggesting that older age is associated with less neuroplastic change (Blumenfeld-Katzir et al., 2011; van Praag et al., 2005).

A meta-analysis (including 13 studies) indicated that WM training in healthy older adults produced both large near- and far-transfer transfer effects, i.e., that training improved performance on similar as well as different cognitive tasks (Karbach & Verhaeghen, 2014). However, others replicated the meta-analysis but included only studies which compared the trained group to active controls and controlled for baseline differences, and found less beneficial effects of training than reported in the original meta-analysis (see Melby-Lervåg et al., 2016). In this thesis, I investigated potential strategy use to explore how WM training might result in better task-performance, in younger and older adults (see Chapter 5).

1.3.6 Non-Cognitive Theories of Cognitive Decline

Some less well-known non-cognitive theories of age-related decline in cognition have also been proposed. For instance, older adults' comparatively weaker memory performance could be due to factors such as lower motivation, more considerable test anxiety, more time since schooling, less overall time spent in formal education, and poorer general health (see Park & Festini, 2017). Such factors appear unlikely as be the sole explanation for age-related memory decline, since different memory tasks appear deficient to different extents (i.e., implicit memory, semantic memory; see Burke & Light, 1981; and also Johnson et al., 2010), and age differences in memory have been observed in samples closely matched on health and education levels (see Kausler, 2012). Therefore, theories regarding the role of non-cognitive factors have not been prominent in the literature. Nevertheless, such factors are likely to contribute to age-differences in memory performance.

1.4 Individual differences in Cognitive Abilities: Do they decline at the same rate with age?

In the 1980s, new models of the ageing mind were developed relying on non-experimental individual differences and structural equation modelling (e.g., Hertzog, 1985; Horn, 1989). Such work indicated that the mechanisms underlying both memory and ageing were multifactorial, which promoted multi-causal views on age-related memory decline (see Park & Festini, 2017). For instance, some findings indicated that both processing abilities and knowledge were important for memory performance and that their relative contributions varied as a function of task and age (Hedden, Lautenschlager, & Park, 2005). Wilson et al. (2002) assessed cognitive decline in nearly 700 older adults over six years. They found that different cognitive abilities declined at different rates with increasing age and that different individuals were affected to differing extents with age. Hence, both the extents and natures of cognitive impacts of growing older appeared to differ among people, rather than producing some general, overall decline.

Other research suggested that cognitive resources which are separate (i.e., specialised) in younger adults develop into one general resource as we grow older, known as *dedifferentiation* (see Balinsky, 1941; Baltes & Lindenberger, 1997; Li, Lindenberger, & Sikstrom, 2001; Li & Lindenberger, 1999). Similar patterns emerged from neuroimaging studies. For instance, Payer et al. (2006) observed less neuronal specialisation for passive viewing of face and house stimuli in older adults than younger adults (see also Cabeza, 2001; Park et al., 2001; Reuter-Lorenz et al., 2001). Behavioural evidence supporting this proposition has been mixed (e.g., DeFrias, Lovden, Lindenberger, & Nilsson, 2007; Juan-Espinosa et al., 2002; Salthouse, 2000; Tucker-Drob & Salthouse, 2008; Zelinski & Lewis, 2003). For instance, Park et al. (2002) found little behavioural evidence for dedifferentiation in the cognitive architecture of memory (including visuospatial and verbal WM, short-term and long-term memory processes) across the lifespan, among 345 adults. However, they suggested that behavioural dedifferentiation might be obscured by compensatory neural recruitment – in line with the Scaffolding Theory mentioned above (see Park et al., 2001; Reuter-Lorenz et al. 2000).

More broadly, research on dedifferentiation highlighted that younger and older adults might recruit different cognitive mechanisms to perform the same task. This notion was central to the research in this thesis. Specifically, I hypothesised that older adults might rely more on verbal strategy use in visual WM task, based on evidence of a differential age-related decline in visual as opposed to verbal abilities with age, which I discuss next.

1.4.1 Do Visual Abilities Decline more than Verbal Abilities with Age?

Park et al. (2002) found that while WM and processing speed exhibited a steady decline starting in the early 20s, forward digit span (which allows verbal rehearsal) declined only modestly, and verbal knowledge continued to increase until participants were in their 70s. Similarly, fluid intelligence appears to decline more with age than crystallised intelligence (Horn & Masunaga, 2000). The notion of relatively more intact verbal abilities was also supported by a recent meta-analysis which investigated shared variance in longitudinal cognitive change from 22 different datasets, containing data from more than 30,000 participants (Tucker-Drob, Brandmaier, & Lindenberger, 2019). They found that processing speed, spatial ability, and reasoning declined significantly more than the grand mean estimate across domains, while verbal knowledge displayed significantly less decline than the grand mean estimate.

In particular, age-related WM deficits for visuospatial material (matrix patterns or coloured shapes) have been shown to be more severe than those for verbal material; such as words, letters, and digits (e.g., Jenkins, Myerson, Hale &

Fry, 1999; Leonards, Ibanez & Giannakopoulos, 2002; Myerson, Hale, Rhee & Jenkins, 1999). For instance, more pronounced decline in visual than verbal WM abilities was found in a set of experiments with 16 younger and 16 older adults. Jenkins, Myerson, Joerding, and Hale (2000) found that age-related slowing was more pronounced in visuospatial than verbal tasks and that younger-older differences were more substantial for location memory than for letter memory. Finally, while older adults were more impaired overall than younger adults when learning new information, this deficit was substantially larger when learning visuospatial than verbal information. Similar patterns were later found in a sample of 1,050 participants, aged between 20–89 years, as spatial memory performance declined significantly more with age than verbal memory, as measured by forward digit span (Myerson, Emery, White, & Hale, 2003).

Johnson et al. (2010) conducted a large online study (N > 95,000; 8 to 80 years old). Participants completed a set of cognitive tests, including forward digit span (verbal memory), and the visual patterns task (requiring memory for visual pattern matrix). Strikingly, visual memory appeared to decline quite markedly from age 20, while the verbal memory task (digit span) declined much less, and much later (see Figure 1.3).

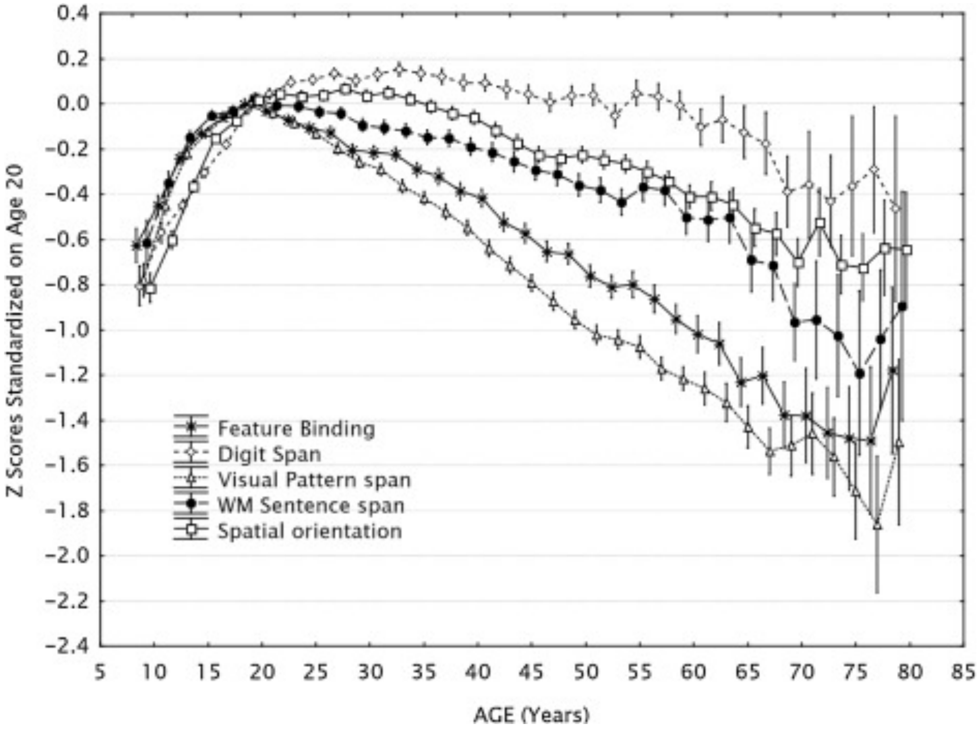


Figure 1.3. Figure reproduced from Logie, Horne, and Petit (2015). Scores plotted as *Z*-scores based on the 20-year old participants' means and standard deviations.

Johnson et al. (2010) also explored patterns of common variance in different tasks. They found that task-specific variance (residual variance) in the digit span task increased with age. They interpreted this as indicating that older people relied more heavily on a specific cognitive ability (such as verbal WM) in this task than younger adults. Conversely, task-specific variance in the spatial orientation and visual patterns memory tasks was higher in younger than older adults. The authors concluded that older people appeared to rely more on a specific cognitive ability to perform the verbal WM task (digit span), but rely on a more general cognitive ability for the visual memory tasks. In contrast, younger adults appeared to rely on a more specialised system for the visual WM tasks. Taken together, these findings that verbal knowledge/abilities appeared less affected by increasing age led to me to hypothesise that older adults may rely on such relatively intact verbal abilities when faced with challenging visual memory tasks. This hypothesis was central to the research described in the experimental chapters below, and is also discussed in more detail in section 1.8, *Development of Thesis Research Rationale.*

1.5 Strategy Use: Verbalisation of visual information?

Both younger and older participants appear to apply strategies to WM tasks (e.g., Atkinson, Baddeley, & Allen, 2017; Brown, Forbes, & McConnell, 2006; Dunning & Holmes, 2014; Laine et al., 2018). Specifically, some studies with younger adults found that memory for visual information was better when it was easy to verbalise, or name. Brown, Forbes, and McConnell (2006) investigated the role of verbal memory in the visual patterns task, typically assumed to be a visual short-term memory task. In this task, an array of squares is presented, a random half of which are filled while the rest are blank. The array is then removed and replaced with an array with all blank squares, and participants are asked to recall which squares were previously coloured. Brown et al. (2006) asked younger adult participants to generate verbal labels for the square patterns and found that some patterns resulted in a higher mean number of labels than others. Excluding these 'more nameable' patterns from the memory task resulted in significantly lower visual memory performance than in the original version of the task. This illustrated that visual information may be remembered via verbal labels – and potential rehearsal of such labels. Similarly, allowing or encouraging verbal labelling (i.e., by instructing participants to say colour names out loud) drastically improved memory performance in younger adults in a colour-wheel delayed estimation task (Souza & Skóra, 2017; Exp. 5 in this thesis was a conceptual replication of this study). These studies illustrate that while visual and verbal WM may rely on separate constructs (as proposed by the Multiple-Component model), and decline at different rates with age, verbal WM can be used to rehearse visually presented stimuli.

Translating visual input into verbal labels is well established as a default —even sometimes unavoidable — tendency (Conrad, 1964; Postle, D'Esposito, & Corkin, 2005; Postle & Hamidi, 2006; Shulman 1971; Simons, 1996). Memoranda in everyday life as well as in memory experiments may be remembered via verbal codes or visual traces, or both. For example, remembering which jumper to purchase after trying on several at the store could be achieved using a verbal description ("the blue one"), as well as a visual representation of it. Neuroscience evidence also suggests that participants may generate visual, phonological and semantic mental codes while viewing visual stimuli (Lewis-Peacock et al., 2014). Despite this tendency to translate visual representations into verbal codes, visual and verbal WM are typically measured separately. Tasks are assumed to measure one or the other, despite evidence that both verbal and visual codes might be stored for visually presented material (e.g., Logie, 2018; Logie, Saito, Morita, Varma, & Norris, 2016; Saito, Logie, Morita, & Law, 2008; Paivio, 1971), and that such codes may interfere with one another in complex ways (e.g., Morey & Cowan, 2004).

When required to maintain representations in WM, participants appear able to adaptively choose between attentional refreshing (i.e., directing attention to an item held in WM; see Camos et al., 2018) and verbal rehearsal. When remembering phonologically similar materials impeded effective verbal rehearsal, participants favoured attentional refreshing – which reduced the detrimental impact of the phonological similarity effects – and when attentional capacity was constrained due to task demands, they appeared to favour rehearsal, which is less attentionally demanding (Camos et al., 2011). Others found that verbal rehearsal appeared more prominent in simple WM tasks (where participants recall a set of items in their correct serial order) than complex span tasks (requiring participants to e.g., solve math problems while also remembering words), as evidenced by a greater detrimental effect of articulatory suppression in simple tasks (Unsworth & Engle, 2007). This increased reliance on verbal rehearsal in simple tasks is likely because it is not possible to verbally rehearse items while also performing some secondary task, such as solving equations.

I was interested in whether older adults may use such a verbal rehearsal strategy more. Arenberg (1977) found that older adults benefitted more than younger adults from auditory descriptions of salient stimulus features in the non-verbal Benton Visual Retention Task. Arenberg suggested that the auditory description may have boosted older adults' performance by providing rehearsable input as well as additional retrieval cues. Hartley et al. (2001) found that in younger adults, WM for verbal information (name identity), visual objects, and their spatial location were dissociable. However, while both younger and older adults showed involvement of name identity in the object task, older adults also showed this involvement in a spatial task. They speculated that this could either reflect verbal strategy use by older adults in the spatial task, or an inability to inhibit verbal memory in the spatial task.

Fox and Charness (2010) investigated the impact of 'thinking aloud' in Raven's Matrices (an inductive reasoning test which requires participants to decide which of eight choices best completes a matrix of abstract figures). In two studies, they found that older adults performed significantly better on the Raven's Matrices while thinking aloud. The improvement corresponded to a near one-standard-deviation increase in fluid intelligence scores. While some research has focused on strategic approaches by older adults in WM (such as focusing only on a subset of items; Atkinson, Baddeley, & Allen, 2017, or the neural mechanisms of verbal to visual code switching training; Osaka, Otsuka, & Osaka, 2012), there is not a lot directly investigating verbal strategy use (although see Horne, 2015). For instance, verbal rehearsal of visual features was proposed as a potential confound in a large visual feature-binding study by Brockmole and Logie (2013, p. 4), but was not tested directly. In this thesis, I examined verbal strategy use by older adults, using the framework outlined below.

1.6 Rejecting the ‘Dull Hypothesis’

This thesis was inspired by calls to move beyond testing 'The Dull Hypothesis' (Perfect & Maylor, 2000), i.e., the idea that older adults are merely like poorly performing younger adults. Confirming the 'Dull Hypothesis' entails comparing younger and older adults' performance on some WM task, and presumably finding that older adults perform worse. Instead, I focus on evidence that some cognitive functions may remain relatively intact with age, and that older adults may preferentially recruit such functions – and thus approach cognitive tasks differently from younger adults.

Knowledge of relatively age-invariant abilities – and whether they are recruited to perform other tasks – may help correctly target support for older adults, and avoid helping them with tasks they can already do (e.g., Gonçalves et al., 2017). Similarly, it may help us avoid designing online environments requiring the use of abilities that do decline more with age. For instance, if older adults remember verbalisable information better, digital application icons with words may be better than purely visual icons.

However, differential task-approaches could also be problematic for research on age-related cognitive decline. If younger and older adults approach an identical cognitive task differently, then arguably, the two groups are not performing the same task. Hence, comparing age-group differences without understanding how participants are performing tasks may not provide an accurate representation of cognitive decline as we age.

Consider a non-cognitive metaphorical illustration of this notion. The outcome variable is how fast one can drive a mile on a motorcycle. This may depend on a variety of factors, such as the maximum velocity of the bike, the width and height of the driver (creating wind-resistance), and the drivers' aptitude for using the gear shift and maneuvering turns in the road without reducing velocity. A better driver on a slower bike would be victorious if the race took place on narrow, curvaceous roads where the opponent was not able to use their bike's top speed. However, on the highway, the highest-velocity bike would win, and driving ability (beyond basic bike-maneuvering) would not matter. Overall speed could be seen as analogous to the proportion of correct answers in a memory test; one score may reflect the recruitment of several different abilities (see Logie, 2018 for similar reasoning). In some visual WM tasks, excellent performance might be achieved either by retaining a visual representation of what the items looked like or by verbally labelling and rehearsing item names. If older adults' verbal rehearsal mechanism is relatively intact, and the visual WM task permits use of verbal rehearsal, younger adults may not perform much better even if their visuospatial memory abilities are superior. Let us return to the metaphor. If younger adults typically beat older adults in motorcycle races, we could test whether this overall difference is caused by age group differences in bike velocity or their ability to manage sharp turns, by letting younger and older drivers race on a straight, as well as a winding road. If older adults were significantly worse on the straight road, but more similar to younger adults on the winding road, it would suggest that their bikes were slower, while their maneuvering skills were relatively intact. This would reject the 'Dull Hypothesis' that older adults are simply worse at everything. I used a similar experimental approach to investigate age differences in WM performance in this thesis, which I describe further below.

1.7 The Experimental Approach

I tested whether younger adults and older adults approach tasks differently using an experimental approach. This was based on 'ANOVA' logic: including age as a between-group factor, and some within-participant task manipulation, to test if the interaction between age-group and the other factor was significant. If so, this is taken to indicate that the *difference in the differences* between the two experimental conditions was larger in one age group than the other; rejecting the 'Dull Hypothesis' for that particular manipulation. Although I have used a variety of paradigms and analysis techniques, all six studies in this thesis relied on this logic.

However, I acknowledge that finding such an interaction does not always lend itself to a straightforward conclusion. For instance, if the difference between two conditions is more substantial in a given age group, we cannot be sure if this is because they are comparatively better in task A or worse in task B – or a bit of both. Also, there are issues regarding statistical power (see Salthouse, 2000 for a discussion regarding interpretation problems of interpretation of age by condition; and Allen, Hitch, Mate, & Baddeley, 2012 for evidence suggesting that evidence for interactions appeared inconsistent using different measures of recognition). Still, testing an age-group by condition interaction is arguably valuable to establish whether the impact of age is equal across conditions.

1.8 Development of Thesis Research Rationale

The research in this thesis was designed to address specific discrepancies in key areas of the literature on how WM declines with age. I explored whether younger and older adults appeared to approach WM tasks differently, and if this could explained inconclusive patterns of results in the visual feature-binding literature (Exp. 1 – 4). Next, I tested the impact of verbal labelling on colour precision memory (Exp. 5), and a paradigm used in WM training (the N-back task; Exp. 6). To avoid repetition, I describe background literature and details of these paradigms within their specific chapters. Below, I outline the four building blocks behind the general rationale.

I. Visual and Verbal WM may be separate components. A wide body of research – often closely associated with the Multiple-component model of WM (see Section 1.2 above) – suggests that visual and verbal WM may rely on different components. Verbal rehearsal of visual information is acknowledged as a special mechanism in most models of WM.

II. Visual and Verbal WM decline at different rates with age. As reviewed above, the ability to hold visual information in WM appears to decline more than the ability to hold verbal information (see Section 1.4.1 above).

III. Cognitive tasks are not process-pure. Some argue that no cognitive task measures only the specific cognitive function the researcher is attempting to measure (e.g., Engle et al., 1999). Specifically, 'visual' WM tasks can be approached with verbal strategies (e.g., see Brown et al., 2006, see Section 1.5 above).

IV. People approach cognitive tasks differently. Participants vary in how they approach any given task (see Laine et al., 2018; Logie, 2018; Logie et al., 1996;

Siegler, 1987; Unsworth & Engle, 2007, see Sections 1.5 and 1.6 above). Furthermore, there might be systematic differences, e.g., groups of younger adults may tend to use one approach, and older adults another.

Combined, the idea that visual and verbal WM abilities are separate (I) and verbal WM abilities remain more intact (II) and can be used in visual tasks (III), and people approach tasks differently (IV), led to my central research question: do older adults rely on different abilities to approach the same task? More specifically; do older adults use verbal strategies more than younger adults in visual WM tasks?

More broadly, this general notion is supported by findings in different research areas, such as differential variation in large-scale correlational studies (see Section 1.4.1 above), and neuro-cognitive findings and theories of how individuals may compensate for brain atrophy by compensatory, additional recruitment in other areas (i.e., the Scaffolding Theory, see Section 2.3.4). Although suggested as a potential explanation for results in a variety of visual WM tasks (e.g., Brockmole & Logie, 2013, p. 4), behavioural evidence is scarce, and the cognitive mechanisms underpinning different approaches by participants of different ages remain poorly understood.

1.9 Overview of Current Research

The overarching aim of the research presented in this thesis was to identify differential approaches by younger and older adults, to see if it might explain discrepancies in the literature on visual feature-binding, visual memory precision, and cognitive training literature. The broader contributions will be a better understanding of how older (compared to younger) adults approach the same WM tasks, and of the role strategic approaches can play in such tasks.

In the first series of experiments (presented in Chapters 2 to 4), I relied on the idea that verbal rehearsal processes rely on a different mechanism, which we can prevent or encourage experimentally. In Experiments 1 to 4, I investigated verbal strategy use in visual feature binding; and whether this might be more beneficial in the single-feature (as opposed to the binding condition), by looking at age differences in the consequences for overall performance of preventing it. Then, I tested the role of explicit verbal labelling in colour memory precision using a mixture modelling approach to test if younger and older adults approach this task differently (Exp. 5). Looking at behavioural effects of meddling with the verbal rehearsal mechanism may allow us to understand differences in how younger and older adults approach tasks spontaneously. If disrupting verbal rehearsal is more detrimental for older adults than younger – compared to spontaneous performance – it suggests older adults were more *reliant* on verbal rehearsal for the overall performance level.

In the final experiment (Chapter 5), I used a slightly different approach, to be able to address an important broader aim of cognitive ageing research - can cognitive abilities be trained to resist the general decline? Here, I tested if an instructed visualisation task – previously found to improve performance in younger adults – would also be beneficial in older adults. I also investigated the role of self-generated strategies. Finally, the conclusion provides a summary, outlines some limitations of the present research, and identifies areas for further research, which could provide more insight into age difference. A final disclaimer: while the perspective of several different abilities – varying between individuals but also between age groups – contributing WM performance is perhaps most closely related to the Multiple-component model, the work in the thesis is not bound to a specific model of WM. I applied this principle to test important questions regarding cognitive ageing but refer to various models throughout.

Chapter 2: Aging and feature-binding in Visual Working Memory: The role of verbal rehearsal

At the time of writing, the following paper is under review after revise and resubmit in Psychology and Aging (hence the American Spelling).

Abstract

Age-related decline in ability to bind and remember conjunctions of features has been proposed as an explanation for the pronounced decline of visual Working Memory (WM) in healthy aging. However, evidence that older adults exhibit greater visual feature binding deficits than younger adults has been mixed. Binding deficits in older adults are often observed using paradigms with easy-to-label features. Labeling and rehearsing single features may result in apparent binding deficits if older adults rely on comparatively intact verbal memory to compensate for declining visual WM. This strategy would be more useful for single features (e.g., 'red'), than for conjunctions of features (e.g., 'red triangle') which are more cumbersome to rehearse, and thus visual feature-binding paradigms which do not prevent verbal strategies may unintentionally measure verbal load differences. Across three experiments (total N = 150), we investigated the role of verbal rehearsal by manipulating ease of stimulus labeling for visually presented single features and conjunctions of two features.

Overall, visual memory for difficult-to-label, non-categorical, visual information appeared especially limited for older adults, likely because it impedes engagement of other systems, such as verbal WM or long-term memory. Therefore, comparing younger- and older-adult task performance may not straightforwardly reveal age-related visual WM decline, but instead reflect applications of different strategies that tap different cognitive mechanisms. We discuss implications for the feature-binding literature, and the wider visual WM literature.

Introduction

Working memory (WM) refers to cognitive functions that support the ready availability of a small amount of information on a temporary basis while undertaking ongoing actions and mental activities (e.g., Logie & Cowan, 2015). Along with other aspects of cognition, WM has been shown to be poorer in groups of older than in groups of younger healthy adults (e.g., Babcock & Salthouse, 1990; Bowles & Salthouse, 2003; Craik, Luo, & Sakuta, 2010; Gazzaley, Cooney, Rissman, & D'Esposito, 2005; Johnson, Logie, & Brockmole, 2010; Park, Lautenschlager, Hedden, Davidson, & Smith, 2002). This age-related decline has practical importance because WM is believed to underpin effective operation of other cognitive functions, such as perception and problem-solving (e.g., Ma, Husain, & Bays, 2014), and to be related to general intelligence (e.g., Unsworth, Fukuda, Awh, & Vogel, 2014) and reasoning ability (e.g., Kyllonen & Christal, 1990; Conway, Kane, & Engle, 2003). The ability to retain visual features of stimuli in working memory appears to be particularly sensitive to age-related cognitive decline (e.g., Bowles & Salthouse, 2003; Gazzaley et al., 2005; Johnson et al., 2010). Two potential components of age-related decline in visual working memory have been proposed: First, reduction in the number of items that can be stored, and second, decreased ability to retain associations (*bindings*) between different object features (Bopp & Verhaeghen, 2009; Brockmole, Parra, Sala, & Logie, 2008; Cowan, Naveh-Benjamin, Kilb, & Saults, 2006; Mitchell, Johnson, Raye, & D’Esposito, 2000b; Olson et al., 2004; Parra, Abrahams, Logie, & Sala, 2009; Sander, Werkle-Bergner, & Lindenberger, 2011). This distinction has been useful for understanding the marked decline of episodic memory with age (for a review, see Shing et al., 2010), where *associative deficits* (impairments when required to remember associations between items over and above any deficit exhibited for those items individually) have been demonstrated across a variety of stimuli (Naveh-Benjamin, 2000; Naveh-Benjamin, Hussain, Guez, & Bar-On, 2003; Spencer & Raz, 1995; see Old & Naveh-Benjamin, 2008 for a review and meta-analysis). However, the role of age-related deficits in temporary binding in visual working memory appears less straightforward.

Typically, memory for such short-term feature-bindings has been measured experimentally by comparing temporary memory for specific features, such as color, shape or location, individually or bound together (e.g., a colored shape in a particular location in an array). In these experiments, the same small sets of features are presented repeatedly in different combinations from trial to trial. For example, on one trial participants might be asked to remember a briefly presented array comprising a green circle, a red square and a blue triangle with the test of memory one or two seconds later. On the next trial, the memory array would consist of different combinations of colors and shapes. Variations of this general paradigm have been used extensively in the study of object perception and attention (e.g., Hu, Allen, Baddeley, & Hitch, 2016; Tas, Luck, & Hollingworth, 2016; for reviews of earlier research see Zimmer, Mecklinger, & Lindenberger, 2006), and in the study of the impact of age on working memory for visual features (e.g., Brockmole & Logie, 2013; Brown, Niven, Logie, Rhodes, & Allen, 2017; Cowan et al., 2006; Kessels, Hobbel, & Postma, 2007; Mitchell, Johnson, Raye, Mather, & D'Esposito, 2000a; Rhodes, Parra, & Logie, 2016). Several studies have observed that while short-term memory for individual colors and shapes (and sometimes locations) was relatively preserved in older adults, it was significantly impaired for combinations of colors, shapes and locations (i.e., bindings) compared with younger adults (Mitchell et al.,, 2000a; Mitchell et al., 2000b; Brockmole & Logie, 2013; Kessels, Hobbel, & Postma, 2007; Chalfonte & Johnson, 1996). In contrast, other studies have reported no evidence for age-related binding deficits (e.g., Brockmole, Parra, Della Sala, & Logie, 2008; Brown et al., 2017; Chen & Naveh-Benjamin, 2012; Cowan et al., 2006; Parra, Abrahams, Fabi, Logie, Luzzi, & Della Sala, 2009a; Parra, Abrahams, Logie, & Della Sala, 2009b; Rhodes et al., 2016; for a review see Allen, Brown, & Niven, 2013). Feature-binding deficits are of practical importance for pathological aging, since simple visual WM binding tasks have distinguished pathological cognitive decline from that associated with healthy aging. Specifically temporary color-shape binding has been found to be unimpaired in healthy older people, but specifically impaired in individuals suffering from Alzheimer's disease (e.g., Parra et al., 2009a; Parra, Della Sala, Abrahams, Logie, Méndez, & Lopera, 2011).

Moreover, memory for temporary bindings between colors and shapes were found to be impaired in people with genetic mutations that result in early-onset Alzheimer's disease, when these individuals were otherwise asymptomatic and up to ten years before they would be expected to develop the disease (Parra, Abrahams, Logie, Méndez, Lopera, & Della Sala, 2010). However, debate about which type of paradigm best distinguishes healthy from pathological aging has emerged due to recent inconsistent observations regarding whether or not there are age-related binding deficits in the WM literature (see Liang et al., 2016; Logie, Parra, & Della Sala, 2016; Parra et al., 2016).

In this paper, we identify and test a potential reason behind these inconsistencies: the studies reporting age-related binding deficits included common, easy-to-label stimuli (e.g., common shapes like triangles, or colors like red) and did not attempt to prevent verbal rehearsal. In contrast, the studies where older adults did not show greater binding deficits were designed to reduce opportunities for verbal strategy use, by including difficult-to-label features (e.g., irregular hexagons or complex fractals) or requiring articulatory suppression, which requires participants to repeat an irrelevant pair of digits or short word aloud while viewing the stimulus arrays and until responding. See Table A.20 in Appendix A for a summary of different paradigms (and stimulus types) used to measure age-related feature-binding deficits. One of the largest studies on age-related deficits, including over 55,000 online participants, used features that could easily be labelled, and observed a significant age-related feature-binding deficit with age as a continuous variable (Brockmole & Logie, 2013), and with memory tested by reconstruction of the feature combinations. Similarly, Kessels, Hobbel, and Postma (2007) observed a binding deficit in older adults with stimuli consisting of easy-to-label objects presented in a grid for 3 seconds, requiring an immediate response, using reconstruction. Also, more recently developed delayed-estimation precision paradigms used to study binding also rely on a reconstructive procedure (Peich, Husain, & Bays, 2013; Pertzov, Heider, Liang, & Husain, 2015). Therefore, in the present study, we used a reconstruction binding paradigm. The method used to quantify binding likely contributes to discrepancies regarding age-related binding deficits. However, we did not directly compare different types of binding paradigms, but focused instead on the effect of permitting verbalization in one paradigm (reconstruction).

Indeed, most memoranda – in memory experiments as well as in everyday life – can be remembered either via verbal codes or visual memory traces, or both (Lewis-Peacock, Drysdale, & Postle, 2014; Morey & Cowan, 2004). For example, remembering which glass was yours after putting it down at a party could be achieved using a verbal description ("the champagne flute"), as well as a visual representation of what the specific glass looked like. Such translation of visual representations into verbal code has been found to improve visual memory performance in younger adults (Brown, Forbes, & McConnell, 2006; Souza & Skóra, 2017). Despite this, tasks are often assumed to measure either visual or verbal WM, perhaps incorrectly (e.g., Logie, 2018). For example, Saito, Logie, Morita, & Law (2008) showed that participants used both visual and verbal codes to retain visually presented letter and word sequences (see also Logie, Saito, Morita, Varma, & Norris, 2016). When a visual stimulus is translated into a verbal code it can be maintained in memory via *sub-vocal rehearsal*, i.e., silent repetition of verbal labels for material to be recalled (see Logie, Della Sala, Laiacona, Chalmers, & Wynn, 1996; Wang, Logie, & Jarrold, 2016). Sub-vocal rehearsal is an essential feature of the 'phonological loop' (Baddeley, 1986, 1992; Baddeley, Lewis, & Vallar, 1984), part of the multi-component model of working memory (Baddeley & Hitch, 1974; Baddeley, 1986; Baddeley & Logie, 1999). While other conceptualizations of WM do not emphasize domain-specific stores, sub-vocal rehearsal of verbal material is generally recognized as a separate mechanism (Cowan, 1992; 2005; Oberauer, 2013, Camos, Lagner, & Barrouillet, 2009), while the presence of a visuospatial rehearsal mechanism is more contentious (see Hanley & Young, 2018; Logie, 2003; Logie et al., 2016; Morey, 2018).

Verbal rehearsal could be problematic in paradigms used to measure visual feature-binding, because such rehearsal is likely comparatively more effective for recalling single features (which requires maintaining, for example, three or four shapes in memory), than bound features (which requires rehearsing six or eight features, and crucially, which of them belong together). Moreover, the time available for rehearsal is typically limited, and is the same for single and binding trials. Therefore, if older adults have a greater tendency to employ verbal rehearsal and do so more successfully with single features, this could create an apparent age-related binding deficit – i.e., relatively preserved performance on single-feature than on binding trials, when statistically compared with the difference between single and binding trials in younger adults – in paradigms which allow verbal rehearsal.

The proposal that older adults may rely more on verbal rehearsal is supported by the broader research literature on cognitive aging, which suggests that not all cognitive functions decline with age to the same degree (for reviews see Logie & Morris, 2014; Perfect & Maylor, 2000). For instance, healthy older adults appear to have relatively spared verbal working memory. Studies (e.g., Johnson et al., 2010) and meta-analyses (e.g., Jenkins, Myerson, Joerding, & Hale, 2000) have indicated that visuospatial cognition declines more with aging than does verbal cognition. In particular, working memory deficits for visuospatial material have been shown to be more severe in older participants than those for verbal material (e.g., Jenkins, Myerson, Hale, & Fry, 1999; Leonards, Ibanez, & Giannakopoulos, 2002; Myerson, Hale, Rhee, & Jenkins, 1999, but see Park et al., 2002; Salthouse, 1995). In general, in tasks that permit verbal rehearsal (e.g., digit span tasks) older adults are often observed performing as well as younger adults (Fisk & Warr, 1996), whereas age differences are large for visual material (e.g., remembering visual patterns; Johnson et al., 2010).

Furthermore, in an online study with over 95,000 participants aged 18 to 90 performing a variety of memory tasks, Johnson et al. (2010) found that the factor structures of performance on various WM tasks varied among age groups. In other words, the relative magnitudes of shared variance among the tasks differed for different age groups. Visual-pattern memory was more correlated with performance on the other measures among the older participants than among the younger participants. Hence, for the older participants, visual pattern memory seemed more related to some general cognitive capacity, but in younger people, it seemed to reflect a specific capacity. The opposite was found for verbal memory, measured by memory for number sequences (digit span). This could suggest that older adults are compensating for decline in function, for instance in brain regions supporting specific visuospatial WM subsystems, by making greater use of verbal strategies (Reuter-Lorenz et al., 2000). For instance, when faced with a visual memory task, they might attempt to support their impaired visual memory ability by applying verbal labels to the visually presented materials and use their relatively intact verbal memory abilities to rehearse those verbal labels, thereby performing better than if they had relied on their visual memory abilities alone. Indeed, some evidence suggests that older adults – despite capacity deficits – can use strategies (such as focusing on a subset of important information) as successfully as younger adults in both the verbal (e.g., Castel, Benjamin, Craik, & Watkins, 2002) and visuospatial (e.g., Siegel & Castel, 2018) memory domains. However, other research suggests that younger adults are more likely than older adults to engage in verbal rehearsal to improve WM performance (e.g., Peterson & Naveh-Benjamin, 2016), and may be more likely than older adults to initiate strategies to support long-term memory (see Craik & Rose, 2012). Despite this, to the extent that older adults become (at least subconsciously) aware of failing visual memory relative to verbal, they may be particularly likely to supplement visual memory with verbal strategies, thus offsetting the general tendency of younger adults to do so more readily under certain conditions. Identifying strategies people use to maintain daily function in old age is essential to understanding cognitive decline, and how to measure it experimentally as well as clinically. If older adults' performance is more negatively affected than younger adults when verbal rehearsal is prevented – compared to circumstances where it is allowed – this would suggest that a greater proportion of older adults' successful 'visual' WM memory is supported by verbal rehearsal. This proposed instance of how a relatively intact capacity may be recruited to compensate for a declining one has potential implications for numerous paradigms used to investigate memory decline across the lifespan. Successful compensatory use of verbal rehearsal strategies by older adults could inform understanding of general circumstances which may alleviate older adults' decline on tasks that are assumed to involve visual memory.

In three experiments we manipulated the likelihood of verbal rehearsal by manipulating stimulus labeling difficulty to test whether older adults show relatively better performance for easy-to-label stimuli, and used articulatory suppression to test whether that performance gain could be attributed to verbal rehearsal. Considering evidence of relatively spared verbal WM memory with age, we hypothesized that older adults would perform more similarly to younger adults when items were easier to label, but show an age-related decline for difficult-to-label materials, which are difficult to rehearse verbally. Secondly, binding deficits are typically quantified by comparing mean performance on single-feature trials with that on binding trials. If participants approach visual feature-binding tasks verbally, the crucial comparison between memory for single and bound features can be thought of as a (verbal) load manipulation, requiring twice as many items in the binding condition. This should impede both development of effective labels and ability to rehearse them. Because successful verbal rehearsal hinges on having sufficient time to rehearse the to-be-remembered words, such a strategy should be suitable for rehearsing three or four single features for a couple of seconds, but not be as useful when asked to retain twice the number of features in the binding condition. Therefore, we hypothesized that the opportunity for verbalization would produce age-related apparent feature-binding deficits by enabling older adults to perform well in the single-feature condition but being much less helpful in the binding condition. This could explain discrepancies in the literature, as outlined above.

Experiment 1

In Experiment 1 we investigated whether making verbal labeling difficult resulted in the appearance of a greater age-related decline in temporary memory for visual features. Recall was tested using a reconstruction procedure in which participants responded by selecting features from arrays of individual features. We presented some stimuli to which it was easy to assign verbal labels, and others for which it was more difficult, measuring Shape and Color Recall separately, similar to Experiment 3 in Brockmole et al. (2008). For some trials, participants remembered only single features (either colors or shapes), and for the other trials, they remembered bound features (integrated objects consisting of a shape of some color). We hypothesized that when task features were easy to label older adults would perform similarly to younger adults in single-feature conditions, but perform more poorly than younger adults when asked to remember bound features, i.e., an age-related feature-binding deficit (in line with Brockmole & Logie, 2013). In contrast, when features were difficult to label, we anticipated that verbal rehearsal would be less feasible and all participants would rely on retaining visual representations rather than verbal labels. Thus we did not expect an age-related deficit for bound, as compared to individual features, consistent with previous studies where verbal rehearsal was prevented (Parra et al., 2009; Rhodes, Parra, & Logie, 2016).

Methods

Participants. We recruited 51 participants, all native speakers of English. Twenty-five University of Edinburgh students (three male and one participant who did not identify as either male or female) aged 18 – 27 (*M* = 22.3, *SD* = 2.1) years received £8.50 in return for participation. Twenty-six older adults (six male), all from the University of Edinburgh psychology research community volunteer panel, aged 66 – 75 (*M* = 69.7, *SD* = 2.8) years, were each given £10 in return for participation. One older adult was excluded for not completing the memory task. The final sample size of 50 participants was determined prior to data collection, based on recent studies' sample sizes addressing similar questions (e.g., Rhodes, Parra, & Logie, 2016; Rhodes, Parra, Cowan, & Logie, 2017; Brown & Brockmole, 2010). Years of education did not differ significantly between the age groups (older: *M* = 15.0, *SD* = 3.7; younger: *M* = 16.2, *SD* = 1.6); *t*(32.27) = 1.40, *p* = .170, *d* = 0.41. Providing years of education was optional, and was given by 20 younger adults, and 24 older adults.

Prior to participating in the main experiment, all participants completed an on-screen version of the Dvorine pseudo-isochromatic plates (Dvorine, 1963) to assess color-vision. More than five errors are indicative of color-vision deficits (Dvorine, 1963), and no one was excluded on this basis. All older adults scored

86 or above (*M* = 96.7, *SD* = 3.2) on Addenbrooke's Cognitive Examination (ACE-III; Hodges, 2012), completed at the very end of their session only. A score lower than 82 is considered indicative of cognitive impairments (Mioshi, Dawson, Mitchell, Arnold, & Hodges, 2006). After completing the memory task, participants were asked to name each stimulus twice by typing it in a computerized naming survey, for a measure of 'Label-ability' and word-length in easy- and difficult-to-label stimuli. All participants completed the National Adult Reading Test (NART; Nelson, 1982) for an estimate of verbal IQ. Estimated verbal IQ scores were significantly higher in the older adult group (*M* = 123.9, *SD* = 4.1) than the younger adult group (*M* = 118.3, *SD* = 4.1), $t(47.9) = 4.69$, $p < .001$, $d = 0.37$.[1] The study was approved by our local Ethics Committee.

Stimulus and Apparatus. We established relative ease of labeling of visual features by asking 15 participants (aged *M* = 25.0, *SD* = 5.9, range 19 to 42 years, 5 male) who did not take part in the main experiment to name each color or shape three times (see Appendix A). This guided our selection of 32 features. The easy-to-label stimuli were shapes such as triangles and squares, and prototype versions of common colors such as red and green. The other stimuli were more difficult to label, such as irregularly-sided shapes and blends of common prototype colors. The eight difficult-to-label shapes were identical to those used by Brockmole et al. (2008) and by Rhodes, Parra, Cowan, and Logie

---

[1] Cohen's *d* was calculated using the population level Standard Deviation for IQ (*SD* = 15) for this and all subsequent Verbal IQ comparisons.

(2017). To select these items, we considered features difficult to label if a) many participants failed to generate labels for them, b) labels for the same feature were not consistent among and within participants, and c) verbal labels were longer, such as combinations of two color labels (e.g., 'greenish-yellow'), presumably making such verbal labels more difficult to rehearse successfully within the experimental time frame (consistent with Baddeley, Thomson, & Buchanan, 1975, word-length effect).

Easy-to-label features were those that generated the same, single-word label consistently among and within the 15 participants[2]. The two complete sets of shapes and the color RGB values are given in Appendix A. Difficult-to-label stimuli were defined as such relative to the easy-to-label stimuli, but they were not impossible to label, as participants could creatively label uncommon colors and shapes. We asked participants to name all items after completing the memory task, so that we could compare 'label-ability' and word-length in easy- and difficult-to-label stimuli.

For each trial, three or four memory items were presented on the computer screen with a grey background, randomly in eight possible locations

---

[2] The final set of easy-to-label colors was named by 100% of surveyed pilot participants with 90% within-participant consistency (average word length 5.4 letters). Difficult-to-label colors were named in 87.2% of instances, with within-participant consistency of 39.2% (average word length 7.6 letters). Easy-to-label shapes were named in 99.8% of instances with 91.9% within-participant consistency (average word length 7.2 letters). Difficult-to-label shapes were named by 54.3%, within-participant consistency 45.5% (average word length 12.1 letters).

around an invisible circle, 4.5 cm from the center of the screen. We combined item colors and shapes randomly without replacement, with the restriction that all features in each trial were either easy or difficult to label. Each stimulus image was about 2.2 cm^2 (visual angle approximately 2.10°) and viewing distance was approximately 60 cm. Stimuli were presented using PsychoPy v1.82.01 (Peirce, 2007) and displayed on a 22" LCD Monitor, with a diagonal of 20.6", and a screen resolution of 1680 × 1050.

Design and Procedure. An example of the memory task procedure is shown in Figure 2.1. Participants indicated recalled colors and shapes with mouse clicks. The experiment consisted of two single- and two dual-feature blocks. One block tested memory for color only (each test object was a ‘blob’ of a single color), one for shape only (all objects were black), and two blocks tested memory for both, with these blocks differing in which response (i.e., color or shape) was required first. We used color blobs and black shapes for our single-feature conditions rather than colored shapes to prevent participants from automatically encoding both features of the objects even when the task was to remember just one feature, based on evidence suggesting that color encoding is automatic (Ecker, Maybery, & Zimmer, 2013). We randomized block order across participants.

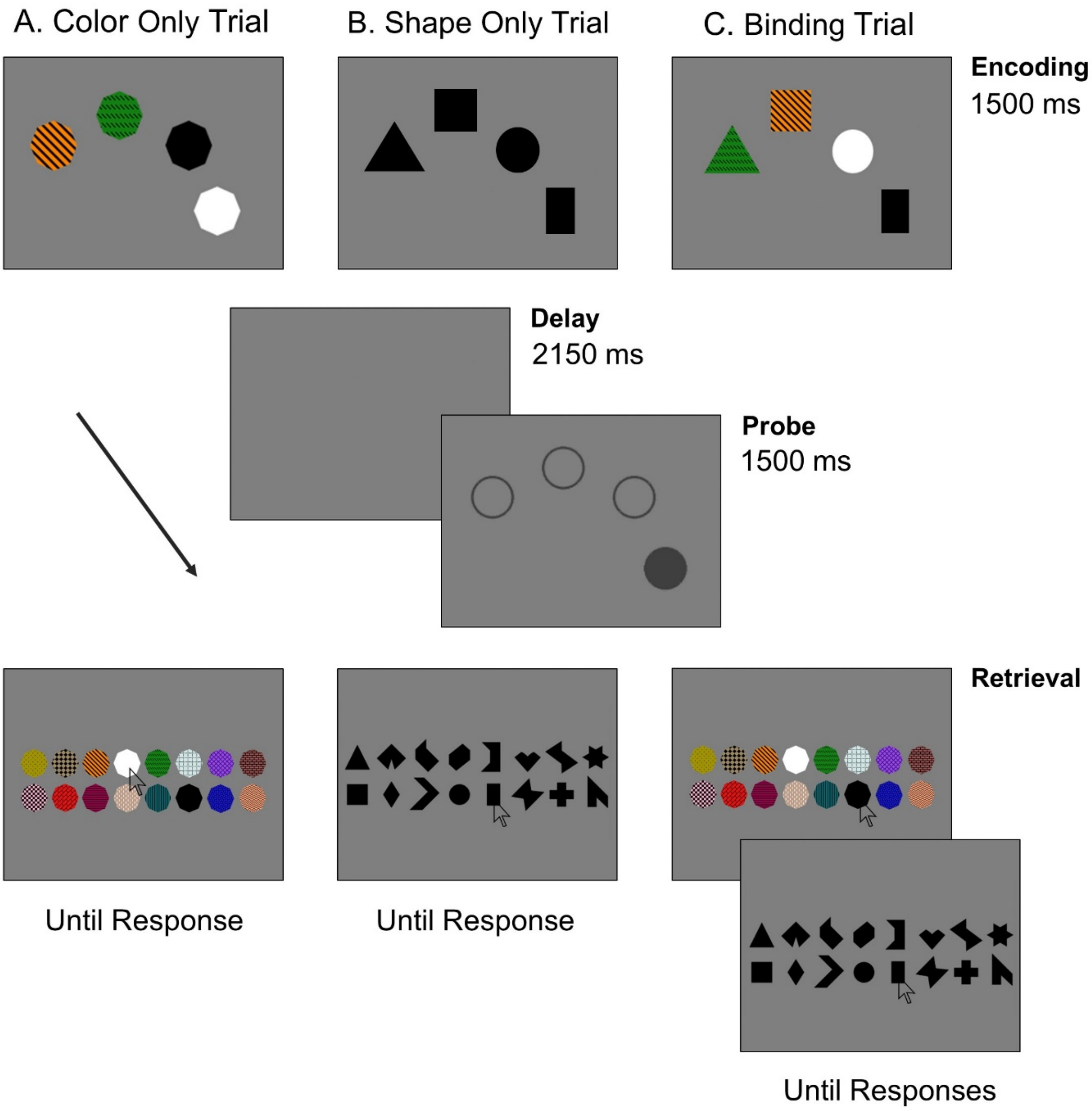


Figure 2.1. Illustration of the trial sequence Experiment 1. *A. Color only trial.* Participants remember colored blobs, and respond with a mouse-click from a selection of 16 color options. *B. Shape only trial.* Participants remember black shapes. *C. Binding trial.* Participants see colored shapes, and need to remember both the color and shape of the probed item. Participants did one binding block of trials where color was probed first (as illustrated), and another block where shape was probed before color. Mouse cursors represents correct responses. Participants had unlimited time to respond. *Note*. Different fill patterns represent different colors and items are not drawn to scale.

In each trial, participants viewed the test array (1500 ms), then there was a delay (2150 ms), and then probe circles outlined in dark grey appeared in all positions occupied by studied items (1500 ms) to offer contextual support for memory (see Figure 2.1). To indicate the randomly selected object to be remembered, one of these probe circles was filled in dark grey. Thereafter, the participant was asked to mouse-click the probed object's originally-displayed color and/or shape from a range of 16 shapes and 16 colors on the screen. The response screens consisted of all the 16 colors (eight easy-to-label and eight difficult-to-label) or 16 shapes (eight easy-to-label and eight difficult-to-label), probing color or shape memory, respectively, as illustrated in Figure 2.1. The presentation position for each color or shape was the same throughout the whole experiment for each individual participant but varied randomly among participants. This was to facilitate responses and minimize time spent searching for a specific color or shape to make responses. Set size varied randomly across trials, such that each trial presented either three or four items, selected based on previous studies indicating that three to four items generate performance levels below ceiling and above chance (Cowan, 2010; Luck & Vogel, 1997). We included two different set sizes as a precaution against floor or ceiling performance within age groups and/or conditions. Within each block, each participant completed 17 trials for each combination of set size and stimulus label-ability, resulting in a total of 68 trials per block, and a total of 272 trials[3]. Trial numbers were selected to ensure a practically reasonable session length (the full task took up to 65 minutes to complete).

Analysis. To analyze the data, we used a model comparison approach based on Bayes factors, also used by Rhodes et al. (2016; see also, Brown et al., 2017; Rhodes et al., 2017), implemented with the BayesFactor package in *R* (see Morey, Rouder, & Jamil, 2015 and R Core Team, 2015). Bayesian statistics arguably provide a better foundation for probabilistic inference than null hypothesis significance testing (Kruschke, 2011; Raftery, 1995; Wagenmakers, 2007). In our implementation, Bayes factors (*B*) reflect the weight of evidence in favor of omitting a particular component from a model containing all relevant available variables. We used the default settings of the *anovaBF* function (*R*; the BayesFactor package), with the modification that 'whichModels' was set to 'top', to compare linear versions of the full model (M_f), including all main effects and interactions, with each different model in which a given experimental parameter was omitted (M_1). The anovaBF function was used with its default settings ("medium" prior scale for fixed effects, and "nuisance" prior scale for the random effect); as recommended by Rouder, Morey, Speckman, and Province (2012) to obtain Bayes factors. This family of priors was designed to be

---

[3] Except two younger participants who did a shorter version of 56 trials, in total 224 trials, due to a computer error.

invariant with respect to linear transformations of measurement units as well as general and broadly applicable (Rouder et al., 2012), and found to be more conservative than conventional ANOVAs (Rouder et al., 2009; Wetzels et al., 2011), and is commonly considered suitable for Bayesian ANOVAs in working memory research (e.g., Oberauer & Eichenberger, 2013; Rhodes et al., 2017). We specified 50,000 MCMC iterations[4], and we ran an additional 10,000 iterations until the proportional error associated with each Bayes factor was less than 5%, similar to Rhodes et al. (2016).

The *anovaBF* function quantifies the strength of evidence *B* in favor of a reduced model (M_1) relative to the comparison full model (M_f) in light of the data, returning the Bayesian likelihood ratio of M_1 and M_f. In our analyses, the output is interpreted as follows: the observed data is *B* times more likely under the reduced model (M_1) than under the full model (M_f). So, $B < 1$ indicates evidence that an omitted parameter was important, while $B > 1$ indicates evidence it was not. *B* can range from 0: indicating overwhelming support for the full model that includes the parameter (M_f), to 1: indicating equal support for both models, to infinity: providing overwhelming support for the reduced model

---

[4] Markov chain Monte Carlo iterations (MCMC) is a stochastic simulation technique commonly used to compute inferential quantities (see Green, 1995; Han & Carlin, 2001), and used to integrate the likelihood with respect to the priors on parameters to compute Bayes factors.

that omits the parameter (M_1; Dienes, 2012). By symmetry, 1/*B* provides evidence for retaining the parameter in the model.

Bayes factors cannot conclusively be interpreted using threshold cut-off points; subjective judgmental interpretation is necessary. Typically, *B* = 1 is considered ‘no evidence’, *B* between 1 and 3 is considered ‘anecdotal’ (Wetzels & Wagenmakers, 2012) or ‘not worth more than a bare mention’ (Jeffreys, 1961), *B* greater than 3 is considered ‘substantial’, between 10 and 30 ‘strong’, 30 – 100 ‘very strong’, and over 100 ‘decisive’ evidence (Jeffreys, 1961; Wetzels & Wagenmakers, 2012). Symmetrically, if *B* is less than 0.33, we may consider the evidence against including its parameter to be at least ‘substantial’. However, these labels are subjective (see Morey, 2015), so we apply them only tentatively and urge readers to evaluate the strength of evidence provided by the *B* values for themselves.

Results

We analyzed Color and Shape Recall separately. In both analyses, the full Bayesian ANOVA model included main effects of Age (young vs. old), Trial Type (single feature vs. binding) and Label-Ability (easy-to-label vs. difficult-to-label) and all possible interactions between these main effects. Recall accuracy was the dependent variable. To reduce error due to participant attentional lapses, we excluded trials with reaction times over 10 seconds from all analyses (color trials:

1.46% excluded from the younger adults, 1.44% from the older. Shape: 2.06% excluded from the young, 4.53% from the older).

In the color analysis, the main effect of trial type was obtained by comparing color-only trials with color-and-shape trials where color was probed first (i.e., the binding condition). The Color Recall accuracies for younger and older adults for more and less easy-to-label stimuli in the different conditions are presented in Table 2.1. See Appendix A for Mean values, *SD*s and Cohen's *d* for all main effects.

Table 2.1
Color accuracy (proportion correct) by age groups and experimental factors in Experiment 1.

			Mean	*SD*
Younger	Single	Difficult	.79	.10
		Easy	.92	.08
	Binding	Difficult	.69	.12
		Easy	.86	.07
Older	Single	Difficult	.58	.08
		Easy	.90	.09
	Binding	Difficult	.51	.11
		Easy	.73	.09

Note. Difficult = Difficult-to-label; Easy = Easy-to-label.

The younger participants performed better than the older (proportion correct younger*:* M = .82, SD = .27, older: M = .68, SD = .33), and our analysis indicated strong evidence in favor of retaining age group in the model (1/B =

31.59). Easy-to-label colors (*M* = .85; *SD* = .25) were remembered better than difficult-to-label (*M* = .64, *SD* = .34), and the evidence for retaining stimulus Label-Ability in the model could be considered 'decisive' (1/*B* = 2.5×10^{98}). Overall memory for color only (*M* = .80, *SD* = .28) was better than in the binding condition when Shape Recall was also required (*M* = .70, *SD* = .32), (1/*B* = 1.4×10^{21}). Overall, the evidence did not indicate an age-related binding deficit; i.e., the *difference* in memory accuracy for color only and color when shape was also remembered did not differ between the age groups (1/*B* = 0.31). Moreover, we found no evidence for an overall Trial Type × Label-Ability interaction (1/*B* = 0.25). However, we observed a larger performance drop between easy-to-label and difficult-to-label colors for older adults than younger (1/*B* = 2.3×10^{6}), which differed between single and binding trials (evidence for keeping the Age × Trial Type × Label-Ability interaction in the model: 1/*B* = 12.82), see Figure 2.2. In other words, the difference in accuracy between easy- and difficult-to-label trials was greater for the older adults than the younger. Furthermore, this interaction was modulated by trial type (i.e., single- feature or binding). The difference in differences in old and young was greater in the single-feature condition because the groups performed similarly in the easy-to-label color-only condition.

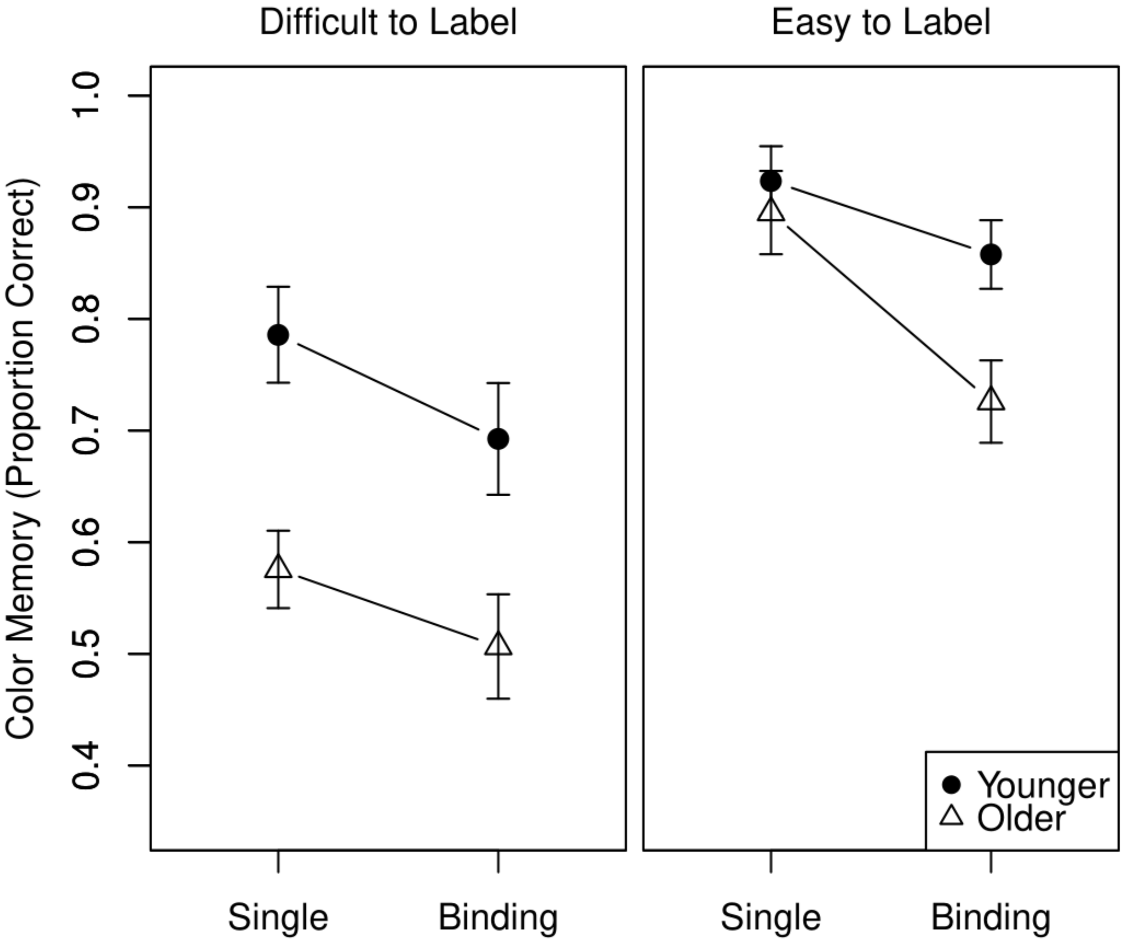


Figure 2.2. Color Memory Accuracy (Proportion correct) in Experiment 1, by Age Group, Label-Ability and Binding condition. Error bars are within-subjects 95% confidence intervals, adjusted values calculated using method from Morey (2008).

We conducted similar analyses with Shape Recall as the dependent variable (see Table 2.2 for mean accuracies across age groups and experimental conditions). We found evidence in favor of a performance difference between younger (*M* = .76, *SD* = .30) and older (*M* = .57, *SD* = .35) participants (1/*B* = 299.13), and memory for easy-to-label shapes (*M* = .74, *SD* = .31) was better than for difficult-to-label shapes (*M* = .59, *SD* = .35, 1/*B* = 3.6 × 10^{47}). There was also ‘decisive’ evidence that overall memory accuracy was better in the single-feature condition (shape only; *M* = .73, *SD* = .31), than in the binding condition (*M* = .60, *SD* = .35; 1/*B* = 5.5 × 10^{30}), but no evidence for a binding deficit in older adults (no evidence for keeping the Trial Type × Age Group interaction in the model: 1/*B* = 0.90). There was strong evidence for keeping the interaction between Label-Ability and trial type in the model (1/*B* = 24.7), but no evidence in favor of an Age-Group × Label-Ability interaction (1/*B* = 0.11), nor was this modified by trial type (evidence of retaining Age Group × Label-Ability × Trial Type interaction in the model; 1/*B* = 0.054). See Appendix A for the complete outputs of the Bayes Factor analyses.

Table 2.2

Shape accuracy (proportion correct) by age groups and experimental factors in Experiment 1.

			Mean	SD
Younger	Single	Difficult	.72	.10
		Easy	.90	.10
	Binding	Difficult	.66	.11
		Easy	.77	.09
Older	Single	Difficult	.54	.09
		Easy	.75	.11
	Binding	Difficult	.42	.13
		Easy	.56	.11

Note. Difficult = Difficult-to-label; Easy = Easy-to-label.

The results of the naming survey completed after the memory task confirmed that participants were able to name easy-to-label colors and shapes more often than difficult-to-label ones, and for named items, difficult-to-label items were described using more characters. See Appendix A for details about these analyses. We also report the proportion of ‘in-array errors’ by condition (i.e. how often participants incorrectly selected an un-probed item from the original memory array) and information about memory for shapes when colors were probed first, for all experiments, in Appendix A.

Discussion

Overall, younger adults performed better than older, and memory for easy-to-label items was better than for difficult-to-label items, both for colors and shapes. For Color Recall, our analysis indicated that older adults' performance was relatively preserved for easy-to-label colors and that the *difference* in accuracy between easy- and difficult-to-label stimuli was greater in older adults than in younger (see Figure 2.2). This was consistent with our hypothesis that older adults may depend more on sub-vocal rehearsal of verbal labels to remember colors than their younger counterparts, as suggested by Logie et al. (2015). Crucially, this older-adult 'label-ability boost' was substantially larger for the single feature condition than the binding condition, consistent with the proposal that our experimental time frame (3650 ms from item disappearance to response request) was suitable to rehearse three or four color labels, but too short to rehearse six or eight features to remember the bound objects successfully.

These observations were consistent with our proposal that use of verbal strategy can 'mask' decline in visual single-feature WM in healthy older adults quite efficiently when colors are easy to label, and the number of features is small (three to four). Hence, studies using this paradigm could find evidence for and against an age-related feature-binding deficit depending on whether the stimuli were easy or difficult to label, respectively. Indeed, we would have interpreted the significantly greater gap between memory accuracy for single- and dual-feature trials observed for the older adults compared to the younger adults in our color trials as evidence for feature-binding deficits in the older adults had we not included the difficult-to-label color condition.

However, it is possible that older adults were especially good at remembering the easy-to-label colors in the single-feature condition because the colors were familiar or easily distinguishable, rather than because they could be labeled and rehearsed. Color familiarity, label-ability, and distinctiveness are inherently entangled because as any language evolves more common, salient colors are more likely to receive linguistic labels. To test this alternative explanation and investigate the role of sub-vocal rehearsal directly, we applied articulatory suppression in Experiment 2. Suppression requires participants to repeat irrelevant phonemes aloud throughout task performance and is thought to prevent sub-vocal rehearsal (e.g., Baddeley, Lewis, & Vallar, 1984; Murray, 1965).

However, we did not find an interaction between stimulus label-ability and age group for shape memory. This attenuated support for our hypothesis that older adults rely more or more successfully on verbal strategies to remember visual stimuli, as it is unclear why older adults would benefit from verbal labels in remembering colors but not shapes. Memory failure may result from failure during encoding, maintenance during the inter-stimulus interval, or failure to respond – or some combination of these (Mitchell et al., 2000a). It is possible that the colors were more visually discriminable than the shapes during the encoding stage. Indeed, studies examining early, low-level visual processing in younger adults have suggested that color is salient in the pre-recognition stage (Callaghan, 1984; Cavanagh, 1987; Troscianko & Harris, 1988), and that colored objects are recognized faster than monochrome objects (Humphrey, Goodale, Jakobson, & Servos; 1994, Wurm, Legge, Isenberg, & Luebker, 1993). It is therefore possible that presentation time in this experiment (1500 ms) was insufficient to allow older adults in particular to encode and label the shapes in a way that facilitated recall, thus preventing them from benefitting from sub-vocal rehearsal during the maintenance phase. We addressed this possibility in Experiment 3.

Experiment 2

To investigate whether the observed age differences for color memory in Experiment 1 were due to the sub-vocal rehearsal of verbal labels, in Experiment 2 we applied suppression to half the trials, using the reconstruction paradigm of Experiment 1. We hypothesized that if reliance on sub-vocal rehearsal enabled the older adults' memory 'boost' for easy-to-label colors in the single-feature condition, suppression should reduce performance in that condition but not affect performance in the other conditions, especially not for the difficult-to-label colors, where we do not expect performance to rely on sub-vocal rehearsal.

Alternatively, if the boost was not due to sub-vocal rehearsal, but because the easy-to-label, common colors were more salient – or more clearly and accessibly encoded in long-term memory – the older adults' performance boost should still be observed during suppression. Memory for color was the only dependent variable in this experiment.

Methods

Participants. We recruited 52 new participants in the same way as in Experiment 1, specifying that they not have taken part in that earlier experiment. They included 25 younger adults (8 female) aged 18 – 25 (M = 21.5, SD = 2.0) years, and 27 older adults (10 male), aged 63 – 76 (M = 70.0, SD = 4.7) years. Two older adults were excluded due to color vision error scores over the cut-off point of 5 errors (6 and 13 errors, respectively). Years of education did not differ significantly between the groups (older adults: M = 14.8, SD = 3.6, younger adults: M = 15.5, SD = 1.7 years, t(33.8) = .92, p = .37, d = 0.25). Providing years of education was optional, N = 20 for the younger adults, N = 24 for the older. All older adults scored above the recommended cut-off score indicating potential cognitive impairment of 25 on the ACE-III mini-score (Hsieh et al., 2015; M = 28.7, SD = 1.1), completed at the very end of the testing session NART-predicted verbal IQ scores were significantly higher in the older group (M = 123.7, SD = 3.3) than the younger group (M = 115.7, SD = 3.7), t(48) = 8.15, p <.001, d = 0.53. The study was approved by our local Ethics Committee.

Stimulus and Apparatus. The stimuli were identical to those used in Experiment 1, displayed using the same equipment.

Design and Procedure. The procedure was identical to that used in Experiment 1, with the following modifications. The four blocks were: 1. Color-only, 2. Color-only with Suppression, 3. Binding, 4. Binding with Suppression. Since the Suppression manipulation adds time and can be tiring, the total number of trials was reduced to 60 trials per block, resulting in a total of 240 trials. Suppression started prior to the encoding of items and continued throughout the encoding and testing phases, to minimize the use of verbal strategies as much as possible. Participants initiated each trial by pressing the space bar. We instructed participants that at the start of each suppression trial two randomly generated digits would be displayed in the center of the screen for 2 seconds. For example, if '1 – 2' appeared, participants repeated 'one, two'. We instructed participants to start repeating these two numbers aloud immediately at a rate slightly faster than one digit per second, and to continue to repeat it during a blank screen for 2 seconds, to minimize potential interference created by initiating the suppression while encoding the memory items. Participants were instructed to continue suppression until they had responded. The experimenter was present to make sure suppression was sustained. No participant was reminded to maintain suppression more than three times throughout the session.

Results

We used a Bayes Factor ANOVA model-comparison analysis similar to that in Experiment 1, with Color Recall as the dependent variable, and suppression included as another factor in addition to Age, Trial Type, and Label-Ability. As in Experiment 1 we excluded trials with reaction times over 10 seconds (Color: 0.41 % of trials excluded from the younger adults, 0.50 % from the older. Shape: 1.02 % trials excluded from the young, 0.70 % from the older). Mean Color Recall accuracies for younger and older adults in the different conditions are presented in Table 2.3.

Table 2.3

Color accuracy (proportion correct) by age groups and experimental factors in Experiment 2.

With Suppression				
			Mean	SD
Younger	Single	Difficult	.80	.10
		Easy	.92	.08
	Binding	Difficult	.73	.12
		Easy	.84	.06
Older	Single	Difficult	.60	.10
		Easy	.77	.12
	Binding	Difficult	.50	.09
		Easy	.71	.09
Without Suppression				
Younger	Single	Difficult	.79	.08
		Easy	.95	.07
	Binding	Difficult	.69	.11
		Easy	.86	.07
Older	Single	Difficult	.57	.13
		Easy	.91	.12
	Binding	Difficult	.48	.09
		Easy	.78	.09

Note. Difficult = Difficult-to-label; Easy = Easy-to-label.

The Bayes factor analysis indicated 'decisive' evidence for retaining age in the model ($1/B = 1.0 \times 10^3$), as the younger adults (proportion correct: $M = .83$, $SD = .27$) performed better than the older adults ($M = .67$, $SD = .33$). Overall, easy-to-label colors ($M = .84$, $SD = .26$) were better recalled than difficult-to-label ($M = .65$, $SD = .38$; $1/B = 3.2 \times 10^{150}$). Color-only memory ($M = .79$, $SD = .29$) was better than memory for color when bound to shapes ($M = .70$, $SD = .32$; $1/B = 2.7 \times 10^{29}$), and Suppression had less than a moderate effect on recall overall ($1/B = 2.55$; without Suppression; $M = .76$, $SD = .30$, with Suppression; $M = .73$, $SD = .31$). We found no evidence that that Suppression affected the age groups differently overall (Suppression × Age Group; $1/B = 0.62$), ruling out the alternative explanation that older adults performance was generally more adversely affected by Suppression. Younger-adult performance in the easiest condition (no suppression, easy-to-label, single feature) might indicate near-ceiling performance by the majority (mean accuracy of .95). However, there was 'decisive' evidence for retaining the interaction between Suppression and Label-Ability ($1/B = 4.2 \times 10^6$), such that Suppression reduced performance for easy-to-label colors, but had little effect on difficult-to-label colors (see Figure. 2.3). This fits with previous research suggesting that Suppression does not impair memory for difficult-to-label, abstract, visual stimuli (Luria, Sessa, Gotler, Jolicoeur & Dell'Acqua, 2010; Morey & Cowan, 2004; 2005; Sense, Morey, Heathcote, Prince & Morey, 2017).

These results did not strongly replicate results of Experiment 1, where older adults exhibited a feature-binding deficit for easy-to-label colors, but not for difficult-to-label colors. Here, there was no evidence for retaining Age Group × Trial Type × Label-Ability; $1/B = 0.041$, and anecdotal evidence against including the four-way interaction of Age Group × Trial Type × Label-Ability × Suppression ($1/B = 0.51$).

However, we replicated decisive evidence for the comparatively larger performance drop between easy-to-label and difficult-to-label colors observed for older adults in Experiment 1 ($1/B = 8.5 \times 10^{11}$). There was some evidence suggesting that Suppression modulated this effect; evidence for retaining the Age Group × Suppression × Label-Ability interaction ($1/B = 2.96$), (see Figure 2.3).

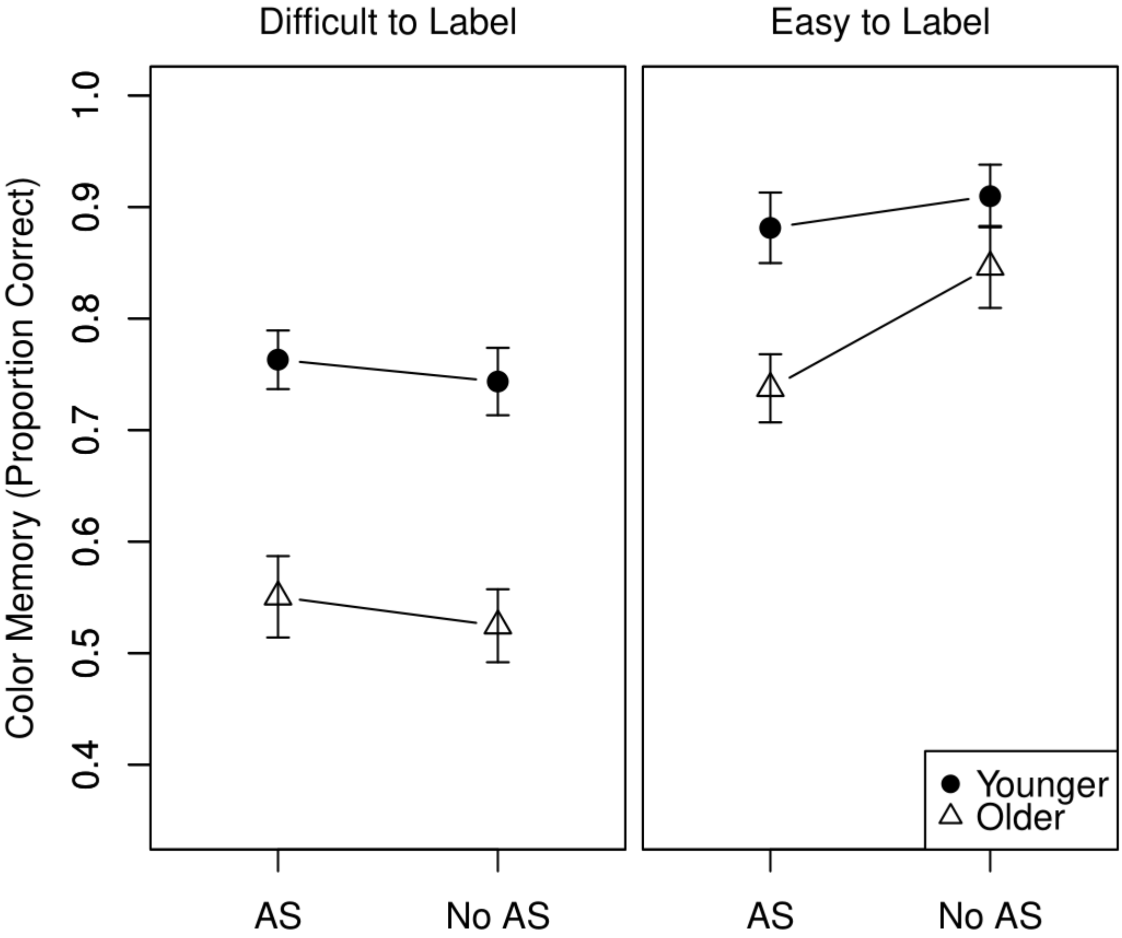


Figure 2.3. Color Memory Accuracy (Proportion correct) in Experiment 2, by Age Group, Label-Ability and AS (Articulatory Suppression). Error bars are within-subjects 95% confidence intervals, adjusted values calculated using method from Morey (2008).

However, the evidence regarding this differential impact of Suppression was rather weak: although BFs close to 3 are often interpreted as 'substantial' (Jeffreys, 1961) this practice is problematic (see Morey, 2015). To follow up on this inconclusive three-way interaction, we looked at the data without Suppression, where evidence for retaining the Age Group × Label-ability interaction was 'decisive' ($1/B = 1.7 \times 10^{11}$). In comparison, for trials with Suppression, it was comparatively smaller (Age Group × Label-Ability: $1/B = 7.7$). This suggests that Suppression did impair the older adults comparatively more for the easy-to-label colors, suggesting that their memory performance was more reliant on sub-vocal rehearsal. However, given the rather weak evidence for retaining the three-way interaction - and potential concerns about near-ceiling performance by younger adults in the easiest condition - this requires replication with a larger sample. See Appendix A for the complete analysis output.

Discussion

We replicated the key findings from Experiment 1: Younger adults performed better than older adults overall, and easy-to-label colors were better-remembered than difficult-to-label colors, overall. Importantly, we again found 'decisive' evidence ($1/B = 8.53 \times 10^{10}$) that the performance drop between easy- and difficult-to-label colors was larger for the older adults. Crucially, while Suppression had a strong negative effect on memory for the easy-to-label colors it did not impair performance for the difficult-to-label colors in either age group. This provided an important manipulation check, because it suggested that participants used verbal labels for our intended easy-to-label colors but not the difficult-to-label ones.

In contrast to Experiment 1, these results did not strongly support our hypothesis that older adults' reliance on verbal rehearsal produces the appearance of feature-binding deficits. *B* for including the four-way interaction (Age Group × Trial Type × Label-Ability × Suppression) was close to 1. Hence, our data did not provide strong evidence either for accepting or rejecting our prediction that older adults would be comparatively more impaired for single-feature easy-to-label colors under suppression. Thus, replication with more participants or trials is required to test this hypothesis adequately. However, the older adults in the second experiment performed a bit better in the binding condition than those in the first one. Perhaps suppression made verbal rehearsal more salient so that older adults attempted to apply it in the binding as well as the color-only condition, despite our prediction that rehearsal would be less effective for the bound condition. This merits further investigation in future studies.

However, we found that older-adult performance differed much more between difficult-to-label and easy-to-label colors than did younger-adult performance, replicating the strong evidence observed in Experiment 1. This effect was a lot weaker during concurrent suppression, but not completely abolished. Thus, our hypothesis that the older adults' benefit for easy-to-label colors was due to sub-vocal rehearsal was partially supported (see Figure 2.3). These results indicate that older adults appeared to benefit greatly from easy-to-label colors. Sub-vocal rehearsal appeared to play a role in driving this benefit, but may not be the only explanation, suggesting that other aspects of color 'commonness' may also benefit older adults, or that suppression does not completely prevent rehearsal. However, these results failed to provide clear indications regarding whether this influenced performance on single or binding trials differently in the two age groups.

Experiment 3

Following two experiments providing decisive evidence for a comparatively larger performance gap between easy-to-label and difficult-to-label colors in older adults, we investigated why a similar effect was not observed for shapes in Experiment 1. In Experiment 3, we examined the possibility that the shapes might have been presented too briefly for older participants to generate the shape labels, in line with evidence that reduced processing speed contributes to older adults' reduced visual WM capacity (Brown, Brockmole, Gow, & Deary, 2012), especially when required to remember multiple objects (Guest, Howard, Brown, & Gleeson, 2015). If so, longer encoding time should result in similar observations for shapes to those observed for colors in Experiments 1 and 2. We included Suppression to test whether it would eliminate any effects of Label-Ability, thus suggesting they might be due to sub-vocal rehearsal. Hence Experiment 3 was identical to Experiment 2, except that Shape Recall was the dependent variable and the stimulus presentation time was 2500 ms for older adults (this encoding duration did not result in ceiling performance for single shape memory in older adults at set size three; Rhodes, Parra, & Logie, 2015). Stimulus presentation was 1500 ms for younger adults to avoid ceiling-level performance.

Methods

Participants. We recruited 52 new participants. Two older adults were excluded (one due to not understanding the task, one due to color-vision deficiency; 8 errors). The final sample consisted of 25 younger adults (8 male) aged 18 – 24 (*M* = 20.1, *SD* = 1.8) years, and 25 older adults (8 male), aged 62 – 77 (*M* = 69.4, *SD* = 4.7) years. The older adults had *M* = 15.8, *SD* = 3.8 years of education, the younger adults *M* = 14.7*, SD* = 1.8 years, which did not differ significantly between the groups *t*(30.26) = 1.24, *p = .*23. Three younger and three older adults did not provide years of education. All older adults completed the ACE-III mini at the very end of the testing session, and no older adult was excluded based on the recommended cut-off score of 25 (Hsieh et al., 2015; *M* = 28.8, *SD* = 1.6). NART-predicted verbal IQ scores were significantly higher in the older group (*M* = 123.4, *SD* = 3.5) than the younger group (*M* = 114.5, *SD* = 3.5), *t*(48) = 8.89, *p* <.001, *d* = .59. The study was approved by our local Ethics Committee.

Stimulus and Apparatus. The stimuli were identical to those used in Experiments 1 and 2, displayed using the same equipment.

Design and Procedure. The procedure was identical to that used in Experiment 2, except that presentation time for older adults was 2500 ms but 1500 ms for younger adults. No participant was reminded to maintain suppression more than twice throughout the session, thus rendering it unlikely to have impacted performance. As in the previous experiment, there were four blocks of 58 trials (total 232 trials): 1. Shape-only, 2. Shape-only with Suppression, 3. Binding, 4. Binding with Suppression.

Results

We used the same Bayes Factor ANOVA model-comparison analysis as in Experiment 2, with Shape Recall as the dependent variable, and Age, Trial Type, Label-Ability and Suppression as the factors of interest. We excluded trials with reaction times over 10 seconds (Color: 0.19% of trials excluded from the younger adults, 0.19% from the old. Shape: 0.33% trials excluded from the young, 1.22% from the old). Mean Shape Recall accuracies for younger and older in the different conditions are presented in Table 2.4.

Table 2.4

Shape accuracy (proportion correct) by age groups and experimental factors in Experiment 3.

With Suppression				
			Mean	*SD*
Younger	Single	Difficult	.64	.11
		Easy	.85	.09
	Binding	Difficult	.61	.10
		Easy	.76	.09
Older	Single	Difficult	.60	.13
		Easy	.77	.10
	Binding	Difficult	.50	.12
		Easy	.62	.08
Without Suppression				
Younger	Single	Difficult	.68	.11
		Easy	.88	.10
	Binding	Difficult	.60	.13
		Easy	.81	.09
Older	Single	Difficult	.62	.09
		Easy	.81	.12
	Binding	Difficult	.52	.12
		Easy	.61	.12

Note. Difficult = Difficult-to-label; Easy = Easy-to-label.

Younger adults performed slightly better (*M* = .73, *SD* = .31) than older adults (*M* = .63, *SD* = .34; 1/*B* = 1.67), despite older adults' longer stimulus presentation time. We observed strong effects of binding condition, such that shape only (*M* = .73, *SD* = .31) was better than shape when also remembering color (*M* = .63, *SD* = .34; 1/*B* = 1.0 × 10^{32}) and Label-Ability (easy-to-label shapes: *M* = .76, *SD* = .30, difficult-to-label shapes: *M* = .60, *SD* = .35; 1/*B* = 1.5 × 10^{88}), but not of Suppression (without Suppression: *M* = .69, *SD* = .33, with Suppression: *M* = .67, SD = .33; 1/*B* = 0.54). Unexpectedly, there was no evidence for interaction between Label-Ability and Suppression (1/*B* = .03). A large age-related binding deficit appeared (1/*B* = 399.69), but we found no clear evidence that this age-related binding deficit was modified by Label-Ability (1/*B* = 0.18), or by Suppression (1/*B* = 0.05). However, there was evidence for overall interaction between age group and Label-Ability (1/*B* = 6.32).

These interactions should be interpreted with caution because younger and older participants performed slightly different tasks due to the difference in encoding time. Therefore, to follow up on the effects of Label-Ability and binding in the different age groups, we also analyzed them separately. For the younger adults, the effect of binding did not differ with Label-Ability (1/*B* = 0.05), but for older adults, there was an interaction between label-ability and binding (1/*B* = 4.31), such that there was a larger binding deficit for the easy-to-label shapes (similar to Experiment 1). Hence, by extending the stimulus presentation time for older adults, we produced an apparent feature-binding deficit for older adults for easy-to-label shapes but not for difficult-to-label shapes (see Figure 2.4), similar to the results for color memory in Experiment 1. However, Suppression did not appear to reduce this binding deficit in the older adults (Suppression × Trial Type × Label-Ability: $1/B = 0.10$), undermining the inference that verbal rehearsal was the reason for the relatively good performance when shapes were easy to label. See Appendix A for the complete analysis output.

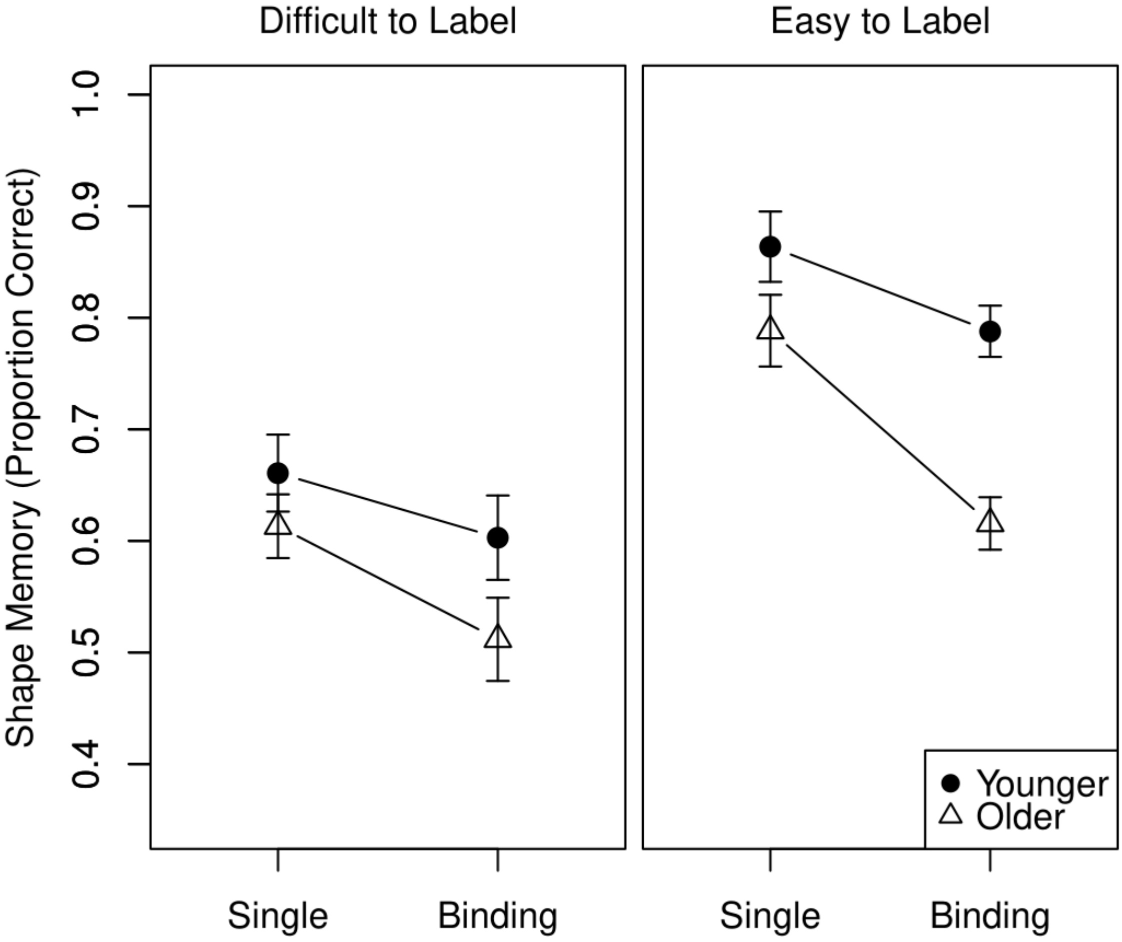


Figure 2.4. Shape Memory Accuracy (proportion correct) in Experiment 3, by Age Group, Label-Ability and Binding Condition. Error bars are within-subjects 95% confidence intervals, adjusted values calculated using method from Morey (2008).

Discussion

These results were consistent with the idea that, given sufficient encoding time, older-adults' recall for single, easy-to-label shapes can benefit, creating an apparent binding deficit for such trials compared to younger adults (similar to results for color memory in Experiment 1). This indicated that this effect was not color-specific. However, curiously, suppression did not modify this effect for shapes, or affect either older or younger adults' recall, regardless of stimulus label-ability. This implied that the relatively good recall for easy-to-label shapes was not due to sub-vocal rehearsal of shape labels, unless the extended encoding time for older adults enabled some verbal labeling despite concurrent suppression. Hence, although the results of Experiment 2 suggested that sub-vocal rehearsal contributed to the older adults' better recall for label-able colors, the results of Experiment 3 indicated that rehearsal did not play a similar role for shape memory.

Performance for label-able stimuli was better than that for less-label-able stimuli, both with and without suppression in both experiments. This suggested that either some sub-vocal rehearsal was still possible despite suppression, or that label-able features were better remembered for other reasons. Older adults may have been able to benefit comparatively more from easy-to-label single shapes than younger adults when given sufficient time to encode them. Instead of being driven by sub-vocal label rehearsal, some other feature of those easy-to-label, common shapes – perhaps that they were more familiar and/or easier to process – seems to have driven this benefit. Even though Brockmole, Parra, Della Sala, and Logie (2008) found no differences in discriminability of the difficult-to-label shapes (identical to those in our experiments) in younger and older adults in a preliminary search task, it is likely that the difficult-to-label shapes were more visually complex than the easy-to-label ones, which might for instance increase visual search rate (Alvarez & Cavanagh, 2004; Eng, Chen, & Jiang, 2005). Therefore, differences between easy- and difficult-to-label shapes in our study may have been driven by perceptual differences.

General Discussion

We observed that while, overall, participants of both age groups remembered label-able stimuli better than less label-able, impairment for difficult-to-label stimuli appeared to be exacerbated in older adults. This was replicated in two experiments for color memory but only observed for single-item shape memory when older adults were given extra time to encode items. Hence, while visual working memory was impaired with age, something about the label-able colors allowed older adults to overcome this deficit partially, and perform more similarly to younger adults. These results were in line with recent findings suggesting that older adults can compensate for visuospatial declines by using strategies during encoding (e.g., Siegel & Castel, 2018). Our results also fit with the Scaffolding Theory of Aging (Park & Reuter-Lorenz, 2009), suggesting that older adults may employ compensatory recruitment such as relying on more active strategies that draw in other brain regions to compensate for deteriorating visual memory with age. This compensatory recruitment is not necessarily associated with improved visual task performance (Reuter-Lorenz, Stanczak, & Miller, 1999). However, in our study we observed substantial performance differences in an otherwise identical visual task simply by manipulating how easily colors could be labeled, indicating that in some circumstances, older adults can successfully compensate for declining visuospatial memory.

However, it is not clear by which mechanism the compensation we observed occurred. Our original hypothesis was that relatively better performance for easy-to-label items in older adults would be due to greater reliance on verbal encoding and rehearsal of labels for visual stimuli, considering that verbal WM is comparatively intact with age (see Jonides et al., 1996 for a review). The results of Experiment 2 (suppression impaired older adults' performance for easy-to-label colors, but had no effect on difficult-to-label colors, see Figure. 2.3) supported the hypothesis that compensation depended at least partly on strategic sub-vocal rehearsal. However, there was no effect of suppression on easy-to-label shape memory in Experiment 3. It is possible that suppression blocked some verbal rehearsal but not all. Even in studies with verbally presented material, suppression does not completely demolish all memory traces, which might explain why we did not observe a stronger effect. For shapes, the lack of effect of suppression in the older adults might be explained by the extended stimulus presentation (2500 ms), which may have provided sufficient time for some labeling despite suppression.

In all three experiments, older adults had significantly higher predicted verbal IQ scores, measured using the NART (National Adult Reading Test), in line with research suggesting relatively intact verbal abilities with age but also potentially indicating higher peak-adult general cognitive abilities. These higher predicted verbal IQ scores could also suggest that the older adults may not have been as representative of the general population as the younger adults. Despite similar levels of reported education in younger and older adults, reaching that level of education was quite a bit less common in the older adults' generation, reinforcing this possibility. Moreover, we analyzed naming data and found that older adults were better able to provide names for the difficult-to-label colors and shapes than the younger adults (see Appendix A). Hence, we can rule out the important potential alternative explanation for older adults' relative impairment for difficult-to-label colors: that they were less able to label uncommon stimuli than younger adults. However, their higher verbal IQ scores introduced another potential explanation for their overall performance.

However, based on the unclear evidence regarding the role of sub-vocal rehearsal we also consider additional processes which may contribute to this strong boost in memory for easy-to-label features in older adults. First, it is possible that our set of easy-to-label colors was better remembered by older adults because such colors are easier to distinguish during encoding, as they can be readily and even possibly automatically categorized. For instance, some case-study evidence has suggested that maintenance of color categories does not rely on language. For example, Haslam, Wills, Haslam, Kay, Baron, and McNab (2007) described a patient with semantic dementia who was able to categorize different colors consistently, despite a near-complete loss of color language. Categorizing colors may depend on basic perceptual features, separate from verbal labels (Haslam et al., 2007). The universality of the basic color categories in most human languages also supports this. Although an effort was made to ensure the average luminance of the two color sets was relatively similar (see Appendix A) other perceptual differences between our easy and difficult-to-label colors cannot be ruled out. However, the large performance reduction associated with suppression specifically for easy-to-label colors in Experiment 2 (regardless of age group), was consistent with the idea that verbal rehearsal does play a role, and that perceptual differences do not provide the whole explanation.

Second, visual memory for difficult-to-label information may be limited because it affords less opportunity to activate long-term memory representations. Souza and Skóra (2017) found that labeling colors (overtly or sub-vocally) improved younger-adult memory performance by activating long-term memory representations, rather than simply by adding verbal memory traces. If older adults' memory for easy-to-label colors was also comparatively 'boosted' because such colors automatically activated lexical, semantic representations that helped compensate visual memory, this would explain why this special benefit for easy-to-label colors in older adults was not completely abolished under suppression. For instance, while suppression strongly disrupts memory for individual letters – thought to rely on a phonological code – in younger adults (Chein & Fiez, 2010, Toppino & Pisegna, 2005), it does not seem to disrupt memory for words (Souza & Oberauer, 2018). Olsson and Poom (2005) proposed that 'pure' visual memory relies on what can be held in the focus of attention (central to some models of WM; see Cowan et al., 2005; Oberauer, 2013), after observing that memory was much poorer for objects that did not belong clearly to different categories, such as ovals with varying aspect ratios and color mixtures along the natural boundaries of established color labels, compared to easy-to-categorize objects such as a red square. While their participants could remember a mean of 2.5 easily categorized objects, for objects without a clear category they had an average memory capacity of one item. They suggested that if an initially-presented object does not belong to a clear category and attention is directed to a new object, the initially-presented one is overwritten, leaving memory capacity of one item. Similar perceptual 'overwriting' processes have been observed in other studies (Enns & Di Lollo, 2000; Lakha & Wright, 2004, see also Logie, 1995; Phillips & Christie, 1977; Wilson, Scott, & Power, 1987). Furthermore, complex recognizable items were associated both with better memory precision and appeared supported by a richer range of neural representations than unrecognizable objects*,* suggesting that recognizable objects evoked richer and more variable contextual associations (Veldsman, Mitchell, & Cusack, 2017).

In our study, the larger age-related deficit for difficult-to-label colors was consistent with the established finding that 'pure' visual WM is impaired in old age (e.g., Johnson et al., 2010), but indicated that older adults can benefit significantly when opportunities for either verbal or semantic representations are available (easy-to-label colors). This supported theories about compensatory memory strategies in older adults (Logie, 2018; Park & Reuter-Lorenz, 2009), and might have implications for numerous memory phenomena. For instance, age-related dual-task deficits (see Jaroslawska & Rhodes, 2018 for a review) may be greater if the secondary task disrupts such strategies. These results also added to evidence that older adults generally benefit from opportunity for elaboration, i.e., strengthening memory traces by adding more information (Kitagami, 2000; Osaka, Otsuka, & Osaka, 2012), and benefit proportionally more than younger adults when to-be-remembered information is consistent with past experience (Hess & Slaughter, 1990). Since we used a limited stimulus set in the present study it is possible that participants were able to generate a LTM entry for difficult-to-label stimuli throughout the study – future studies should consider this possibility. We now discuss the implications of using easy- or difficult-to-label stimuli when investigating age-related feature-binding deficits.

In Experiment 1, when colors were easy to label, older adults' performance was comparable to that of younger adults in the single-feature condition, suggesting they may have successfully used sub-vocal rehearsal to compensate for age-related declines in visual memory. Because their performance was poorer in the binding condition, this appeared as a visual feature-binding deficit. Hence, it is possible that some previous findings of age-related binding deficits observed using reconstruction paradigms may be explained similarly – perhaps deficits would not have been found if verbal strategies had been prevented. Specifically, conditions in several studies where age-related binding deficits were observed were similar to ours; they used easy-to-label stimuli, and did not impede verbal strategies (e.g., Brockmole & Logie, 2013; see Table 2.1 for an overview).

In the present study, the younger adults in Exp. 2 performed at near-ceiling level in the easy-to-label, color-only condition (proportion correct: .95). Near-ceiling performance might explain the Age Group × Label-ability interaction (i.e., why the difference between easy- and difficult-to-label colors was smaller for younger than for older adults). However, this interaction was also observed in Exp. 1 (where younger adults performed further from ceiling;

.92), suggesting that near-ceiling performance alone did not cause this interaction. Younger adults performing close to ceiling (over .95 proportion correct) in some conditions has been observed both in studies where older adults were found to have a binding deficit (e.g., Brown & Brockmole, 2010, Exp. 2; Chalfonte & Johnson, 1996), and where they did not (e.g., Brown & Brockmole, 2010, Exp. 1; Rhodes et al., 2016). Nevertheless, the role of ceiling effects in younger adults should be considered when measuring feature-binding deficits.

We initially hypothesized better performance in the single feature condition because three single features can be verbalized twice as fast as three bound features. However, the alternative explanation that older adults' performance benefits from easy-to-label colors because they enable activation of semantic representations may produce a similar pattern of results. While common features like 'green' (or 'circle'), should have accessible semantic repetitions, arbitrary *combinations* of common features in the binding condition, e.g., 'green circle' would likely have less accessible, or at least less rapid and/or routinely familiar semantic representations.

The results of our three experiments suggested that older adults' better performance for easy-to-label stimuli was because such stimuli enabled both sub-vocal rehearsal and activation of semantic representations. Particularly, puzzling differences between how the opportunity to label influenced memory for color and shapes, respectively, suggested that the type of feature measured might also play a complex role. Visual complexity differences between easy- and difficult-to-name shapes may have been a confounding factor. Still, our findings strongly indicated that reconstruction paradigms with easy-to-label stimuli should be used with caution, as they may introduce strategy-related confounds between age groups. For instance, this might explain discrepancies in recently developed delayed-estimation precision paradigms, where an age-related mis-binding deficit was observed for colored bars (Peich, Husain, & Bays, 2013), but not for the location of complex fractals (Pertzov, Heider, Liang, & Husain, 2015). Such paradigms quantify feature-binding deficits via mis-binding errors, i.e., incorrectly reporting a feature that was part of the memory array, but was not the target item. In the present study, we did not analyze the proportion of in-array errors due to the small number of errors in some conditions. However, it appeared that more such errors were made for easy- than for difficult-to-label colors, while no clear pattern appeared for shapes (see Appendix A). This suggests that verbalization might also contribute to mis-binding errors.

Different types of binding tasks may draw on different cognitive processes (e.g., Delvenne & Bruyer, 2004; Ecker et al., 2013; Shimi & Logie, 2018; Wheeler & Treisman, 2002). Some include location as a feature, others as a cue to probe items, others simply as what inherently binds features into an object (see Kovacs & Harris, 2019 for discussion on how separate visual features become integrated objects via a mutual location). Older adults may struggle specifically with binding items to locations, more than remembering which items were present (Thomas, Bonura, Taylor, & Brunyé, 2012). This might exacerbate binding deficits in some paradigms. The extent to which change detection, reconstruction and delayed estimation (misbinding) paradigms measure the same binding process is controversial – and beyond the scope of this paper. However, all visual *conjunctive bindings* (i.e., surface feature bindings, such as shape and color) are by definition created by features coinciding in space (but see studies comparing such conjunctive bindings with *relational bindings,* i.e., a shape and a colored blob presented side by side but joined by an arrow. Van Geldorp, Parra, and Kessels, 2015 found that healthy aging affected both these types of binding similarly). While we only used one type of paradigm here, our results added to this debate by highlighting that some discrepancies in the literature may depend on whether stimuli in the single-feature condition allow and/or facilitate activation of additional verbal and/or semantic representations. The usefulness of such representations may vary depending on how feature memory is probed, and it is unclear whether similar effects would be observed using change detection or other paradigms. For instance, Sense, Morey, Prince, Heathcote, and Morey (2017) observed that verbalization did not improve visual change detection performance in younger adults.

Simple feature-binding tests may help distinguish healthy from pathological aging, although there is some debate regarding about which type of paradigm to use (see Liang et al., 2016; Parra et al., 2016). Pathological aging is identified by comparing patient binding deficits to those observed in healthy aging, and it is therefore important that paradigms not produce ‘false’ visual binding deficits in the healthy comparison group. Correspondingly, pathological binding-deficits observed in paradigms with easy-to-label, common stimuli could reflect greater patient reliance on either verbal rehearsal or semantic representations boosting visual single-feature performance – processes which are less efficient in the binding condition. This should be carefully distinguished from visual feature-binding memory deficits.

Taken together, our results suggested that to ensure that participants groups do not use different cognitive mechanisms in different experimental conditions (i.e., single-feature and binding conditions), using difficult-to-label, uncommon items is a more reliable approach than blocking verbal rehearsal with suppression. Unexpectedly, participants appeared more likely to sub-vocally rehearse color than shape stimuli. These results highlighted the importance of considering differences in spontaneous strategy use not just between younger and older adults, but also among the specific stimuli used. Experimental manipulations of stimulus label-ability and suppression may help detect patterns in such differences.

Conclusion

Our results fit with previous evidence indicating that visual memory for non-categorical information is very limited (e.g., to only one item; *Olsson & Poom, 2005*), likely because it is more difficult to engage other systems, such as verbal WM or long-term memory*.* Crucially, our results showed that older participants were more adversely affected by difficult-to-label colors, indicating an increased reliance on rehearsal of verbal labels or other types of elaboration to maintain visually presented stimuli with age. This possibility should be considered when designing future memory studies because age-related visual memory decline may be more accurately captured when such elaboration is prevented. This has interesting implications for the wider visual WM literature beyond feature-binding, as it suggests that comparing younger- and older-adult task performance may not straightforwardly reveal the age-related decline in visual WM*,* but instead applications of different strategies that tap different cognitive mechanisms, to varying degrees (for a discussion see Logie, 2018).

We found some evidence that differential application of verbal strategies or opportunity for activation of semantic representations might account for some of the literature's inconsistencies regarding age-related feature-binding deficits. This highlighted that observing binding 'deficits', depending as it does on statistical differences in differences, may also depend on many experimental and sampling parameters. These include participant tendency and ability to label stimuli and procedural availability to rehearse labels, which may transact in creating required statistical interactions (i.e., binding deficits). Identifying and understanding the roles of procedure and participant characteristics in this could also be useful in establishing appropriate experimental paradigms for using healthy older adults as controls to identify symptoms of pathological aging (e.g., Parra et al., 2009a).

More broadly, our results highlighted how stimuli that can easily be labeled or categorized may have qualitatively different effects on participants of different age groups, and stressed the importance of considering the interplay between visual and verbal memory (Souza & Skóra, 2017), and the importance of considering alternative strategies that may be used for performing the same task (Logie, 2018) both of which appear crucial for understanding how older adults perform every day – as well as experimental – cognitive tasks.

Chapter 3: Change-Detection

Aims

As illustrated in Table 1, several different paradigms have been used to measure age-related visual feature-binding deficits. In Chapter 2, I measured such deficits using a reconstruction paradigm (e.g., Pertzov, Dong, Peich, & Husain, 2012; Ueno, Mate, Allen, Hitch, & Baddeley, 2011; Brockmole & Logie, 2013). However, many studies on feature-binding deficits have employed change-detection paradigms (e.g., Brown, Niven, Logie, Rhodes, & Allen, 2017; Rhodes et al., 2017; Rhodes et al., 2016). Indeed, use of different paradigms may also explain discrepancies in the literature.

In the change-detection paradigm, participants need to remember a set of items - typically shapes of different colours (e.g., a red circle and a blue square) - and indicate if a subsequent probe item matches one of the original items. Probe items are *different* if features have recombined (e.g., a red square), or the *same* when identical to one of the original items (e.g., a red circle). Therefore, remembering all the individual features without remembering which shape was of which colour (the *bindings*) would result in chance memory performance. Different types of binding tasks may draw on different cognitive processes (e.g., see Delvenne & Bruyer, 2004; Ecker et al., 2013; Shimi & Logie, 2018; Wheeler & Treisman, 2002). This notion is supported by studies comparing the effect of repeatedly presenting identical memory arrays in change-detection or reconstruction binding tasks, in younger adults. Logie, Brockmole, and Vandenbroucke (2009) found that participants benefitted much less from the repeated trials in the change-detection task, while such repetition greatly improved performance across trials in the reconstruction paradigm (see also Shimi & Logie, 2018). This led to the suggestion that change-detection involved visual cache memory (Logie, 1995; 2003; 2011; Logie et al., 2009, see Section 1.2.1.2), which retains feature-bindings only for the duration of a trial, while producing little or no long-term episodic memory traces. In contrast, memory performance in the reconstruction paradigm appeared less reliant on the visual cache, as information in one trial ‘carried over' to subsequent trials. This indicates that the act of reproducing the memory item (in the reconstruction paradigms) involves a different mechanism than simply indicating whether the probed feature is different from that in the memory array (in change-detection).

These paradigm differences are important in relation to our research question regarding verbal strategy use by older adults in visual feature-binding tasks, as they indicate that participants may approach change-detection and reconstruction paradigms differently. Hence, change-detection and reconstruction paradigms may also differ in the extent to which verbalisation tends to be used and/or influence performance. For instance, Sense et al. (2017) found that opportunity for verbalisation did not improve visual change-detection performance in younger adults. However, even though numerous studies on feature-binding deficits in healthy (Brown et al., 2017; Rhodes et al., 2017; Rhodes et al., 2016) as well as pathological ageing (e.g., Parra et al., 2011) have used change-detection, the role of verbal rehearsal by older adults in this paradigm had not been tested directly.

The aim of this experiment was therefore to test if a similar stimulus label-ability and age group interaction (as observed in Exp. 1) would be found in this other commonly used feature-binding paradigm; i.e., change-detection. More detailed comparisons between change-detection and reconstruction measures of feature-binding would be useful to understand age-related decline and discrepancies in findings. However, in this thesis, I focused on the potential use of verbal rehearsal in these respective paradigms, and did not test other differences between the paradigms, or explore other differences in the cognitive mechanisms used in the respective tasks explicitly. Similarly, I did not attempt to discern which approach 'better' measures feature-binding.

The aim of this study was to test if stimulus label-ability affected memory performance differently in younger and older adults, and if so, if this would differ between single-feature and binding trials. I used the same easy- and difficult-to-label items as in Experiments 1, 2 and 3, to explore the potential role of verbal strategies by older adults.

Introduction

In Experiment 4, I investigated the role of stimulus label-ability on memory accuracy using a change-detection paradigm similar to that of Rhodes et al. (2015) in which a single probe was compared with an item in the stimulus array. Using a single probe without the presence of un-probed items is recommended for addressing potential feature binding deficits in healthy aging (Read et al., 2016; Rhodes et al., 2017). If participants use verbal codes to boost memory in a change-detection paradigm, we would expect better performance for the easy-to-label items. If older adults rely more on this strategy, they should perform similarly to younger adults in single-feature conditions for common colours, but exhibit a feature-binding deficit because rehearsing the labels of all eight features should not be as useful a strategy. For uncommon features, where verbal codes may be less likely to be used (or used successfully) by participants in either age group, we would expect no evidence of age-related binding deficits. In contrast, if verbalization plays less role in change-detection (e.g., see Sense et al., 2017), there should be no age-related label-ability effect.

Also, in my previous experiments (see Chapter 2) the single-feature conditions consisted of coloured blobs or black shapes. This methodological choice aimed to prevent potential age-differences in the ability to only encode relevant features (i.e., perhaps participants would automatically encode both the shape and colour, despite knowing that only colour memory would be probed –

in line with findings that colouring a shape can affect shape-recognition performance even when colour is irrelevant to the recognition task, e.g., Ecker et al., 2013). However, this approach appeared uncommon in change-detection paradigms, so I opted against it in the current experiment. Thus, participants saw four coloured shapes in each condition, regardless of which feature they needed to remember. While this choice facilitated comparison of the present results to previous change-detection studies with younger and older adults, it complicated comparison of these results to those in Chapter 2.

Method

Participants. I recruited 51 new participants, all native speakers of English. One older adult did not complete the task, and was excluded. Twenty-five younger adults (7 male; *M* = 20.7 *SD* = 2.3 years) from the student population of the University of Edinburgh received £7 in return for participation. Twenty-five older adults (7 male), all from the community-based University of Edinburgh psychology research volunteer panel (*M* = 68.3, *SD* = 3.4 years), were each given £10 in return for participation. Colour vision was assessed as in the previous experiments; no one was excluded. NART-based verbal IQ scores in the older adult group were *M* = 122.8 (*SD* = 3.9), and *M* = 114.5 (*SD* = 3.9) in the younger adult group, and differed significantly *t*(44) = 7.6, *p* <.001, *d* = 0.55. Mean years of education in older (*M* = 15.4, *SD* = 3.6) and younger adults (*M* =

14.7, *SD* = 2.6) did not differ significantly *t*(44) = .75, *p* >.05 (providing years of education was optional, N = 24 for the younger adults, N = 22 the older). All older adults scored higher than 82 (*M* = 98.04, *SD* = 2.0) on Addenbrooke's Cognitive Examination (ACE-III; Hodges, 2012).

Stimuli. The stimuli were identical to those used in Experiments 1, 2 and 3, displayed using the same equipment.

Design and Procedure. The general memory task procedure is shown in Figure 3.1.

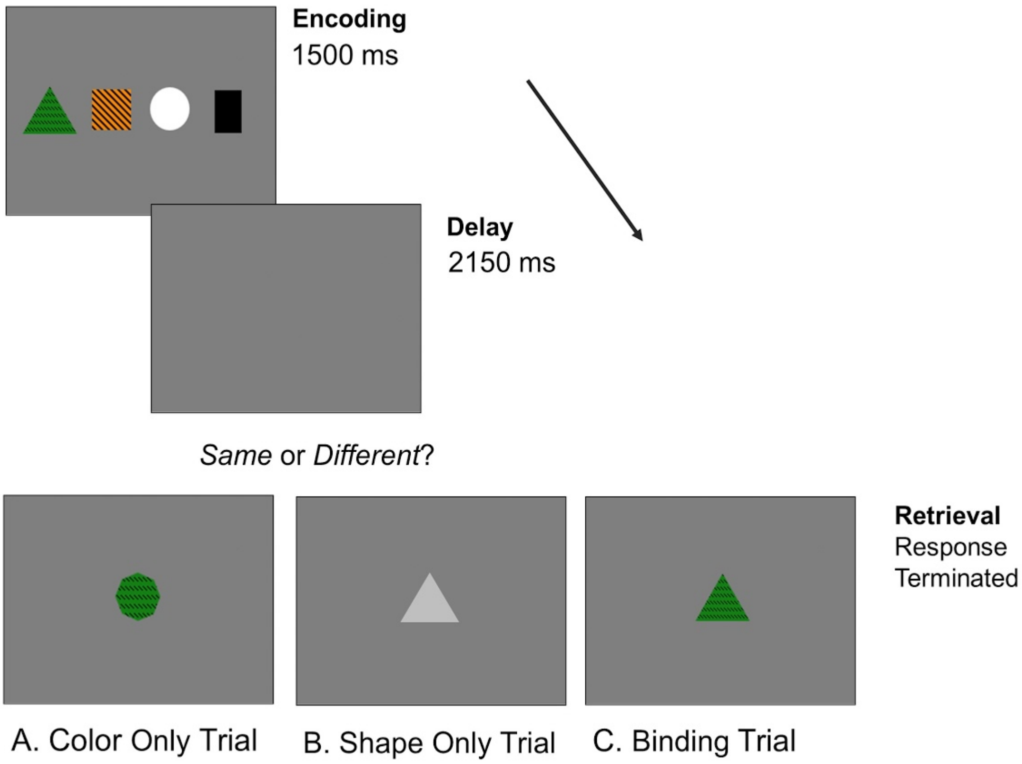


Figure 3.1. Illustration of the trial sequence Experiment 4. *A. Colour only trial.* Participants indicated whether the colour (presented as a blob) was present in the memory array. *B. Shape only trial.* Participants indicated whether the shape (presented in grey) was present in the memory array. *C. Binding trial.* Participants indicated whether this exact colour-shape combination was one of the four items. In the illustration, the correct response in all three conditions would be 'Same'. Participants had unlimited time to respond. *Note*. Different fill patterns represent different colours and items are not drawn to scale.

First, each participant completed two practice trials in each experimental condition (colour-only, shape-only and binding). The task instructions included "Please be as accurate as possible, speed is not important". In each trial, participants saw four stimulus items presented in a row. Trials were divided into three blocks, one in which participants only had to remember the colours of the items, one only the shapes, and one both colours and shapes (*bindings*). Block order was randomised across participants. In each trial, participants viewed the test array (1500 ms), followed by a delay (2150 ms), and then a single probe item appeared on the screen. Participants indicated whether the probe item was the same as, or different from, one of the four studied items, using keys labelled 'same' or 'diff.'. In the binding trials, 'same' indicated that the probe object was identical to one of the four studied objects, and 'different' indicated that the probe had the shape of one of the four studied objects but the colour of another one. In shape-only trials, the probe was always light grey (colour did not matter), and participants had to indicate whether the probe shape was one of the four original items in the study display ('same'), or a new shape (''different'). In colour-only trials, participants responded to whether a 'blob' shape was of the same – or different – colour as a memory target item. Each block contained 100 trials. Half of these were common, half uncommon, and within each, half were change ('diff.') trials, the other half no-change trials ('same'). The task took between 30 to 50 minutes to complete.

Analysis

Change-detection task performance can be expressed in several different ways, and analysis of different measures may result in different conclusions, especially regarding interactions (see Allen et al., 2012). Here, I included a measure of proportion correct to facilitate comparison with the results presented in Chapter 2. However, proportion correct does not account for potential participant response bias. If participants are biased towards the 'different' response this would increase the hit rate (i.e., correctly detecting when the probe item differed from the memory items), but also reduce the correct rejection rate (i.e., produce more false alarms; incorrectly claiming that there was a change). Younger and older adults could have different levels of response bias in the different memory conditions. Therefore, I also used a common measure of discriminability (or sensitivity); *d'*. This measure originated in a signal detection theory model based on the premise that change and no-change trials form two Gaussian equal variance distributions along a familiarity continuum. No-change trials are generally thought to result in higher levels of familiarity. The separation of the change and no-change distributions, *d'*, indicates the sensitivity of the observer[5].

Results

I conducted two separate Bayes Factor ANOVA model comparison analyses (similar to those in Experiments 1 – 3 in Chapter 2). The dependent variables in

---

[5] To avoid calculation problems created by hit rates of 1 and false alarm rates of 0 I performed a standard correction. Assuming that the true false alarm rate would lie between 0 and 1/N (where N = the maximum number of lures), I used 1/(2N) as the false alarm rate, instead of 0 (i.e., essentially adjusting zero false alarms to half a false alarm). Similarly, for hit rates of 1, I used 1 - 1/(2N), where N = the number of possible targets.

the two analyses were response accuracy (proportion correct) and *d′*. Again, I excluded trials with reaction times over 10 seconds (a total of 0.15 % of trials were excluded from the young adults, 0.28 % from the older). Proportion correct and sensitivity as indexed by *d′* across the age-groups and experimental conditions are shown in Figure 3.2. I report the results of the Bayes Factor ANOVA model comparison analyses for proportion correct and *d′* below.

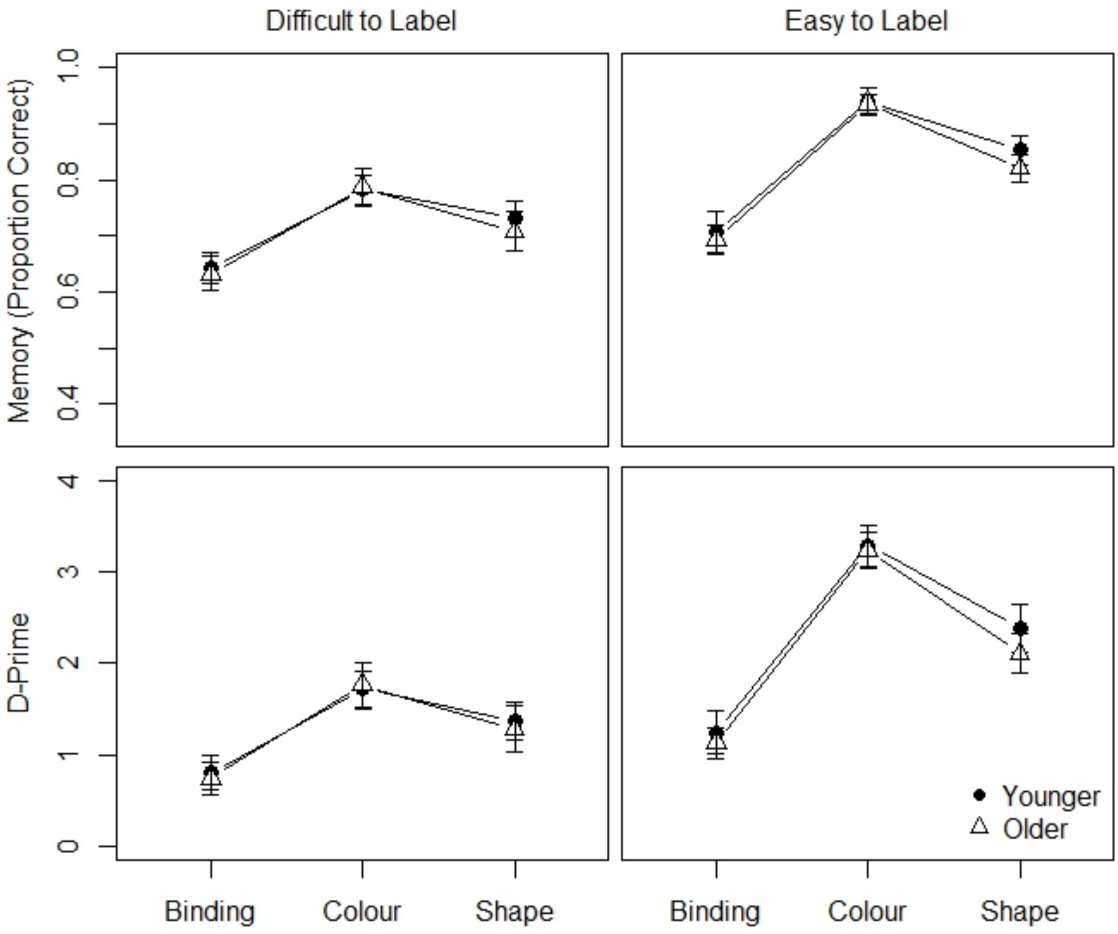


Figure 3.2. Memory Accuracy (proportion correct) in Experiment 4, by Age Group, Label-Ability and Binding Condition. Error bars are within-subjects 95% confidence intervals, adjusted values calculated using method from Morey (2008).

I observed no clear evidence that overall memory accuracy differed between younger ($M = .78$, $SD = .42$) and older participants (proportion: $M = .76$, $SD = .42$; $1/B = 0.075$, d': $1/B = 0.30$; i.e., memory for either shapes, colours or bound objects). Memory for easy-to-label items (proportion: $M = .83$, $SD = .38$) was better than that for difficult-to-label items (proportion: $M = .72$, $SD = .45$; $1/B = 4.00 \times 10^{58}$, d': $1/B = 1.84 \times 10^{36}$). There was evidence for an effect of condition (proportion: colour; $M = .88$, $SD = .35$, shape: $M = .78$, $SD = .41$, binding: $M = .67$, $SD = .47$, $1/B = 2.00 \times 10^{117}$, d': $1/B = 5.99 \times 10^{51}$). However, there was no clear support for a binding deficit in older adults (proportion: Trial Type × Age Group: $1/B = 7.71 \times 10^{-3}$, d': $1/B = 0.12$), nor that the effect of binding condition was modified by label-ability (proportion: $1/B = 3.33 \times 10^{-3}$, d': $1/B = 0.12$). However, there was strong evidence (proportion: $1/B = 7.29 \times 10^{3}$, d': $1/B = 1.11 \times 10^{9}$) for retaining the Label-ability × Trial Type interaction, indicating that label-ability had less impact on accuracy in the binding condition than in the single-feature condition (see Table 3.1).

Table 3.1

Change-Detection accuracy (proportion correct) by age groups and experimental factors in Experiment 4

			M	*SD*
Binding	Younger	Difficult	.64	.06
		Easy	.71	.09
	Older	Difficult	.63	.07
		Easy	.69	.06
Colour Only	Younger	Difficult	.78	.07
		Easy	.94	.06
	Older	Difficult	.79	.08
		Easy	.94	.04
Shape Only	Younger	Difficult	.73	.07
		Easy	.85	.06
	Older	Difficult	.71	.08
		Easy	.82	.06

Note. Difficult = Difficult-to-label; Easy = Easy-to-label.

This suggested that label-ability played a larger role in the single-feature conditions than in the binding condition. However, this effect was similar for participants in the two age groups, so these results did not suggest that older adults – compared to younger – relied more on sub-vocal rehearsal of easy-to-label stimuli in a change-detection task with a single probe. See Appendix B for complete outputs of the Bayes Factor ANOVA models, and mean values, SDs and Cohen's *d* for the main effects on memory accuracy.

Discussion

Unexpectedly, I observed no overall effect of age on memory accuracy. While overall memory for difficult-to-label stimuli was lower than for easy-to-label stimuli, this was observed to equal extents in the younger and older adults. Similarly, memory for bound features was generally poorer than that for single features, but this was also observed to equal extents for both younger and older adults, with no differences in degree between shape and colour. This lack of evidence for an age-related binding deficit was consistent with previous research using colour-shape change-detection paradigms (Brockmole et al., 2008; Brown et al., 2017; Peterson & Naveh-Benjamin, 2016; Rhodes et al., 2016). The absence of an overall age effect was unexpected, and differed from Experiments 1, 2, and 3. As outlined in Chapter 1, usually older adults perform worse than younger adults on visuospatial tasks (e.g. Johnson et al., 2010, but see Olson et al., 2004). The average NART scores (providing predicted verbal IQ scores) were significantly higher in the older adult group, which might suggest that the older adults were of higher mental ability. However, this was observed in the reconstruction experiments as well.

The lack of strong evidence for an age effect in Exp. 4 (c.f., that observed with the reconstruction paradigm in Chapter 2) might be because recall tasks place greater demands on memory compared to recognition tasks, and require greater self-initiated processing, which tends to reveal greater age-related memory decline (e.g. Craik, 2006). However, Rhodes et al. (2015) used a similar change-detection task, and a very similar (even possibly overlapping) demographic of participants recruited from the same volunteer panel, and found a main effect of age on performance. Our paradigm differed from that of Rhodes et al. (2016) in that we used four rather than three items in the memory arrays. This was to prevent ceiling effects in the younger adults in the label-able single-feature condition, based on their excellent performance in Experiments 1 to 3. While four items might be beyond some participants' capacity regardless of age (see Cowan, 2010; Luck & Vogel, 1997), it is unclear how this could have obscured an age effect.

Interestingly, no specific effect of colour label-ability was apparent in older adults in the change-detection paradigm, in contrast to its strong effect in the reconstruction paradigm. This indicated that change-detection tasks may be less prone to confounds of verbal strategy use by older adults. In our reconstruction tasks participants selected the specific colours or shapes by mouse-clicks. This required finding their response amongst other options, which might have made verbal labels more salient and useful than in the "same or different" judgement this change-detection task. For instance, it is possible that verbal labels helped older adults maintain the memory traces despite the visual interference introduced by the response array in the reconstruction paradigms. However, the strong evidence that label-ability influenced memory performance more for the single-feature blocks than the binding blocks (see Figure 3.2), suggested that verbal strategies may be used in this paradigm – but seemingly to equal extents by younger and older adults. However, since articulatory suppression was not applied in this study, it is unclear whether the benefit for easy-to-label single features was driven by sub-vocal rehearsal.

These results add to findings suggesting that results cannot be generalised across different types of feature-binding measures, such as between identity and location, features within the same dimension (such as colour-to-colour binding), or between features drawn from different locations, as they may all draw on different cognitive processes (e.g., Delvenne & Bruyer, 2004; Ecker et al., 2013; Niven, Brown & Allan; Shimi, & Logie, 2018; Wheeler & Treisman, 2002; Xu, 2002). Thus, while label-ability may produce 'fake' age-related binding deficits for easy-to-label items in reconstruction tasks, it appeared to be less of a problem in change-detection paradigms – at least under the specific circumstances in this study. Indeed, it appears that the specific experimental circumstances where older adults may benefit from applying verbal strategies vary depending on stimulus type and response mode.

Chapter 4: Cognitive Aging and Verbal Labelling in Continuous Visual Memory

Aims

In Experiments 1 to 4 I explored the idea of differential approaches to tasks measuring visual feature-binding, to see if older adults' verbal rehearsal explained discrepancies in the literature. Results regarding the implications for the feature-binding literature were mixed. A key interesting outcome was the indication that older adults' memory was comparatively more boosted for common, easy-to-label colours, in the reconstruction paradigm. However, the evidence that sub-vocal rehearsal of colour labels drove this boost was not entirely convincing (1/*B* for the Age Group × AS × Label-Ability interaction was 2.96). The research presented in this thesis has so far measured verbal rehearsal by comparing task performance in silence with that under suppression (for stimuli hypothesised to be either easy- or difficult-to-label), and the *difference* between silence and suppression was taken to indicate sub-vocal labelling and/or rehearsal of stimulus labels. In contrast, in the present study I tested the role of *overt* (instructed) labelling of colours in younger and older adults, using a paradigm used in younger adults by Souza and Skóra (2017). Theoretically, such overt verbal rehearsal would also have been an informative experimental manipulation in Experiments 1 to 4, but it was not practical to ask participants to name aloud difficult-to-label colours and shapes, nor the six or eight features meant to be retained in the binding condition, in the limited time interval.

In this study, in addition to directly instructing overt labelling, I also used a precision paradigm of colour memory. As mentioned in the introduction (Section 1.2.5), precision measures of visual WM require precise reproduction of the memory item along a continuous scale, such as selecting the specific shade of colour from a continuous colour-wheel (e.g., Bae, Olkkonen, Allred, Wilson, & Flombaum, 2014) or the orientation of an arrow (e.g., Fallon, Mattiesing, Muhammed, Manohar, & Husain, 2017). Such precision measures may produce stronger experimental paradigms by avoiding misleading cut-off points in performance (i.e., ‘remembered’ or ‘forgotten’), and distinguishing between participants who were just slightly wrong, and those who appeared to guess randomly. For example, consider a hypothetical grid task requiring participants to reproduce which cells in the grid were coloured. Neither of two hypothetical participants remembers a certain filled-in cell perfectly; one participant has a general idea that it was in the top right of the stimulus display area, but incorrectly selects the cell right next to it. The other participant may have a much more limited representation and just guess, randomly selecting any cell. Precision paradigms enable quantification of differences between these two incorrect responses, and help distinguish between two such hypothetical participants, who might attain identical scores if their performance was quantified using a binary (non-precision) measure. Precision paradigms have been used to explore memory decline in healthy ageing (Peich, Husain & Bays, 2013; Pertzov, Heider, Liang, & Husain, 2015), as well as pathological ageing (Liang et al., 2016).

However, recent observations of categorical responding in continuous visual WM tasks such as the colour-wheel task (e.g., Hardman et al., 2017) suggested that perhaps participants rely on colour categories or names, rather than merely remembering the visual representation of the specific shade. Hardman et al. (2017) found that their model that included a mechanism for remembering rough categories (e.g., ‘purple’) outperformed the continuous representation-only model. Similarly, Bae et al. observed that colour values remembered with higher precision within and across participants were shades most commonly selected as prototypical (e.g., the ‘bluest shade of blue’) by participants. Hence, assigning and sub-vocally rehearsing a label (e.g., ‘blue’) might be a likely mechanism for remembering colours categorically (Donkin et al., 2015).

The work in Chapter 4 was a conceptual replication of Souza and Skóra (2017), who investigated the effect of overt labelling on continuous visual colour memory in younger adults. Similar to them, I compared colour memory performance when participants *had* to label, with that when they could not (suppression), with spontaneous performance in silence, and used Hardman et al.’s (2017) model to separate continuous and categorical responses. I contrasted the amount of categorical and continuous responses in the age groups in silence, compared to during overt labelling and suppression, to test whether spontaneous performance in silence was more similar to that during overt labelling or suppression. The aim was to explore the extent to which younger and older adults seemed to use verbal labelling in silence, and how they benefitted from being able to label. *At the time of writing, the following paper is under review in Memory and Cognition (hence the American Spelling).*

Abstract

The decline of Working Memory is an essential feature of general cognitive decline, and visual and verbal temporary memory abilities appear to decline at different rates with age. Visual material may be remembered via verbal codes or visual traces, or both. Souza and Skóra (2017) found that labeling boosted memory in younger adults by activating categorical visual Long-Term memory knowledge. Here, we replicated this and tested whether it held in healthy older adults. We compared spontaneous performance (silence), instructed overt labeling, and articulatory suppression (preventing labeling) in the delayed estimation paradigm.

Labeling improved memory performance in both age groups, but older adults appeared spontaneously to rehearse verbal labels sub-vocally to maintain coarse, categorical representations more than did younger adults. However, older adults did not appear to use verbal memory to ‘compensate’ for declining visual abilities; younger adults were actually comparatively more impaired when such labeling was prevented. Older adults also appeared to benefit from labels differently than younger adults. In younger adults labeling appeared to improve visual, continuous memory, suggesting that labels activated visual LTM representations. However, for older adults, labels did not appear to enhance

visual representations, but instead boosted memory via additional verbal (categorical) memory traces.

These results highlighted the importance of controlling differences in age-related strategic preferences in visual memory tasks, and that such differences may be detected using mixture modeling combined with explicit labeling manipulations. They challenged the assumption that visual memory paradigms measure the same cognitive ability in younger and older adults.

Introduction

Visual Working Memory (WM) – maintaining visual information in memory during a short interval when it is no longer present but needed for an upcoming task – declines steeply with age (Babcock & Salthouse, 1990; Bowles & Salthouse, 2003; Craik, Luo, & Sakuta, 2010; Johnson, Logie, & Brockmole, 2010; Gazzaley, Cooney, Rissman, & D'Esposito, 2005; Jost, Park, Lautenschlager, Hedden, Davidson, & Smith, 2002; Jost, Bryck, Vogel, & Mayr; 2010; Reuter-Lorenz & Sylvester, 2005). This age-related decline has practical importance as WM is believed to underpin effective operation of other cognitive functions, such as perception and problem-solving (e.g., Ma, Husain & Bays, 2014), and to be related to general intelligence (e.g., Unsworth, Fukuda, Awh, & Vogel, 2014) and reasoning ability (e.g., Conway, Kane, & Engle, 2003; Kyllonen & Christal, 1990). However, there is disagreement about whether WM is best conceptualized as a unitary construct (e.g., Cowan, 2005; Oberauer, 2013) or as consisting of different components. For instance, the Multi-Component Model of WM (Baddeley, 1986; 2012; Baddeley & Logie, 1999; Logie, 2011) is based on the postulate that visuospatial and verbal information are stored separately in dedicated storage buffers, which may also rely on separate rehearsal mechanisms (Baddeley, 2012; Logie, 2011), possibly supported by different brain networks (Jonides et al., 1996; Gruber, 2001), but see also; D'Esposito & Postle, 2015). Furthermore, visual and verbal temporary memory abilities appear to decline at different rates with age (e.g., Johnson et al., 2010), supporting the notion that they do not rely on the same mechanisms. In this paper, we addressed this debate by investigating whether younger and older adults use verbal labels for retaining visually presented stimuli to the same extent and whether such labels have the same effect on memory in the different age groups.

The nature of the relation between language and cognition has been debated for decades (Hunt & Angoli, 1991; Watson, 1924; Whorf, 1956). Verbalization — translating visual perceptual input into phonologically-based verbal code — has a central role in this debate. Although translating visual input into verbal labels is well established as a default — perhaps sometimes even unavoidable — tendency (Conrad, 1964; Postle, D’Esposito, & Corkin, 2005; Postle & Hamidi, 2006; Shulman 1971; Simons, 1996), its impact on cognition is unclear (Lewis-Peacock, Drysdale & Postle, 2014). For instance, verbalization might be detrimental to cognitive tasks such as decision-making (Wilson & Schooler, 2001), analogical reasoning (Lane & Schooler, 2004) and visual imagery (Brandimonte, Hitch, & Bishop, 1992). This well-established phenomenon is known as *verbal overshadowing,* because resources are thought to be allocated to the verbal label at the expense of the original task (Schooler & Engstler-Schooler, 1990), but the mechanisms behind it are still disputed (Chin & Schooler, 2008; Hatano, Ueno, Kitagami, & Kawaguchi, 2015).

Indeed, many memoranda – in everyday life as well as in memory experiments – may be remembered via verbal codes or visual traces, or both, among other possibilities. For example, remembering which jumper to purchase after trying on several at the store could be achieved using a verbal description ("the blue one"), as well as a visual representation of it. Neuroscience evidence supports the notion that individual participants generate visual, phonological and semantic mental codes when viewing visual stimuli (Lewis-Peacock et al., 2014). Despite this tendency to translate visual representations into verbal codes, visual and verbal WM are typically measured separately, and a given task is assumed to measure one or the other, despite evidence that both verbal and visual codes might be stored for visually presented material (e.g., Logie, 2018; Logie, Saito, Morita, Varma, & Norris, 2016; Saito, Logie, Morita, & Law, 2008; Paivio, 1971). In the WM literature, there is little doubt that perceptual input results in different types of mental codes – both within and between individuals

– which may interfere with one another in complex ways (e.g., Morey & Cowan, 2004). Moreover, while domain-specific stores are not emphasized in unitary conceptions of WM, they are not explicitly rejected (Cowan, 2005; Cowan, Saults, & Blume, 2014; Oberauer, 2013). It is generally agreed that sub-vocal rehearsal of verbal material is a separate mechanism that can support memory; however the presence of a visuospatial rehearsal mechanism (Logie, 1995) is more contentious (Morey & Mall, 2012; Morey & Miron, 2016). Hence, the limits of visual WM and its decline with age are often investigated while attempting to prevent verbal rehearsal of labels using concurrent articulatory suppression, i.e. repeating nonsense syllables out loud during the encoding and/or retention period (Allen, Baddeley, & Hitch, 2006; Hollingworth & Rasmussen, 2010; Logie, Brockmole, & Vandenbroucke, 2009; Matsukura & Hollingworth, 2011; van Lamsweerde & Beck, 2012), or assumed to be prevented by presenting items very briefly.

The past decade has seen a new way to measure visual WM. *Delayed estimation paradigms* provide precise, continuous measures of memory, in line with the idea that WM resources are allocated among items in memory and remembering more items leads to loss of precision as resources are spread more thinly (Ma et al., 2014). In these paradigms, participants reproduce features in memory on a continuous report scale (Prinzmetal, Amiri, Allen, & Edwards, 1998; Wilken & Ma, 2004; Zhang & Luck, 2008), which enables analysis of the distribution of the magnitudes of recall errors. These can be characterized by mathematical models that estimate both WM precision and the proportions of items participants remember (Wilken & Ma, 2004; Zhang & Luck; 2008). For example, participants recall colors by selecting among different color shades arranged around a color-wheel continuum after a brief delay (e.g., Bays, Catalao, & Husain, 2009; Bays, Wu, & Husain, 2011; Emrich & Ferber, 2012; Fougnie & Alvarez, 2011; Fougnie, Asplund, & Marois, 2010; Fougnie, Suchow, & Alvarez, 2012; van den Berg, Shin, Chou, George, & Ma, 2012; Wilken & Ma, 2004; Zhang & Luck, 2008, 2009, 2011; Peich, Husain, & Bays, 2013). Crucially, this paradigm differs from traditional memory tasks where the to-be-remembered items belong to limited sets of categories (e.g., 'red', 'blue') and rough categorical retention alone is sufficient to perform the task perfectly (e.g., Allen, Baddeley, & Hitch, 2006; Cocchini, Logie, Della Sala, MacPherson, & Baddeley, 2002; Kane et al., 2004; Saults & Cowan, 2007). Initially, researchers measuring color memory precision assumed that all colors were stored as visual, continuous representations (Zhang & Luck, 2008). Others later questioned this assumption. Several studies indicated that even continuous color values were stored in WM based on categorization, as participants' responses clustered closely around specific, prototypical color values instead of being evenly distributed along the color-wheel continuum (Bae, Olkkonen, Allred, Wilson, & Flombaum, 2014; Bae, Olkkonen, Allred, & Flombaum, 2015; Donkin, Nosofsky, Gold, & Shiffrin, 2015; Hardman, Vergauwe, & Ricker, 2017; Olsson & Poom, 2005; Souza & Skóra, 2017).

Hardman et al. (2017) found that their model that included a mechanism for remembering rough categories (e.g., "purple") outperformed the continuous representation-only model. Others have partially addressed what produces such categorical responding. Categorization of colors is not necessarily verbal – it occurs in perceptual tasks (Bae et al., 2015) and when labeling is unlikely (see Bae et al., 2014). However, Bae et al. observed that color values remembered with higher precision within and across participants were shades most commonly selected as prototypical (e.g., the "bluest shade of blue") by an independent sample of participants. Based on such findings, assigning and sub-vocally rehearsing a label (e.g., "blue") was proposed as a likely mechanism for remembering colors categorically (Donkin et al., 2015). To investigate this, Souza and Skóra (2017) manipulated participants' use of verbal labels in the color-wheel paradigm. They used Hardman et al.'s (2017) model to separate continuous and categorical responses, and found that labeling increased both the number of categorical and continuous representations and the precision with which continuous representations were remembered in healthy young adults. Hence, verbal labels did not appear to boost memory representations by merely adding verbal memory representations. Instead, Souza and Skóra suggested that labeling boosted memory by activating categorical visual LTM knowledge, which enabled participants to rely on two visual representations: a temporary visual representation of what the color looked like (independent of labeling), and a representation of the given visual category in LTM. As most research using the color-wheel paradigm has been done using younger-adult samples and there is evidence that older adults may perform better in tasks which allow verbal rehearsal of labels (e.g. Johnson et al., 2010), we investigated how verbal labeling impacted visual memory in healthy older adults.

We hypothesized that older adults might rely more on verbal labels in visual tasks to support or compensate for declining visual memory (Baltes & Baltes, 1990; Park & Reuter-Lorenz, 2009; Reuter-Lorenz & Park, 2014), thus relying on a different cognitive ability than younger adults to perform the same task. This hypothesis derived from literature suggesting that while older adults' WM for verbal material remains relatively intact, they show more severe deficits for visuospatial material (for a meta-analysis, see Jenkins, Myerson, Joerding, & Hale, 2000, and also, Johnson et al., 2010; Leonards, Ibanez, & Giannakopoulos, 2002; Logie & Maylor, 2009; Myerson, Hale, Rhee, & Jenkins, 1999; Park et al., 2002). Johnson et al. (2010) found evidence supporting the notion that younger and older adults may rely on different cognitive abilities, investigating the factor structures of performance on various WM tasks in 95,000 online participants of different ages. They found that for older participants, visual memory seemed more related to some general cognitive capacity, but in younger participants, it seemed to reflect a more specific capacity. The reverse age contrast appeared to apply to verbal short-term memory, supporting the idea that older people might rely on their verbal memory ability to perform a visual memory task. Additionally, Reuter-Lorenz et al. (2000) observed lateral organization of the WM system in young participants, whereas older participants showed considerable activity in both left and right frontal sites for both verbal and visual WM, suggesting that younger and older adults may engage different brain areas to different extents when performing the same task (Reuter-Lorenz, 2002).

Merely testing the simple hypothesis that older adults are worse at various cognitive tasks than younger adults (i.e., the 'Dull Hypothesis'; Logie, Horne, & Petit, 2015; Perfect & Maylor, 2000) arguably does little to further our understanding of how or why cognition declines with age. Instead, identifying and investigating different sub-processes which decline at different rates and how this might drive older adults to recruit relatively spared cognitive functions to compensate for cognitive functions that decline with age is likely a more informative approach to understanding what changes in healthy aging. If visual tasks allow verbal rehearsal of stimuli, and younger and older adults differ in the extents to which they rely on such rehearsal, problematic confounds likely occur – especially if visual WM paradigms used to measure a given phenomenon differ in the extent to which verbalization is possible. For instance, age-related differences in verbal recoding could be problematic in paradigms measuring visual feature-binding if single features lend themselves to such verbal rehearsal and bound objects do not (Brockmole & Logie, 2013). Indeed, age-related binding deficits in delayed estimation tasks were observed in some experimental settings (memory for color and orientation of bars; Peich et al., 2013) but not others (locations of complex, hard-to-name fractal objects; Pertzov, Heider, Liang, & Husain, 2015). Furthermore, if older adults favor verbal, categorical representations over continuous ones, this could cause reduced precision in older adults (Peich et al., 2013) since on average, categorical responses may be further from the correct shades than responses based on precise visual representations. Differential reliance on verbal rehearsal in participants of different age groups would likely go unnoticed in the majority of visual WM paradigms. Here, we tested whether older adults support a declining visual WM system by relying on verbal rehearsal of labels using a relatively intact phonological loop system (Baddeley & Hitch, 2018) by preventing verbal labeling in the delayed estimation paradigm. We conducted a conceptual replication of Souza and Skóra (Exp. 4, 2017), including both younger and older participants, to investigate the following three questions:

(1) Do older adults spontaneously use verbal labels more when free to do so? We compared the proportion of categorical vs. continuous responses by the two age groups performing the task in silence, with instructed labeling, or under suppression. If participants generally sub-vocally label, the number of visual vs.

verbal representations should be similar to that during instructed labeling. If they do not spontaneously label, it should be similar to that under suppression.

(2) Does preventing labeling and rehearsal of labels impair older adults' memory performance more? If older adults' 'visual' memory performance depends on verbal rehearsal of labels (i.e., a verbal strategy) to compensate for poor visual memory, their memory should be more impaired while verbal rehearsal is disrupted. To test this, we compared how overall memory performance in the two age groups was affected by articulatory suppression.

(3) Do older adults benefit from labels in the same way as younger adults? This can be tested by comparing instructed labeling with disrupted labeling in two age groups. Souza and Skóra (2017) found that – in younger adults – labeling was associated with increased categorical and continuous memory representations, and improved precision, consistent with the labeling benefit being due to activated visual LTM representations. However, if verbal labels overwrite continuous representations in older adults, that would be consistent with *the Verbal Recoding Hypothesis* (see Schooler & Engstler-Schooler, 1990). Alternatively, labeling might be beneficial because it adds a verbal representation to the visual WM trace (*the Dual-Trace (Visual +Verbal) Hypothesis).* Then, the number of visual representations would be the same, but there would be additional, verbal representations. Evidence suggesting that older adults benefit from having two traces (Osaka, Otsuka, & Osaka, 2012)

might support this hypothesis. If one of these alternative hypotheses better explains the labeling benefit in older adults, this would indicate that being allowed to label impacts performance via separate processes in younger and older adults.

Researchers comparing color memory precision in younger and older adults typically assume that the same cognitive ability (visuospatial WM) is measured in all participants. With our three research questions, we explicitly tested this assumption. We examined if preventing or instructing verbal labeling would affect various aspects of participants' memory performance in the different age groups differently. Specifically, we distinguished continuous and categorical responding (Hardman, 2017) and used an explicit labeling paradigm to separate spontaneous performance (Silence), instructed overt labeling, and articulatory suppression to prevent labeling (similar to Souza & Skóra, 2017) to test the following three hypotheses, which were pre-registered via the Open Science Framework [osf.io/m64px].

H_1: Do younger and older adults differ in the probability of storage in WM (of the to-be-remembered colors)? If older adults depend more on verbal labels for their visual WM performance, *preventing labels* should impair their memory performance comparatively more than it does for the younger adults.

H_2: Do younger and older adults differ in the probability that the representation in memory is continuous as opposed to categorical? If older adults spontaneously use verbal labels more in silence, their relative performance in silence to that in the two verbalization manipulations (labeling, suppression) should differ from that of the younger adults. Also, if older adults benefit differently from verbal labels, labeling (compared to suppression) may influence the types of representations differently.

H_3: Do younger and older adults differ in the imprecision of the continuous representation in memory? Applying a continuous/categorical model to delayed recall data from older adults is useful to test the possibility that older adults' favoring of categorical representations over continuous may cause reduced precision in older adults (Peich et al., 2013), since on average, categorical responses are further from the correct shades than are responses based on precise visual representations. Here, we investigated continuous precision in the age groups without categorical representations using Hardman's (2017) model to separate these types of responses.

Methods

Participants

To reach our target sample size of 60 participants[6], we recruited 32 younger adults and 33 older adults, and excluded and replaced two younger and

---

[6] We based this sample size on Exp. 3 in Souza and Skóra (2017), since measures of effect sizes are less straightforward to obtain from Bayesian analysis (see Bayarri & Berger, 2004). This experiment was most similar to our design since each participant performed three different conditions, with a set size of 4 items, with 50 trials in each. In 30 younger adults, they could detect a credible difference between these conditions (e.g. between suppression and overt labeling).

three older adults for not completing all trials, per our pre-registered exclusion criteria. The final sample consisted of 30 younger adults (18 – 27 years old, *M* = 22.0, *SD* = 2.7, 9 male), and 30 older adults (62 – 78 years, *M* = 68.6, *SD* = 4.9, 10 male). All participants reported having normal or corrected-to-normal vision, and were without color-vision deficits (indicated by less than five errors on the on-screen version of the Dvorine pseudo-isochromatic plates; Dvorine, 1963). All older adults scored above the recommended cut-off point for cognitive impairments of 25 on the ACE-III-Mini (Hsieh et al., 2015; *M* = 28.5, *SD* = 1.5), completed after the color memory task. Younger (*M* = 16.07, *SD* = 2.07) and older adults (*M* = 16.22, *SD* = 4.07) did not differ in how many years of full-time education they had completed (*t* (58) = -.18, *p* = .856, *d* = -0.047). Younger adults received £7.50 for their participation and older adults £10, as their sessions differed in length because older adults completed the ACE-III-Mini. Our methods and analysis plan were pre-registered at osf.io/m64px.

Materials and Procedure

The study had a mixed design, with *Age Group* as a between-subject factor (Younger or Older), and a within-subject factor of *Verbalization*: silence, overt labeling (labeling colors out loud) and suppression (repeating ‘ba-ba-ba’ out loud). All participants completed one block in each Verbalization condition, and each block consisted of 50 trials, resulting in a total of 150 trials per participant.

Block order was counterbalanced among participants, such that five participants in each age group performed each of the six possible block order combinations. For the two blocks requiring vocalization we instructed participants to speak at normal conversational volume. Participants were tested individually. The experimenter remained in the room to ensure that participants followed instructions. All participants performed three practice trials in whichever condition they started with, before starting the experiment. If they did not understand the task, they did further practice.

Each trial started with an instruction text appropriate for the current verbalization block: '*Trial Starting*', '*Name Colours*' or *'Ba-Ba-Ba'*. The participant pressed a key to start each trial. Then, four colored circles appeared simultaneously for 930 ms, on a grey background. This presentation time was longer than the 250 ms used by Souza and Skóra (2017), to ensure that our older adult participants would be able to perceive and label all four colors. See Figure 4.1 for an outline of a typical trial. The circles were evenly spaced at 90°, 180°, 270° and 360° angles around a larger imaginary circle (radius = 150 pixels) and each circle had a radius of 30 pixels (corresponding roughly to Souza & Skóra's circle radius; 1.6° visual angle, and imaginary circle radius; 6.65° for our screen size).

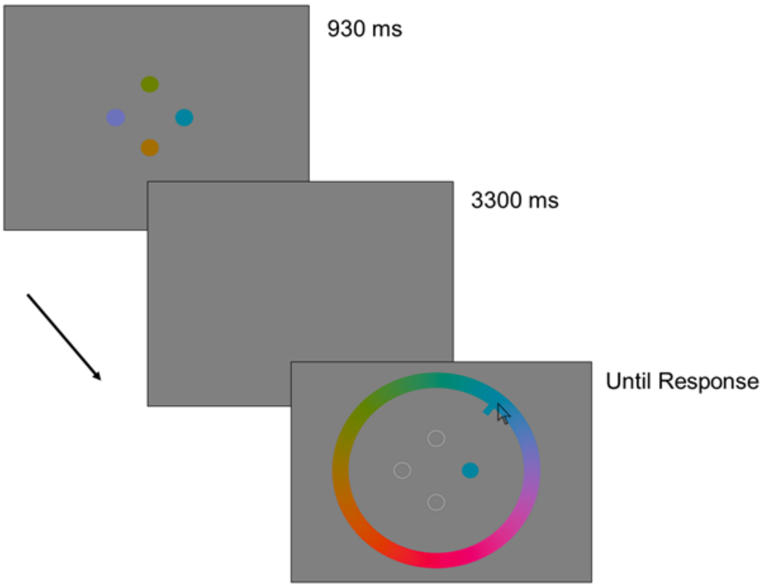


Figure 4.1. Outline of a Typical Trial.

The colors of the circles presented in the memory task were randomly chosen on each trial, selected from 360 possible color values. The 360 color values were evenly distributed in a circle in the CIELAB color space[7], centered in the color space with L = 50, a = 20, b = 20, radius = 60, and then converted to RGB values, trimming nonsense values (see Appendix C). Next, they saw a grey inter-stimulus-interval (ISI) screen of 3300 ms, followed by the color-wheel and outlines of the four circles. One of the circles was filled in dark grey, probing memory for the first item. Participants responded by clicking the mouse cursor

---

[7] The CIELAB color space is a system for the organization of colors, which expresses color as three numerical values, *L** for the lightness and *a** and *b** for the green–red and blue–yellow color components.

on the shade in the color-wheel they recalled having been in that circle. A second color was then probed, and so on, until participants had recreated all four original colors from memory, probed in a random order. On each trial, the color-wheel rotated randomly (e.g., the pink end of the spectrum might be at the top in one trial, and somewhere else in the next trial). In the suppression condition, participants were instructed to say "ba-ba-ba” while they saw the colored circles and during the ISI, but to respond in silence to ensure articulatory demands during the response phase were similar between conditions (see Souza & Skóra, 2017). During overt labeling, we told participants to label the colors out loud as soon as they appeared on the screen and to use whatever labels they found suitable.

Stimuli were presented using PsychoPy2 v1.82.01 (Peirce, 2007) and displayed on a 22" LCD Monitor, with a diagonal of 20.6", and a screen resolution of 1680 × 1050. Participants sat at approximately 50 cm unconstrained viewing distance from the screen. After completing the silent block, we asked participants: “Did you use a strategy to help you remember the colors? If so, could you please describe that strategy?”. I scored strategies including either 'naming', 'labeling', 'saying', 'repeating' or 'mumbling' as using a labeling strategy. Strategies which did not include these terms were not scored as labeling. The rater (AF) was aware of the study hypothesis.

Data analysis. We used a Categorical-Continuous mixture model (CatContModel, Hardman et al., 2017) to analyze the data. This model estimates the proportions of colors remembered categorically (participants remember coarse representations such as “red” that tend to cluster around a few canonical values) vs. continuously (participants remember the specific shade of the color) in the delayed-response color-wheel paradigm. Specifically, responses in this task are assumed to be informed either by memory or guessing. If the studied color is a light shade of pink, responses based primarily on categorically labeling it ‘pink’ should cluster around a specific number of canonical values (see Figure 4.2B.).

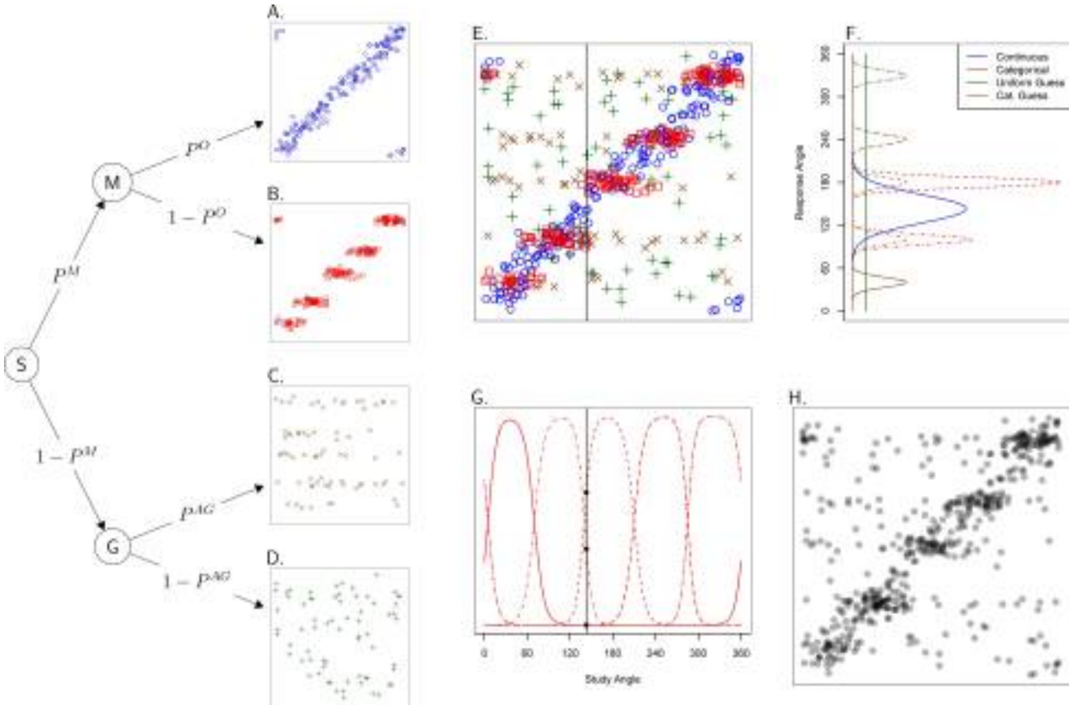


Figure 4.2. Multinomial process tree for the model and related plots for Hardman et al.'s (2017) categorical-continuous model. For all scatterplots, the x-axis represents the studied color-hue and the y-axis the response hue. Panel A. Continuous memory: responses vary linearly with the studied hue. The width of the diagonal line indicates continuous imprecision. Panel B. Categorical memory: for a range of studied hues, the same categorical response is provided. The width of the categorical bands reflects categorical imprecision. Panel C. Categorical guessing: guessing is distributed over categories. Panel D. Random guessing. Panel E. show the points in panels A to D combined. Panel F shows response densities for the four different response types for a single study angle (indicated by vertical line in Panel E). Panel G show the function giving the probability that a given study angle will be assigned to the given category. Panel H shows the same points as Panel E, but without information about the type of the response. Reprinted from Hardman, K. O., Vergauwe, E., & Ricker, T. J. (2017). Categorical working memory representations are used in delayed estimation of continuous colors. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(1), 30.

Alternatively, responses could be informed by continuous, more fine-grained representations of the specific hue, varying linearly with the studied colors (see Figure 4.2A). This representation would include information about the particular studied color, for instance, that it was a lighter pink. Storage of continuous information can be more or less fine-grained, and this memory precision is measured by the *Continuous Imprecision* parameter, represented by the width of the diagonal line in Figure 4.2A (equivalent to the imprecision parameter proposed by Zhang & Luck, 2008). In contrast, if responses are not informed by memory, they are classified as guessing. Guesses can be random (Continuous Guessing, see Figure 4.2D), or in accordance with certain categories (Categorical Guessing, see Figure 4.2C). The CatContModel classifies responses into these categories based on probabilistic mixture modeling. This is done by estimating the number of categories and their mean values for each participant using their overall response patterns. Figure 4.2H shows all responses from an imaginary participant. This imaginary participant has five color categories. The heights of the distributions in Panel F show the likelihood that different response angles would be chosen for a specific study angle, for each of the four response types. Panel E illustrates all the responses classified into the four different categories. Figure 4.2 also shows the multinomial process tree for the model (see the left part of the figure). S represents the start node, and the first branch depends on whether the participant had the tested item in WM, which happens with probability P^M. If so, they reach node M (Memory). If not, they reach node G (Guessing). Remembered items can be stored with continuous information – which happens with probability P^O – corresponding to the response distribution illustrated in Panel A. In contrast, the probability that the memory representation was categorical equals $1 - P^O$. When the item is not remembered, the model assumes that the participant will guess (probability of $1 - P^M$). The response distribution of categorical guessing is illustrated in panel C, and uniform (continuous) guessing in panel D. Both guessing distributions are independent of the study angle, while the response distributions are not. Thus, over all responses given by all participants, the model can estimate the following three parameters:

1. The probability that responses were informed by memory (P^M)
2. The probability that memory information was continuous (P^O)
3. The precision of the continuous information in memory (σ^O)

Simply put, P^M is the estimated probability of remembering (either categorical or continuous responses) as opposed to guessing. P^O is the probability of responding using continuous representations rather than categorical (i.e. informed by precise visual memory representation rather than clustering around a category center), and σ^O is the estimated precision of the responses classified as continuous. Here, we test how our experimental manipulations (silence, labeling or suppression) affected these three parameters in the two age groups. Furthermore, we used these parameters to calculate estimates of WM capacity (K). Regular K is a measure of WM Capacity where capacity (K) represents the total items in memory:

$$Total\ K = \ P^M \times \ Set\ Size$$

If capacity is truly four items, P^M for four items would equal 1, but when shown five items P^M would equal 0.8 ($4 = P^M \times 5$). In this study, Set Size was always four items. Categorical and Continuous K can be calculated by combining P^M; probability of storage, and P^O; probability that representation was continuous (rather than categorical; $1 - P^O$). Thus,

$$Continuous\ K \ = \ P^M \ \times \ P^O \ \times \ Set\ Size$$

$$Categorical\ K \ = \ P^M \ \times \ [1 \ - \ P^O] \times \ Set\ Size.$$

These measures allow us to distinguish whether verbalization manipulations caused *shifts* from one type of representation to the other (i.e. the capacity for one decreased, and the other increased) from a scenario where the manipulation increased one type of representation while the other remained the same.

We implemented the CatCont models in a Bayesian Hierarchical Framework (i.e., a model written in multiple levels that estimates the parameters of the posterior distributions using the Bayesian method). All parameter values were determined through Bayesian Markov Chain Monte Carlo (MCMC) sampling techniques. MCMC iteration is a method for obtaining information about posterior distributions in Bayesian inference (van Ravenzwaaij, Cassey & Brown, 2018), commonly used to compute inferential quantities (see Green, 1995; Han & Carlin, 2001). Hierarchical models reflect an assumption that a participant's parameter values in a given experimental condition are drawn from a population-level normal distribution. We could thus also obtain population-level parameter estimates for each experimental condition and age group, to assess whether they differed. We used Bayesian inference to investigate if there was an effect of age group and verbalization on memory. We based inferences on Bayesian hypothesis testing, which combines prior knowledge about the parameter space (the "prior") with knowledge about the parameter space after seeing the data (the "posterior"). Hence, Bayesian inference is not based only on the mean parameter estimate, but also its uncertainty (Kruschke, 2011). Parameter uncertainty is described by the 95% credible interval of the posterior distribution, in addition to the mean parameter value. To compare conditions, we used the Savage-Dickey density ratio, a method of obtaining the Bayes factor by dividing the height of the posterior by the height of the prior at the point of interest (Dickey, 1971; Gamerman & Lopes, 2006; Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010). Specifically, this provides a ratio of the likelihood of one hypothesis relative to some other hypothesis (e.g., the alternative hypothesis, cf. the null hypothesis: BF_{10}). For more details on how BFs are computed for between-subjects designs, see Hardman (2017).

Bayes factors cannot conclusively be interpreted using threshold cut-off points; therefore some subjective interpretation is inevitable when describing the result. Typically, $B = 1$ is considered 'no evidence', B between 1 and 3 is considered 'anecdotal' (Wetzels & Wagenmakers, 2012) or 'not worth more than a bare mention' (Jeffreys, 1961), and B greater than 3 is considered 'substantial'[8], between 10 and 30 'strong', 30 - 100 'very strong', and over 100 'decisive' evidence (Jeffreys, 1961; Wetzels & Wagenmakers, 2012). However, these labels are arbitrary (see Morey, 2015), so we applied them tentatively and encourage readers to evaluate the strength of evidence for themselves.

Results

The following analyses were pre-registered via the Open Science Framework [osf.io/m64px]. No participant had an average error distance more than 90 degrees on the color-wheel circle – which would indicate chance performance – hence no one was excluded and replaced for that reason. Verbal labeling strategies in the silence block were not reported to different extents by younger (76%) and older adults (70%); X^2 (1, $N = 60$) = 0.34, $p = .56$.)

---

[8] The common meaning of 'substantial' has changed over time (Morey, 2015).

Mixture Modeling

Model fitting. All our CatCont-models had age group (younger or older) and verbalization condition (silence, overt labeling or suppression) as factors. As specified in the pre-registration, we conducted separate models including either the error distance of (1) only the first-probed memory item or (2) all four items. The first-item analysis is similar to traditional visual WM tasks, only probing one item. In contrast, including all items tested the impact of labels despite interference and decay caused by previous responding. Due to word limit constraints, we focus on the traditional analysis in this paper (as it was of most interest for our hypotheses), while the analyses for all four items are presented in Appendix C.

We fit all models with three parallel chains of 10,000 iterations each, with a burn-in of 2000 iterations. Before running the models, we ensured that all Metropolis-Hastings acceptance rates were in the acceptable range (about 0.4 to 0.6; see Hardman, 2017), by adjusting the Metropolis-Hastings tuning values and re-running the parameter estimation. How colors were assigned to categories (category selectivity, σS) and imprecision with which participants selected categories (σA); accounting for motor noise were fixed across verbalization conditions (Souza & Skóra, 2017). However, we allowed these parameters to vary between the age groups. Similarly, the probability of categorical guessing (P^{AG}) was fixed across verbalization conditions (similar to Souza & Skóra, 2017), but allowed to vary between age groups, to allow for the possibility that younger and older adults may rely on such guessing to different extents. Souza and Skóra (2017) collected information about the numbers of categories participants used by recording participants labeling out loud. They compared using that maximum category number with letting the model freely estimate the number of categories for each participant and found similar results. Here, we did not record labels, so we let the model freely estimate the number of categories and their means for each individual (using the default maximum number of 16 categories; Hardman, 2017).

First, we assessed the fits of the two types of CatCont models, the *between-item* model variant (models an individual response as based on *either* a continuous or categorical representation) and the *within-item* model variant (models both kinds of representations as available and combined to produce responses; see also Bae et al., 2015; Donkin et al., 2015). We compared model fit of CatContModel variants using the Watanabe-Akaike Information Criterion (WAIC), as recommended by Hardman (2017). The between-item model had a smaller WAIC than the within-item model (Δ = -190.7), indicating a better fit to the data. Therefore, we only discuss the results of this model. See Appendix C for output from the within-item models.

Memory Performance: Parameter Estimates

Figure 4.3 shows the group-level probability that the first-presented items were in memory (panel A), the group-level probability that they were stored continuously (panel B), and the group-level imprecision of continuous representations (panel C), by age group and verbalization condition. See Table 4.1 for BF_{10}s for the factors and interactions, and Table 4.2 for BF_{10}s for all subset comparisons.

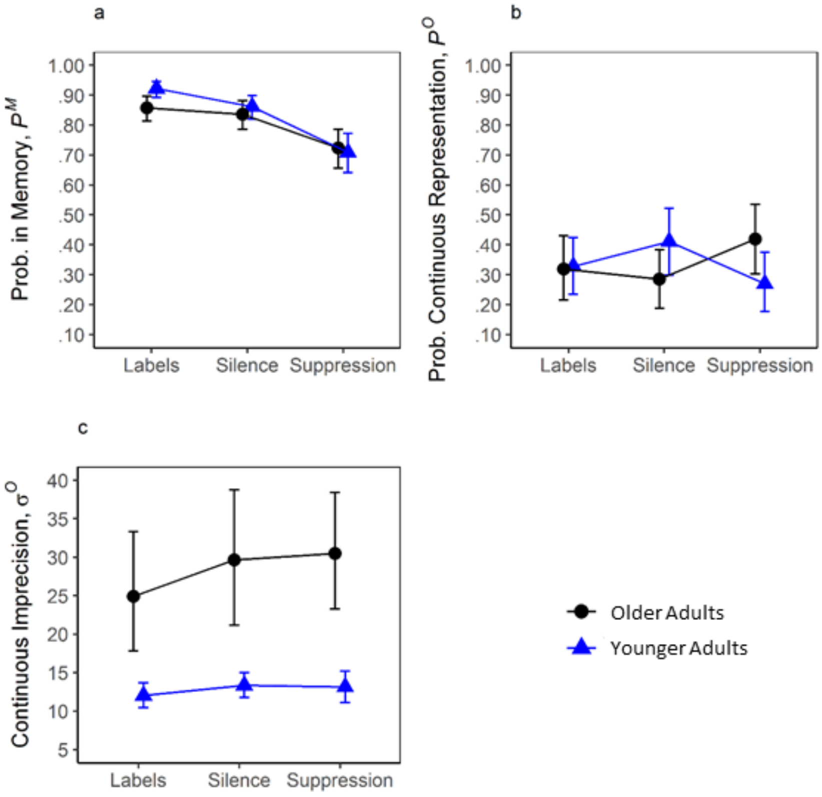


Figure 4.3. Memory for the First Item Only. Panel a. The group-level probability of having the probed item in memory. Panel b. The group-level probability that memory representation is continuous. Panel c. The imprecision of the group-level continuous memory representation.

Table 4.1

BF_{10}s for the effects of the experimental factors.

Predictor	**Parameter**		
	Probability memory (P^M)	**Probability continuous (P^O)**	**Continuous imprecision (σ^O)**
First-presented item only			
Age Group	.16	0.090	4.35×10^3
Verbalization	4.84×10^5	.018	2.58
Age Group × Verbalization	1.53	3.50	.354

Table 4.2.

BF_{10}s for the effects of subset analyses of the verbalization manipulation.

Predictor	**Parameter**		
	Probability memory (P^M)	**Probability continuous (P^O)**	**Continuous imprecision (σ^O)**
First response only			
Silence vs. Suppression			
Age Group	.079	.10	7.90×10^3
Verbalization	9.98×10^6	.13	.38
Age Group × Verbalization	.045	2.10×10^2	.15
Silence vs. Labeling			
Age Group	0.39	0.13	407.05
Verbalization	42.03	0.079	3.02
Age Group × Verbalization	.14	0.053	0.20
Suppression vs. Labeling			
Age Group	.208	.18	5.76×10^3
Verbalization	3.11×10^{11}	.16	5.55
Age Group × Verbalization	29.06	.90	1.85

There was no evidence here that the probability that the items were remembered (P^M) differed in the age groups. There was a large effect of verbalization and weak ‘anecdotal’ support for an Age Group × Verbalization interaction. However, our hypothesis that older adults would be comparatively more impaired under suppression was not supported, as suppression impaired the younger adults’ performance comparatively more (see Figure 4.3; also confirmed by subset analysis contrasting suppression with labeling; Age Group × Verbalization $BF_{10} = 29.06$; see Table 4.2). There was no main effect of age or verbalization on the probability of continuous (as opposed to categorical) responding. However, there was some evidence for an interaction between age group and verbalization ($BF_{10} = 3.50$), suggesting that the verbalization instructions affected the proportions of continuous responding differently in the two age groups. There was a 'decisive' main effect of age on precision ($BF_{10} = 4.35 \times 10^3$), such that the older adults' responses were less precise. The effect of verbalization on precision was 'anecdotal', with no evidence for an interaction with age.

WM Capacity: Categorical vs. Continuous K

We also calculated estimated categorical vs. continuous K, for each age group and verbalization condition. Categorical K ($P^M \times [1 - P^O] \times Set\ Size$) is the estimated capacity for Categorical representations, while Continuous K ($P^M \times P^O \times Set\ Size$) is the estimated memory capacity for continuous information, in a given condition. With these estimates, we tested whether labeling was associated with greater continuous or categorical capacity – or both – and distinguished among different hypotheses outlined above regarding the labeling benefit in the two age groups: (1) *The categorical visual LTM hypothesis*: labeling increases both continuous and categorical Ks, as well as total K. (2) *The* *verbal recoding hypothesis:* labeling results in recoding of visual information to a verbal trace, which is used instead of the visual representation, resulting ideally in all categorical K and no continuous K, with no change in total K. (3) *The Dual-Trace (Visual + Verbal) Hypothesis*: labeling provides a second, categorical and verbal trace, resulting in increased categorical and total K, but no change in continuous K. See Fig 7 for estimates of total, categorical and continuous K. Posterior differences in categorical and continuous K by verbalization condition and age group are presented in Figure 4.4. To test differences in labeling in the age groups compared to silence, we compared *Prevented Labeling* (Silence vs. Suppression*)* and *Enforced Labeling* (Silence vs. Labeling). Finally – to avoid potential confounds of age-related differences in spontaneous sub-vocal rehearsal of labels when performing the task in silence we quantified *The Labeling Benefit* by comparing Labeling vs. Suppression.

Preventing Labeling (Silence vs. Suppression)

There was a 'decisive' main effect of suppression on the probability of remembering (P^M), such that participants remembered better when doing the task in silence than under suppression ($BF_{10} = 9.98 \times 10^6$, see Table 4.2). This was true for both age groups, with no evidence of interaction. There was no clear main effect of suppression on the probability of having a continuous representation (P^O), but there was 'decisive' evidence for an interaction with age group ($BF_{10} = 210.36$). Specifically, in silence, younger adults were more likely to have continuous representations than older adults, but under suppression, this pattern was reversed (see Figure 4.3). Categorical memory representations in older adults' were credibly reduced by suppression (*M* = - 0.71 items), but not in younger adults (*M* = - 0.04 items). Continuous K did not change credibly under suppression in older adults (*M* = + 0.26 items), but in younger adults it was credibly reduced under suppression (*M* = - 0.65; see Figure 4.4). Suppression did not have a conclusive effect on precision (BF_{10} = .38), and did not appear to affect precision differently in younger and older adults’ (BF_{10} = .15).

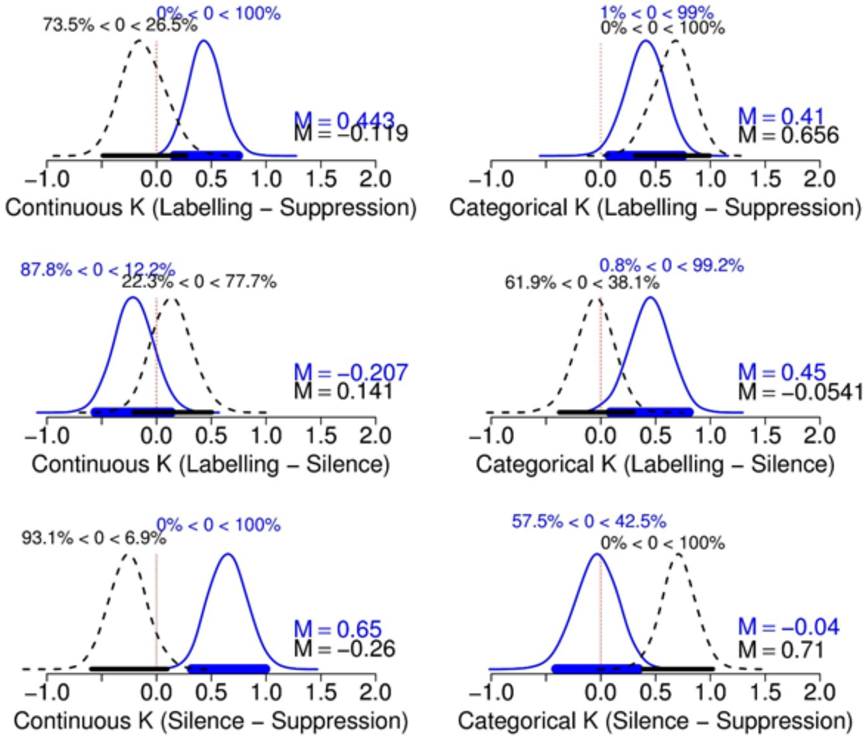


Figure 4.4. Memory for the First item only. Posterior differences in continuous and categorical K for specified comparisons. Mean values (M) larger than 0 for condition (A – B) indicates larger estimates in condition A than B. Each panel presents the percentages of the curves that are above and below 0 (null effect), the means (M), and the 95% credible intervals of the means (bars underneath each curve). Older adults in dotted black, younger in solid blue.

Enforcing Labeling (Silence vs. Labeling)

Overt labeling improved memory (P^M) compared to performing the task in silence (BF_{10} = 42.03), in both age groups (there was no evidence for an interaction with age; BF_{10} = .14). It did not influence the probability of having a continuous representation of the first item (P^O), regardless of age group (see Table 4.2). Labeling did not produce a credible change in continuous K in either age group (see Figure 4.4).

However, compared to spontaneous performance in silence, younger adults' categorical memory representations increased credibly when instructed to label (*M* = + 0.45 items). In contrast, older adults' categorical memory capacity under instructed labeling did not differ credibly from their performance in silence (see Figure 4.5). Furthermore, there was some evidence that overt labeling increased precision (σ^O) for the first item (BF_{10} = 3.02), but not to different extents in the age groups.

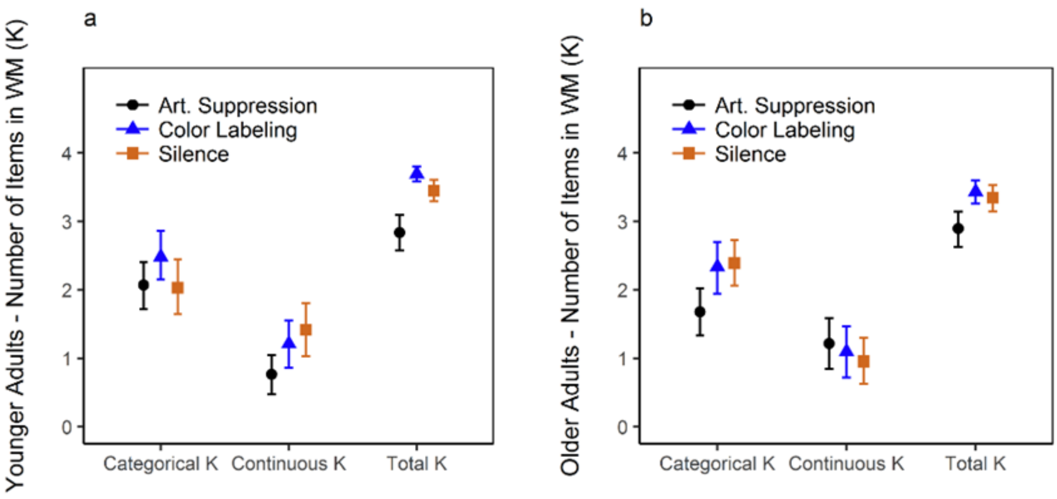


Figure 4.5. Memory for the first item only. Categorical, Continuous and Total K, by age group and verbalization condition.

The Labeling Benefit (Labeling vs. Suppression)

Overt labeling improved memory (P^M) in both age groups compared to suppression (BF_{10} = 3.11 × 10^{11}). There was no evidence of any effect of labeling on the probability of continuous responding (P^O). However, the benefit associated with labeling was due to a boost in categorical K in both age groups (younger $M = +0.41$, older $M = +0.65$ items). However, while this was the only source of the boost in older adults, the younger adults’ continuous K also increased credibly ($M = +0.44$ items). Thus, similar to Souza and Skóra (2017), our younger adults’ labeling benefit fit best with *the Categorical Long-Term Memory* hypothesis. In contrast, the older adults’ gain fit best with *the Dual-Trace hypothesis*. Furthermore, there was some evidence that suppression reduced precision (σ^O) for the first item ($BF_{10} = 5.55$), but unclear whether this occurred to different extents in the age groups (Age Group × Verbalization; $BF_{10} = 1.85$).

Consistency Check

Finally, we analyzed data from only those participants who completed the silence block before being introduced to the overt labeling manipulation ($N = 32$), to test whether results were driven by participants changing how they performed the task in silence after exposure to instructed labeling. Results generally appeared similar, apart from no evidence for age difference in precision, and less evidence for an Age Group × Verbalization interaction for Suppression vs. Silence comparison for continuous responding (P^O). However, categorical vs. continuous K comparisons were similar to the original results (see Appendix C), suggesting that being exposed to the labeling instruction may have increased age group differences, but differences were still present in participants who were unaware of the labeling instruction.

Discussion

Following evidence that verbal labeling improved visual WM performance by boosting the number of categorical and continuous representations, as well as precision of continuous representations in young adults (Souza & Skóra, 2017), we investigated if labeling would have a similar effect in healthy older adults. Evidence of relatively intact verbal WM in older adults (e.g., Jenkins et al., 2000; Park et al., 2002) and suggestions that participants of different age groups may rely on different cognitive capacities to perform the same task (Johnson et al., 2010; Reuter-Lorenz, 2002) made us question the 'Dull Hypothesis' that older adults perform just like younger adults, but more poorly. We addressed the following questions: Do older adults (1) Spontaneously use verbal labels more when performing the task in silence? (2) Depend more on verbal labels for visual memory performance? (3) Benefit from verbal labels in the same way as younger adults?

Spontaneous Use of Verbal Labels

We tested if younger and older adults differed in the extents to which they spontaneously applied sub-vocal rehearsal to this visual WM task using the following logic: if participants used verbal labels in silence, performance during silence and labeling should be similar, but different under suppression (when labels cannot be used). Hence, if the proportion of categorical and continuous representations differs from that in silence under either verbalization instruction (labeling or suppression), it suggests that the manipulated condition differed from spontaneous performance.

Evidence that preventing sub-vocal labeling (Silence vs. Suppression) affected the age groups' probabilities to respond continuously differently was 'decisive' ($BF_{10} = 207.94$). Specifically, compared to spontaneous performance in silence, younger adults' categorical memory representations increased credibly when instructed to label, but were not reduced by suppression – suggesting that they were not consistently labeling sub-vocally in silence. In contrast, older adults' categorical memory capacity in silence did not differ credibly from their performance under instructed labeling (see Figure 4.4) but was poorer under suppression – consistent with spontaneous sub-vocal labeling in silence. In sum, the older adults lost categorical representations under suppression, while the younger adults gained categorical representations when instructed to label, compared to silence. Combined, these observations suggested that older adults spontaneously (i.e., in the silence condition) used verbal labels to maintain coarse, categorical representations more than the younger adults, and furthermore that these representations were maintained via sub-vocal rehearsal, since suppression reduced them.

These different tendencies to rely on different types of representations in silence were not detected when comparing the overall memory performance between age groups (P^M; which combines both continuous and categorical representations; Table 4.1). Moreover, similar proportions of younger and older adults reported using verbal strategies in the silence condition. Hence, these age differences likely go unnoticed in visual WM tasks.

Previous research has found that younger participants can control the trade-off between quality and quantity in some visual WM tasks via verbal encoding (e.g. Ramaty & Luria, 2018; Zhang & Luck, 2011). If reliance on such verbal encoding differs systematically between age-groups – as our results indicate – this could be problematic for age-comparisons in a variety of visual WM tasks, for instance paradigms measuring visual feature-binding, if remembering individual features lends itself to such labeling and remembering bindings does not (Brockmole & Logie, 2013).

More broadly, endeavors to measure neural states that correspond to mental codes are central to hypotheses in cognitive psychology (Haxby et al., 2001; Haynes & Rees, 2006; Lewis-Peacock & Postle, 2012; Lewis-Peacock et al., 2014). However, tasks based on the same visual stimuli have been observed to elicit different activity depending on which strategy participants were instructed to use (Decety et al., 1997). Older adults appear to activate less, more, or even different neural structures than younger adults when performing a memory task (see Cabeza, 2002; Park, Polk, Mikels, Taylor, & Marshuetz, 2001), thought to reflect compensatory recruitment (Cabeza, 2002; Cherry, Park, & Donaldson, 1993; Park et al., 2001). Our results highlighted the importance of establishing the extents to which differences are driven by age-related strategic preferences in approaches to visual memory tasks, and that such differences may be detected using mixture modeling combined with explicit labeling manipulations.

Do Older Adults Depend More on Verbal Labels for Visual Memory?

We had hypothesized that older adults' memory (P^M) would be comparatively more impaired when we prevented labeling, i.e., during concurrent suppression compared to the two other conditions. This was not confirmed. Instead, we observed strong evidence that suppression impaired the younger adults' performance comparatively more when contrasting it with labeling. This Age × Labeling comparison leaves it unclear whether younger adults were able to benefit more from labeling, or were comparatively more disrupted by suppression. Either way, these results contradicted our hypothesis that older adults depend on an intact verbal store to compensate for reduced visual memory capacity, since if so, their overall performance should have deteriorated more than that of the younger adults under suppression.

Moreover, older adults' continuous representations were less precise than the younger adults'. Our results indicated that age-related decline in precision was not simply due to greater reliance on categorical, verbal representations in older adults, since we observed a substantial age-related decline in the precision of continuous representations even when categorical representations were separated out – supporting the notion that declining visual WM precision is an important feature of cognitive aging (Peich et al., 2013). However, we did not find strong evidence for an age effect on precision in the consistency check analysis, which only included participants who had not been exposed to the overt labeling condition prior to the silence block. Bayes factors close to 1 may suggest insufficient data in this reduced sample size (see Dienes, 2014).

Do Older and Younger Adults Benefit from Verbal Labels in the Same Way?

To measure memory gain associated with labeling in the two age groups without potential confounds of personal or age-related preferences, we compared enforced labeling with prevented labeling (suppression) and found that older adults appeared to benefit differently from verbal labels. We compared the influence of labeling on three aspects of memory performance: categorical representations, continuous representations, and increased precision of continuous representations. Souza and Skóra (2017) observed that verbal labels improved all three in young adults. They concluded that labels boosted memory by activating categorical visual LTM, rather than simply improving memory by providing extra, verbal traces. Several other studies have also suggested that Categorical LTM (i.e. pre-existing visual representations; Brady, Konkle, & Alvarez, 2011) can boost visual (continuous) WM (Olsson & Poom, 2005; Alvarez & Cavanagh, 2004; Curby, Glazek, & Gauthierm, 2009; Buttle & Raymond, 2003; Curby & Gauthier, 2007; Sørensen & Kyllingsbæk, 2012, but see also; Pashler, 1988). For younger adults, we replicated these observations. In contrast, older adults did not appear to gain continuous representations when labeling. Instead, the additional information associated with labeling was primarily categorical, suggesting that older adults' gains associated with labeling were verbal in nature (consistent with *the Dual-Trace (Visual + Verbal) Hypothesis*).

However, relative to silence as a baseline, we found that overt labeling improved categorical capacity while suppression reduced continuous capacity in younger adults (see Figure 4.5). This highlighted a potential alternative explanation behind the labeling benefit in younger adults. Instead of verbal labels activating categorical visual LTM (Souza & Skóra, 2017), the greater continuous contribution to performance associated with labeling could be because suppression reduced visual memory capacity during encoding, perhaps by draining a general resource (Cowan, 2005; Ma et al., 2014). In contrast, according to the MCM framework, Suppression would not be considered a 'visual' WM task, and should only have a negative impact on visual WM performance to the extent that it prevents recoding of visual information into verbal (and the opportunity to rehearse such information). This highlights the difficulty associated with attempting to prevent labeling such that processing demands are equal between conditions, and is a limitation of this paradigm. However, referring to performance in silence when investigating the labeling benefit is problematic. Instructed labeling might disrupt other processes occurring in unrestrained conditions. For example, perceptual grouping based on what people might label as 'warm' or 'cold' colors might be one such process, which has been found to influence memory for individual items even in randomly selected to-be-remembered arrays of colors (Alvarez, 2011; Brady & Alvarez, 2011). Overt labeling disrupting some other process would explain why labeling did not improve continuous memory compared to silence in our younger adults. Either way, it appeared that labeling – despite being very beneficial for overall memory in both age groups – affected the types of representations held in memory differently.

Limitations

All parameter estimates presented in this paper depend on the assumptions of the CatCont Model (Hardman, 2017). It is possible that other processes which contribute to responses in the delayed estimation task (e.g. perceptual grouping processes) may not be adequately captured by model. This is a limitation of this research. However, the CatCont model appears to more adequately fit data generated by participants than models which assume that all responses are continuous (see Hardman et al., 2017), and it enabled us to compare categorical vs. continuous representations in a way that would not be possible using a set of fixed, categorical to-be-remembered items.

General Discussion

A range of studies have shown that people can retain many mental codes in parallel (the ‘Multiple Encoding’ Hypothesis; e.g. Lewis-Peacock et al., 2014; Logie et al., 2016; Paivio, 1971; Wickens, 1973), and many cognitive theories explicitly model the multidimensional nature of memory representations. Researchers debate how flexibly resources can be shared across the visual and verbal/auditory modalities in memory tasks, as well as whether these modalities are actually distinct in memory: observations of clear capacity costs due to cross-modal sharing of resources (Morey & Cowan, 2004, 2005; Morey, Cowan, Morey, & Rouder, 2011; Saults & Cowan, 2007; Vergauwe, Barrouillet, & Camos, 2010) conflict with studies in which such costs were not found (e. g., Cocchini et al., 2002; Fougnie & Marois, 2011).

Introduction of delayed estimation tasks enabled fidelity-measures of visual WM representations and started intense debates about whether visual WM is best characterized as an information-limited system (Alvarez & Cavanagh, 2004; Wilken & Ma, 2004), or as limited by a pre-determined, fixed item limit

(Luck & Vogel, 1997; Zhang & Luck, 2008). WM capacity has usefully been conceptualized as both the number of items that can be stored and the precision with which those items are stored – analogous to storing images on a USB-drive: you can store more images with low resolution or fewer images with very high resolution, given its finite volume (Brady et al., 2011). Our research adds to the body of research highlighting that verbal representations, either as ‘audio files’ or as ‘file names’ categorizing visual representations, also need to be acknowledged, and that they may be connected to visual items in complex ways. Participants may use one or other of these forms of representation according to their preferences and ability to use each of them (e.g., Logie, 2018). Indeed, preferences for the type of mental codes (or ‘file formats’) appeared to vary with age. In younger adults, verbal labels may be better conceptualized as file names, useful to open specific files from the hard drive (LTM knowledge), thus activating representations (e.g. how the color red looks), which can support this limited visual storage system (e.g. Olsson & Poom, 2005). In older adults, the benefit of verbal information appeared to act more like an audio file (i.e., maintaining a coarse representation, but not necessarily supporting activation of a stored visual representation). Our results highlighted challenges associated with attempting to study the visual system in isolation (e.g., separate from verbal labels) when comparing younger and older adults, and suggest that these challenges might be addressed by explicitly comparing instructed labeling with suppression, and modeling categorical and continuous responses.

Conclusion

At first glance, our results appeared consistent with the Dull Hypothesis (Perfect & Maylor, 2000); that older adults performed the memory task like less precise younger adults. Indeed, while we found no strong effect of age on the overall probability of remembering (continuous and categorical representations taken together), there were differences in precision, emphasizing its usefulness as a more fine-grained measure of the effects of aging. Older adults' overall memory was not more impaired when we prevented verbal labeling, which provided evidence against our hypothesis that older adults' visual memory performance *depends* more on verbal rehearsal.

However, older adults appeared to rely more on coarse, categorical representations than younger adults when spontaneously doing the task in silence. Furthermore, these representations appeared to be supported by sub-vocal rehearsal, since they were specifically reduced under suppression. Finally, while we replicated Souza and Skóra’s finding that labeling benefitted younger adults via activated visual categorical LTM, older adults did not appear to benefit in the same way.

People likely use different kinds of mental codes (e.g., visual and verbal) interchangeably to retain information in memory to navigate day-to-day situations. Different types of representations may rely on different stores or different rehearsal mechanisms. These results suggested that there are age differences which are not apparent when looking only at overall memory performance, which may be important in understanding age differences in more nuanced visual WM phenomena, such as feature-binding or brain activity, in future research.

Chapter 5: Strategic Mediation in Working Memory Training in Younger and Older Adults

Aims

So far, results in this thesis have suggested that older adults may approach certain visual memory tasks by using verbal rehearsal, and/or relying on categorical representations, to a greater extent than younger adults. The 'ANOVA approach' (as outlined in Section 1.7) has appeared useful to reject the 'Dull Hypothesis'.

For the final study, I stepped back from strategy differences in visual WM, and returned to my initial, broader research question: "Do older adults approach WM tasks differently?". In this final study, I explored strategic approaches by younger and older adults in a paradigm relevant to the WM training literature – a research area with substantial practical implications. Arguably, the main purpose behind cognitive ageing research is to *prevent* age-related decline (see Section 1.3.5); cognitive training, if it works, is a direct approach to this end. Moreover, older adults are perhaps more likely be exploited by commercial training paradigms, since they might worry about age-induced cognitive decline (e.g., Federal Trade Commission, 2016).

As outlined in Section 1.3.5; initially training seemed promising, but subsequent findings have curbed the initial enthusiasm regarding its effectiveness. There are suggestions that younger adults improve more with training than older adults (Burki et al., 2014; Dahlin et al., 2008; Heinzel et al., 2014; Zinke et al., 2013). Larger gains in younger adults are perhaps consistent with animal models suggesting that older age is associated with fewer neuroplastic changes (Blumenfeld-Katzir et al., 2011; van Praag et al., 2005). However, gains of similar magnitude on trained tasks in younger and older adults are sometimes found (e.g., Bürki et al., 2014; Li et al., 2008; Richmond et al., 2011; von Bastian et al., 2013; Zając-Lamparska & Trempała, 2016).

The potential role of strategy use in cognitive training was increasingly recognised while I was conducting research for this thesis. For instance, Laine et al. (2018) instructed participants to use a specific visualisation strategy during a single 30-minute N-back training session. This strategy was associated with significant N-back task improvements, compared to participants who were not instructed to use a strategy. However, learning about the strategy did not result in improved performance on other (structurally dissimilar) WM digit tasks. This highlighted that strategic changes – rather than increased WM capacity – may underpin successful WM training outcomes in younger adults (see also Dunning & Holmes, 2014; Soveri et al., 2017). Here, I tested whether this strategy would also be beneficial for older adults, and explored the role of age-related strategic differences in a training paradigm. Some research has found that instructing participants to apply a mnemonic technique or strategy appeared more beneficial for younger than for older adults (e.g., Brehmer, Li, Müller, von Oertzen, & Lindenberger, 2007; Lindenberger, Kliegl, & Baltes, 1992; Lövdén, Brehmer, Li, & Lindenberger, 2012; Verhaeghen & Marcoen, 1996; Verhaeghen et al., 1992; but see Gross et al., 2012). However, since the strategy required visualisation of stimuli, and visual abilities may decline more with age than verbal abilities (see Section 1.4.1, for a summary), one might expect older adults to benefit less. In contrast, other evidence suggests that older adults can benefit from switching from verbal to visual codes (e.g., Osaka, Otsuka, & Osaka, 2012).

This study was a conceptual replication of Laine et al. (2018), also including healthy older adults. The aims of this replication was to see if their findings regarding the impact of instructing this visualisation strategy in younger would replicate, and to see if older adults would benefit from the strategy. Also, it was a test of age-differences in spontaneous strategic approaches in the N-back paradigm, using a method which allowed us to classify the type of strategies (and how detailed they were), and test associations with memory performance. *At the time of writing, the following paper is under review in the Quarterly Journal of Experimental Psychology. The paper was produced in collaboration with Dr Daniel Fellman and Prof. Matti Laine. They provided the experimental program, an analysis script to replicate their original analysis, and comments and suggestions regarding the design and the write up of the manuscript.*

Abstract

Working Memory (WM) training is thought to improve cognitive capacity and general cognitive abilities (the Capacity Hypothesis of training). Here, we tested the rival Strategy Mediation Hypothesis of WM training, by examining whether post-training improvements in the trained task paradigm were influenced by strategies. This study was a systematic replication Laine et al. (2018), to test the validity of their results in a different sample of younger participants, and explore if the effect of strategy would generalise to healthy older adults.

Participants (N = 120) completed a set of WM tasks followed by a 30-minute N-back training session some days later. Half the participants were instructed to use a visualisation strategy, the others received no instruction. Participants then performed the WM tasks again. The pre-posttest battery encompassed a criterion task (digit N-back), two untrained tasks N-back tasks (letters and colours), and three structurally different WM tasks.

The instructed visualisation strategy significantly boosted N-back performance in participants of both age groups, although the strategy appeared more difficult to implement for older adults. However, the strategy did not improve performance on structurally different WM tasks. We also found significant associations between N-back performance and the type of and level of detail of self-generated strategies in the uninstructed participants, as well as age group differences in reported strategy types. WM performance appeared to

partly reflect the application of strategies, and strategic mediation should be considered to understand the mechanisms of WM training. Claims of efficient training should demonstrate useful improvement beyond task-specific strategies.

Introduction

Working Memory (WM) refers to cognitive functions that support the ready availability of a small amount of information on a temporary basis while we undertake ongoing actions and mental activities (e.g., Logie & Cowan, 2015). WM is viewed as a core mechanism underpinning higher-order cognitive abilities such as perception and problem-solving (Ma, Husain, & Bays, 2014), and is related to fluid intelligence (Kane, Hambrick, & Conway, 2005; Unsworth, Fukuda, Awh, & Vogel, 2014), reasoning ability (Conway, Kane & Engle, 2003; Kyllonen & Christal, 1990), and measures of cognitive control (Conway, Cowan, & Bunting, 2001; Kane & Engle, 2003; Redick, Calvo, Gay, & Engle, 2011). WM also suffers pronounced, linear decline during adult ageing (Bopp & Verhaeghen, 2005; Borella, Carretti, & De Beni, 2008; Park & Payer, 2006), although some aspects of WM decline faster than others; verbal WM appears least susceptible, and visuo-spatial most susceptible to age-related decline (Johnson, Logie & Brockmole, 2010; Park et al., 2003). Functioning of WM abilities is important for the autonomy and wellbeing in older adults (Tomaszewski Farias et al., 2009).

Hence, when early studies suggested that repeated adaptive WM training could protect older adults from cognitive decline (e.g., Brehmer, Westerberg & Bäckman, 2012), there was great interest (von Bastian & Oberauer, 2014; Green & Bavelier, 2008; Klingberg, 2010; Lövdén, Bäckman, Lindenberger, Schaefer, & Schmiedek, 2010; Morrison & Chein, 2011), due to the potential benefits to public health and well-being.

The idea of improving cognition is not new (see Logie, 2012). In the late 19th century, Ebbinghaus (1885/1964) demonstrated that repeated learning of lists of non-words improved memory performance for the materials being learned. Numerous studies since have investigated the impact of cognitive training (e.g., Fabiani et al., 1989; Gopher et al., 1975; Wickens & Weingartner, 1985). For example, one participant (Faloon) practiced remembering strings of digits several hours a week for 20 months and was eventually able to remember up to 79 digits (Ericsson, Chase & Faloon, 1980). However, despite this extreme improvement in remembering digits, the participant's ability to remember letters remained at six letters. It appeared that the participant's general ability to remember strings of information did not improve despite the extensive digit training (for a recent example, see Ericsson et al., 2017). The likely reason emerged when he reported developing his ability to chunk digits by relating them to semantic knowledge of distance-running times. The scientific consensus was that while practising resulted in improvements on practised tasks, it did not improve short-term memory or WM ability per se, or generalise to other domains of cognition.

Cognitive training was revisited several decades later, based on evidence of neurological plasticity related to cognition in both younger and older adults (e.g., Hertzog et al., 2008). The brain was likened to muscles, growing physically larger and stronger when repeatedly challenged at close to maximum currently manageable difficulty (i.e., adaptivity). Based on this analogy, researchers proposed that such challenging training of WM increases WM capacity (e.g., Morrison & Chein, 2010) by eliciting functional and anatomical changes in the brain (Dahlin et al., 2008). Such changes, they suggested, may help preserve brain integrity as we age, and produce lasting improvements in fluid intelligence, if WM and fluid intelligence rely on a shared capacity constraint (Halford, Cowan, & Andrews, 2007). The attractive idea of increased WM capacity as a result of training has been referred to as the *Capacity Hypothesis* of WM training (Peng & Fuchs, 2017). If training improves cognitive functioning (capacity) beyond a specific task – such as memorising digits – training benefits should generalise to other cognitive tasks due to the strong relationship between WM and other cognitive activities (e.g., Daneman & Merikle, 1996). The distinction between *near-* and *far-transfer* (see the taxonomy proposed by Noack, Lövden, Schmiedek & Lindenberger, 2009) is therefore crucial to the debate on the efficacy of WM training. *Near-transfer* indicates improvements on tasks very similar to the trained task itself. In contrast, to demonstrate *far-transfer,* training should improve performance on for instance measures of fluid intelligence, or reasoning tasks that, crucially, are quite unlike the trained task. Recently, some authors (e.g., de Simoni & von Bastian, 2018) have separated the near transfer domain into two categories according to the similarity of the tasks to the trained WM task, namely *task-specific near-transfer* and *task-general near-transfer*. Task-specific near transfer refers to improvements in WM tasks sharing the same task paradigm with the trained task, whereas task-general near transfer refers to improvements in WM tasks that are structurally dissimilar to the trained task. Failure to separate these two types of near transfer might make near-transfer effects seem broader than they actually are (see Soveri, Antfolk, Karlsson, Salo, & Laine, 2017 for a re-analysis of Melby-Lervåg et al., 2016), or may obscure task-specific near-transfer effects.

The new generation of cognitive training studies focused specifically on WM tasks and made more extensive use of computerised methods. Initially, this new WM training appeared promising. Several early studies found that training improved performance even on untrained, quite different cognitive tasks in healthy adults (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008) and children with ADHD (Klingberg, Forssberg, & Westerberg, 2002). In addition, several commercial companies have promoted WM training software as a scientifically supported approach to increase IQ (Mindsparke, 2011), improve grades (Jungle Memory, 2011), and reduce day-to-day lapses of attention (Cogmed, 2011).

However, subsequent research in healthy children and younger adults moderated these claims. With more appropriate experimental controls, it appeared that WM training typically improved performance on the trained task itself, and performance on other verbal and visuospatial WM tasks, whereas far-transfer effects to reasoning, or fluid intelligence were at most small (for comprehensive meta-analyses see; Melby-Lervåg & Hulme, 2013; Melby-Lervåg, Redick, & Hulme, 2016; Schwaighofer, Fischer & Buhner, 2015; Weicker, Villringer, & Thöne-Otto, 2016). Evidence regarding the effects of training in older adults is also mixed. A meta-analysis of 13 studies indicated that WM training in healthy older adults produced both large near- and far-transfer transfer effects (Karbach & Verhaeghen, 2014). However, when Melby-Lervåg et al. (2016) replicated the meta-analysis only including studies which compared the trained group to active controls and controlled for baseline differences, they found much smaller effects of training than originally reported. Moreover, in a recent meta-analysis of the commonly used N-back WM training by Soveri et al. (2017), the only reliable effects following WM training were seen in task-specific near transfer measures, that is, in tasks that were structurally similar to the trained WM task(s). In general, meta-analyses with less stringent inclusion criteria typically find both near- and far-transfer effects in older adults (e.g., Chiu et al., 2017). It has been difficult to reach consensus regarding the effects of cognitive training due to variations in training paradigms and in what is considered an appropriate control group (see Melby-Lervåg & Hulme, 2013; Melby-Lervåg, Redick, & Hulme, 2016; Morrison & Chein, 2011; Shipstead, Redick, & Engle; 2010; Shipstead, Redick, & Engle, 2012). Underpowered studies, confirmation bias, bias for publishing statistically significant (especially positive) results, and placebo effects may all inflate the apparent effect of training (see also Sala, Aksayli, Semir, Gondo, & Gobet, 2018; Sala & Gobet, 2017). Indeed, one review suggested that training effect size was inversely related to the rigor of study design (Simons et al., 2016).

In addition to methodological inconsistencies, different theoretical perspectives may contribute to confusion in the literature. Some theories propose that online cognition is limited by the capacity of a domain-general attentional resource or WM system (Engle & Kane, 2004), which can be increased by WM training, thus enhancing general cognitive abilities (Jaeggi et al., 2008). For example, the amount of information WM can retain and manipulate is thought to constrain 'fluid' intelligence, as measured by Raven's Progressive Matrices (Jaeggi et al., 2008). According to the *Capacity Hypothesis* of WM training (Peng & Fuchs, 2017) cited above, WM training should improve a general mental WM workspace, and thus perhaps result in improved performance on such measures of 'fluid' intelligence.

In contrast, other theories view WM as involving a variety of cognitive systems, among which participants select according to task demands (Baddeley & Logie, 1999; Logie, 2011; Logie & Niven, 2012). For instance, one system may retain phonological codes, another visual codes. When tasked to remember sets of digits, participants may remember them phonologically, by their visual shapes, or using a semantic memory strategy. Therefore, performance may reflect use of different cognitive resources in different participants (Logie, Della Sala, Laiacona, Chalmers, & Wynn, 1996, Logie, Pernet, Buonocore, & Della Sala, 2011; Logie, 2018; Johnson et al., 2010; Thurstone, 1931), and crucially, participants may change how they approach a task as they see how well any approach works with repeated exposure, or as a result of explicit instruction. Training thus might improve one particular cognitive skill, or lead to strategic recruitment of a different cognitive mechanism, with potentially different implications for transfer to other tasks. Based on studies that had indicated improved Raven's Matrices performance following training with the commercial Cogmed WM training program (Roughan & Hadwin, 2011), Shipstead et al. (2012) suggested that this might occur because this test used to measure fluid intelligence requires visual processing and matching very similar to the tasks trained with Cogmed. Thus, WM training may improve specific abilities, rather than improving some underlying intelligence 'capacity'.

Typically, adaptive training (i.e., tasks get harder as the participant improves) is associated with significantly better performance improvement than non-adaptive training (i.e. performing the task at a consistent level of difficulty; e.g. Holmes, Gathercole, & Dunning, 2009; Klingberg et al., 2005; Olson & Jiang, 2004; Thorell et al., 2009), and is seen as a key ingredient of effective training. Interestingly, some evidence suggests that adaptive training may also affect strategy use. Post-training interviews following Cogmed training indicated that participants in an adaptive training group reported using grouping strategies significantly more than did active and passive control group participants. This was associated with larger performance gains in some of the post-tests (Dunning & Holmes, 2014). This suggested that adaptive training may be comparatively more beneficial because participants are encouraged to develop new strategies as the task gets more challenging.

Laine et al. (2018) proposed and explicitly tested one aspect of this, the *Strategy Mediation Hypothesis* of WM training: that task-specific near-transfer gains are driven by developing and using a task-specific strategy during training. In younger adults, they used the N-back training paradigm (Kirchner, 1958) in which participants see an ongoing string of individual stimuli (e.g., digits) stream on a computer screen. They indicate whether each stimulus is identical to that presented *n* items back. Laine et al. (2018) instructed some participants to use a particular visualisation strategy during a single 30-minute N-back training session. This strategy instruction resulted in significant improvement in the trained N-back task (i.e., digits), and in two untrained N-back tasks using different stimuli such as letters or colours, compared to participants not instructed to use any particular strategy. Furthermore, the level of detail and type of self-generated N-back strategies reported by the uninstructed participants was significantly related to their post-test N-back performance. The results in Laine et al. (2018) provided strong evidence for the Strategy Mediation Hypothesis, supporting suggestions that strategic changes rather than increased WM capacity may underlie successful WM training outcomes (Dunning & Holmes, 2014; Soveri et al., 2017).

However, the Strategy Mediation and Capacity Hypotheses are not mutually exclusive. While associations of performance gains with strategies provide support for the Strategy Mediation Hypothesis, they do not rule out the possibility that training increases actual capacity of some sort. Laine et al.'s (2018) finding that practising with a strategy for 30 minutes resulted in gains equivalent to those typically observed after five weeks of N-back training did indicate that for training studies to be taken seriously, they should also demonstrate that trained participants developing a task-specific strategy cannot alone explain improved performance. For instance, the strategy of visualising digits used by Laine et al. may be unlikely to improve general reasoning or prevent age-related cognitive declines, but it did appear to boost N-back performance greatly. Establishing the mechanisms behind training-induced performance improvements is crucial to determining whether the intended cognitive improvement has occurred, and what factors might have led to any such improvement.

Moreover, important findings should be replicated, ideally in a different lab and with a different participant sample (see Simons, 2014). Therefore, in the present study, we conducted a systematic replication of Laine et al. (2018) in a different country, using an online methodology, and unlike that previous study, also recruited healthy older adults. Evidence suggests that older adults are not merely like poorly performing younger adults (e.g., Perfect & Maylor, 2000; Rabbitt, 2005). Instead, as noted earlier, different cognitive abilities appear to decline at different rates, and younger and older adults may use different cognitive resources when performing the same cognitive task (e.g., Johnson et al., 2010). Therefore, it is unclear whether Laine et al.'s (2018) visualisation strategy would be equally efficient in older adults, and whether non-instructed older adults would make different strategic choices than younger adults. However, healthy older adults are a target group for training, and, given that they might be worried about cognitive decline (e.g., Federal Trade Commission, 2016), are perhaps more likely to be targets for commercial training packages. Similar to the original study, our purpose was not to falsify the Capacity Hypothesis. Instead, we tested the Strategy Mediation Hypothesis by investigating the roles strategy use can play in these tasks, in order to further explore its role as one possible source for WM training outcomes. Specifically, our research question was: what are the effects of instructed and self-generated strategy use on WM updating performance, in healthy younger and older adults? These were our hypotheses:

H_1: Explicit instruction of a visuospatial grouping and comparison strategy in a digit N-back task will improve N-back performance in (1) the trained task, and (2) in untrained N-back tasks employing different stimuli (letters, colours) in younger adults (directional; replication of findings in Laine et al., 2018)

H_2: Explicit strategy instruction will affect post-test performance in healthy older adults, to the same extent as in younger adults (see H_1)[9]

H_3: Reported self-generated strategies (in the non-instructed group) will be associated with better memory performance on (1) the trained N-back task, (2) in untrained N-back tasks employing different stimuli (letters, colours) in younger adults (directional; replication of findings in Laine et al., 2018)

H_4: Similar effects of self-generated strategies will be observed in the older adults as in younger adults (see H_3)

---

[9] Both H_2 and H_4 were phrased as Null (or 'Dull') Hypotheses due to lacking background evidence; the strategy instruction was novel and to my knowledge, older adults' ability to visualise and compare strings of information in this context had not been tested before.

The hypotheses, methods and analyses were pre-registered via the Open Science Framework [https://osf.io/npzkc].

Method

Participants

Our pre-registered target sample size was 60 younger and 60 older adults. These numbers ensured power of at least .95 to detect a medium effect of strategy condition on the trained N-back digit task, and power of .80 to detect near-transfer to other N-back tasks, determined by a power analysis using G*power (Faul, Erdfelder, Lang, & Buchner, 2007), based on effect sizes in the study we aimed to replicate (Laine et al., 2018)[10]. We recruited a total of 136 participants: 74 younger adults who were students or former students at the University of Edinburgh, and 62 older adults who were members of a Participant Volunteer Panel, or a life-long learning group. Two older and 13 younger adults were excluded and replaced for failing to complete all three sessions. We excluded one younger participant who reported using pen and paper in the memory tasks, and one who completed the first session twice. The final sample consisted of 60

---

[10] Power analysis based on the reported effect sizes in Laine et al. (2018): main effect of strategy condition on the trained digit task; η_p^2 = 0.23. The weakest significant η_p^2 = 0.15 in the post-test (untrained letter N-back). To replicate the former (0.95 power) we need 46 younger adults, and the latter (0.80 power), need 47 younger adults. We recruited 60 participants in each age group to increase power for age comparisons as much as possible within research budget limitations.

younger adults (*M* = 22.50, *SD* = 3.50 years), and 60 older adults (*M* = 69.30, *SD* = 5.46 years). All older adults had either scored above the recommended threshold for cognitive impairments (Addenbrooke's Cognitive Examination; ACE-III; Hodges & Larner, 2017; Mioshi, Dawson, Mitchell, Arnold & Hodges, 2006) within two years prior to participating, or scored over the recommended threshold for their ages on the TICS™ (Telephone Interview for Cognitive Status™; Brandt & Folstein, 2003) within two weeks of participating in this study. Before starting the study, all participants did a red-green colour vision test. See Table 5.1 for participant demographics. No participants were excluded for being multivariate outliers at pre-test (using Mahalanobis distance value; Tabachnick & Fidell, 2007). The PPLS Research Ethics committee approved this research and participants received £15 each for participating.

Table 5.1

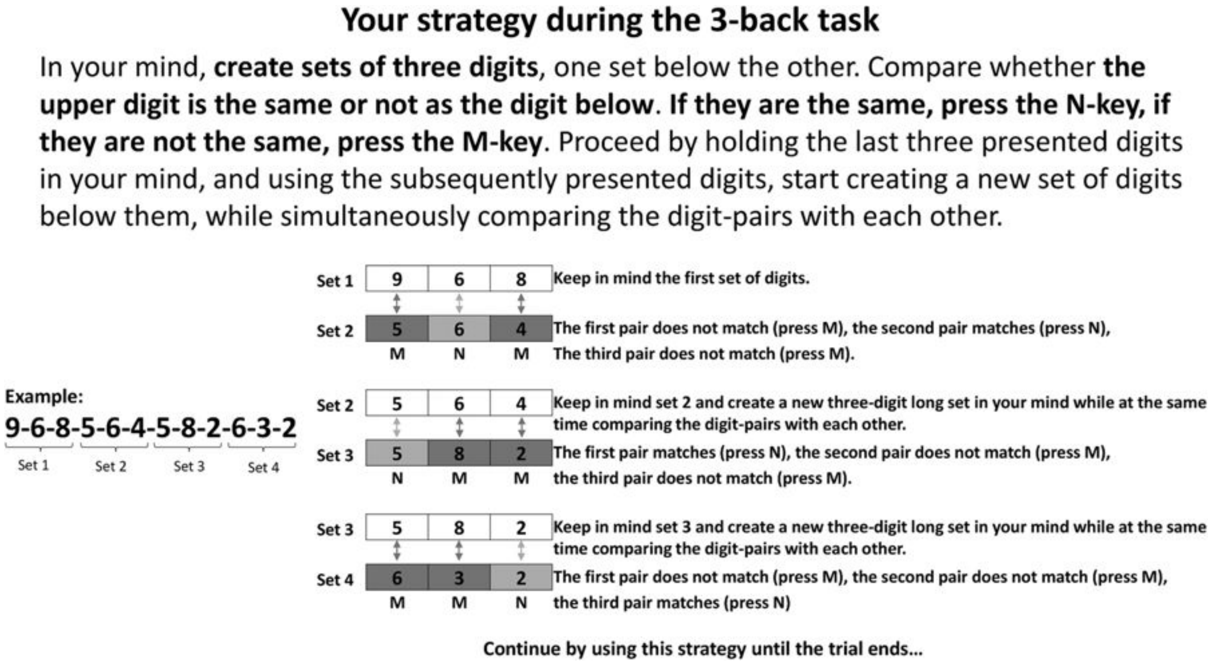
Demographics

	Younger Adults			Older Adults		
	Control	Strategy	*p*	Control	Strategy	*p*
N	29	25		30	19	
Age	23.0 (3.96)	22.3 (3.22)	0.497	70.3 (5.69)	66.6 (3.82)	0.015
Gender F/M	21/8	19/6	1	20/10	12/7	1
Education	16.2 (2.81)	15.9 (2.68)	0.715	15.5(3.43)	15.95 (2.5)	0.588
Pre-training N-back	0.28 (4.99)	- 0.2 (5.62)	0.747	0.61 (4.88)	-1.4 (5.29)	0.197

Note. Values in parentheses are standard deviations. *P*-values were calculated from t-tests for continuous variables and χ^2 test for gender. The N-back composite scores were the summed values of the z-transformations of the average and maximum level accuracy in the adaptive digit n-back task, and d-prime values and RTs for correct responses in the letter and colour N-back tasks.

Procedure

We used a mixed pre- and post-test intervention design. First, participants completed a set of cognitive tasks (taking 1 - 1.5 hours) to assess baseline abilities. Two days later, they did a 30-minute adaptive N-back task (*training session*). Half the participants from each age group were instructed to use a visualisation strategy (see Figure 5.1) during this training session (i.e., the strategy group), and the others performed the training without a strategy instruction (control group).



Your strategy during the 3-back task

In your mind, **create sets of three digits**, one set below the other. Compare whether **the upper digit is the same or not as the digit below. If they are the same, press the N-key, if they are not the same, press the M-key**. Proceed by holding the last three presented digits in your mind, and using the subsequently presented digits, start creating a new set of digits below them, while simultaneously comparing the digit-pairs with each other.

Example:

9-6-8-5-6-4-5-8-2-6-3-2

Keep in mind the first set of digits.

The first pair does not match (press M), the second pair matches (press N),
The third pair does not match (press M).

Keep in mind set 2 and create a new three-digit long set in your mind while at the same time comparing the digit-pairs with each other.

The first pair matches (press N), the second pair does not match (press M),
the third pair does not match (press M).

Keep in mind set 3 and create a new three-digit long set in your mind while at the same time comparing the digit-pairs with each other.

The first pair does not match (press M), the second pair does not match (press M),
the third pair matches (press N)

Continue by using this strategy until the trial ends...

Figure 5.1. The visualisation strategy instructions for participants in the strategy groups during training.

Two days later, participants completed the same set of cognitive tasks as on day one. All participants were instructed to complete the pre-test session on a Monday, the training session two days later, and the post-test on Friday in the same week. They received instructions and an access link by email each night before the next session. At least 24 hours elapsed between sessions, and we did not exclude participants who completed sessions on slightly different days. Participants were not aware of the purpose of the study, nor that some were instructed to use a strategy and others not. They were instructed not to discuss study details with others who may also wish to take part. When they had completed the study, participants filled out a strategy questionnaire, reporting if they had used strategies and if so to describe those strategies. Participants were then informed about the purpose of the study, and the existence of the different groups.

Our procedure differed from that of Laine et al. (2018) as follows. (1) In contrast to Laine et al. (2018), we did not include a passive control group that did not perform any training between pre- and post-test, because the central question concerned the presence or absence of strategy instruction. (2) While all their participants were younger-adult university-level students, we also included a group of older adults. (3) Their participants performed pre- and post-test sessions in the laboratory while our participants completed all sessions online. (4) Our instructions and tasks were in English, theirs in Finnish. (5) We did not screen participants for health conditions. Apart from these differences, our study was identical to theirs. We chose an online methodology because WM training software promoted by companies are typically intended for independent use with home computers or smartphones, and it enabled us to test a larger number of participants. However, there was a possibility of less attentive or compliant participants. To minimise the impact of this, we screened for outliers and asked participants if they used external tools (e.g. writing things down) when performing the tasks.

Working memory: Training task. The strategy and control groups performed the same digit N-Back training task, but the strategy group was instructed to use the strategy illustrated in Figure 5.1. Participants saw digits (1-9) displayed one at a time, in the centre of the screen. They responded to each digit with the N or M key on their keyboard (meaning Yes or No, respectively) to indicate whether the current digit corresponded to the digit presented *n* items back in the sequence. After receiving task instructions, the actual training task started. Each sequence began with a blank screen (450 ms), followed by a digit (1500 ms). Responses were recorded while the digit was on display or during the blank interval that followed. Hence, participants had a total of 1950 ms to respond to each digit.

Each participant completed 20 blocks of 20 + *n* trials. All participants started at the 1-back level. However, the training was adaptive, so if 18 to 20 responses in a block were correct, *n* increased by one in the next block. If 15 to 17 responses were correct *n* remained the same, but following less than 15 correct responses, *n* decreased by one (or remained at one) in the next block. Each block contained randomised digit sequences with the constraint that each sequence included six targets (i.e., the digit was the same as the one displayed *n* digits back) and 14 non-targets. To prevent responses based on familiarity – enabling correct rejection based on not seeing that digit recently – four items out of the 14 non-targets were lures, i.e., they were identical to a digit presented *n* ± 1 digits back (not applied to the 1-back condition). The maximum possible level was 9-back.

Strategy instruction. The strategy instruction taught participants to visualise the incoming *n* items as parallel digit strings (see Figure 5.1). For a 3-back sequence of 1-8-3-2-8-6, they would visualise 1-8-3 on top and 2-8-6 underneath. This strategy permitted visualised comparison of the upper and lower three digits, to judge whether they were identical. After comparing the two strings of digits the upper string would be discarded, and new digits were to be visualised as a new string, underneath. Participants in the strategy group were reminded of this strategy before each new block started.

Expectations. Prior to starting the training session, participants reported how much they thought they would improve on the training task during the session, using a 10-point Likert scale (1 = "*No improvement at all*", 10 = "*A large improvement*"). Participants in the strategy group were informed about the strategy prior to giving these ratings, to capture differences in expectations associated with the instructed strategy. They also rated how much better they thought they would perform each of the tasks in the post-test session using a 1-10 Likert scale (1 = "*The same performance as in the pre-test*, 10 = "*A much better performance compared with the pre-test*").

Motivation & Alertness. Before the training session, participants rated their motivation to perform the tasks and alertness on scales from 1-5.

Pre- and post-test measures. The following six cognitive tasks made up both the pre- and post-training test sessions, and were thus completed by each participant twice, to compare performance improvement in participants who trained using the visualisation strategy with that observed in the control, no strategy group.

Criterion Training Task.

Digit N-back. This was a shortened version of the adaptive training task described above, including ten blocks instead of 20. Dependent variables were: (1) the maximum digit level the participant had reached, and (2) their average N-back level.

Task-specific near-transfer measures.

Letter N-back. This was a non-adaptive letter N-back task (2-back and 3-back), in which participants saw sequences of letters, and responded whether a given letter was identical to one presented 2 or 3 letters back. Participants did one block of the 2-back, one of the 3-back (order randomised) each containing 48 letters. Among these, 16 were targets, 32 non-targets, and half of the non-targets were lures (i.e., a letter identical to the letter presented next to the letter participants should base their response on; 8 $n + 1$ lures, 8 $n - 1$ lures). Each letter was shown for 1500 ms, followed by a blank screen for 450 ms. Dependent variables: (1) accuracy (d-prime; Stanislaw & Todorov, 1999) and (2) mean reaction time (RT) on correct target responses.

Colour N-back. This was identical to the Letter N-back task, but coloured squares were shown instead of letters.

Task-general near-transfer measures.

Selective updating of digits. In this WM updating task (Murty et al., 2011), five digits between 1 and 9 were displayed on the screen in a row of five squares. Participants attempted to memorise the digit sequence. Then, a new row of five squares replaced the initial sequence. Two of the new squares contained digits, and three were empty. Participants were to replace the old digits with the new digits while maintaining the unchanged digits in memory. Each participant completed ten trials with three such updating stages (i.e., new digits replaced original ones) and also ten trials without updates. Participants saw the original five-digit sequence (4000 ms), followed by a blank screen

(100 ms), and the first updating stage (2000 ms). At the end of each trial, participants reported the final five-digit sequence by clicking on the relevant digits in a recall grid with horizontally aligned squares containing numbers 1 to 9. All digit sequences followed these rules: (1) digit updates never occurred in adjacent squares, (2) adjacent digits deviated with more than one from each other (e.g. '2' could not be next to '1' or '3'), and (3) the two updated digits were never identical. Trial order was randomised between participants. The dependent variable was the percentage of the correctly recalled digits (in the right order) in the updating trials.

Forward simple span. Participants remembered sequentially presented digit sequences containing between 4 and 10 digits (one trial of each length) in order of appearance. Trial order was randomised for each participant. First, a fixation cross was shown in the middle of the screen (500 ms), followed by a digit (1000 ms) and this procedure continued until all digits in the sequence had been presented. Then, participants recalled the digits by clicking on the correct digits (in the right order), displayed in horizontally aligned squares containing all possible digits (1 to 9). The dependent variables were: (1) total number of correctly recalled digits in the correct serial position, and (2) maximum span; i.e., highest span length where all digits were recalled in the right order.

Running memory. Participants were instructed to report the final four digits of sequences containing between 4 and 11 items. A total of eight trials – one trial per sequence length – appeared in random order. First, a fixation cross appeared on the screen (500 ms), then a digit (1000 ms), until the sequence ended. Participants then selected the final four digits in the same order as they had been presented, using a recall grid with horizontally aligned squares containing numbers 1 to 9. The dependent variable was the total number of correctly recalled items, in the correct position.

The strategy questionnaire. After completing all cognitive tasks in the post-training test session, participants filled out a questionnaire about their strategy use in each task they completed in the pre- and post-training test sessions, respectively. First, they responded to whether they had used a strategy (yes or no) for each specific task during the pre-test. If yes, they described the strategy. They then indicated whether their strategy had changed between pre- and post-training tests (yes or no). If yes, they described their post-training test strategy.

Results

Exclusions

We excluded one younger adult in the control group who reported using pen and paper in the majority of the tasks. Also, one younger adult in the strategy group used pen and paper in one task and was excluded from that specific analysis. We excluded five participants with five or more errors on the Ishihara colour vision test from the colour N-back analyses and four participants from specific tasks due to missing data. See Table 5.2 for a summary of all exclusions by age and strategy group.

Table 5.2

Participant exclusions by age and strategy group

Reason for Exclusion	Younger Adults		Older Adults	
	Control	**Strategy**	**Control**	**Strategy**
Excluded from all analyses				
Cheating	1			
Non-compliant	-	6	-	11
Excluded from specific analyses				
Cheating		1[i]		
Missing data		2[ii]		2[iii]
Extreme outliers				
Multivariate outliers				
Colour Vision[iv]	1	1	1	2

[i] One excluded from the training analysis

[ii] Post-test N-back digit (1), RTs in pre-test 2-back colours (1).

[iii] Missing data in both N-back colours and RTs in pre-test 2-back letters (2)

[iv] Colour-blind participants were excluded from Colour N-back task.

Our results differed from Laine at al.'s (2018) in a way we had not anticipated – many of our strategy-group participants reported that they did not use the instructed strategy during training. In the original study only 3 of 37 (8%)

strategy-group participants failed to comply with the instruction, and non-compliant participants were not removed. In the present study, 6 of 31 (19%) younger adults and 11 of 30 (37%) older adults in the strategy group reported not using the instructed strategy. We had not specified in our pre-registration how we would handle non-compliant participants. However, the aim was to replicate Laine et al. (2018) with a different sample and test the effect of the instructed strategy in older adults. Hence, including non-compliant participants may lead to the trivial explanation that results did not replicate because too many of our participants did not use the strategy. Excluding non-compliant participants left 49 older and 54 younger adults, resulting in a power of .95 to detect the main effect on digit N-back performance observed by Laine et al. (2018) and power of at least .80 to replicate the effects on untrained letter and colour N-back tasks. Therefore, we focused on results from compliant participants. For transparency, we present output from analyses including all participants in Appendix D and point out the differences. We also conducted exploratory analyses to confirm that non-compliant participants were not a less motivated or capable subset by comparing pre-test composite scores in younger and older compliant and noncompliant strategy group participants (no significant differences; see Appendix D). We performed all analyses in the R environment version *3.5.1*, and the script and data are available via the OSF [https://osf.io/bwtuy].

Background and Pre-Test Characteristics

The control and compliant strategy groups did not differ significantly in years of education, gender distribution or pre-test N-back composite performance in either age group (Table 5.1). However, there was a significant age difference between control and strategy groups in older adults, such that participants in the strategy group were younger ($t(47) = -2.76$, $p = 0.01$). When non-compliant older adults in the strategy group were included, there were no age differences (see Appendix D), suggesting that the non-compliant older adults tended to be older.

Alertness, Motivation, and Expectations

We assessed expected training-session improvement in participants in the strategy and control groups after the strategy group participants had learned the strategy, but before starting the training. This was to check whether expectations were higher in the strategy groups, which might signal a placebo effect. There was no difference in expectations between control and strategy group participants in younger ($t(51) = 0.23$, $p = .82$) or older adults ($t(47) = 0.86$, $p = .39$). Similarly, improvement expectations between pre-test and post-test did not differ for any of the tasks in either age group (all p-values ≥ .07). Self-reported alertness and motivation – assessed upon completion of the training session – also did not differ between strategy and control groups (all p-values ≥ .13).

These measures were taken to test whether the strategy made the training more engaging. Similar results were observed when including non-compliant strategy participants (see Appendix D).

Training Session Data

Figure 5.2 shows performance over the 20 N-back blocks during the 30-minute training session in the control and strategy groups in younger (panel A) and older adults (panel B). While Laine et al. (2018) found that participants using the instructed strategy outperformed control group participants already in the fourth training block, we found no differences in the fourth block in our younger adults ($t(51) = -0.08$, $p = .94$; controls $M = 3.10$ digits, strategy $M = 3.08$). However, among the older adults, the control group performed significantly better on the fourth N-back block than the strategy group ($t(47) = -2.48$, $p = .02$; controls $M = 2.53$ digits, strategy $M = 1.93$). To capture the curvilinear performance pattern in increases across the 20 training blocks (see Figure 5.4), we performed an exploratory linear mixed effects analysis using second-order orthogonal polynomials. The R packages lme4 (Bates, Maechler & Bolker, 2012) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2015) were used in the model computation. Age Group, Strategy Group, and Block (coded both as a linear and a quadratic term), as well as all possible interactions were entered as fixed effects into the model. As random effects, we had participants' individual intercepts. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. Relations between performance levels in the two strategy groups across the training session did not differ between the age groups (Group × Age: *t*(102) = 1.09, *p* = .280). However, overall, strategy participants improved more across the training session than those training without a strategy, as evidenced in a Group × Block interaction both in the linear term (Estimate = -0.99, *SE* = 0.29, *p* = .001), and in the quadratic term (Estimate = 1.24, *SE* = 0.29, *p* < .001). Also, younger adults improved more across the training session than older adults, manifesting in a significant Block × Age interaction in the linear term (Estimate = -3.07, *SE* = 0.37, *p* < .001) as well as in the quadratic term (Estimate = -1.67, *SE* = 0.37, *p* < .001). There was no evidence for a three-way interaction (Group × Age × Block) in the linear term (Estimate = 0.20, *SE* = 0.42, *p* = .624). However, the quadratic term showed a statistically significant three-way interaction (Estimate = -0.95, *SE* = 0.42, *p* = .023), indicating that the relative effects of strategy across time differed between younger and older adults. Because the N-back training task was adaptive in its nature, with most of the participants managing the easiest levels, it is likely that only the quadratic term succeeded to capture the increased learning rates among the younger strategy group, potentially stemming from increased demands on WM resources towards the end of the training session.

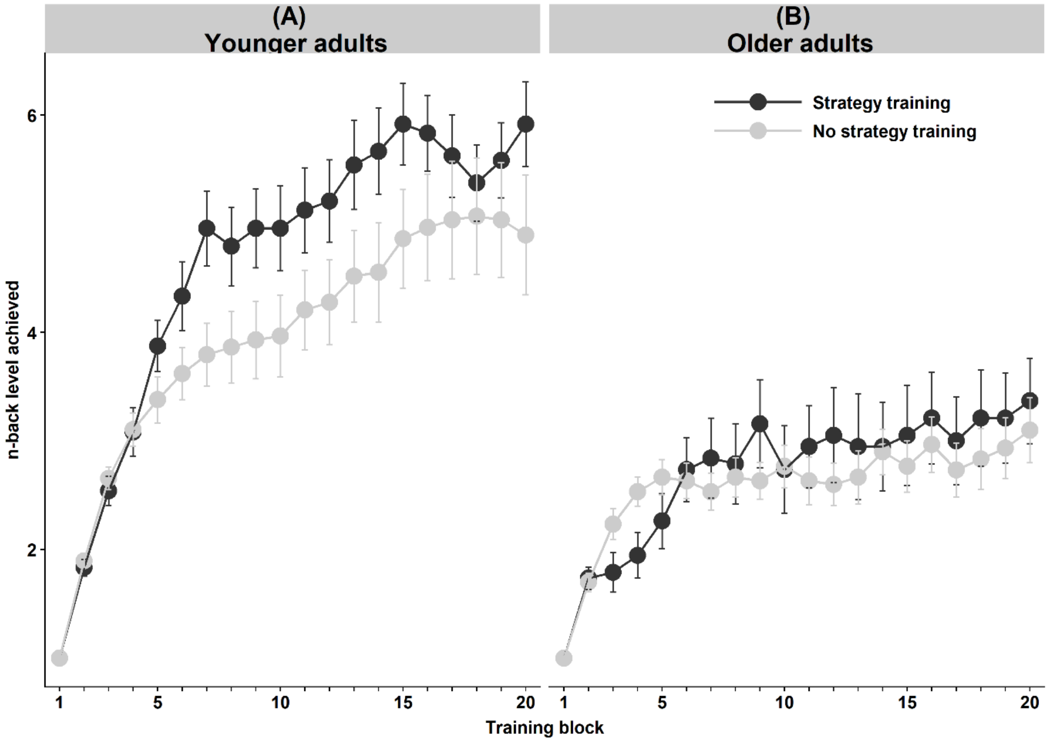


Figure 5.2. Performance across the 20 N-back digit training blocks, in the control and strategy groups in (A) younger and (B) older adults. Error bars represent standard errors of means.

The Effects of Training: Pre- versus Post-Test Performance

We tested whether training with the instructed strategy improved performance from pre- to post-training sessions on the various tasks to similar extents in the two age groups. Post-test performance was the dependent variable, pre-test performance the covariate, and strategy and age groups were between-subjects factors. See Tables 5.3 and 5.4 for pre- and post-training descriptives (means, standard deviations, pre-post correlations, and effect sizes) for each group, and Table 5.5 for ANCOVA statistics. To adjust for multiple comparisons, we applied Benjamini-adjusted *p*-values for group comparisons on each pre-post outcome measure (Benjamini & Hochberg, 1995).

Table 5.3

Mean values (standard deviations) for the pre-post measures per group at pre- and post-test, for younger adults

	Control Group (N = 29)				**Strategy Group (N = 25)**			
	Pre	**Post**	***r***	***d***	**Pre**	**Post**	***r***	***d***
Trained Digit N-back								
Maximum level	4.28 (1.71)	5.52 (2.16)	0.66	0.76	4.04 (1.49)	6.75 (1.59)	0.32	1.5
Average level	2.72 (0.91)	3.41 (1.02)	0.70	0.91	2.67 (0.94)	4.13 (0.81)	0.35	1.45
Task-specific near transfer								
Letter 2-back (d-prime)	2.25 (0.94)	2.48 (0.96)	0.71	0.32	2.19 (1.05)	3.01 (0.85)	0.38	0.77
Letter 3-back (d-prime)	1.19 (0.76)	2.00 (1.15)	0.55	0.83	1.12 (1.10)	2.67 (0.91)	0.40	1.40
Colour 2-back (d-prime)	2.03 (0.78)	2.54 (0.93)	0.39	0.54	2.14 (1.08)	2.85 (1.03)	0.62	0.78
Colour 3-back (d-prime)	0.90 (0.82)	1.69 (1.22)	0.52	0.75	1.03 (0.59)	2.53 (0.96)	0.47	1.74
Letter 2-back RT (ms)	803.85 (108.04)	686.08 (127.44)	0.51	-1.00	784.43	636.02 (151.96)	0.40	-0.96
Letter 3-back RT (ms)	802.58 (120.04)	676.99 (100.28)	0.36	-1.00	787.27	623.85 (132.05)	0.38	-0.82
Colour 2-back RT (ms)	811.47 (119.21)	696.87 (115.34)	0.26	-0.80	811.88	661.41 (146.82)	0.29	-0.93
Colour 3-back RT (ms)	857.86 (129.50)	721.56 (106.42)	0.24	-0.93	817.49	661.63 (150.60)	0.20	-0.64
Task-general near transfer								
Selective updating of digits	32.38 (8.14)	33.00 (7.08)	0.78	0.12	35.32 (8.53)	37.24 (7.60)	0.69	0.30
Digit span (correct items)	34.52 (10.00)	34.10 (8.83)	0.73	-0.06	35.16 (9.33)	37.76 (7.47)	0.22	0.24
Digit span (maximum span)	6.79 (2.06)	7.28 (1.53)	0.70	0.33	7.36 (2.00)	7.88 (1.54)	0.26	0.24
Running memory	25.31 (4.49)	26.28 (5.32)	0.49	0.19	24.92 (4.97)	27.20 (4.38)	0.53	0.50

Note. Values in parentheses are standard deviations. *r* = correlation between pre- and post-test. Cohen's *d* represents effect sizes for correlated samples. Exclusions to specific analyses apply.

Table 5.4

Mean values (standard deviations) for the pre-post measures per group at pre- and post-test, for older adults

	Control Group (N = 30)				**Strategy Group (N = 19)**			
	Pre	**Post**	***r***	***d***	**Pre**	**Post**	***r***	***d***
Trained Digit N-back								
Maximum level	3.10 (0.92)	3.83 (1.29)	0.51	0.64	2.79 (0.71)	3.95 (1.35)	0.39	0.92
Average level	1.94 (0.55)	2.55 (0.76)	0.58	0.98	1.96 (0.45)	2.56 (0.79)	0.50	0.87
Task-specific near transfer								
Letter 2-back (d-prime)	1.85 (0.79)	2.31 (0.86)	0.63	0.64	1.84 (0.72)	2.19 (0.86)	0.40	0.39
Letter 3-back (d-prime)	0.76 (0.48)	1.28 (0.88)	0.45	0.65	0.80 (0.65)	1.45 (1.01)	0.43	0.70
Colour 2-back (d-prime)	1.81 (0.75)	2.09 (0.83)	0.53	0.36	1.36 (0.86)	2.08 (1.01)	0.32	0.65
Colour 3-back (d-prime)	0.77 (0.58)	0.94 (0.76)	0.16	0.19	0.44 (0.46)	0.94 (0.71)	0.25	0.67
Letter 2-back RT (ms)	1017.30	869.92 (178.94)	0.82	-1.42	1021.58 (177.47)	915.14	0.49	-0.64
Letter 3-back RT (ms)	1002.24	936.59 (167.59)	0.70	-0.49	983.38 (158.56)	922.80	0.46	-0.38
Colour 2-back RT (ms)	1013.51	909.24 (160.31)	0.64	-0.75	1050.99 (168.84)	951.99	0.84	-1.08
Colour 3-back RT (ms)	1071.56	959.52 (199.16)	0.54	-0.64	1010.70 (176.35)	1026.44	0.45	0.09
Task-general near transfer								
Selective updating of digits	24.63	30.43 (11.33)	0.75	0.72	25.79 (11.59)	29.53 (10.50)	0.75	0.48
Digit span (correct items)	33.23 (8.24)	34.37 (7.91)	0.64	0.17	32.37 (8.54)	32.74 (7.78)	0.74	0.06
Digit span (maximum span)	6.93 (1.36)	7.23 (1.36)	0.18	0.17	6.79 (1.65)	6.74 (1.63)	0.64	-0.04
Running memory	24.33 (4.33)	23.80 (5.29)	0.51	-0.11	23.32 (5.63)	24.37 (4.19)	0.59	0.23

Note. Values in parentheses are standard deviations. *r* = correlation between pre- and post-test. Cohen's *d* represents effect sizes for correlated samples. Exclusions to specific analyses applied.

Table 5.5

ANCOVA results for the trained task and for the transfer measures

		F	***p***	**d /** η_p^2
Trained Digit N-back				
Maximum level	**Strategy**	**8.73**	**.004**	**0.61**
	Age	**20.57**	**<.001**	**0.88**
	Interaction	3.45	.066	0.034
Average level	**Strategy**	**6.53**	**.015**	**0.53**
	Age	**21.25**	**<.001**	**0.87**
	Interaction	**6.49**	**.011**	**0.065**
Task-specific near transfer				
Letter 2-back (d-prime)	Strategy	2.27	.204	0.33
	Age	3.76	.111	0.32
	Interaction	5.21	.066	0.051
Letter 3-back (d-prime)	Strategy	5.75	.055	0.50
	Age	**16.85**	**<.001**	**0.78**
	Interaction	2.40	.204	0.024
Colour 2-back (d-prime)	Strategy	1.95	.235	0.29
	Age	3.95	.109	0.42
	Interaction	0.01	.924	<.001
Colour 3-back (d-prime)	**Strategy**	**6.96**	**.033**	**0.57**
	Age	**25.98**	**<.001**	**1.00**
	Interaction	2.26	.204	0.024
Letter 2-back (RT in ms)	Strategy	0.01	.924	<.001
	Age	8.11	.021	-0.57
	Interaction	2.64	.198	0.026
Letter 3-back (RT in ms)	Strategy	1.17	.356	- 0.23
	Age	**48.32**	**<.001**	**-1.47**
	Interaction	0.71	.483	0.007
Colour 2-back (RT in ms)	Strategy	0.06	.889	-0.07
	Age	**20.43**	**<.001**	**-0.98**
	Interaction	1.44	.312	0.015
Colour 3-back (RT in ms)	Strategy	0.42	.59	0.09
	Age	**44.02**	**<.001**	**-1.36**
	Interaction	4.78	.075	0.049

Task-general near transfer				
Selective updating of digits	Strategy	0.04	.987	0.06
	Age	0.38	.715	-0.18
	Interaction	2.47	.309	0.025
Digit span (correct items)	Strategy	0.67	.624	0.19
	Age	1.00	.55	0.14
	Interaction	2.94	.309	0.029
Digit span (maximum span)	Strategy	0.01	.987	<.001
	Age	3.58	.309	0.33
	Interaction	2.35	.309	0.023
Running memory	Strategy	1.72	.385	0.26
	Age	5.34	.276	0.47
	Interaction	< .001	.987	< .001

Note. To adjust for multiple comparisons, Benjamini-Hochberg adjusted *p*-values were applied for group comparisons on each pre-post outcome measure. Cohen's *d* is presented for the group comparisons, η_p^2 for the interactions.

The trained N-back task with digits. A 2 (Group) × 2 (Age) between-subjects ANCOVA of maximum post-test N-back performance that controlled maximum pre-test N-back performance indicated significant main effects of strategy (*F*(4, 97) = 8.57, *p* = 0.006, *d* = 0.61) and age group (*F*(4, 97) = 20.57, *p* < .001, *d* = 0.88), but no significant interaction (*F*(4, 97) = 3.45, *p* = 0.111, η_p^2 = 0.03). For average digit N-back performance, there were also significant effects of strategy (*F*(4, 97) = 6.53, *p* = .015, *d* = 0.53) and age group (*F*(4, 97) = 21.25, *p* < .001, *d* = 0.87, as well as a significant interaction (*F*(4, 97) = 6.49, *p* = .011, η_p^2 = .06). The strategy manipulation appeared more beneficial in younger adults (see Figure 5.3). Younger adults in the control group gained on average 0.69 digits from pre-test to post-test, compared to 1.46 digits in the instructed strategy group (see Table 5.3). However, older adults benefitted less (control group gained 0.61 digits; strategy group, 0.60). When including non-compliant participants, no effect of strategy group was observed for maximum digit level in either age group, however there was a significant interaction between age group and strategy level for average digit N-back performance (see Appendix D).

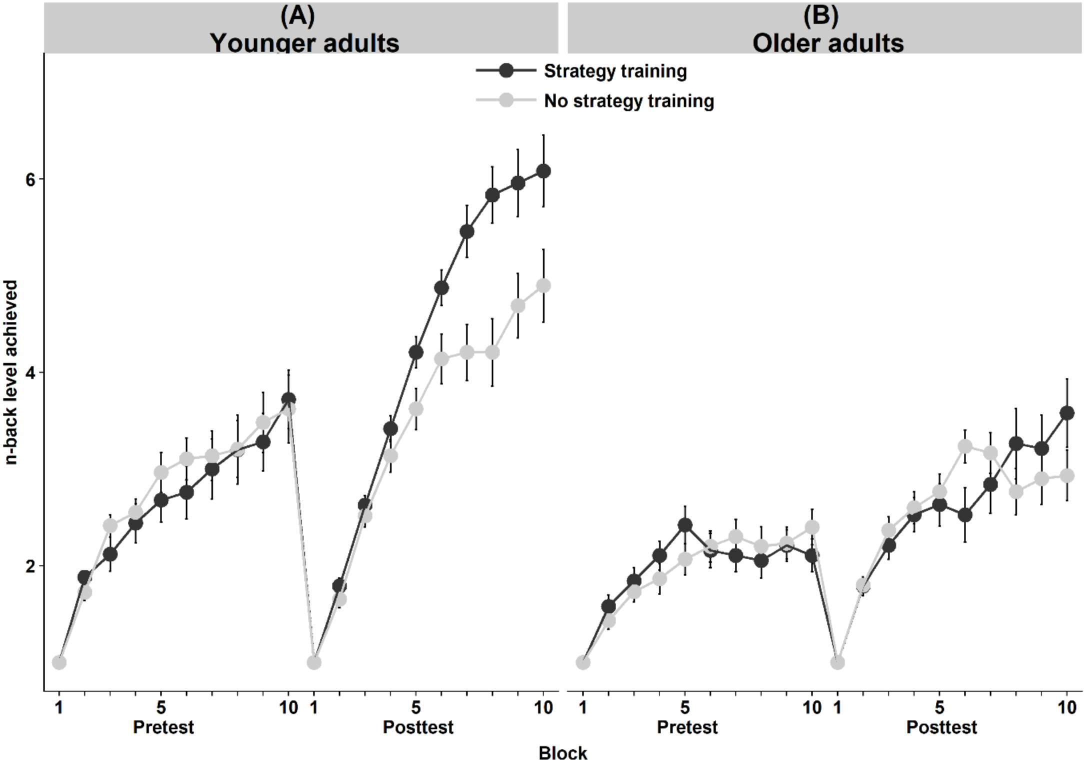


Figure 5.3. Average performance across the 10 blocks of the trained N-back task at pre- and post-test in the control and strategy groups, in (A) younger and (B) older adults. Error bars = standard errors of means.

Letter N-back. There was no significant effect of age or strategy group on d-prime in the Letter 2-back, and no interaction (all *p*'s ≥ .066). There was no significant main effect of strategy in the more demanding 3-back condition, (*F*(4, 98) = 5.75, *p* = .055, *d* = 0.50), despite a medium effect size. This was the only instance where our results regarding the strategy manipulation deviated from Laine et al.'s (2018). We observed a statistically significant main effect of age group (*F*(4, 98) = 16.85, *p* < .001, *d* = 0.78) but our strategy × age interaction was non-significant (*F*(4, 98) = 2.40, *p* = .204, η_p^2 = 0.02).

There were significant effects of age on RTs in both the 2-back (*F*(4, 98) = 8.11, *p* = .021, *d* = - 57), and the 3-back tasks (*F*(4, 98) = 48.32, *p* <.001, *d* = - 1.47), but no effects of strategy group, nor any interactions between strategy and age group (all *p*'s≥ .198). Results were similar when including non-compliant participants (see Appendix D).

Colour N-back. We excluded five participants with five or more errors on the Ishihara colour vision test from these analyses. There was no significant main effect of strategy group for the 2-back d-prime (*p* = .29), but strategy group showed more improvement on the more demanding 3-back task (*F*(4, 93) = 6.96, *p* = .033, *d* = 0.57). Correspondingly, we observed a significant main effect of age in the 3-back (*F*(4, 93) = 25.98, *p* < 0.001, *d* = 1.00) but not the 2-back (*F*(4, 93) = 3.95, *p* = 0.109, *d* = 0.42). There were no interactions between age and strategy (both *p*'s ≥ 0.204). The older adults were significantly slower in both the 2-back (*F*(4, 93) = 20.42, *p* < .001 *d* = - 0.98) and 3-back tasks (*F*(4, 93) = 44.02, *p* <.001, *d* = - 1.36) but there were no effects of strategy group, nor any interactions between strategy and age group (all *p*'s ≥ .075). When including non-compliant participants results were similar, but no effect of strategy group in the 3-back task was observed (see Appendix D).

Task-general near-transfer. There were no significant main effects either of age or strategy group nor any interactions for selective updating of digits, running memory, or either forward digit span measure (correctly recalled digits, or maximum span), all *p*'s ≥ .276. The same patterns of results were found including non-compliant participants (see Appendix D).

Self-Generated Strategies and Performance

We tested whether (1) the types of reported self-generated strategies and (2) the reported levels of detail of those strategies were associated with better post-test N-back performance in control group participants. Only control participants were used to obtain a 'pure' measure of spontaneously generated strategies in participants who were not exposed to any strategy instruction. One older adult was excluded due to missing strategy data for N-back letters and colours. Thus, the final sample of controls included 58 participants. The types of strategies and level of detail reported in the two age groups are presented in Figure 5.4.

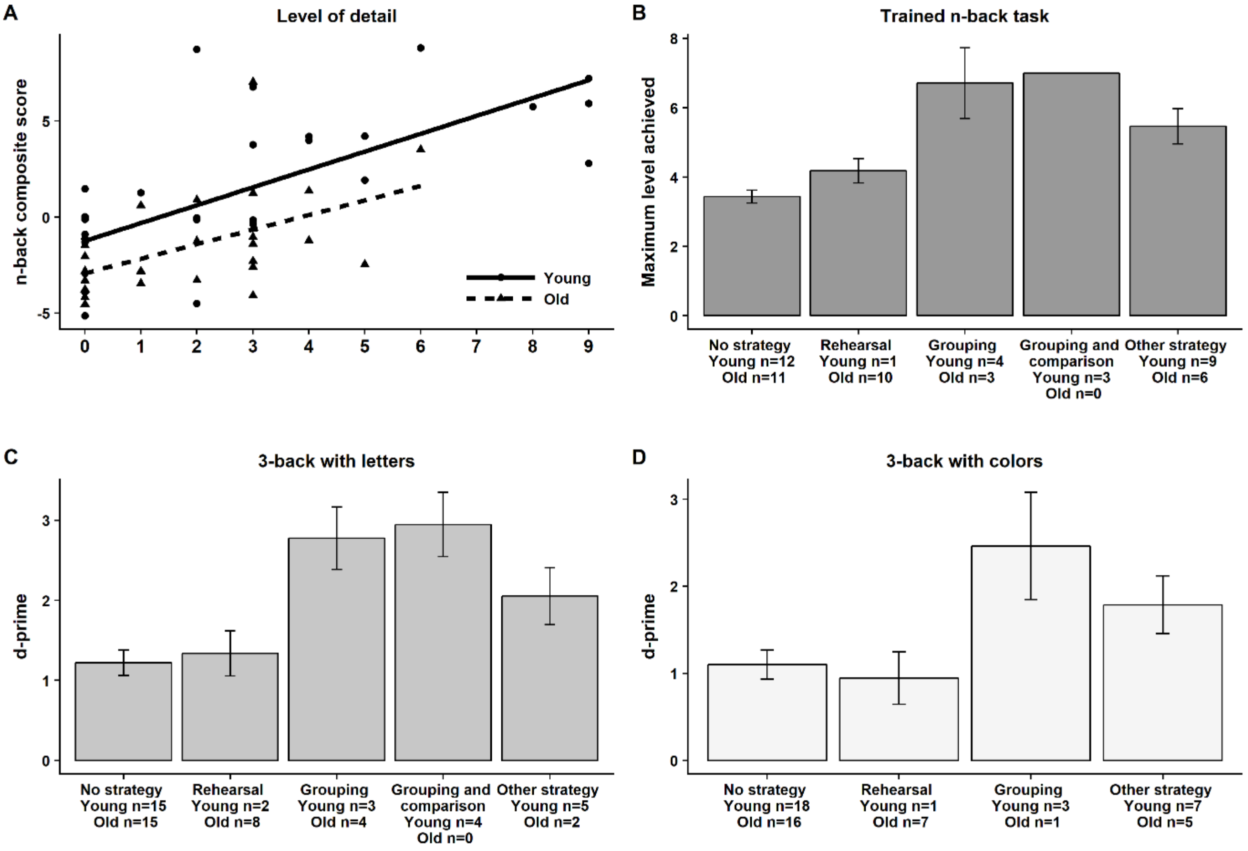


Figure 5.4. (A) Regression plot with level of detail of reported strategies (9 = maximum level of detail) as the independent variable (X-axis) and the N-back composite score (Y-axis) as the dependent. The N-back composite summed up post-test z-values of average and maximum N-back level reached in the trained digit N-back task, and the d-prime values in the untrained letter and colour 3-back tasks. (B) Strategy type and performance in the trained N-back digit task at post-test. (C) Strategy type and performance in the untrained letter N-back task at post-test. (D) Strategy type and performance in the untrained colour N-back task at post-test. Whiskers in panels B C, and D represent standard errors of means. The three participants using Grouping and Comparison in the Trained N-Back task all reached the same level, hence no error bar.

Self-generated strategies: Type. We classified self-generated post-test strategies according to Laine et al.’s (2018) classification scheme, based on categories used by Morrison, Rosenbaum, Fair, and Chein (2016). Two independent raters classified each strategy report into one of these categories: Rehearsal, Grouping, Updating, Grouping and Comparison, Semantics, Phonology, Imagery, Familiarity, Guessing, Other Strategy, or No Strategy (see Appendix D). Initial inter-rater reliability (unweighted Cohen's kappa) for the three N-back tasks was consistent and good: trained digit N-back (κ = .79, 95% CI [.72, .86]), letter N-back (κ = .81, 95% CI [.74, .88]) and colour N-back (κ = .81, 95% CI [.73, .88]). The raters then resolved discrepancies through discussion consensus, producing the final strategy type classifications used in the analysis. Strategies reported by less than 5% of participants were grouped as "Other Strategy" (see Appendix D for the distributions of strategy types used in the three N-back tasks at post-test)._The final list comprised five categories for the digit and letter N-back (No Strategy, Rehearsal, Grouping, Grouping and Comparison, and Other Strategy) and four categories for the colour N-back (No Strategy, Rehearsal, Grouping, and Other Strategy). We tested if N-back performance differed by strategy type using one-way ANOVAs. No strategy was the baseline. In each model, the dependent variable was N-back post-test performance and strategy type was the between-subjects factor. Figure 5.4 shows N-back post-test performance as a function of strategy type at post-test for each N-back task. We did not include age as a factor given the limited number of observations but see Figure 5.4 for usage by age group.

Digit N-back (Maximum Level). Reported strategy-use was associated with significantly better performance than not using a strategy (*F*(4, 54) = 9.75, *p* = <.001, η_p^2 = 0.419). Participants in the No Strategy group were outperformed by participants who reported using Grouping (*t*(28) = 4.49, *p* < .001, *d* = - 2.22), and Grouping and Comparison (*t*(24) = 2.39, *p* = .02, *d* = -4.16). However, those not reporting using strategies did not differ in performance from those using Rehearsal (*t*(32) = 1.32, *p* = .192, *d* = - 0.76), or Other Strategy (*t*(36) = 0.27, *p* = .79, *d* = -1.44).

Letter 3-back (d-prime). Using a strategy was significantly better than not using a strategy (*F*(4, 53) = 7.17, *p* < .001, η_p^2 =0.35). Again, participants in the No Strategy group were outperformed by participants using Grouping (*t*(35) = 3.96, *p* < .001 , *d* = -1.72), and Grouping and Comparison (*t*(32) = 2.45, *p* = .018, *d* = -1.99), but not by those using Rehearsal (*t*(38) = 0.36, *p* = .721, *d* = - 0.13), or Other Strategy (*t*(35) = -0.05, *p* = .964, *d* = -0.94).

Colour 3-back (d-prime). Again, using a strategy was better than not using a strategy (*F*(3**,** 54) = 3.39, *p* = .025, η_p^2 = 0.16). Participants using Grouping performed significantly better than those using No Strategy (*t*(36) = 2.61, *p* = .012, *d* = - 1.37). There was no difference between No Strategy and Rehearsal (*t*(40) = - 0.35, *p* = .729, *d* = 0.16, or between No Strategy and Other Strategy (*t*(44) = 0.77, *p* < .444, *d* = -0.67).

Verbal Rehearsal in older adults: Exploratory analyses*.* Perhaps Rehearsal was not associated with better performance compared to No Strategy because Rehearsal was primarily used by older adults, who may generally perform worse than younger adults. To test this possibility, we performed exploratory analyses comparing older adults using Rehearsal with older adults using No Strategy, for the three different N-back tasks.[11] For the Letter N-back (3-back d-prime) there were no differences ($t(21) = 0.11$, $p = .92$, $d = -0.05$), nor for the colour N-back (3-back d-prime) ($t(20) = -0.98$, $p = .34$, $d = -0.47$). However, for the digit N-back (maximum level), Rehearsal was associated with better performance than No Strategy ($t(19) = -2.21$, $p = .04$, $d = -0.96$).

Self-generated strategies: Level of detail. We tested whether the level of detail of the reported strategy during post-test was associated with post-test N-back performance in controls. The same raters as above scored the reported strategies based on the criteria used by Laine et al. (2018) on a scale from 0 to 3. Zero meant that participants did not report using a strategy. One point was given to a vague, non-specific strategy (e.g., “I memorised the digits in my mind”) and two points for a clear strategy with at most one detail (“I memorised the digits in pairs, such as 52–48”). Scorers gave three points for

---

[11] Since only four younger adults reported using rehearsal across the three N-back tasks, we did not include younger adults in these analyses.

clearly described strategies with at least two details (e.g., "I split the digits into different series, and compared those to each other"). The raters scored the three N-back varieties (digit, letter, and colour), such that each participant had a total N-back level-of-detail score between 0 and 9.

There was good interrater reliability analysis between the two independent raters for this scoring procedure (linearly weighted kappa analysis; $\kappa_w = .83$, 95% CI [.80, .86]; Cohen, 1968). The raters then discussed and reached consensus on all discrepant scores, producing a final level of detail score for each control group participant. To test if these scores predicted general N-back post-test performance, we calculated an N-back composite score including: (1) for the trained digit N-back task: summed values of the z-transformations of the post-test average and maximum level reached, and (2) post-test d-prime variables in the Letter and Colour 3-back tasks.

We performed a multiple regression analysis with the N-Back composite score serving as the dependent variable, and level of strategy detail and age group serving as predictors. The results showed a significant regression equation ($F(3, 52) = 18.15$, $p < 001$) with an adjusted R^2 of 0.48. Level of detail was significantly associated with post-test N-back composite performance ($\beta = 0.621$, $t = 5.557$, $p < .001$), whereas age group was not (β = -

0.570, t = -1.655, p = .104), and there was no evidence for an interaction (β = -0.115, t = - 0.507, p = .614). See Figure 5.4, panel A.

Discussion

The present study tested the Strategy Mediation Hypothesis of WM training via external (i.e., instructed) and internal (i.e., spontaneously self-generated) strategy use in a single session of adaptive N-back training. It was a systematic replication of Laine et al. (2018) to test the validity of their results for younger adults in a different sample of participants (see Simons, 2014). We also explored potential implications of strategy use in N-back training in healthy older adults, given that they are often targeted by commercial training programmes (e.g., Federal Trade Commission, 2016).

The instructed N-back strategy was associated with greater performance improvement during the training session across the 20 training blocks in younger adults, and was associated with significantly better performance on the trained N-back digit task a few days later, during the post-test session. We found some evidence that the older adult strategy group performed worse than controls in the initial blocks (see Figure 5.2), resulting in reduced average performance, which we discuss below. Instructed strategy was also associated with significantly more accurate performance on the more difficult version of the untrained colour N-back task (3-back) in both age groups, without improved reaction times – similar to transfer patterns typically seen after weeks of ordinary adaptive WM training (Soveri et al., 2017), and similar to Laine et al.'s. (2018) observations. However, even though the effect size of the strategy (i.e., Control group vs. Strategy group) was moderate following training in the untrained letter N-back (*d* = 0.50), after correcting for multiple comparisons, we did not replicate the beneficial effect of strategy on the untrained letter N-back. This is difficult to interpret. Perhaps including older adults, who appeared to struggle with implementing the strategy across earlier blocks for the digit N-back tasks increased variability in our ANCOVA models. As expected, there was no effect of strategy group on any of the structurally different WM tasks (i.e., no task-general near transfer). These tasks tested memory for digits – like the trained task – but did not require comparison, making the instructed strategy inapplicable.

These results indicate that learning to use a specific strategy – which is unlikely to improve general reasoning ability or prevent age-related cognitive decline – can produce significant N-back performance gains, and have several implications for the training literature. Firstly, our results were in line with the notion that much of N-back training is task-specific (Soveri et al., 2017). Before encouraging members of the public to spend time and money on cognitive training, it should be established that improvements are not limited to some task-specific strategic approach – which is probably nearly useless in the individuals' lives. Some training programs keep users engaged via task-improvement feedback, suggesting that better performance implies improved working memory ability. However, our findings of significant strategy-induced task-specific near transfer without task-general near transfer, along with those from many other studies, suggest that such claims are vastly overstated.

Strategy-induced improvements raise further questions regarding whether training strategies can be applied to outcome variables claimed to reflect far transfer. If so, perhaps some types of training are associated with far transfer improvement because trained participants develop a strategy which generalises to the outcome measure. Further research should explore whether strategies developed during training are applied to seemingly unrelated outcome measures. For instance, tests assumed to measure fluid intelligence (e.g., Raven's Matrices) are often used as measures of far-transfer training gains. Cogmed's visual processing and matching training is similar to Raven's Matrices (Shipstead et al., 2012). Using a speeded-up version of Raven's Matrices (e.g., Jaeggi et al., 2008) may even increase these similarities (Chuderski, 2013). Moreover, some evidence suggests that opportunity to practice may improve performance on Raven's Progressive Matrices (e.g., Blieszner, Willis, & Baltes, 1981; Denney & Heidrich, 1990; Klauer, Willmes, & Phye, 2002). Thus, training control groups on a different task can be misleading if it differs in terms of structural similarity from outcome measures. If a WM training paradigm only improves performance on one specific reasoning measure, strategic mediation in far transfer measures needs to be ruled out. Arguably, transfer should generalise to several structurally different outcome tasks, before transfer to for instance fluid intelligence is asserted.

However, evidence that strategy use improves performance on trained tasks does not ‘falsify’ the Capacity Hypothesis of WM training; it is still possible that training also usefully improves cognitive capacity. According to the Capacity Hypothesis, training works by challenging the cognitive system, and working at one’s capacity limits is considered a prerequisite for the sorts of plastic changes in the brain considered to reflect increased capacity (e.g., see Klingberg, 2010). If strategies reduce cognitive load by making the task easier, this might prevent capacity-increasing change and therefore prevent broader transfer. Strategy use may, therefore, produce problematic confounds in training studies either by making possible improvements without meaningfully increasing cognitive capacity or by preventing optimally ‘broad’, efficient training.

The assumption that online cognition is limited by the capacity of a domain-general attentional resource or WM system (Engle & Kane, 2004) which can be ‘trained’ and thus improve cognitive abilities more broadly (Jaeggi et al., 2008) underlies the Capacity Hypothesis. The finding that a visualisation strategy was associated with improved memory performance might fit better with theories of WM as containing a variety of cognitive systems among which participants may choose according to task demands (Baddeley & Logie, 1999;

Logie, 2011; Logie & Niven, 2012). Encouraging participants to use other sub-components of the cognitive system (e.g. visualising the strings of digits) appeared to boost performance significantly, as suggested by Logie (2012). Strategic 'off-loading' from a general resource to another system might be useful by freeing up its cognitive resources (McNamara & Scott, 2001). This would not imply that a general resource cannot be trained at all, but it suggests that this resource was not necessarily trained as was intended in many training studies.

While our results suggest that instructed strategies can play a significant role in WM performance, strategies arguably only have implications for the training literature if participants spontaneously use them during adaptive training (e.g., Dunning & Holmes, 2014), which needs to be demonstrated. Our results suggested that participants did generate and use strategies spontaneously. Both strategy type and level of detail (i.e., how elaborate the strategy was) were associated with higher performance on all three N-back tasks at post-test (see Figure 5.4). However, the categories used in our study did not capture all strategies (16.1% classified as 'Other' across the three tasks). Strategies classified as 'Other' were not associated with improved performance in either N-back tasks (compared to not using a strategy). This suggests that a substantial proportion of participants applied potentially inefficient strategies. The implications of such strategies for the training literature are unclear, and more detailed research into the causes – and consequences – of these 'Other' self-generated strategies may help design better training paradigms.

Moreover, the beneficial effects of spontaneous self-reported strategies on performance may be inflated. For instance, strategies may be used more by high-capacity individuals, who have more cognitive resources available for generating effective strategies while performing the task (Dunlosky & Kane, 2007) and who may also be more likely to reap training benefits regardless of strategy use. As well, reports of strategy use could be influenced by general task motivation, if participants who tried their best on the task are also keener to provide detailed descriptions of their approaches. Therefore, explicitly manipulating strategy use via instructed strategies that participants can and do use is important to ensure that associations between performance and strategies are not driven by such confounds. Our instructed strategy manipulation suggested that most participants can benefit from using a strategy – but an unexpected limitation was the relatively large proportion of non-compliant participants, whom we excluded from the main analyses. While WM capacity appeared similar in compliant and non-compliant participants (indicated by no significant differences in pre-test composite scores), we cannot infer whether noncompliant participants were unable to apply the strategy or preferred not to. However, despite these limitations regarding the causes of strategy application, our results suggest that both internally-generated and externally-instructed strategies boosted N-back performance.

Strategy Training in Healthy Older Adults

We specifically included healthy older-adult participants to compare their strategy use with that of younger adults, noting both similarities and differences. The older adults in the strategy group appeared to benefit less during the training session than the younger-adult strategy group (see Figure 5.2). In the post-test, younger and older adults both benefitted from the strategy in the untrained N-back colour 3-back, and in the maximum digit N-back score. However, in the average digit N-back level attained, the older adults benefitted less, reflecting that on average, the control group outperformed the strategy group until block eight of ten (see Figure 5.3).

Some previous studies instructing participants to apply mnemonic techniques or strategies have found more substantial training gains in younger than in older adults (e.g., Lövdén, Brehmer, Li, & Lindenberger, 2012; Verhaeghen & Marcoen, 1996; Verhaeghen, Marcoen, & Goossens, 1992, but see Gross et al., 2012). Taken together, our results suggested that while both age groups benefitted from the strategy, older adults appeared to benefit more slowly, as implementing the new strategy reduced performance during early trials. If participants develop spontaneous strategies during uninstructed, regular training and younger participants generate and effectively apply them more quickly, our results might be consistent with these observations of initially larger gains in younger adults, followed by comparable improvements in both age groups in the final weeks (e.g., Brehmer et al., 2012). Furthermore, a large proportion of our older adults (11 of 30) did not use the instructed strategy, possibly indicating they found it difficult to implement. Perhaps if implementing a strategy is generally more challenging for older than younger adults, it is also more beneficial once they learn how to do it effectively. For instance, cognitive training using an episodic memory strategy task was associated with less age-related decline in white matter microstructures in healthy older adults compared to a control group, after 40 weeks (de Lange, Bråthen, Rohani, Grydeland, Fjell, & Walhovd, 2017).

Also, it is possible that older adults struggled to implement the strategy because it was visually based – some previous research suggests that visual WM declines more in healthy ageing that verbal WM (e.g., Johnson et al., 2010). Similarly, more older than younger adults in our uninstructed control group reported using a sub-vocal Rehearsal strategy; i.e., silent repetition of verbal labels for material to be recalled (see Logie, Della Sala, Laiacona, Chalmers, & Wynn, 1996; Wang, Logie, & Jarrold, 2016). Specifically, four younger and 25 older adults used this strategy in the three N-back tasks combined (see Figure 5.4), supporting previous suggestions that older adults may rely more on verbal rehearsal even in visual WM tasks (Forsberg, Johnson, & Logie, *under review*). More severe working memory deficits for visuospatial material than for verbal material have been observed in older adults (e.g., Jenkins, Myerson, Hale, & Fry, 1999; Leonards, Ibanez, & Giannakopoulos, 2002; Myerson, Hale, Rhee, & Jenkins, 1999), and perhaps sub-vocal rehearsal can be used to compensate for declining visual memory. Rehearsal benefitted older adults in our digit N-back task (compared to those not using a strategy), in line with observations that older adults' WM benefitted from verbal encoding strategies (Bailey, Dunlosky, & Hertzog, 2014). However, it was not beneficial in the letter or colour N-back tasks. Verbal rehearsal might have been most useful for the digit task because the letter set likely produced more phonological similarity effects (Salamé & Baddeley, 1986), and colour names are longer, thus less efficient to rehearse (Schweickert, Guentert, & Hersberger, 1990). Also, the digit task was adaptive (maximum levels reached by older adults: control group *M* = 3.83, *SD* = 1.29; strategy group *M* = 3.95, *SD* = 1.35) – in contrast to the letter and colour tasks, which only tested accuracy at 2- and 3-back levels. Thus, rehearsal may not have efficiently improved 2 and 3-back accuracy, but allowed participants to reach 4-back levels and beyond, when such improvement was allowed in the digit N-back task.

In the broader training literature, younger adults often improve more than older adults (Burki et al., 2014; Dahlin et al., 2008; Heinzel et al., 2014; Zinke et al., 2013) – however, gains of similar magnitude on trained tasks in younger and older adults are also sometimes observed (e.g., Bürki et al., 2014; Li et al., 2008; Richmond et al., 2011; von Bastian et al., 2013; Zając-Lamparska & Trempała, 2016). Larger training gains in younger adults are thought to be consistent with animal models suggesting that older age is associated with less neuroplastic change (Blumenfeld-Katzir et al., 2011; van Praag et al., 2005). Our results suggest an alternative explanation: perhaps younger adults appear to benefit more from training because they are more adept at developing strategies. Furthermore, age differences in training gains between paradigms may be driven by differences in strategy effectiveness (e.g., visual versus verbal). The observed age differences in the effectiveness of the instructed visualisation strategy, and the use of spontaneous verbal rehearsal strategies fit with literature suggesting that not all cognitive functions decline with age to the same degree (for reviews see Logie & Morris, 2014; Perfect & Maylor, 2000). In sum, these results support the notion that overall N-back performance may reflect use of different cognitive resources in different participants (Johnson et al., 2010; Logie, 2018; Thurstone, 1931).

To conclude, our results supported Laine et al.’s (2018) conclusion that using a visualisation strategy during training improved N-back performance in younger adults. Furthermore, the strategy also improved performance in older adults. The results provided support for the Strategy Mediation hypothesis of training, and suggest that strategies can enable more efficient use of a limited WM capacity, which may have various implications for the training literature and industry. Commercial training programmes need to demonstrate useful improvement beyond task-specific strategies, which are unlikely to benefit the user in their everyday life. Also, confirming that the trained task and outcome measures are structurally different – ideally by demonstrating far-transfer to several different reasoning and intelligence measures – is needed to ensure that transfer effects are not strategy-specific.

Furthermore, older adults may benefit more slowly when attempting to apply a visual strategy – indeed, we found some evidence that implementing the strategy was initially associated with worse performance. While the instructed strategy did appear to benefit those older adults who were able to apply it overall, our results did not generalise to the substantial proportion of older adults who chose not to implement (or perhaps were unable to implement) the instructed strategy. Furthermore, older adults spontaneously applied verbal strategies more than did younger adults (with varied success) which suggests differences in spontaneous strategies used by younger and older adults. While our paradigm could not determine if this was driven by preference or ability, it did indicate that perhaps the same training paradigm – or cognitive task, more broadly – is not always measuring the same cognitive capacity in younger and older adults.

The present results highlighted that measures of performance and capacity may largely reflect the extent to which participants apply appropriate strategies, rather than domain-general underlying constructs. Investigating strategies and accounting for individual variability (see Logie, 2018), as well as for systematic, age-related variabilities during real, long-term training, and how specific task strategies may generalise to outcome measures in unintended ways may be essential to resolving discrepancies in the cognitive training literature.

Chapter 6: Conclusion

The overarching aim of the research presented in this thesis was to test if younger and older adults approached WM tasks differently. Specifically, I tested whether older adults appeared to rely more on verbal rehearsal in visual WM tasks (Chapters 2 – 4), and examined age differences in memory performance in relation to instructed and self-reported strategic approaches in a WM training paradigm (Chapter 5). In this final chapter, I discuss the implications of the empirical work reported in the previous chapters, in relation to the central research question: are older adults simply like more poorly performing younger adults? I present the main findings and conclusions in Section 6.1. Thereafter, I discuss the implications of this research for our understanding of the literature on age-related deficits in visual feature-binding, continuous colour memory, and cognitive training, and finally discuss the broader implications for the 'Dull Hypothesis' of cognitive ageing (Section 6.2). In Section 6.3 I discuss some limitations of the current research and finally, the conclusion (Section 6.4) provides a summary of the key outcomes.

6.1 Summary of Results

The first part of this section focuses on experimental work exploring the role of verbal rehearsal in visual feature-binding tasks (Experiments 1 – 4, Chapters 1 and 2). To my knowledge, this was the first set of studies to manipulate the opportunity for verbal rehearsal strategies in visual feature-binding paradigms, in younger and older adults. All four studies found that memory was enhanced for easy- as opposed to difficult-to-label stimuli. This benefit appeared comparatively larger for older adults' colour memory (Exp. 1 and 2), and created a 'fake' feature-binding deficit (i.e., a comparatively larger difference between the single-feature and binding condition) in older adults for easy-to-label colours (Exp 1.) which did not replicate in Exp. 2. Hence, while older adults' visual WM generally appeared worse, something about the label-able colours allowed them to partially overcome this deficit, and perform more similarly to younger adults. The older adults' memory advantage for easy-to-label colours appeared attenuated by concurrent suppression in Exp. 2. However, the evidence that suppression affected younger and older participants differently was not very strong, and requires replication. It seemed like sub-vocal rehearsal could only partly explain the older adults' boost for the easy-to-label colours. Intriguingly, a corresponding greater age difference for easy- as compared to difficult-to-label items was not found for shape memory in Exp. 1. I explored this further in Exp. 3, reasoning that maybe older adults failed to benefit from easy-to-label shapes because they did not have sufficient time to perceive them (the encoding time in Exp. 1 was 1500 ms), thus preventing efficient labelling. Indeed, recognition of coloured objects appears faster than recognition of monochrome objects (e.g., Humphrey et al., 1994; Wurm et al., 1993), which might explain why older adults were able to benefit from labelling of easy-to-label colours, but not for the shapes (presented in black). I extended encoding time for older adults in Exp. 3 to 2500 ms, also including a condition with articulatory suppression. While older adults appeared comparatively better for easy-to-label shapes with this longer encoding time (evidence for keeping the Age Group × Label-ability interaction in model, $1/B = 6.32$), this was not mediated by suppression, suggesting that sub-vocal rehearsal by older adults was not driving this benefit. I concluded that verbalisation *combined* with some other aspect of the easy-to-label stimuli, such as semantic properties, might make such items comparatively easier to remember for older adults. Thus, younger and older adults seemed to approach the recall task somewhat differently. However, potential implications for the feature-binding literature are unclear. In the change-detection study in Chapter 3 younger adults did not appear to outperform older adults. While overall performance was better for easy-to-label items, this effect was similar in both age groups. Interestingly, the effect of label-ability was much greater in the single-feature conditions than in the binding condition, which could indicate that participants of both age groups may have sub-vocally rehearsed the easy-to-label single features. However, this study did not require suppression, so the mechanism behind the benefit for easy-to-label single features is unclear. The different patterns of results in the change-detection and reconstruction paradigms fit with evidence suggesting that the two paradigms draw on different cognitive mechanisms (e.g., Delvenne & Bruyer, 2004; Ecker et al., 2013; Shimi & Logie, 2018; Wheeler & Treisman, 2002). Taken together, our results highlighted that differences between the paradigms might be especially salient when looking at age differences; older adults appeared to benefit comparatively more than younger adults from easy-to-label, common items in the reconstruction paradigm (or be comparatively more impaired for the difficult-to-label colours), but not in the change-detection paradigm.

In the next experiment (Chapter 4), I investigated the effect of overt, instructed labelling (participants were instructed to say colour names out loud) in younger and older adults in a continuous visual WM task. I explored three questions: Do older adults seem to (1) spontaneously use verbal labels more when performing the task in silence? (2) depend more on verbal labels for visual memory performance? (3) benefit from verbal labels in the same way as younger adults? The study followed up on evidence that in younger adults, overt verbal labelling seemed to improve visual WM performance by boosting the number of categorical and continuous representations, as well as the precision of continuous representations (Souza & Skóra, 2017). I used a similar paradigm (see Souza & Skóra, 2017), and Hardman et al.'s (2017) model to separate categorical and continuous memory representations. My results suggested that older adults spontaneously used verbal labels more when performing the task in silence. While the younger adults' categorical memory representations increased when instructed to label, they were not reduced by suppression. This suggested that younger adults were not consistently labelling sub-vocally in silence. In contrast, older adults' categorical memory capacity in silence did not differ credibly from the older adults’ performance under instructed labelling, but was poorer under suppression – consistent with spontaneous sub-vocal labelling in silence. However, our hypothesis that older adults’ overall memory performance (as quantified by the P^M parameter; the overall proportion of responses that were classified as being informed by memory, as opposed to guessing) would be comparatively more impaired by suppression (compared to that of younger adults) was not supported. Instead, suppression reduced younger adults’ memory comparatively more.

However, our hypothesis that older adults’ overall memory performance would be comparatively more impaired by suppression was not supported. Instead, suppression reduced younger adults’ memory comparatively more. Thus, older adults did not appear to depend on a relatively intact verbal memory rehearsal mechanism to support their visual WM performance in this task. Finally, I explored *how* participants benefitted from labels (i.e. via increased memory precision, or more categorical or continuous representations). In our younger adults verbal labelling seemed to boost memory by improving all three of these (similar to Souza & Skóra, 2017). However, labelling did not seem to boost the number of continuous representations in older adults. Instead, the additional information associated with labelling appeared mainly categorical, suggesting that older adults’ labelling gains were verbal in nature.

Finally, I looked at the impact of strategies on WM training performance (Chapter 5) by instructing both younger and older participants to use a visualisation strategy in the N-back task. This strategy instruction was associated with greater performance improvement across the 20 training blocks in participants in both age groups. The strategy was also associated with significantly better performance on the trained N-back digit task a few days later. However, some evidence suggested that older adults struggled to implement the strategy. For instance, the older adults in the strategy group performed worse than those in the control group in the earlier training blocks. Similarly, while older adults in the strategy group reached a significantly higher maximum N-back level in the follow-up session a few days later, their average digit N-back score did not improve to the same extent as that in the younger adults, suggesting more effortful strategy implementation.

6.2 Implications

The outcomes of this thesis have implications for understanding and identifying age differences in how participants approach WM tasks. Generally, an important overarching implication is that such age differences may be observable in WM performance patterns, for instance by comparing memory for easy- and difficult-to-label stimuli (Chapters 2 and 3), the effect of articulatory suppression

(Chapters 2 and 4), categorical/continuous mixture modelling (Chapter 4), and instructed and self-reported strategy (Chapter 5). Below I summarise the key implications of age-specific strategy differences for the different research areas studied in this thesis.

6.2.1. Implications for the Feature-Binding Literature

My original prediction was that paradigms which allow verbal rehearsal would create an apparent visual feature-binding deficit in older adults, because such rehearsal is more efficient in the single-feature conditions (requiring memory for three or four features), but not the binding condition (requiring memory for six or eight features). This prediction was supported for colour memory in Exp. 1, but not in Exp. 2. For shape memory, the prediction was only weakly supported when providing longer encoding time for older adults in Exp. 3. Also, I found no evidence for a specific boost for easy-to-label single features in older adults in the change-detection study (Chapter 3). Combined, these results did not provide clear evidence that feature-binding paradigms which allow verbal rehearsal will produce ‘binding deficits’ in older adults. Rather, the results showed that verbal rehearsal might create such deficits under specific circumstances (e.g., for colour memory in a reconstruction paradigm).

The finding that older participants were comparatively more adversely affected by difficult-to-label colours in the reconstruction paradigm, but not in the change-detection paradigm, indicated that change-detection and reconstruction may draw on different processes. However, while there was no evidence that the label-ability factor was modulated by age in the change-detection paradigm, label-ability varied substantially between single and binding trials for all participants. The greater difference between the easy- and difficult-to-label stimuli in single feature conditions could be because the easy-to-label, common items allow engagement of other systems, such as verbal WM or long-term memory (see Olsson & Poom, 2005). My slightly longer-than-average encoding duration (1500 ms) and delay (2150 ms; see Table 1 for an overview of different studies) might have increased the opportunity to engage such other systems. Still, it would be interesting to investigate the effect of verbalisation in the single-feature conditions (in younger and older adults) further. Moreover, while simple feature-binding tests appear useful to distinguish healthy from pathological aging, there is some debate regarding which type of paradigm to use (see Liang et al., 2016; Parra et al., 2016). Since pathological ageing is identified by a comparison with healthy older adults, it is important that paradigms not produce ‘false’ visual binding deficits in the healthy comparison group. As our results partly suggest, verbal rehearsal or semantic representations may boost visual single-feature performance, but appear less efficient in the binding condition. This should be carefully distinguished from visual feature-binding memory deficits in different participant groups.

6.2.2. Implications for continuous WM literature

The results presented in Chapter 4 suggested that verbal labelling greatly improved continuous, visual WM performance. This study adds to the evidence suggesting that participants rely on categorical and/or verbal representations in this supposedly visual WM task (see Bae et al., 2014; Bae et al., 2015; Donkin et al., 2015; Hardman et al., 2017; Souza & Skóra, 2017). The findings suggested that people may use different kinds of mental codes (e.g., visual and verbal) interchangeably to retain information in memory.

To my knowledge, this study was the first to distinguish categorical from continuous colour-wheel responses in older adults. Importantly, age-related decline in memory precision did not appear to be driven simply by to greater reliance on categorical, verbal representations in older adults, since we observed a substantial age-related decline in the precision of continuous representations even when categorical representations were separated out – supporting the notion that declining visual WM precision is an important feature of cognitive aging (Peich et al., 2013).

Hardman et al.'s (2017) model combined with explicit manipulations of verbal strategies was useful to estimate age differences in the types of memory representations used to inform responses (i.e., continuous or categorical). These differences were not distinguishable when looking only at overall memory performance. Indeed, while overall memory performance appeared similar between age groups, using mixture modelling to explore the types of representations may indicate whether participants reach similar levels of performance using different underlying mechanisms. The indication that older participants appeared to rely more on verbal or categorical codes when performing the task spontaneously should be considered when measuring age-related visual WM decline using the colour-wheel paradigm.

6.2.3. Implications for the Cognitive Training Literature

The results presented in Chapter 6 supported the theory that strategies may play an important role in cognitive training paradigms (Laine et al., 2018). The findings also highlighted that measures of memory capacity likely partly reflect the extent to which participants apply appropriate strategies, and thus do not necessarily measure the same domain-general underlying constructs in all participants. Investigating strategy use and accounting for individual variability in such strategies (see Logie, 2018), appears a useful avenue to better understand – and hopefully resolve – discrepancies in the cognitive training literature. Similarly, accounting for strategy use seems important for investigating whether WM capacity can be ‘trained’ in a way that is meaningful beyond specific tasks.

Importantly, I found that the beneficial effect of the instructed visualisation strategy also generalised to an older adult sample. However, there were also some indication that the older adults may have found the visualisation strategy slightly harder to implement, and older adults reported spontaneously using verbal rehearsal strategies more in the N-back tasks. This indicated that training improvements could reflect different strategic approaches in younger and older adults.

The significant performance boost associated with the instructed strategy, implemented during a single 30-minute training session, and the relatively large proportion of participants in the control group who reported using some sort of strategy, suggested that strategies play a large role in such paradigms. Therefore, improvements on a trained task itself should arguably not be described as an increase in cognitive capacity unless strategy development has convincingly been ruled out. Future work is needed to investigate whether strategies developed during training are also applied to seemingly unrelated outcome measures, such as Raven's Matrices (a measure of fluid intelligence, which may be similar to aspects of Cogmed's commercial training paradigms, see Shipstead et al., 2012). Based on the observed substantial impact of strategies in such a short training session, I would argue that training transfer effects should generalise to several structurally different outcome tasks before 'training gains' are sold to members of the public as useful.

6.2.4. Implications for the 'Dull Hypothesis'

Several experimental results reported in this thesis suggested that older adults are not simply like poorly performing younger adults, thus providing evidence against the 'Dull Hypothesis'. Some results suggested that older adults may have relied more on verbal strategies – although further testing is required to understand how such strategies contribute to memory performance. Let us return to the metaphor of cognitive ability as the overall time taken to travel a mile on a motorcycle. For instance, older adults appeared able to 'drive as fast' as younger adults (i.e., perform equally well) when memory items consisted of easy-to-label single-feature colours, but lost this advantage when colours were difficult-to-label (see Chapter 2). These results were in line with recent findings suggesting that older adults may compensate for visuospatial declines by using strategies during encoding (e.g., Siegel & Castel, 2018). Our results also fit with the Scaffolding Theory of Aging (Park & Reuter-Lorenz, 2009, see Section 1.3.4), suggesting that older adults may employ compensatory recruitment such as relying on more active strategies that draw in other brain regions to compensate for deteriorating visual memory with age. Such compensatory recruitment is not necessarily associated with improved visual task performance (Reuter-Lorenz, Stanczak, & Miller, 1999). However, the substantial performance differences in an otherwise identical visual task when manipulating how easily colours can be labelled (i.e., colour commonness) indicated that – in some circumstances – older adults may successfully compensate for declining visuospatial memory. In Experiment 5 (see Chapter 4, overt labelling in continuous colour memory), older adults also seemed to spontaneously rely more on verbal, categorical representation when remembering colours in silence. However, older adults' overall memory performance was actually comparatively *less* impaired when suppression prevented verbal rehearsal, suggesting that verbal strategies may to some extent be a preference, rather than a ‘necessity’ to compensate for declining visual abilities.

Taken together, the results suggested that comparing younger and older adults’ task performance may not straightforwardly reveal age-related decline in (visual) WM, but instead applications of different strategies that tap different cognitive mechanisms, to varying degrees (for a discussion see Logie, 2018). This shows the importance of considering the interplay between visual and verbal memory (Souza & Skóra, 2017), and the importance of alternative strategies which may be used for performing the same task. These factors appear important to understand how older adults perform every day – as well as experimental – cognitive tasks. Even if overall performance levels in older adults are generally found to be poorer than those in younger adults, our results suggest that older adults’ responses to tasks *despite* cognitive decline is far from ‘dull’.

6.3 Limitations

All memoranda in this thesis consisted of arbitrary lab tasks, and thus may not generalise to how we encode visual scenes in real life, and this is a general limitation of this research. Replication using a broader range of stimuli could improve ecological validity. However, I experimentally manipulated opportunity for strategy use to see how this would influence performance in a given cognitive task, so in that sense the arbitrary tasks and stimuli were appropriate. Similarly, the role of ensemble statistics (i.e., how items may be encoded in relation to one another), known to influence memory performance (see Section 1.2.6) was not explicitly accounted for. It is possible that younger and older adults differed in the extent to which they used such statistics.

Moreover, it would be interesting for future studies to explore whether individual differences in various cognitive skills, such as inhibition, processing speed, prospective memory and attentional control – as well as measures of visual and verbal short-term memory capacity – may predict strategic approaches. This could situate this work in relation to other theories of cognitive decline and aid understanding of *why* individuals use certain strategies. For instance, perhaps a greater detrimental effect of preventing verbal rehearsal strategies in visual WM tasks would be found for older adults with especially poor visual WM abilities. The studies in this thesis were too limited in sample size to test such effects.

Relatedly, it is possible that older adults used verbal rehearsal more in some tasks because the tasks were more demanding for them than for the younger adults. Camos et al., (2011) found that when younger adults had to remember phonologically similar materials, they favoured attentional refreshing, which reduced the detrimental impact of the phonological similarity effects.

However, when task-demands drained attentional capacity, they appeared to favour rehearsal, which is less attentionally demanding. This would still result in younger and older adults approaching the same task differently, but attempts to equate task-difficulty (for instance by titrating the number of memory items based on each participants' capacity) could be useful to explore whether younger adults would appear to exhibit similar strategic adaptations if they found the task equally challenging.

All experiments involved cross-sectional comparisons of different groups of younger and older adults. Therefore, age-group differences as measured do not necessarily only reflect the effect of age on specific WM tasks, but also cohort effects, education effects and lifestyle factors. For instance, older adults had significantly higher predicted verbal IQ scores (using the NART; National Adult Reading Test in Experiments 1 – 4). This is in line with research suggesting that verbal, crystallised abilities are relatively intact with age (see Section 1.4.1). While the younger and older adults generally reported similar levels of education, that level might have been less common in the older adult cohort. Similarly, older adults who are motivated - and healthy enough - to participate in research may constitute a select sample. More broadly, the younger and older adults may volunteer for different reasons. For example, if younger adults tend to participate mainly for the financial reward, they may be less concerned about performing the task as well as possible. Thus, the strategic differences observed could be driven by differences in task-motivation. Alternatively, greater tendencies to use verbal rehearsal by older adults could be driven by more verbal-based education in that cohort, compared to greater tendency for visuospatial processing (e.g., driven by more computer based-learning and general exposure to computers and video games) in younger adults. While it would be useful to attempt to compare such factors and make groups as equal as possible, some of these differences appear difficult to control. Also, since our questions were about how experimental participants in the two age groups approached the tasks (e.g., to address previous discrepancies in the literature), controlling for factors such as exposure to computers and motivation could result in ‘throwing the baby out with the bathwater’, since it is likely that previous results were driven by similar confounding factors. However, cohort differences in younger and older adults limit our understanding of ‘pure’ cognitive decline, and there is no way to know whether results would replicate if we waited 40 years to test the younger adults again.

Using computerised tasks introduces another limitation. Although I did take care to ensure instructions were as clear as possible, and that all participants practiced each task before the experimental session, factors such as poorer computer-mouse skills may slow down reaction times in older adults (participants responded with a mouse-click in Exp. 1-3, and 5, some tasks in Exp. 6). Physical slowing could also contribute to such effects (see Section 1.3.1 on processing speed theory of memory decline with age). These inherent factors of ageing seem difficult to circumvent – if responses were recorded using pen and paper, older adults might write more slowly. Slower response execution results in a longer period for potential decay and interference in memory (see Salthouse, 1996). It is possible that older adults' response slowness explained increased use of or benefit from verbal traces, if such traces are more resistant to decay than perceptual traces (see Donkin et al., 2015). The work in this thesis did not explore response speed as a potential mechanism behind this effect. Future research controlling response rates by slowing the rate at which the mouse cursor can move for younger adults to mimic average response times in older adults might give some insight.

Moreover, it is possible that older adults were less able to judge their own performance, and how it was affected by strategies, as suggested by Fox and Charness (2010). Indeed, subjective measures of memory may become less accurate with age (e.g., Bunnell et al., 1999; Crumley, Stetler, & Horhota, 2014, but see Dunlosky & Hertzog, 1997; Halamish, McGillivray, & Castel, 2011; Rabinowitz, Ackerman, Craik, & Hinchley, 1982). It would be interesting to investigate if performance feedback influenced the use of strategies across trials in younger and older adults.

Experiments 1-5 did not include a way to distinguish the generation of a label (which might result in the activation of a long-term Memory representation) from its rehearsal (which might rely on verbal WM mechanisms). Articulatory suppression was initiated prior to item presentation *and* during the retention interval, thus presumably minimising both the generation and rehearsal of labels. Applying suppression selectively to the memory encoding phase and/or the retention interval might distinguish prevention of label generation from its rehearsal, and thus help understand the mechanism(s) by which older adults benefitted comparatively more from being easy-to-label colours.

Other limitations in the present research relate more directly to problems within particular studies. These issues are discussed in the relevant chapters, so only a brief summary is provided here.

Chapters 2 and 3. I used two sets of easy- and difficult-to-label colours and shapes. The sets were rather limited (eight items in each), and differed in more ways than being easy- or difficult-to-label. For instance, the difficult-to-label shapes appeared more visually complex than the easy-to-label shapes. The easy-to-label colours may have been more salient than their difficult-to-label counterparts. I did attempt to ensure overall luminance levels were similar between the sets, but this was not perfect (see Appendix A). I attempted to ensure that perceptual differences did not explain the older adults larger performance differences for the two colour (and later shape) sets by including articulatory suppression. The benefit for easy-to-label shapes was not abolished under suppression, suggesting that those items were probably better remembered due to other reasons. I discussed these perceptual differences, as well as the potential role of long-term memory representations of easy-to-label items in the relevant chapter.

Chapter 4. All results in Chapter 4 are contingent on the estimation of continuous and categorical memory representations using the CatCont Model (Hardman et al., 2017). While the model has been used by others, there are no guarantees that it does not over- or underestimate representations in either category, nor that it can account for potential age-differences in for instance motor skills – this might appear as decreased memory precision in older adults, or result in classification of continuous responses as categorical.

Chapter 5. A large proportion of participants were noncompliant (i.e., they were instructed to use the visualisation strategy but reported failing to comply). We cannot be sure whether those participants were not able to apply the strategy, tried to implement it but failed and therefore gave up, or simply preferred not to use it. Potential removal of participants who were not able to use the strategy could have muddled our age-comparison. For instance, the older adults group might have appeared less able to benefit from the strategy if those with impaired abilities had not been removed from the sample. Also, some of the pre-registered analyses looking at the association between self-reported strategy types and performance suffered from very small sample sizes in certain strategy groups. I used the same strategy grouping classification as Laine et al. (2018), aiming to replicate their results. However, a large proportion of strategies were classified as 'Other' (around 16%), which suggests that adding some more strategy types might fit our data better.

6.4 Theoretical Implications?

As stated in the introduction (Chapter 1), the present work was not designed to distinguish between models of WM, nor to investigate whether verbal information and visual information are stored and processed within a single system or in multiple, domain-specific systems. However, the emphasis on verbal rehearsal and the possible recruitment of separate resources by participants of different age groups in this thesis was arguably most aligned with the Multiple-Component model framework, in its emphasis on visual and verbal WM as distinct capacities. Similarly, the quest to 'reject the Dull Hypothesis' of cognitive ageing – to the extent that this endeavour focuses on certain abilities which are more intact than others – when applied WM is perhaps inherently 'modular' (in contrast, if WM capacity is considered constrained by some general attentional capacity, the notion that memory for 'visual' and 'verbal' information may decline differently with age is arguably less of a next-logical-step). Thus, the way our research questions (and hypotheses) are framed, may be most compatible with the Multiple-Component model. However, observations that older adults may be comparatively 'more worse' when we prevent verbal rehearsal (see Chapter 2) are not incompatible with other conceptions of WM, since the importance of verbal rehearsal is also recognized in such frameworks (e.g., Cowan, Saults, & Blume, 2014; Camos, Lagner, & Barrouillet, 2009). Some models may regard memory for 'verbal' information as part of a shared, central resource. Verbal rehearsal may be seen as analogous to writing information down on a notepad – this would presumably improve WM capacity considerably, but the act of writing would not been seen as a fundamental, theoretical WM process. Others may regard verbal rehearsal as part of the WM system itself. Arguably, such model discrepancies are less about the fundamental principles of cognition (since there is agreement that verbal rehearsal is used), and more about what we as researchers decide to include into the construct we label "Working Memory". The ability to rehearse verbally can be seen as an experimental confound; something we need to control (e.g., with Articulatory Suppression, difficult-to-name stimuli, or very brief retention intervals) to truly measure (visual) WM – similar to preventing participants from writing information down. On the other hand, if the ability to rehearse verbally differs between people, and influences WM performance, it appears important to consider for the overall output of the cognitive system (such as good memory performance). Some of our results indicate that there may be systematic differences in the use of verbal rehearsal process between younger and older adults, and thus support the notion that verbal/visual strategic processes are a) important to acknowledge, and b) may provide insights into the operation of moment-to-moment retention of information as we age. In sum, our results highlight the importance of considering the impact of verbal rehearsal in research on visual WM in ageing, regardless of where and how such processes fit within different theoretical WM frameworks.

6.5 Final Conclusions

Researchers often attempt to isolate and measure a specific feature of cognition by reducing it to a simple cognitive task. For instance, visual WM is researched by measuring participants' ability to remember a set of coloured shapes on a computer screen. This reduction allows us to conduct experiments. However, real-life memory depends on an interaction between numerous abilities. The work in this thesis was based on the notion that a no experimental task is 'process pure', and that overall performance may reflect a variety of factors, depending on which strategy we use. I attempted to increase our understanding of how strategic approaches in WM tasks may differ between younger and older adults, with a focus on the potential 'problems' this can cause if a given task is assumed to measure one specific ability. The studies in Chapters 2 – 4 explicitly explored whether the opportunity to rehearse single and bound visual features verbally would affect participants of different ages differently. Taken together, the results suggested that it might.

Differential effects on performance in younger and older adults when certain strategic options were 'blocked' via experimental manipulations, and some differences in self-reported strategy use, suggested systematic group differences in how younger and older adults approached the same WM task. Arguably, this is the most important contribution of this thesis. However, my research also highlighted that just like strategies may differ between younger and older adults, they also differ between (and likely within) individuals, as well as between tasks, and between specific conditions within a task (e.g., remembering single or bound features). In general, common (and thus verbalise-able) materials appeared to afford more opportunity for elaboration (perhaps via verbal rehearsal or some other process; such as activation of long-term memory representation), and influenced younger and older adults' overall performance differently – across different tasks and task conditions. The research in this thesis illustrated that differences between younger and older adults use of verbal rehearsal in visual WM task are not straightforward. However, using stimuli and paradigms presumed to minimise verbal strategy use appeared useful to reduce confounds of such age-related strategic differences.

The indications of differential WM task approaches by younger and older adults presented in thesis highlight the extreme challenges facing researchers who attempt to understand cognitive aging. However, such differential approaches can also be seen as opportunities for developing environments to support older adults' memory, and for understanding of how resilient older adults – and the cognitive system – may be when presented with challenging cognitive tasks, despite age-related changes in abilities.

References

References

Adam, K. C., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology*, *97*, 79-97.

Aine, C. J., Woodruff, C. C., Knoefel, J. E., Adair, J. C., Hudson, D., Qualls, C., ... & Stephen, J. M. (2006). Aging: compensation or maturation?. *NeuroImage*, *32*(4), 1891-1904.

Alala, B., Mwangi, W., & Okeyo, G. (2014). Image representation using RGB color space. *International Journal of Innovative Research and Development*, 3(8).

Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2006). Is the binding of visual features in working memory resource-demanding?. *Journal of Experimental Psychology: General*, *135*(2), 298.

Allen, R. J., Brown, L. A., & Niven, E. (2013). Aging and visual feature binding in working memory. *Working memory: Developmental differences, component processes, and improvement mechanisms*, 83-96.

Allen, R. J., Hitch, G. J., Mate, J., & Baddeley, A. D. (2012). Feature binding and attention in working memory: A resolution of previous contradictory findings. *The Quarterly Journal of Experimental Psychology*, *65*(12), 2369-2383.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in cognitive sciences*, *15*(3), 122-131.

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.

Arenberg, D. (1977). The effects of auditory augmentation on visual retention for young and old adults. *Journal of Gerontology*, *32*(2), 192-195.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological science*, *12*(2), 157-162.

Atkinson, A. L., Baddeley, A. D., & Allen, R. J. (2017). Remember some or remember all? Ageing and strategy effects in visual working memory. *The Quarterly Journal of Experimental Psychology*, 1-41.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. 2, pp. 89-195). Academic Press.

Awh, 2018, Conference Talk at Psychonomics Amsterdam, Amsterdam, The Netherlands, May, 2018.

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in cognitive sciences*, *5*(3), 119-126.

Baars, B. J. (2002). The conscious access hypothesis: origins and recent evidence. *Trends in cognitive sciences*, *6*(1), 47-52.

Babcock, R. L., & Salthouse, T. A. (1990). Effects of increased processing demands on age differences in working memory. *Psychology and aging*, *5*(3), 421.

Baddeley, A. (1986). Oxford psychology series, No. 11. Working memory. New York, NY, US.

Baddeley, A. (1992). Working memory. Science, 255(5044), 556-559. 10.1126/science.1736359

Baddeley, A. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology Section A*, *49*(1), 5-28.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews neuroscience*, *4*(10), 829.

Baddeley, A. (2007). *Working memory, thought, and action* (Vol. 45). OUP Oxford.

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual review of psychology*, *63*, 1-29.

Baddeley, A. D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *The Quarterly journal of experimental psychology*, *18*(4), 302-309.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47-89). Academic press.

Baddeley, A. D., & Hitch, G. J. (2018). The phonological loop as a buffer store: An update. *Cortex*.

Baddeley, A. D., & Lieberman, K. (2017). Spatial working memory. In *Exploring Working Memory* (pp. 206-223). Routledge.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). New York, NY, US: Cambridge University Press.

Baddeley, A. D., Grant, S., Wight, E., & Thomson, N. (1975). Imagery and visual working memory. *Attention and performance V*, 205-217.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of verbal learning and verbal behavior*, *14*(6), 575-589.

Baddeley, A., & Wilson, B. A. (2002). Prose recall and amnesia: Implications for the structure of working memory. *Neuropsychologia*, *40*(10), 1737-1743.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological review*, *105*(1), 158.

Baddeley, A.D. & Logie, R.H. (1999). Working memory: The multiple component model. In A. Miyake & P. Shah (eds.) *Models of Working Memory*, pp 28-61. New York: Cambridge University Press.

Baddeley, A.D., Lewis, V.J. & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology*, 36, 233-252.

Bae, G. Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, *144*(4), 744.

Bae, G. Y., Olkkonen, M., Allred, S. R., Wilson, C., & Flombaum, J. I. (2014). Stimulus-specific variability in color working memory with delayed estimation. *Journal of Vision*, *14*(4), 7-7.

Bailey, H. R., Dunlosky, J., & Hertzog, C. (2014). Does strategy training reduce age-related deficits in working memory?. *Gerontology*, *60*(4), 346-356.

Balinsky, B. (1941). An analysis of the mental factors of various age groups from nine to sixty. *Genetic Psychology Monographs*.

Baltes, P. B., & Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: a new window to the study of cognitive aging?. *Psychology and aging*, *12*(1), 12.

Baltes, P.B. & Baltes, M.M. (1990). Psychological perspectives on successful aging: The model of selective optimization with compensation. In P.B. Baltes & M.M. Baltes

(Eds.) *Successful Aging: Perspectives from the Behavioral Sciences*. Cambridge University Press, Cambridge, UK, pp1-34.

Banich, M. T. (1998). The missing link: the role of interhemispheric interaction in attentional processing. *Brain and cognition*, *36*(2), 128-157.

Barnard, P. J. (1999). Interacting cognitive subsystems: Modelling working memory phenomena with a multi-processor architecture. In "Models of Working Memory" Miyake, A. and Shah, P.

Barrouillet, P. & Camos, V. (2015). *Working Memory: Loss and Reconstruction*. Hove, UK: Psychology Press.

Barrouillet, P., & Camos, V. (2007). The time-based resource-sharing model of working memory. *The cognitive neuroscience of working memory*, *455*, 59-80.

Barrouillet, P., & Camos, V. (2015). Essays in cognitive psychology.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*(1), 83.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes.

Bayarri, M. J., & Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science*, 58-80.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, 321(5890), 851-854.

Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of vision*, *9*(10), 7-7.

Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia*, *49*(6), 1622-1631.

Belopolsky, A. V., & Theeuwes, J. (2009). When are attention and saccade preparation dissociated?. *Psychological Science*, *20*(11), 1340-1347.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289-300.

Birren, J. E. (1965). Age changes in speed of behavior: Its central nature and physiological correlates. *Behavior, aging, and the nervous system*, 191-216.

Blieszner, R., Willis, S. L., & Baltes, P. B. (1981). Training research in aging on the fluid ability of inductive reasoning. *Journal of applied developmental psychology*, *2*(3), 247-265.

Blumenfeld-Katzir, T., Pasternak, O., Dagan, M., & Assaf, Y. (2011). Diffusion MRI of structural brain plasticity induced by a learning and memory task. *PloS one*, *6*(6), e20678.

Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *60*(5), P223-P233.

Bopp, K. L., & Verhaeghen, P. (2009). Working memory and aging: separating the effects of content and context. *Psychology and aging*, 24(4), 968.

Borella, E., Carretti, B., & De Beni, R. (2008). Working memory and inhibition across the adult life-span. *Acta psychologica*, *128*(1), 33-44.Bowles, R. P., & Salthouse, T. A. (2003). Assessing the age-related effects of proactive interference on working memory tasks using the Rasch model. *Psychology and aging*, *18*(3), 608.

Bowles, R. P., & Salthouse, T. A. (2003). Assessing the age-related effects of proactive interference on working memory tasks using the Rasch model. *Psychology and aging*, *18*(3), 608.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological science*, *22*(3), 384-392

Brady, T. F., & Alvarez, G. A. (2015). Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision*, 15(15), 6-6.

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological review*, *120*(1), 85.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of vision*, *11*(5), 4-4.

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. (1992). Verbal recoding of visual stimuli impairs mental image transformations. *Memory & Cognition*, *20*(4), 449-455.

Brandt, J., & Folstein, M. F. (2003). *TICS, telephone interview for cognitive status: Professional manual*. Psychological Assessment Resources.

Brehmer, Y., Westerberg, H., & Bäckman, L. (2012). Working-memory training in younger and older adults: training gains, transfer, and maintenance. *Frontiers in human neuroscience*, *6*, 63.

Broadbent, D. (1958). E.(1958). *Perception and communication*. Elmsford, NY, US.

Brockmole, J. R., & Logie, R. H. (2013). Age-related change in visual working memory: a study of 55,753 participants aged 8–75. *Frontiers in psychology*, *4*, 12.

Brockmole, J. R., Parra, M. A., Della Sala, S., & Logie, R. H. (2008). Do binding deficits account for age-related decline in visual working memory?. *Psychonomic Bulletin & Review*, 15(3), 543-547.

Bromley, D. B. (1958). Some effects of age on short-term learning and remembering. *Journal of Gerontology*.

Brown, L. A., Brockmole, J. R., Gow, A. J., & Deary, I. J. (2012). Processing speed and visuospatial executive function predict visual working memory ability in older adults. *Experimental Aging Research*, 38(1), 1-19.

Brown, L. A., Forbes, D., & McConnell, J. (2006). Limiting the use of verbal coding in the Visual Patterns Test. *The Quarterly journal of experimental psychology*, 59(7), 1169-1176.

Brown, L. A., Niven, E. H., Logie, R. H., Rhodes, S., & Allen, R. J. (2017). Visual feature binding in younger and older adults: Encoding and suffix interference effects. *Memory*, *25*(2), 261-275.

Bunnell, J. K., Baken, D. M., & Richards-Ward, L. A. (1999). The effect of age on metamemory for working memory. *New Zealand Journal of Psychology*, *28*(1), 23.

Burke, D. M., & Light, L. L. (1981). Memory and aging: The role of retrieval processes. *Psychological bulletin*, *90*(3), 513.

Bürki, C. N., Ludwig, C., Chicherio, C., & de Ribaupierre, A. (2014). Individual differences in cognitive plasticity: an investigation of training curves in younger and older adults. *Psychological Research*, *78*(6), 821-835.

Butler, M., McCreedy, E., Nelson, V. A., Desai, P., Ratner, E., Fink, H. A., ... & Davila, H. (2018). Does cognitive training prevent cognitive decline?: a systematic review. *Annals of internal medicine*, *168*(1), 63-68.

Buttle, H., & Raymond, J. E. (2003). High familiarity enhances visual change detection for face stimuli. *Perception & psychophysics*, *65*(8), 1296-1306.

Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: the HAROLD model. *Psychology and aging*, *17*(1), 85.

Cabeza, R., Grady, C. L., Nyberg, L., McIntosh, A. R., Tulving, E., Kapur, S., ... & Craik, F. I. (1997). Age-related differences in neural activity during memory encoding and retrieval: a positron emission tomography study. *Journal of neuroscience*, *17*(1), 391-400.

Callaghan, T. (1984) Dimensional interaction of hue and brightness in preattentive field.

Calvillo, D. P. (2012). Working memory and the memory distortion component of hindsight bias. *Memory*, *20*(8), 891-898.

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

Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language*, 61(3), 457-469.

Camos, V., Mora, G., & Oberauer, K. (2011). Adaptive choice between articulatory rehearsal and attentional refreshing in verbal working memory. *Memory & Cognition*, *39*(2), 231-244.

Castel, A. D., Benjamin, A. S., Craik, F. I., & Watkins, M. J. (2002). The effects of aging on selectivity and control in short-term recall. *Memory & Cognition*, *30*(7), 1078-1085. https://doi.org/10.3758/BF03194325

Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. *The journal of abnormal and social psychology*, *38*(4), 476.

Cavanagh, P. (1987). Reconstructing the third dimension: Interactions between color, texture, motion, binocular disparity, and shape. *Computer Vision, Graphics, and Image Processing,* 37(2), 171-195.

Chalfonte, B. L. & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & cognition*, 24(4), 403-416.

Chein, J. M., & Fiez, J. A. (2010). Evaluating models of working memory through the effects of concurrent irrelevant information. *Journal of Experimental Psychology: General*, *139*(1), 117.

Chen, T., & Naveh-Benjamin, M. (2012). Assessing the associative deficit of older adults in long-term and short-term/working memory. *Psychology and aging, 27*(3), 666.

Cherry, K. E., Park, D. C., & Donaldson, H. (1993). Adult age differences in spatial memory: Effects of structural context and practice. *Experimental Aging Research*, *19*(4), 333-350.

Chin, J. M., & Schooler, J. W. (2008). Why do words hurt? Content, process, and criterion shift accounts of verbal overshadowing. *European Journal of Cognitive Psychology*, *20*(3), 396-413.

Chiu, H. L., Chu, H., Tsai, J. C., Liu, D., Chen, Y. R., Yang, H. L., & Chou, K. R. (2017). The effect of cognitive-based training for the healthy older people: A meta-analysis of randomized controlled trials. *PloS one*, *12*(5), e0176742.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision research*, *43*(4), 393-404.

Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition*, *30*(7), 1086-1095.

Cogmed. (2011). *Frequently asked questions: What is Cogmed all about?* Retrieved from http://www.cogmed.com/faq (archived by webcite at http://www.webcitation.org/625IdpzwG).

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, *70*(4), 213.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British journal of Psychology*, *55*(1), 75-84.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British journal of psychology*, *55*(4), 429-432.

Conway, A. R., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic bulletin & review*, *8*(2), 331-335.

Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in cognitive sciences, 7*(12), 547-552.

Cowan, N. (1992). Verbal memory span and the timing of spoken recall. *Journal of Memory and Language, 31*(5), 668-684.

Cowan, N. (1999). An embedded-processes model of working memory. *Models of working memory: Mechanisms of active maintenance and executive control*, *20*, 506.

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.

Cowan, N. (2005). Working memory capacity limits in a theoretical context. In *Human learning and memory: Advances in theory and application. The 4th Tsukuba international conference on memory* (pp. 155-175). Lawrence Erlbaum Associates, Publishers Mahwah.

Cowan, N. (2010). The magical mystery four how is working memory capacity limited, and why?. *Current directions in psychological science*,19(1), 51-57.

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic bulletin & review*, *24*(4), 1158-1170.

Cowan, N., & Chen, Z. (2008). How chunks form in long-term memory and affect short-term memory limits. In *Interactions between short-term and long-term memory in the verbal domain* (pp. 98-119). Psychology Press.

Cowan, N., & Morey, C. C. (2007). How can dual-task working memory retention limits be investigated?. *Psychological science*, *18*(8), 686-688.

Cowan, N., Chen, Z., & Rouder, J. N. (2004). Constant capacity in an immediate serial-recall task: A logical sequel to Miller (1956). *Psychological science*, *15*(9), 634-640.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive psychology*, *51*(1), 42-100.

Cowan, N., Naveh-Benjamin, M., Kilb, A., & Saults, J. S. (2006). Life-span development of visual working memory: When is feature binding difficult?. *Developmental psychology, 42*(6), 1089.

Cowan, N., Saults, J. S., & Blume, C. L. (2014). Central and peripheral components of working memory storage. *Journal of Experimental Psychology: General*, *143*(5), 1806.

Craik, F. I. (1994). Memory changes in normal aging. *Current directions in psychological science*, *3*(5), 155-158.

Craik, F. I. M., Klix, F., & Hagendorf, H. (1986). Human memory and cognitive capabilities: Mechanisms and performances. *A functional account of age differences in memory. Amsterdam: Elsevier Science*, 395-422.

Craik, F. I., & Byrd, M. (1982). Aging and cognitive deficits. In *Aging and cognitive processes* (pp. 191-211). Springer, Boston, MA.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, *11*(6), 671-684.

Craik, F. I., & Rose, N. S. (2012). Memory encoding and aging: a neurocognitive perspective. *Neuroscience & Biobehavioral Reviews*, *36*(7), 1729-1739.

Craik, F. I., Luo, L., & Sakuta, Y. (2010). Effects of aging and divided attention on memory for items and their contexts. *Psychology and Aging*, *25*(4), 968.

Craik, F. I., Luo, L., & Sakuta, Y. (2010). Effects of aging and divided attention on memory for items and their contexts. *Psychology and Aging*, *25*(4), 968.

Crumley, J. J., Stetler, C. A., & Horhota, M. (2014). Examining the relationship between subjective and objective memory performance in older adults: A meta-analysis. *Psychology and Aging*, *29*(2), 250.

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

Curby, K. M., Glazek, K., & Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 94.

Dahlin, E., Neely, A. S., Larsson, A., Bäckman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science*, *320*(5882), 1510-1512.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic bulletin & review*, *3*(4), 422-433.

De Frias, C. M., Lövdén, M., Lindenberger, U., & Nilsson, L. G. (2007). Revisiting the dedifferentiation hypothesis with longitudinal multi-cohort data. *Intelligence*, *35*(4), 381-392.

De Lange, A. M. G., Bråthen, A. C. S., Rohani, D. A., Grydeland, H., Fjell, A. M., & Walhovd, K. B. (2017). The effects of memory training on behavioral and microstructural plasticity in young and older adults. *Human brain mapping*, *38*(11), 5666-5680.

Deary, I. J., Gow, A. J., Taylor, M. D., Corley, J., Brett, C., Wilson, V., ... & Starr, J. M. (2007). The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC geriatrics*, *7*(1), 28.

Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., ... & Fazio, F. (1997). Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain: a journal of neurology*, *120*(10), 1763-1777.

Della Sala, S., & Logie, R. (1993). When working memory does not work. The role of working memory in neuropsychology. In F. Boller, & H. Spinnler (Eds.), *Handbook of Neuropsychology, Volume 8* (pp. 1-63). Amsterdam: Elsevier B.V..

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: a tool for unwelding visuo–spatial memory. *Neuropsychologia*, *37*(10), 1189-1199.

Delvenne, J. F., & Bruyer, R. (2004). Does visual short-term memory store bound features?. *Visual cognition*, 11(1), 1-27.

Denney, N. W., & Heidrich, S. M. (1990). Training effects on Raven's Progressive Matrices in young, middle-aged, and elderly adults. *Psychology and Aging*, *5*(1), 144.

D'esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual review of psychology*, *66*, 115-142.

Dickey, J. M. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics*, 204-223.

Dienes, Z. (2012). Using Bayes to Interpret Non-significant Results. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 34, No. 34). https://escholarship.org/uc/item/2pm65411

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in psychology*, *5*, 781.

Doherty, J. M., Belletier, C., Rhodes, S., Jaroslawska, A., Barrouillet, P., Camos, V., ... & Logie, R. H. (2018). Dual-task costs in working memory: An adversarial collaboration. Journal of experimental psychology: Learning, Memory, and Cognition.

Donkin, C., Nosofsky, R., Gold, J., & Shiffrin, R. (2015). Verbal labeling, gradual decay, and sudden death in visual short-term memory. *Psychonomic Bulletin & Review*, *22*(1), 170-178.

Dunlosky, J., & Hertzog, C. (1997). Older and younger adults use a functionally identical algorithm to select items for restudy during multitrial learning. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *52*(4), P178-P186.

Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *The Quarterly Journal of Experimental Psychology*, *60*(9), 1227-1245.

Dunning, D. L., & Holmes, J. (2014). Does working memory training promote the use of strategies on untrained working memory tasks?. *Memory & cognition*, *42*(6), 854-862.

Dvorine, I. (1963). Quantitative classification of the color-blind. *The Journal of General Psychology, 68*(2), 255-265. https://doi.org/10.1080/00221309.1963.9920533

Ebbinghaus, H. (1885/1964). *Über das Gedächnis* (Leipzig: Duncker) [translated by H.A. Ruger & C.E. Bussenius]. New York: Dover Publications.

Ecker, U. K., Maybery, M., & Zimmer, H. D. (2013). Binding of intrinsic and extrinsic features in working memory. *Journal of Experimental Psychology: General, 142*(1), 218.

Eisenreich, B.R., Akaishi, R., & Hayden, B.Y. (2017). Control without controllers: Towards a distributed neuroscience of executive control. *Journal of Cognitive Neuroscience*, 29, 10, 1684-1698.

Emrich, S. M., & Ferber, S. (2012). Competition increases binding errors in visual working memory. *Journal of Vision*, *12*(4), 12-12.

Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review*, 12 (6), 1127–1133. https://doi.org/10.3758/BF03206454

Engle, R. W. (2002). Working memory capacity as executive attention. *Current directions in psychological science*, *11*(1), 19-23.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of learning and motivation*, *44*, 145-200.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), Models of working memory: mechanisms of active maintenance and executive control (pp. 102–134). Cambridge, UK: Cambridge University Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of experimental psychology: General*, *128*(3), 309.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking?. *Trends in cognitive sciences, 4*(9), 345-352.

Ericcson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, *208*(4448), 1181-1182.

Ericsson, K. A. (2017). Expertise and individual differences: the search for the structure and acquisition of experts' superior performance. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(1-2), e1382.

Ericsson, K. A., Cheng, X., Pan, Y., Ku, Y., Ge, Y., & Hu, Y. (2017). Memory skills mediating superior memory in a world-class memorist. *Memory*, *25*, 1294–1302.

Estes, W. K. (1973). Phonemic coding and rehearsal in short-term memory for letter strings. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 360-372.

Fabiani, M., Buckley, J., Gratton, G., Coles, M. G., Donchin, E., & Logie, R. (1989). The training of complex task performance. *Acta Psychologica*, *71*(1-3), 259-299.

Fallon, S. J., Mattiesing, R. M., Muhammed, K., Manohar, S., & Husain, M. (2017). Fractionating the neurocognitive mechanisms underlying working memory: independent effects of dopamine and Parkinson's disease. *Cerebral Cortex, 27*(12), 5727-5738.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191.

Federal Trade Commission. (2016). Lumosity to pay 2$ million to settle FTC deceptive advertising charges for its "brain training" program. *Federal Trade Commission*.

Fisk, J. E., & Warr, P. (1996). Age and working memory: the role of perceptual speed, the central executive, and the phonological loop. *Psychology and aging, 11*(2), 316. 10.1037/0882-7974.11.2.316

Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. *Journal of vision*, *11*(12), 3-3.

Fougnie, D., & Marois, R. (2011). What limits working memory capacity? Evidence for modality-specific sources to the simultaneous storage of visual and auditory arrays. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1329.

Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory?. *Journal of vision*, *10*(12), 27-27.

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature communications*, *3*, 1229.

Fougnie, D., Zughni, S., Godwin, D., & Marois, R. (2015). Working memory storage is intrinsically domain specific. *Journal of Experimental Psychology: General*, *144*(1), 30.

Fox, M. C., & Charness, N. (2010). How to gain eleven IQ points in ten minutes: Thinking aloud improves Raven's Matrices performance in older adults. *Aging, Neuropsychology, and cognition*, *17*(2), 191-204.

Fukuda, K., & Vogel, E. K. (2011). Individual differences in recovery time from attentional capture. *Psychological science*, *22*(3), 361-368.

Gamerman, D., & Lopes, H. F. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman and Hall/CRC.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of memory and language*, *28*(2), 200-213.

Gazzaley, A., Cooney, J. W., Rissman, J., & D'Esposito, M. (2005). Top-down suppression deficit underlies working memory impairment in normal aging. *Nature neuroscience*, *8*(10), 1298.

Godijn, R., & Theeuwes, J. (2012). Overt is no better than covert when rehearsing visuo-spatial information in working memory. *Memory & Cognition*, *40*(1), 52-61.

Gold, J. M., Hahn, B., Zhang, W. W., Robinson, B. M., Kappenman, E. S., Beck, V. M., & Luck, S. J. (2010). Reduced capacity but spared precision and maintenance of working memory representations in schizophrenia. *Archives of general psychiatry*, *67*(6), 570-577.

Gonçalves, V. P., de Almeida Neris, V. P., Seraphini, S., Dias, T. C., Pessin, G., Johnson, T., & Ueyama, J. (2017). Providing adaptive smartphone interfaces targeted at elderly people: an approach that takes into account diversity among the elderly. *Universal Access in the Information Society*, *16*(1), 129-149.

Gopher, D., Williges, B. H., Williges, R. C., & Damos, D. L. (1975). Varying the type and number of adaptive variables in continuous tracking. *Journal of motor behavior*, *7*(3), 159-170.

Green, C. S., & Bavelier, D. (2008). Exercising your brain: a review of human brain plasticity and training-induced learning. *Psychology and aging*, *23*(4), 692.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 711-732.

Gross, A. L., Parisi, J. M., Spira, A. P., Kueider, A. M., Ko, J. Y., Saczynski, J. S., ... & Rebok, G. W. (2012). Memory training interventions for older adults: A meta-analysis. *Aging & mental health*, *16*(6), 722-734.

Gruber, O. (2001). Effects of domain-specific interference on brain activation associated with verbal working memory task performance. *Cerebral Cortex*, *11*(11), 1047-1055.

Guest, D., Howard, C. J., Brown, L. A., & Gleeson, H. (2015). Aging and the rate of visual information processing. *Journal of vision, 15*(14), 10-10. doi:10.1167/15.14.10

Halamish, V., McGillivray, S., & Castel, A. D. (2011). Monitoring one's own forgetting in younger and older adults. *Psychology and aging*, *26*(3), 631.

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in cognitive sciences*, *11*(6), 236-242.

Hambrick, D. Z., Oswald, F. L., Darowski, E. S., Rench, T. A., & Brou, R. (2010). Predictors of multitasking performance in a synthetic work paradigm. *Applied cognitive psychology*, *24*(8), 1149-1167.

Han, C., & Carlin, B. P. (2001). Markov chain Monte Carlo methods for computing Bayes factors: A comparative review*. Journal of the American Statistical Association, 96*(455), 1122-1132.

Hanley, J.R. & Young, A.W. (2019). ELD Revisited: A second look at a neuropsychological impairment of working memory affecting retention of visuo-spatial material. *Cortex*, 112, 172-179.

Hanley, J.R., Young, A.W., & Pearson, N. (1991). Impairment of the visuo-spatial sketch pad. *The quarterly Journal of Experimental Psychology, 43*(1), 101-125.

Hardman, K. O., & Cowan, N. (2015). Remembering complex objects in visual working memory: Do capacity limits restrict objects or features?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 325.

Hardman, K. O., Vergauwe, E., & Ricker, T. J. (2017). Categorical working memory representations are used in delayed estimation of continuous colors. *Journal of Experimental Psychology: Human Perception and Performance, 43*(1), 30.

Hardman, K.O. (2017). CatContModel: Categorical and Continuous Working Memory Models for Delayed Estimation Tasks (Version 0.8.0). Retrieved from <https://github.com/hardmanko/CatContModel/releases/tag/v0.8.0>.

Hardman, K.O. (2017). CMBBHT: Cell Means Based Bayesian Hypothesis Tests. R. Retrieved from <https://github.com/hardmanko/CMBBHT> (Original work published March 5, 2017).

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In *Psychology of learning and motivation* (Vol. 22, pp. 193-225). Academic Press.

Hasher, L., Lustig, C., & Zacks, R. T. (2007). Inhibitory mechanisms and the control of attention. *Variation in working memory*, *19*, 227-249.

Haslam, C., Wills, A. J., Haslam, S. A., Kay, J., Baron, R., & McNab, F. (2007). Does maintenance of colour categories rely on language? Evidence to the contrary from a case of semantic dementia. *Brain and language, 103*(3), 251-263.

Hatano, A., Ueno, T., Kitagami, S., & Kawaguchi, J. (2015). Why verbalization of non-verbal memory reduces recognition accuracy: A computational approach to verbal overshadowing. *PloS one*, *10*(6), e0127618.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425-2430.

Haynes, J. D., & Rees, G. (2006). Neuroimaging: decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, *7*(7), 523.

Hedden, T., Lautenschlager, G., & Park, D. C. (2005). Contributions of processing ability and knowledge to verbal memory tasks across the adult life-span. *The Quarterly Journal of Experimental Psychology Section A*, *58*(1), 169-190.

Heinzel, S., Schulte, S., Onken, J., Duong, Q. L., Riemer, T. G., Heinz, A., ... & Rapp, M. A. (2014). Working memory training improvements and gains in non-trained cognitive tasks in young and older adults. *Aging, Neuropsychology, and Cognition*, *21*(2), 146-173.

Hertzog, C. (1985). An individual differences perspective: Implications for cognitive research in gerontology. *Research on Aging*, *7*(1), 7-45.

Hertzog, C., Kramer, A. F., Wilson, R. S., & Lindenberger, U. (2008). Enrichment effects on adult cognitive development: can the functional capacity of older adults be preserved and enhanced?. *Psychological science in the public interest*, *9*(1), 1-65.

Hitch, G. J., Halliday, S., Schaafstal, A. M., & Schraagen, J. M. C. (1988). Visual working memory in young children. *Memory & cognition*, *16*(2), 120-132.

Hodges, J. R. (2012). Addenbrooke's Cognitive Examination-III (ACE-III).

Hodges, J. R., & Larner, A. J. (2017). Addenbrooke's Cognitive Examinations: ACE, ACE-R, ACE-III, ACEapp, and M-ACE. In *Cognitive Screening Instruments* (pp. 109-137). Springer, Cham.

Hollingworth, A., & Rasmussen, I. P. (2010). Binding objects to locations: The relationship between object files and visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(3), 543.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental science*, *12*(4), F9-F15.

Horn, J. L. (1989). Intelligence: Measurement, theory, and public policy: Proceedings of a symposium in honor of Lloyd G. *Humphreys*. University of Illinois Press.

Horn, J. L., & Masunaga, H. (2000). New directions for research into aging and intelligence: The development of expertise.

Hsieh, S., McGrory, S., Leslie, F., Dawson, K., Ahmed, S., Butler, C. R., ... & Hodges, J. R. (2015). The Mini-Addenbrooke's Cognitive Examination: a new assessment tool for dementia. *Dementia and geriatric cognitive disorders*, *39*(1-2), 1-11.

Hu, Y., Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2016). Executive control of stimulus-driven and goal-directed attention in visual working memory. *Attention, Perception, & Psychophysics, 78*(7), 2164-2175.

Humphrey, G. K., Goodale, M. A., Jakobson, L. S., & Servos, P. (1994). The role of surface information in object recognition: Studies of a visual form agnosic and normal subjects. *Perception, 23*(12), 1457-1481.

Hunt, E., & Agnoli, F. (1991). The Whorfian hypothesis: A cognitive psychology perspective. *Psychological Review*, *98*(3), 377.

Hyde, T. S., & Jenkins, J. J. (1969). Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology*, *82*(3), 472.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829-6833.

Jaeggi, S. M., Buschkuehl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & cognition*, *42*(3), 464-480.

James, W. (1890). The principles of psychology, Vol. 2. NY, US: Henry Holt and Company.

Jeffreys, H. (1961). Theory of Probability. Oxford: UK Oxford University Press.

Jenkins, L., Myerson, J., Hale, S., & Fry, A. F. (1999). Individual and developmental differences in working memory across the life span. *Psychonomic Bulletin & Review*, *6*(1), 28-40.

Jenkins, L., Myerson, J., Joerding, J. A., & Hale, S. (2000). Converging evidence that visuospatial cognition is more age-sensitive than verbal cognition. *Psychology and aging*, *15*(1), 157.

Johnson, J. S., Spencer, J. P., Luck, S. J., & Schöner, G. (2009). A dynamic neural field model of visual working memory and change detection. *Psychological science*, *20*(5), 568-577.

Johnson, W., Logie, R. H., & Brockmole, J. R. (2010). Working memory tasks differ in factor structure across age cohorts: Implications for dedifferentiation. *Intelligence*, *38*(5), 513-528.

Jonides, J., Reuter-Lorenz, P. A., Smith, E. E., Awh, E., Barnes, L. L., Drain, M., ... & Schumacher, E. H. (1996). Verbal and spatial working memory in humans. *Psychology of learning and motivation, 35*, 43-88.

Jost, K., Bryck, R. L., Vogel, E. K., & Mayr, U. (2010). Are old adults just like low working memory young adults? Filtering efficiency and age differences in visual working memory. *Cerebral cortex*, *21*(5), 1147-1154.

Juan-Espinosa, M., Garcıa, L. F., Escorial, S., Rebollo, I., Colom, R., & Abad, F. J. (2002). Age dedifferentiation hypothesis: Evidence from the WAIS III. *Intelligence*, *30*(5), 395-408.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of experimental psychology: General*, *132*(1), 47.

Kane, M. J., Conway, A. R., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. *Variation in working memory*, *1*, 21-48.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. (2005). Working memory capacity and fluid intelligence are strongly related constructs: comment on Ackerman, Beier, and Boyle (2005).

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189.

Kane, M. J., Hasher, L., Stoltzfus, E. R., Zacks, R. T., & Connelly, S. L. (1994). Inhibitory attentional mechanisms and aging. *Psychology and aging*, *9*(1), 103.

Kane, R. L., Butler, M., Fink, H. A., Brasure, M., Davila, H., Desai, P., ... & Calvert, C. (2017). Interventions to prevent age-related cognitive decline, mild cognitive impairment, and clinical Alzheimer's-type dementia.

Karbach, J., & Verhaeghen, P. (2014). Making working memory work: a meta-analysis of executive-control and working memory training in older adults. *Psychological science*, *25*(11), 2027-2037.

Kausler, D. H. (1970). Retention–-Forgetting as a Nomological Network for Developmental Research. In *Life-span developmental psychology* (pp. 305-353). Academic Press.

Kausler, D. H. (2012). *Experimental psychology, cognition, and human aging*. Springer Science & Business Media.

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of verbal learning and verbal behavior*, *1*(3), 153-161.

Kessels, R. P., Hobbel, D., & Postma, A. (2007). Aging, context memory and binding: A comparison of "what, where and when" in young and older adults. *International Journal of Neuroscience, 117*(6), 795-810.

Killin, L., Abrahams, S., Parra, M. A., & Della Sala, S. (2018). The effect of age on the FCSRT-IR and temporary visual memory binding. *International psychogeriatrics*, *30*(3), 331-340.

Kinjo, H. (2010). Effects of self-paced encoding and practice on age-related deficits in binding three features. *The International Journal of Aging and Human Development*, *71*(3), 185-208.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of experimental psychology*, *55*(4), 352.

Kitagami, S. (2000). The influence of verbal encoding on the memory of visual information. *Shinrigaku kenkyu: The Japanese journal of psychology, 71*(5), 387-394.

Klauer, K. C., & Zhao, Z. (2004). Double dissociations in visual and spatial short-term memory. *Journal of experimental psychology: General*, *133*(3), 355.

Klauer, K. J., Willmes, K., & Phye, G. D. (2002). Inducing inductive reasoning: Does it transfer to fluid intelligence?. *Contemporary educational psychology*, *27*(1), 1-25.

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in cognitive sciences*, *14*(7), 317-324.

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., ... & Westerberg, H. (2005). Computerized training of working memory in children with ADHD-a randomized, controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*(2), 177-186.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of clinical and experimental neuropsychology*, *24*(6), 781-791.

Kovacs, O., & Harris, I. M. (2019). The role of location in visual feature binding. *Attention, Perception, & Psychophysics*, 1-13.

Kramer, A. F., Humphrey, D. G., Larish, J. F., & Logan, G. D. (1994). Aging and inhibition: beyond a unitary view of inhibitory processing in attention. *Psychology and aging*, *9*(4), 491.

Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science, 6*(3), 299-312.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?!. *Intelligence, 14*(4), 389-433.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?!. *Intelligence*, *14*(4), 389-433.

Laine, M., Fellman, D., Waris, O., & Nyman, T. J. (2018). The early effects of external and internal strategies on working memory updating training. *Scientific reports*, *8*(1), 4045.

Lakha, L., & Wright, M. J. (2004). Capacity limitations of visual memory in two-interval comparison of Gabor arrays. *Vision research, 44*(14), 1707-1716.

Lane, S. M., & Schooler, J. W. (2004). Skimming the surface: Verbal overshadowing of analogical retrieval. *Psychological Science*, *15*(11), 715-719.

Leonards, U., Ibanez, V., & Giannakopoulos, P. (2002). The role of stimulus type in age-related changes of visual working memory. *Experimental Brain Research*, *146*(2), 172-183.

Lewis-Peacock, J. A., & Postle, B. R. (2012). Decoding the internal focus of attention. *Neuropsychologia*, *50*(4), 470-478.

Lewis-Peacock, J. A., Drysdale, A. T., & Postle, B. R. (2014). Neural evidence for the flexible control of mental representations. *Cerebral Cortex*, *25*(10), 3303-3313.

Li, S. C., & Lindenberger, U. (1999). Cross-level unification: A computational exploration of the link between deterioration of neurotransmitter systems and dedifferentiation of cognitive abilities in old age. In *Cognitive neuroscience of memory* (pp. 103-146). Hogrefe & Huber.

Li, S. C., Lindenberger, U., & Sikström, S. (2001). Aging cognition: from neuromodulation to representation. *Trends in cognitive sciences*, *5*(11), 479-486.

Li, S. C., Schmiedek, F., Huxhold, O., Röcke, C., Smith, J., & Lindenberger, U. (2008). Working memory plasticity in old age: practice gain, transfer, and maintenance. *Psychology and aging*, *23*(4), 731.

Liang, Y., Pertzov, Y., Nicholas, J. M., Henley, S. M., Crutch, S., Woodward, F., ... & Husain, M. (2016). Visual short-term memory binding deficit in familial Alzheimer's disease. *Cortex*, *78*, 150-164.

Liesefeld, H. R., Liesefeld, A. M., & Müller, H. J. (2019). Two good reasons to say "change!"–ensemble representations as well as item representations impact standard measures of vwm capacity. *British Journal of Psychology*, *110* (2), 328–356.

Lin, P. H., & Luck, S. J. (2012). Proactive interference does not meaningfully distort visual working memory capacity estimates in the canonical change detection task. *Frontiers in psychology*, *3*, 42.

Lissauer, H., & Jackson, M. (1988). A case of visual agnosia with a contribution to theory. *Cognitive neuropsychology*, *5*(2), 157-192.

Logie, R. (2018). Human cognition: Common principles and individual variation. *Journal of applied research in memory and cognition*, *7*(4), 471-486.

Logie, R. H. (1986). Visuo-spatial processing in working memory. *The Quarterly Journal of Experimental Psychology Section A*, *38*(2), 229-247.

Logie, R. H. (2011). The functional organization and capacity limits of working memory. *Current directions in Psychological science*, *20*(4), 240-245.

Logie, R. H. (2012). Cognitive training: Strategies and the multicomponent cognitive system. *Journal of Applied Research in Memory and Cognition, 1*, 206–207.

Logie, R. H. (2016). Retiring the central executive. *The Quarterly Journal of Experimental Psychology*, *69*(10), 2093-2109.

Logie, R. H., & Cowan, N. (2015). Perspectives on working memory: introduction to the special issue. *Memory & cognition*, *43*(3), 315-324.

Logie, R. H., & Cowan, N. (2015). Perspectives on working memory: introduction to the special issue. *Memory & cognition*, *43*(3), 315-324.

Logie, R. H., & Maylor, E. A. (2009). An Internet study of prospective memory across adulthood. *Psychology and aging*, *24*(3), 767.

Logie, R. H., & Morris, R. G. (Eds.). (2014). *Working memory and ageing*. Hove, UK: Psychology Press.

Logie, R. H., & Niven, E. H. (2012). Working memory: An ensemble of functions in on-line cognition. *From mental imagery to spatial cognition and language: Essays in honour of Michel Denis*, 77-105.

Logie, R. H., & Pearson, D. G. (1997). The inner eye and the inner scribe of visuo-spatial working memory: Evidence from developmental fractionation. *European Journal of cognitive psychology*, *9*(3), 241-257.

Logie, R. H., Brockmole, J. R., & Jaswal, S. (2011). Feature binding in visual short-term memory is unaffected by task-irrelevant changes of location, shape, and color. *Memory & cognition*, *39*(1), 24-36.

Logie, R. H., Brockmole, J. R., & Vandenbroucke, A. R. (2009). Bound feature combinations in visual short-term memory are fragile but influence long-term learning. *Visual Cognition*, *17*(1-2), 160-179.

Logie, R. H., Della Sala, S., Laiacona, M., Chalmers, P., & Wynn, V. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Memory & Cognition*, 24(3), 305-321.

Logie, R. H., Della Sala, S., Laiacona, M., Chalmers, P., & Wynn, V. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Memory & Cognition*, *24*(3), 305-321.

Logie, R. H., Horne, M. J., & Pettit, L. D. (2015). When cognitive performance does not decline across the lifespan. *Working memory and ageing*, 21-47.

Logie, R. H., Saito, S., Morita, A., Varma, S., & Norris, D. (2016). Recalling visual serial order for verbal sequences. *Memory & Cognition*, 44(4), 590-607.

Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychologica*, *75*(1), 55-74.

Logie, R.H. & Della Sala, S. (2005). Disorders of visuo-spatial working memory. In P. Shah and A. Miyake (Eds.) *Handbook of Visuospatial Thinking.* Cambridge University Press: New York, pp 81-120.

Logie, R.H. (1995). Visuo-Spatial Working Memory. Hove, UK: Lawrence Erlbaum Associates.

Logie, R.H. (2003). Spatial and Visual Working Memory: A Mental Workspace. In D. Irwin and B Ross (Eds.) *Cognitive Vision: The Psychology of Learning and Motivation*, Vol 42, pp 37-78. Elsevier Science (USA).

Logie, R.H. (2016). Retiring the Central Executive. *Quarterly Journal of Experimental Psychology,* 69, 2093–2109.

Logie, R.H., Parra, M.A., & Della Sala, S. (2015). From cognitive science to dementia assessment. *Policy Insights from the Behavioral and Brain Sciences,* 2, 81-91.

Logie, R.H., Pernet, C.R., Buonocore, A., Della Sala, S. (2011). Low and High Imagers Activate Networks Differentially in Mental Rotation. *Neuropsychologia*, 49, 3071–3077.

Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., & Schmiedek, F. (2010). A theoretical framework for the study of adult cognitive plasticity. *Psychological bulletin*, *136*(4), 659.

Lövdén, M., Brehmer, Y., Li, S. C., & Lindenberger, U. (2012). Training-induced compensation versus magnification of individual differences in memory performance. *Frontiers in human neuroscience*, *6*, 141.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279.

Lumosity. (2011). *Enhance creativity*. Retrieved from http://www.lumosity.com/how-we-help/enhance-creativity(archived by webcite at http://www.webcitation.org/63nvD8NzR).

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.

Lustig, C., & Jantz, T. (2015). Questions of age differences in interference control: when and how, not if?. *Brain Research*, *1612*, 59-69.

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature neuroscience*, *17*(3), 347-356.

Matsukura, M., & Hollingworth, A. (2011). Does visual short-term memory have a high-capacity stage?. *Psychonomic bulletin & review*, *18*(6), 1098-1104.

McNab, F., Zeidman, P., Rutledge, R. B., Smittenaar, P., Brown, H. R., Adams, R. A., & Dolan, R. J. (2015). Age-related changes in working memory and the ability to ignore distraction. *Proceedings of the National Academy of Sciences*, *112*(20), 6515-6518.

McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & cognition*, *29*(1), 10-17.

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

Melby-Lervåg, M., & Hulme, C. (2016). There is no convincing evidence that working memory training is effective: A reply to Au et al. (2014) and Karbach and Verhaeghen (2014). *Psychonomic Bulletin & Review*, *23*(1), 324-330.

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

Miles, W. R. (1933). Age and human ability. *Psychological Review*, *40*(2), 99.

Milham, M. P., Erickson, K. I., Banich, M. T., Kramer, A. F., Webb, A., Wszalek, T., & Cohen, N. J. (2002). Attentional control in the aging brain: insights from an fMRI study of the stroop task. *Brain and cognition*, *49*(3), 277-296.

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.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). Plans and the structure of behavior.

Milner, B., Corkin, S., & Teuber, H. L. (1968). Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of HM. *Neuropsychologia*, *6*(3), 215-234.

Mindsparke. (2011). *Increase IQ*. Retrieved from http://www.mindsparke.com/increase_iq.php (archived by webcite at http://www.webcitation.org/625JO0RZx).

Mioshi, E., Dawson, K., Mitchell, J., Arnold, R., & Hodges, J. R. (2006). The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences*, *21*(11), 1078-1085.

Mitchell, D. J., & Cusack, R. (2018). Visual short-term memory through the lifespan: Preserved benefits of context and metacognition. *Psychology and aging*, *33*(5), 841.

Mitchell, K. J., Johnson, M. K., Raye, C. L., & D'Esposito, M. (2000b). fMRI evidence of age-related hippocampal dysfunction in feature binding in working memory. *Cognitive brain research, 10*(1), 197-206.

Mitchell, K. J., Johnson, M. K., Raye, C. L., Mather, M., & D'Esposito, M. (2000a). Aging and reflective processes of working memory: binding and test load deficits. *Psychology and aging, 15*(3), 527.

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

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

Morey, C. C. (2018). The case against specialized visual-spatial short-term memory. *Psychological bulletin*, *144*(8), 849.

Morey, C. C., & Cowan, N. (2004). When visual and verbal memories compete: Evidence of cross-domain limits in working memory. *Psychonomic bulletin & review*, *11*(2), 296-301.

Morey, C. C., & Cowan, N. (2005). When do visual and verbal memories conflict? The importance of working-memory load and retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(4), 703.

Morey, C. C., & Mall, J. T. (2012). Cross-domain interference costs during concurrent verbal and spatial serial memory tasks are asymmetric. *The Quarterly Journal of Experimental Psychology*, *65*(9), 1777-1797.

Morey, C. C., & Miron, M. D. (2016). Spatial sequences, but not verbal sequences, are vulnerable to general interference during retention in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(12), 1907.

Morey, C. C., Cowan, N., Morey, R. D., & Rouder, J. N. (2011). Flexible attention allocation to visual and auditory working memory tasks: Manipulating reward induces a trade-off. *Attention, Perception, & Psychophysics*, *73*(2), 458-472.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *reason*, *4*(2), 61-64.

Morey, R. D. (2015, January 30). On verbal categories for the interpretation of Bayes factors [Blog post]. Retrieved from https://richarddmorey.org/2015/01/on-verbal-categories-for-the-interpretation-of-bayes-factors/

Morey, R. D., Rouder, J. N., & Jamil, T. (2015). BayesFactor: Computation of Bayes factors for common designs. R package version 0.9, 9, 2014.

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.

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.

Murray, D. (1965). Vocalization-at-presentation, with varying presentation rates. Quarterly Journal of Experimental Psychology, 17, 47-56.

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology*, *78*(4p1), 679.

Murty, V. P., Sambataro, F., Radulescu, E., Altamura, M., Iudicello, J., Zoltick, B., ... & Mattay, V. S. (2011). Selective updating of working memory content modulates meso-cortico-striatal activity. *Neuroimage*, *57*(3), 1264-1272.

Myerson, J., Emery, L., White, D. A., & Hale, S. (2003). Effects of age, domain, and processing demands on memory span: Evidence for differential decline. *Aging, Neuropsychology, and Cognition*, *10*(1), 20-27.

Myerson, J., Hale, S., Rhee, S. H., & Jenkins, L. (1999). Selective interference with verbal and spatial working memory in young and older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 54*(3), P161-P164.

Myerson, J., Hale, S., Rhee, S. H., & Jenkins, L. (1999). Selective interference with verbal and spatial working memory in young and older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *54*(3), P161-P164.

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1170.

Naveh-Benjamin, M., Hussain, Z., Guez, J., & Bar-On, M. (2003). Adult age differences in episodic memory: further support for an associative-deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(5), 826.

Nelson, H. E. (1982). National Adult Reading Test (NART): For the assessment of premorbid intelligence in patients with dementia: Test manual. NFER-Nelson.

Noack, H., Lövdén, M., Schmiedek, F., & Lindenberger, U. (2009). Cognitive plasticity in adulthood and old age: gauging the generality of cognitive intervention effects. *Restorative neurology and neuroscience*, *27*(5), 435-453.

Nordahl, C. W., Ranganath, C., Yonelinas, A. P., DeCarli, C., Fletcher, E., & Jagust, W. J. (2006). White matter changes compromise prefrontal cortex function in healthy elderly individuals. *Journal of cognitive neuroscience*, *18*(3), 418-429.

Norman, D. A., & Shallice, T. (1986). Attention to action. In *Consciousness and self-regulation* (pp. 1-18). Springer, Boston, MA.

Nosofsky, R. M., & Donkin, C. (2016). Response-time evidence for mixed memory states in a sequential-presentation change-detection task. Cognitive Psychology, 84, 31-62.

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

Oberauer, K., & Eichenberger, S. (2013). Visual working memory declines when more features must be remembered for each object. *Memory & Cognition, 41*(8), 1212-1227. https://doi.org/10.3758/s13421-013-0333-6

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D., Conway, A., Cowan, N., ... & Ma, W. J. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885.

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D., Conway, A., Cowan, N., ... & Ma, W. J. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885.

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: a meta-analysis. *Psychology and aging*, *23*(1), 104.

Olson, I. R., & Jiang, Y. (2004). Visual short-term memory is not improved by training. *Memory & cognition*, *32*(8), 1326-1332.

Olson, I. R., Zhang, J. X., Mitchell, K. J., Johnson, M. K., Bloise, S. M., & Higgins, J. A. (2004). Preserved spatial memory over brief intervals in older adults. *Psychology and Aging, 19*(2), 310.

Olson, I. R., Zhang, J. X., Mitchell, K. J., Johnson, M. K., Bloise, S. M., & Higgins, J. A. (2004). Preserved spatial memory over brief intervals in older adults. *Psychology and Aging*, *19*(2), 310.

Olsson, H., & Poom, L. (2005). Visual memory needs categories. *Proceedings of the National Academy of Sciences*, *102*(24), 8776-8780.

Osaka, M., Otsuka, Y., & Osaka, N. (2012). Verbal to visual code switching improves working memory in older adults: an fMRI study. *Frontiers in human neuroscience*, *6*, 24.

Paivio, A. (1971). Imagery and verbal processes. New York, NY: Holt, Rinheart & Winston.

Paivio, A. 1986. Mental representation: A dual-coding approach.

Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance, 16*(2), 332.

Papagno, C., & Shallice, T. (2019). Introduction to impairments of short-term memory buffers: Do they exist?.

Park, D. C., & Festini, S. B. (2017). Theories of memory and aging: A look at the past and a glimpse of the future. *The Journals of Gerontology: Series B*, *72*(1), 82-90.

Park, D. C., & Reuter-Lorenz, P. (2009). The adaptive brain: aging and neurocognitive scaffolding. *Annual review of psychology, 60*, 173-196.

Park, D. C., & Reuter-Lorenz, P. (2009). The adaptive brain: aging and neurocognitive scaffolding. *Annual review of psychology*, *60*, 173-196.

Park, D. C., & Shaw, R. J. (1992). Effect of environmental support on implicit and explicit memory in younger and older adults. *Psychology and Aging*, *7*(4), 632.

Park, D. C., and Payer, D. (2006). Working memory across the adult lifespan. In *Lifespan Cognition: Mechanisms of Change (pp.* 128–142), New York, US: Oxford University Press.

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.

Park, D. C., Polk, T. A., Mikels, J. A., Taylor, S. F., & Marshuetz, C. (2001). Cerebral aging: integration of brain and behavioral models of cognitive function. *Dialogues in clinical neuroscience*, *3*(3), 151.

Park, D., & Schwarz, N. (2012). *Cognitive aging: A primer*. Psychology Press.

Parra, M. A., Abrahams, S., Fabi, K., Logie, R., Luzzi, S., & Sala, S. D. (2009). Short-term memory binding deficits in Alzheimer's disease. *Brain*, *132*(4), 1057-1066.

Parra, M. A., Abrahams, S., Logie, R. H., & Della Sala, S. (2009). Age and binding within-dimension features in visual short-term memory. *Neuroscience Letters, 449*(1), 1-5.

Parra, M. A., Abrahams, S., Logie, R. H., Méndez, L. G., Lopera, F., & Della Sala, S. (2010). Visual short-term memory binding deficits in familial Alzheimer’s disease. *Brain, 133*(9), 2702-2713.

Parra, M.A., Della Sala, S., Logie, R.H. & Abrahams, S. (2009) Selective impairment in visual short-term memory binding. *Cognitive Neuropsychology,* 26, 583 - 605.

Pashler, H. (1988). Familiarity and visual change detection. *Perception & psychophysics*, *44*(4), 369-378.

Patrick, D. L., Starks, H. E., Cain, K. C., Uhlmann, R. F., & Pearlman, R. A. (1994). Measuring preferences for health states worse than death. *Medical Decision Making*, *14*(1), 9-18.

Payer, D., Marshuetz, C., Sutton, B., Hebrank, A., Welsh, R. C., & Park, D. C. (2006). Decreased neural specialization in old adults on a working memory task. *Neuroreport*, *17*(5), 487-491.

Pearson, D. G., Ball, K., & Smith, D. T. (2014). Oculomotor preparation as a rehearsal mechanism in spatial working memory. *Cognition*, *132*(3), 416-428.

Pearson, D., & Sahraie, A. (2003). Oculomotor control and the maintenance of spatially and temporally distributed events in visuo-spatial working memory. *The Quarterly Journal of Experimental Psychology Section A*, *56*(7), 1089-1111.

Peich, M. C., Husain, M., & Bays, P. M. (2013). Age-related decline of precision and binding in visual working memory. *Psychology and aging*, *28*(3), 729.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods, 162*(1-2), 8-13.

Peng, P., & Fuchs, D. (2017). A randomized control trial of working memory training with and without strategy instruction: Effects on young children’s working memory and comprehension. *Journal of learning disabilities*, *50*(1), 62-80.

Perfect, T. J., & Maylor, E. A. (2000). *Rejecting the dull hypothesis: The relation between method and theory in cognitive aging research*. Oxford University Press.

Persson, J., Nyberg, L., Lind, J., Larsson, A., Nilsson, L. G., Ingvar, M., & Buckner, R. L. (2005). Structure–function correlates of cognitive decline in aging. *Cerebral cortex*, *16*(7), 907-915.

Pertzov, Y., Dong, M. Y., Peich, M. C., & Husain, M. (2012). Forgetting what was where: The fragility of object-location binding. *PLoS One*, *7*(10), e48214.

Pertzov, Y., Heider, M., Liang, Y., & Husain, M. (2015). Effects of healthy ageing on precision and binding of object location in visual short term memory. *Psychology and Aging*, *30*(1), 26.

Peterson, D. J., & Naveh-Benjamin, M. (2016). The role of aging in intra-item and item-context binding processes in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(11), 1713.

Peterson, L. R., & Johnson, S. T. (1971). Some effects of minimizing articulation on short-term retention. *Journal of Verbal Learning and Verbal Behavior*, *10*(4), 346-354.

Phillips, W. A., & Christie, D. F. M. (1977). Interference with visualization. *Quarterly Journal of Experimental Psychology, 29*(4), 637-650.

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, *139*(1), 23-38.

Postle, B. R., & Hamidi, M. (2006). Nonvisual codes and nonvisual brain areas support visual working memory. *Cerebral Cortex*, *17*(9), 2151-2162.

Postle, B. R., D'Esposito, M., & Corkin, S. (2005). Effects of verbal and nonverbal interference on spatial and object visual working memory. *Memory & cognition*, *33*(2), 203-212.

Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998). Phenomenology of attention: I. Color, location, orientation, and spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(1), 261.

Quinn, J. G., & Ralston, G. E. (1986). Movement and attention in visual working memory. *The Quarterly Journal of Experimental Psychology Section A*, *38*(4), 689-703.

R Core Team. A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. http://www.R-project.org/ (2013).

Rabbitt, P. (1965). Age and discrimination between complex stimuli. *Behavior, aging, and the nervous system*, 35-53.

Rabbitt, P.M.A .(ed.) (2005). Cognitive Gerontology: Cognitive Changes in Old Age. Special issue of *Quarterly Journal of Experimental Psychology,* 58A(1).

Rabinowitz, J. C., Ackerman, B. P., Craik, F. I., & Hinchley, J. L. (1982). Aging and metamemory: The roles of relatedness and imagery. *Journal of Gerontology*, *37*(6), 688-695.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological methodology*, 111-163. https://www.jstor.org/stable/271063

Ramaty, A., & Luria, R. (2018). Visual working memory cannot trade quantity for quality. *Frontiers in psychology, 9.*

Raz, N., & Rodriguez, K. M. (2006). Differential aging of the brain: patterns, cognitive correlates and modifiers. *Neuroscience & Biobehavioral Reviews*,*30*(6), 730-748.

Raz, N., Gunning-Dixon, F. M., Head, D., Dupuis, J. H., & Acker, J. D. (1998). Neuroanatomical correlates of cognitive aging: evidence from structural magnetic resonance imaging. *Neuropsychology*, *12*(1), 95.

Read, C. A., Rogers, J. M., & Wilson, P. H. (2016). Working memory binding of visual object features in older adults. *Aging, Neuropsychology, and Cognition*, *23*(3), 263-281.

Redick, T. S., Calvo, A., Gay, C. E., & Engle, R. W. (2011). Working memory capacity and go/no-go task performance: selective effects of updating, maintenance, and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 308.

Resnick, S. M., Goldszal, A. F., Davatzikos, C., Golski, S., Kraut, M. A., Metter, E. J., ... & Zonderman, A. B. (2000). One-year age changes in MRI brain volumes in older adults. *Cerebral cortex*, *10*(5), 464-472.

Rettmann, M. E., Kraut, M. A., Prince, J. L., & Resnick, S. M. (2006). Cross-sectional and longitudinal analyses of anatomical sulcal changes associated with aging. *Cerebral Cortex*, *16*(11), 1584-1594.

Reuter-Lorenz, P. A. (2002). New visions of the aging mind and brain. *Trends in cognitive sciences*, *6*(9), 394-400.

Reuter-Lorenz, P. A., & Cappell, K. A. (2008). Neurocognitive aging and the compensation hypothesis. *Current directions in psychological science*, *17*(3), 177-182.

Reuter-Lorenz, P. A., & Park, D. C. (2014). How does it STAC up? Revisiting the scaffolding theory of aging and cognition. *Neuropsychology review*, *24*(3), 355-370.

Reuter-Lorenz, P. A., & Sylvester, C. Y. C. (2005). The cognitive neuroscience of working memory and aging. *Cognitive neuroscience of aging: Linking cognitive and cerebral aging*, 186-217.

Reuter-Lorenz, P. A., Jonides, J., Smith, E. E., Hartley, A., Miller, A., Marshuetz, C., & Koeppe, R. A. (2000). Age differences in the frontal lateralization of verbal and spatial working memory revealed by PET. *Journal of cognitive neuroscience*, *12*(1), 174-187.

Reuter-Lorenz, P. A., Stanczak, L., & Miller, A. C. (1999). Neural recruitment and cognitive aging: Two hemispheres are better than one, especially as you age. *Psychological Science, 10*(6), 494-500.

Rhodes, S., Cowan, N., Hardman, K. O., & Logie, R. H. (2018). Informed guessing in change detection. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *44* (7), 1023–1035.

Rhodes, S., Parra, M. A., & Logie, R. H. (2016). Ageing and feature binding in visual working memory: The role of presentation time. *The Quarterly Journal of Experimental Psychology, 69*(4), 654-668.

Rhodes, S., Parra, M. A., Cowan, N., & Logie, R. H. (2017). Healthy aging and visual working memory: The effect of mixing feature and conjunction changes. *Psychology and aging*, *32*(4), 354.

Richmond, L. L., Morrison, A. B., Chein, J. M., & Olson, I. R. (2011). Working memory training and transfer in older adults. *Psychology and aging*, *26*(4), 813.

Rodakowski, J., Saghafi, E., Butters, M. A., & Skidmore, E. R. (2015). Non-pharmacological interventions for adults with mild cognitive impairment and early stage dementia: An updated scoping review. *Molecular aspects of medicine*, *43*, 38-53.

Rosi, A., Del Signore, F., Canelli, E., Allegri, N., Bottiroli, S., Vecchi, T., & Cavallini, E. (2018). The effect of strategic memory training in older adults: who benefits most?. *International psychogeriatrics*, *30*(8), 1235-1242.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*(5), 356-374.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review, 16*(2), 225-237.

Rudkin, S. J., Pearson, D. G., & Logie, R. H. (2007). Executive processes in visual and spatial working memory tasks. *The Quarterly Journal of Experimental Psychology*, *60*(1), 79-100.

Sailor, K. A., Schinder, A. F., & Lledo, P. M. (2017). Adult neurogenesis beyond the niche: its potential for driving brain plasticity. *Current opinion in neurobiology*, *42*, 111-117.

Saito, S., Logie, R. H., Morita, A., & Law, A. (2008). Visual and phonological similarity effects in verbal immediate serial recall: A test with kanji materials. *Journal of Memory and Language*, *59*(1), 1-17.

Sala, G., & Gobet, F. (2017). Does far transfer exist? Negative evidence from chess, music, and working memory training. *Current directions in psychological science*, *26*(6), 515-520.

Sala, G., Aksayli, N. D., Semir, K., Gondo, Y., & Gobet, F. (2018). Working memory training does not enhance older adults' cognitive skills: A meta-analysis.

Salame, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of verbal learning and verbal behavior*, *21*(2), 150-164

Salamé, P., & Baddeley, A. (1986). Phonological factors in STM: Similarity and the unattended speech effect. *Bulletin of the Psychonomic Society*, *24*(4), 263-265.

Salat, D. H., Buckner, R. L., Snyder, A. Z., Greve, D. N., Desikan, R. S., Busa, E., ... & Fischl, B. (2004). Thinning of the cerebral cortex in aging.*Cerebral cortex*, *14*(7), 721-730.

Salthouse, T. A. (1985). Speed of behavior and its implications for cognition.

Salthouse, T. A. (1995). Differential age-related influences on memory for verbal-symbolic information and visual-spatial information?. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *50*(4), P193-P201.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological review, 103*(3), 403.

Salthouse, T. A. (2000). Steps toward the explanation of adult age differences in cognition.

Salway, A.F.S. & Logie, R.H. (1995). Visuo-spatial working memory, movement control and executive demands. *British Journal of Psychology,* 86*,* 253-269.

Sander, M. C., Werkle-Bergner, M., & Lindenberger, U. (2011). Binding and strategic selection in working memory: A lifespan dissociation. *Psychology and Aging, 26*(3), 612.

Saults, J. S., & Cowan, N. (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology: General*, *136*(4), 663.

Schneegans, S., & Bays, P. M. (2019). New perspectives on binding in visual working memory. *British Journal of Psychology*, *110* (2), 207–244.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive psychology*, *22*(1), 36-71.

Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist*, *50*(2), 138-166.

Schweickert, R., Guentert, L., & Hersberger, L. (1990). Phonological similarity, pronunciation rate, and memory span. *Psychological Science*, *1*(1), 74-77.

Sense, F., Morey, C. C., Prince, M., Heathcote, A., & Morey, R. D. (2017). Opportunity for verbalization does not improve visual change detection performance: A state-trace analysis. *Behavior research methods, 49*(3), 853-862.

Sewell, D. K., Lilburn, S. D., & Smith, P. L. (2014). An information capacity limitation of visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 40*(6), 2214.

Shallice, T. I. M., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, *114*(2), 727-741.

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *The Quarterly journal of experimental psychology*, *22*(2), 261-273.

Shimi, A., & Logie, R. H. (2018). Feature binding in short-term memory and long-term learning. *Quarterly Journal of Experimental Psychology*, 1747021818807718.

Shing, Y. L., Werkle-Bergner, M., Brehmer, Y., Müller, V., Li, S. C., & Lindenberger, U. (2010). Episodic memory across the lifespan: The contributions of associative and strategic components. *Neuroscience & Biobehavioral Reviews*, *34*(7), 1080-1091.

Shipstead, Z., Lindsey, D. R., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language*, *72*, 116-141.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2010). Does working memory training generalize?. *Psychologica Belgica*, *50*(3), 245-276.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective?. *Psychological bulletin*, *138*(4), 628.

Shulman, H. G. (1971). Similarity effects in short-term memory. *Psychological Bulletin*, *75*(6), 399.

Siegel, A. L., & Castel, A. D. (2018). Memory for important item-location associations in younger and older adults. *Psychology and aging*, *33*(1), 30.

Siegler, S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, *116*, 250–264.

Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science*, *7*(5), 301-305.

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*(1), 76-80.

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

Smyth, M. M., & Scholey, K. A. (1994). Interference in immediate spatial memory. *Memory & Cognition*, *22*(1), 1-13.

Sørensen, T. A., & Kyllingsbæk, S. (2012). Short-term storage capacity for visual objects depends on expertise. *Acta psychologica*, *140*(2), 158-163.

Souza, A. S., & Oberauer, K. (2018). Does articulatory rehearsal help immediate serial recall?. *Cognitive psychology*, 107, 1-21.

Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. *Cognition*, *166*, 277-297.

Soveri, A., Antfolk, J., Karlsson, L., Salo, B., & Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin & Review*, *24*(4), 1077-1096.

Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: a meta-analysis. *Psychology and aging*, *10*(4), 527.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers*, *31*(1), 137-149.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5). Boston, MA: Pearson.

Tas, A. C., Luck, S. J., & Hollingworth, A. (2016). The relationship between visual attention and visual working memory encoding: A dissociation between covert and overt orienting. *Journal of experimental psychology: human perception and performance, 42*(8), 1121.

Team, R. C. (2015). R: A Language and Environment for Statistical Computing (version 3.3. 2). *R Foundation for Statistical Computing, Vienna.*

Thalmann, M., & Oberauer, K. (2017). Domain-specific interference between storage and processing in complex span is driven by cognitive and motor operations. *The Quarterly Journal of Experimental Psychology*, *70*(1), 109-126.

Thomas, A. K., Bonura, B. M., Taylor, H. A., & Brunyé, T. T. (2012). Metacognitive monitoring in visuospatial working memory. *Psychology and Aging*, *27*(4), 1099.

Thorell, L. B., Lindqvist, S., Bergman Nutley, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental science*, *12*(1), 106-113.

Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review*, *38*(5), 406.

Tomaszewski Farias, S., Cahn-Weiner, D. A., Harvey, D. J., Reed, B. R., Mungas, D., Kramer, J. H., & Chui, H. (2009). Longitudinal changes in memory and executive functioning are associated with longitudinal change in instrumental activities of daily living in older adults. *The Clinical Neuropsychologist*, *23*(3), 446-461.

Toppino, T. C., & Pisegna, A. (2005). Articulatory suppression and the irrelevant speech effect in short-term memory: Does the locus of suppression matter?. *Psychonomic bulletin & review, 12*(2), 374-379.

Troscianko, T. and Harris, J. (1988) Phase discrimination in compound chromatic gratings. *Vision Research, 28*(9), 1041–1049.

Tucker-Drob, E. M., & Salthouse, T. A. (2008). Adult age trends in the relations among cognitive abilities. *Psychology and aging*, *23*(2), 453.

Tucker-Drob, E. M., Brandmaier, A. M., & Lindenberger, U. (2019). Coupled cognitive changes in adulthood: A meta-analysis. *Psychological bulletin*, *145*(3), 273.

Turner, G. R., & Spreng, R. N. (2012). Executive functions and neurocognitive aging: dissociable patterns of brain activity. *Neurobiology of aging*, *33*(4), 826-e1.

Ueno, T., Mate, J., Allen, R. J., Hitch, G. J., & Baddeley, A. D. (2011). What goes through the gate? Exploring interference with visual feature binding. *Neuropsychologia*, *49*(6), 1597-1604.

Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language*, *54*(1), 68-80.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological review*, *114*(1), 104.

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, *62*(4), 392-406.

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive psychology*, *71*, 1-26.

Vallar, G., & Baddeley, A. D. (1984). Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 151-161.

Van den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*(22), 8780-8785.

Van Lamsweerde, A. E., & Beck, M. R. (2012). Attention shifts or volatile representations: What causes binding deficits in visual working memory?. *Visual Cognition*, *20*(7), 771-792.

Van Praag, H., Shubert, T., Zhao, C., & Gage, F. H. (2005). Exercise enhances learning and hippocampal neurogenesis in aged mice. *Journal of Neuroscience*, *25*(38), 8680-8685.

Van Ravenzwaaij, D., Cassey, P., & Brown, S. D. (2018). A simple introduction to Markov Chain Monte–Carlo sampling. *Psychonomic bulletin & review*, *25*(1), 143-154.

Veldsman, M., Mitchell, D. J., & Cusack, R. (2017). The neural basis of precise visual short-term memory for complex recognisable objects. *NeuroImage*, *159*, 131-145.

Vergauwe, E., Barrouillet, P., & Camos, V. (2010). Do mental processes share a domain-general resource?. *Psychological science*, *21*(3), 384-390.

Verhaeghen, P., &, A. (1996). On the mechanisms of plasticity in young and older adults after instruction in the method of loci: Evidence for an amplification model. *Psychology and aging*, *11*(1), 164.

Verhaeghen, P., Marcoen, A., & Goossens, L. (1992). Improving memory performance in the aged through mnemonic training: a meta-analytic study. *Psychology and aging*, *7*(2), 242.

Von Bastian, C. C., & Oberauer, K. (2014). Effects and mechanisms of working memory training: a review. *Psychological research*, *78*(6), 803-820.

Von Bastian, C. C., Langer, N., Jäncke, L., & Oberauer, K. (2013). Effects of working memory training in young and old adults. *Memory & cognition*, *41*(4), 611-624.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic bulletin & review, 14*(5), 779-804.

Wagenmakers, E. J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive psychology*, *60*(3), 158-189.

Wang, X., Logie, R. H., & Jarrold, C. (2016). Interpreting potential markers of storage and rehearsal: Implications for studies of verbal short-term memory and neuropsychological cases. *Memory & cognition, 44*(6), 910-921.

Wang, X., Logie, R. H., & Jarrold, C. (2016). Interpreting potential markers of storage and rehearsal: Implications for studies of verbal short-term memory and neuropsychological cases. *Memory & cognition*, *44*(6), 910-921.

Watson, J. B. (1924). The place of kinaesthetic, visceral and laryngeal organization in thinking. *Psychological Review*, *31*(5), 339.

Weicker, J., Villringer, A., & Thöne-Otto, A. (2016). Can impaired working memory functioning be improved by training? A meta-analysis with a special focus on brain injured patients. *Neuropsychology*, *30*(2), 190.

West, R., & Alain, C. (2000). Age-related decline in inhibitory control contributes to the increased Stroop effect observed in older adults. *Psychophysiology*, *37*(2), 179-189.

Wetzels, R., & Wagenmakers, E. J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic bulletin & review, 19*(6), 1057-1064.

Wetzels, R., & Wagenmakers, E. J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic bulletin & review, 19*(6), 1057-1064.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science, 6*(3), 291-298.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131*(1), 48.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*(1), 48.

Whorf, B. L. (1956). Language, Thought, and Reality: selected writings of Benjamin Lee Whorf (ed. JB Carroll), Cambridge, MA: The Massachusetts Institute of Technology Press; trad. It. 1970. *Linguaggio, pensiero e realtà*.

Wickens, C. D., & Weingartner, A. (1985). Process control monitoring: The effects of spatial and verbal ability and concurrent task demand. *Trends in ergonomics and human factors*, *2*, 25ą32.

Wickens, D. D. (1973). Characteristics of word encoding.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of vision*, *4*(12), 11-11.

Wilson, J. L., Scott, J. H., & Power, K. G. (1987). Developmental differences in the span of visual memory for pattern. *British Journal of Developmental Psychology, 5*(3), 249-255.

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: introspection can reduce the quality of preferences and decisions. *Journal of personality and social psychology*, *60*(2), 181.

Woodman, G. F., Vogel, E. K., & Luck, S. J. (2012). Flexibility in visual working memory: Accurate change detection in the face of irrelevant variations in position. *Visual cognition*, *20*(1), 1-28.

Wurm, L. H., Legge, G. E., Isenberg, L. M., & Luebker, A. (1993). Color improves object recognition in normal and low vision. *Journal of Experimental Psychology: Human perception and performance, 19*(4), 899.

Xu, Y. (2002). Encoding color and shape from different parts of an object in visual short-term memory. *Perception & psychophysics*, *64*(8), 1260-1280.

Zając-Lamparska, L., & Trempała, J. (2016). Effects of working memory and attentional control training and their transfer onto fluid intelligence in early and late adulthood.

Zeef, E. J., Sonke, C. J., Kok, A., Buiten, M. M., & Kenemans, J. L. (1996). Perceptual factors affecting age-related differences in focused attention: performance and psychophysiological analyses. *Psychophysiology*, *33*(5), 555-565.

Zelinski, E. M., & Lewis, K. L. (2003). Adult age differences in multiple cognitive functions: differentiation, dedifferentiation, or process-specific change?. *Psychology and aging*, *18*(4), 727.

Zeman, A. Z., Dewar, M., & Della Sala, S. (2015). Lives without imagery-Congenital aphantasia.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233.

Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological science*, *20*(4), 423-428.

Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological Science*, *22*(11), 1434-1441.

Zimmer, H.D., Mecklinger, A., & Lindenberger, U. (2006). Handbook of Binding and Memory. Oxford, UK: Oxford University Press

Zinke, K., Zeintl, M., Rose, N. S., Putzmann, J., Pydde, A., & Kliegel, M. (2014). Working memory training and transfer in older adults: effects of age, baseline performance, and training gains. *Developmental psychology*, *50*(1), 304.

Appendix A: Supplementary data and analyses for Chapter 2: Aging and feature-binding in Visual Working Memory: The role of verbal rehearsal

Table A.1

Results of Experiment 1 (Bayes Factor ANOVA for Color Memory).

	B	Error	$1/B$
Age Group × Trial Type × Label-Ability	0.078	± 4.31%	12.82
Label-Ability × Trial Type	4.07	± 4.04%	0.25
Age Group × Label-Ability	4.3×10^{-7}	± 4.52%	2.33×10^{6}
Age Group × Trial Type	3.22	± 4.43%	0.31
Label-Ability	4.02×10^{-99}	± 4.21%	2.5×10^{98}
Trial Type	7.20×10^{-22}	± 4.06%	1.4×10^{21}
Age Group	0.032	± 3.97%	31.59

Note. In this Bayes Factor model comparison approach B represents strength of evidence in favor of removing the main effect or interaction from the full model (including all other main effects and interactions). So, $B < 1$ indicates evidence that an omitted parameter was important, while $B > 1$ indicates evidence it was not. $1/B$ provides evidence for retaining the parameter in the model.

Table A.2

Results of Experiment 1 (Bayes Factor ANOVA for Shape Memory).

	B	Error	1/*B*
Age Group × Trial Type × Label-Ability	18.43	± 2.31%	0.054
Label-Ability × Trial Type	0.040	± 2.4%	24.72
Age Group × Label-Ability	8.90	± 3.02%	0.11
Age Group × Trial Type	1.12	± 4.68%	0.90
Label-Ability	2.8×10^{-48}	± 3.14%	3.6×10^{47}
Trial Type	1.8×10^{-31}	± 2.59%	5.5×10^{30}
Age Group	0.0033	± 3.57%	299.13

Note. 1/*B* provides evidence for retaining the parameter in the model when greater than 1.

Table A.3

Results of Experiment 2 (Bayes Factor ANOVA for Color Memory).

	B	Error	1/*B*
Age Group × AS × Trial Type × Label-Ability	1.98	± 4.55%	0.51
AS × Trial Type × Label-Ability	11.0	± 4.23%	0.091
Age Group × AS × Label-Ability	0.34	± 4.32%	2.96
Age Group × Trial Type × Label-Ability	23.71	± 4.91%	0.04
Age Group × AS × Trial Type	20.69	± 4.5%	0.05
AS × Label-Ability	2.40×10^{-7}	± 4.4%	4.2×10^{6}
Label-Ability × Trial Type	28.46	± 4.41%	0.04
Age Group × Label-Ability	1.2×10^{-11}	± 5%	8.53×10^{10}
AS × Trial Type	7.72	± 4.51%	0.13
Age Group × AS	1.63	± 4.84%	0.62
Age Group × Trial Type	25.06	± 4.48%	0.04
Label-Ability	3.11×10^{-151}	± 4.7%	3.2×10^{150}
AS	0.39	± 4.4%	2.55
Trial Type	3.7×10^{-30}	± 4.55%	2.7×10^{29}
Age Group	9.9×10^{-4}	± 4.87%	1.0×10^{3}

Note. 1/*B* provides evidence for retaining the parameter in the model when greater than 1. *AS* = Articulatory Suppression.

Table A.4

Results of Experiment 3 (Bayes Factor ANOVA for Shape Memory).

	B	Error	1/*B*
Age Group × AS × Trial Type × Label-Ability	3.97	± 4.33%	0.25
AS × Trial Type × Label-Ability	26.46	± 4.35%	0.038
Age Group × AS × Label-Ability	16.25	± 3.79%	0.062
Age Group × Trial Type × Label-Ability	5.43	± 3.64%	0.18
Age Group × AS × Trial Type	21.51	± 4.38%	0.05
AS × Label-Ability	30.70	± 3.81%	0.033
Label-Ability × Trial Type	0.80	± 3.83%	1.25
Age Group × Label-Ability	0.16	± 4.72%	6.32
AS × Trial Type	22.52	± 3.72%	0.044
Age Group × AS	28.90	± 3.8%	0.035
Age Group × Trial Type	0.0025	± 4.04%	399.69
Label-Ability	6.7×10^{-89}	± 4.04%	1.5×10^{88}
AS	1.85	± 4%	0.54
Trial Type	9.7×10^{-33}	± 3.95%	1.03×10^{32}
Age Group	0.60	± 3.98%	1.67

Note. 1/*B* provides evidence for retaining the parameter in the model when greater than 1. *AS* = Articulatory Suppression.

Table A.5

Pilot data collected to determine Label-Ability of Colors.

Color	Easy-to-Label Colors		
	Named (%)*	Within-participant consistency (%)**	Average letters
Black	100.0	100.0	5.0
White	100.0	100.0	5.0
Red	100.0	93.3	3.4
Purple	100.0	80.0	6.7
Orange	100.0	86.7	6.7
Pink	100.0	80.0	4.9
Green	100.0	93.3	6.0
Blue	100.0	86.7	5.3
Average	**100.0**	**90.0**	**5.4**
	Difficult-to-Label Colors		
	Named (%)*	Within-participant consistency (%)**	Average letters
Color 1	97.8	40.0	7.0
Color 2	95.6	33.3	8.2
Color 3	93.3	60.0	7.6
Color 5	77.8	20.0	6.9
Color 6	73.3	40.0	7.0
Color 7	82.2	40.0	7.8
Color 8	82.2	40.0	8.4
Color 9	95.6	40.0	7.6
Average	**87.2**	**39.2**	**7.6**

Note. * Named (as opposed to using the three-digit number) out of the total 45 viewing instances. ** Named consistently by number of participants out of the total 15.

Table A.6

Pilot data collected to determine Label-Ability of Shapes.

	Easy-to-Label Shapes		
Shape	Named (%)*	Within-participant consistency (%)**	Average letters
Rectangle	100.0	100.0	9.0
Square	100.0	100.0	6.0
Heart	100.0	100.0	5.0
Circle	100.0	93.3	6.2
Star	100.0	86.7	7.6
Triangle	100.0	93.3	10.7
Cross	100.0	86.7	5.8
Diamond	97.8	80.0	7.5
Average	**99.8**	**91.9**	**7.2**
	Difficult-to-Label Shapes		
	Named (%)*	Within-participant consistency (%)**	Average letters
Shape 1	62.2	33.3	14.0
Shape 2	57.8	53.3	9.7
Shape 3	86.7	80.0	11.1
Shape 4	40.0	33.3	13.8
Shape 5	31.1	26.7	11.8
Shape 6	40.0	40.0	15.2
Shape 7	74.3	46.7	9.0
Shape 8	42.2	26.7	12.1
Average	**54.3**	**42.5**	**12.1**

*Note.** Named (as opposed to using the three-digit number) out of the total 45 viewing instances. ** Named consistently by number of participants out of the total 15.

Table A.7

Color Luminance Values.

	Easy-to-Label Colors			
	R	G	B	Luminance
Purple	153	67	255	114.2
Red	225	10	10	74.3
Green	0	128	0	75.1
Blue	25	25	213	46.4
Pink	255	174	201	201.3
White	255	255	255	255.0
Black	0	0	0	0.0
Orange	255	128	0	151.4
Average				**114.7**
	Difficult to Label Colors			
Color 1	210	230	230	224.0
Color 2	153	0	76	54.4
Color 3	153	134	100	135.8
Color 4	190	160	140	166.7
Color 5	255	160	122	184.1
Color 6	154	84	82	104.7
Color 7	161	147	5	135.0
Color 8	11	100	108	74.3
Average				**134.9**

Note. We used this relative luminance formula; Y= (0.299 * R) + (0.587 * G) + (0.114 * B) (e.g. see Alala, B., Mwangi, W., & Okeyo, G., 2014).

Table A.8.

Experiment 2. Effects of the experimental factors on color memory performance, including only trials without suppression.

	B	Error	1/*B*
Age Group × Trial Type × Label-Ability	6.56	± 3.67%	0.15
Trial Type × Label-Ability	12.00	± 3.52%	0.083
Age Group × Label-Ability	5.8×10^{-12}	± 3.49%	1.7×10^{11}
Age Group × Trial Type	15.75	± 3.46%	0.063
Label-Ability	4.8×10^{-118}	± 3.70%	2.1×10^{117}
Trial Type	1.5×10^{-20}	± 3.44%	6.7×10^{19}
Age Group	6.2×10^{-3}	± 4.73%	161.5

Note. In this Bayes Factor model comparison approach *B* represents strength of evidence in favor of removing the main effect or interaction from the full model (including all other main effects and interactions). So, $B < 1$ indicates evidence that an omitted parameter was important, while $B > 1$ indicates evidence it was not. 1/*B* provides evidence for retaining the parameter in the model when greater than 1.

Table A.9

Experiment 2. Effects of the experimental factors on color memory performance, including only trials with suppression.

	B	Error	1/*B*
Age Group × Trial Type × Label-Ability	6.2	± 2.79%	0.16
Trial Type × Label-Ability	21.3	± 2.63%	0.047
Age Group × Label-Ability	0.13	± 4.13%	7.72
Age Group × Trial Type	25.9	± 3.85%	0.039
Label-Ability	7.9×10^{-43}	± 2.82%	1.3×10^{42}
Trial Type	3.6×10^{-10}	± 2.79%	2.8×10^{9}
Age Group	6.2×10^{-4}	± 3.57%	1605.6

Note. 1/*B* provides evidence for retaining the parameter in the model when greater than 1.

Table A.10

Experiment 3. Effects of the experimental factors on shape memory performance, including only data from younger adults.

	B	Error	1/*B*
AS × Trial Type × Label-Ability	6.02	± 3.94%	0.17
Label-Ability × AS	13.55	± 3.48%	0.074
Trial Type × Label-Ability	19.44	± 3.64%	0.051
AS × Trial Type	25.52	± 3.74%	0.039
Label-Ability	2.22×10^{-65}	± 3.05%	4.51×10^{64}
AS	2.25	± 4.26%	0.44
Trial Type	5.14×10^{-7}	± 3.7%	1.95×10^{6}

Note. 1/*B* provides evidence for retaining the parameter in the model when greater than 1. AS = articulatory suppression.

Table A.11

Experiment 3. Effects of the experimental factors on shape memory performance, including only data from older adults.

	B	Error	1/*B*
AS × Trial Type × Label-Ability	9.76	± 3.5%	0.10
Label-Ability × AS	23.68	± 3.68%	0.042
Trial Type × Label-Ability	0.23	± 3.4%	4.31
AS × Trial Type	12.05	± 3.53%	0.083
Label-Ability	3.03×10^{-28}	± 3.31%	3.30×10^{27}
AS	15.45	± 4.94%	0.065
Trial Type	3.79×10^{-27}	± 3.61%	2.64×10^{26}

Note. 1/*B* provides evidence for retaining the parameter in the model when greater than 1. AS = articulatory suppression.

Table A.12

Experiment 1. Means, SDs and Cohen's d for color memory performance, by experimental factors

		Mean	*SD*		Mean	*SD*	*d*
Age Group	Younger	.82	.27	Older	.68	.33	0.32
Trial Type	Single	.80	.28	Binding	.70	.32	0.23
Label-Ability	Easy-to-label	.85	.25	Difficult-to-label	.64	.34	0.48

Note. Cohen's *d* obtained using the overall pooled *SD* = 0.44 (i.e. the SD of all responses included in the given contrast).

Table A.13

Experiment 1. Means, SDs and Cohen's d for shape memory performance, by experimental factors

		Mean	*SD*		Mean	*SD*	*d*
Age Group	Younger	.76	.30	Older	.57	.35	0.40
Trial Type	Single	.73	.31	Binding	.60	.35	0.28
Label-Ability	Easy-to-label	.74	.31	Difficult-to-label	.59	.35	0.32

Note. Cohen's *d* obtained using the overall pooled *SD* = 0.47 (i.e. the SD of all responses included in the given contrast).

Table A.14

Experiment 2. Means, SDs and Cohen's d for color memory performance, by experimental factors

		Mean	*SD*		Mean	*SD*	*d*
Age Group	Younger	.83	.27	Older	.67	.33	0.37
Trial Type	Single	.79	.29	Binding	.70	.32	0.21
Label-Ability	Easy-to-name	.84	.26	Difficult-to-Name	.65	.38	0.44
Suppression	Without AS	.76	.30	With AS	.73	.31	0.069

Note. Cohen's *d* obtained using the overall pooled *SD* = 0.44 (i.e. the SD of all responses included in the given contrast).

Table A.15

Experiment 3. Means, SDs and Cohen's d for shape memory performance, by experimental factors

		Mean	*SD*		Mean	*SD*	*d*
Age Group	Younger	.73	.31	Older	.63	.34	0.21
Trial Type	Single	.73	.31	Binding	.63	.34	0.21
Label-Ability	Easy-to-name	.76	.30	Difficult-to-Name	.60	.35	0.34
Suppression	Without AS	.69	.33	With AS	.67	.33	0.043

Note. Cohen's *d* obtained using the overall pooled *SD* = 0.47 (i.e. the SD of all responses included in the given contrast).

Table A.16

Color accuracy (proportion correct), in-array errors (proportion of total responses), and shape memory, by age groups and experimental factors in Experiment 1.

			Color accuracy	In-array errors	Shape accuracy
Younger	Single	Difficult	.79	.29	-
		Easy	.92	.38	-
	Binding	Difficult	.69	.45	.61
		Easy	.86	.50	.78
Older	Single	Difficult	.58	.29	-
		Easy	.90	.40	-
	Binding	Difficult	.51	.31	.37
		Easy	.73	.52	.53

Note. Difficult = Difficult-to-label; Easy = Easy-to-label.

Table A.17

Shape accuracy (proportion correct), in-array errors (proportion of total responses), and color memory, by age groups and experimental factors in Experiment 1.

			Shape accuracy	In-array errors	Color accuracy
Younger	Single	Difficult	.72	.35	-
		Easy	.90	.23	-
	Binding	Difficult	.66	.37	.60
		Easy	.77	.41	.82
Older	Single	Difficult	.54	.36	-
		Easy	.75	.35	-
	Binding	Difficult	.42	.37	.43
		Easy	.56	.39	.66

Note. Difficult = Difficult-to-label; Easy = Easy-to-label.

Table A.18

Color accuracy (proportion correct), in-array errors (proportion of total responses), and shape memory, by age groups and experimental factors in Experiment 2.

With Suppression					
			Color	In-array errors	Shape
Younger	Single	Difficult	.80	.33	-
		Easy	.92	.48	-
	Binding	Difficult	.73	.40	.58
		Easy	.84	.63	.77
Older	Single	Difficult	.60	.31	-
		Easy	.77	.45	-
	Binding	Difficult	.50	.48	.39
		Easy	.71	.33	.56
Without Suppression					
Younger	Single	Difficult	.79	.25	-
		Easy	.95	.27	-
	Binding	Difficult	.69	.41	.62
		Easy	.86	.49	.79
Older	Single	Difficult	.57	.26	-
		Easy	.91	.37	-
	Binding	Difficult	.48	.33	.39
		Easy	.78	.40	.56

Note. Difficult = Difficult-to-label; Easy = Easy-to-label.

Table A.19

Shape accuracy (proportion correct), in-array errors (proportion of total responses), and color memory, by age groups and experimental factors in Experiment 3.

			Shape	In-array errors	Color
With Suppression					
Younger	Single	Difficult	.64	.37	-
		Easy	.85	.42	-
	Binding	Difficult	.61	.46	.60
		Easy	.76	.34	.80
Older	Single	Difficult	.60	.44	-
		Easy	.77	.44	-
	Binding	Difficult	.50	.42	.39
		Easy	.62	.42	.59
Without Suppression					
Younger	Single	Difficult	.68	.38	-
		Easy	.88	.51	-
	Binding	Difficult	.60	.43	.60
		Easy	.81	.35	.81
Older	Single	Difficult	.62	.44	-
		Easy	.81	.45	-
	Binding	Difficult	.52	.41	.38
		Easy	.61	.44	.67

Note. Difficult = Difficult-to-label; Easy = Easy-to-label.

Table A.20

Overview of previous studies on age-related feature-binding deficits.

Study	Stimulus Type	Set Size	Suppression	Binding Paradigm	Encoding	Retention	Age-related deficit
Brockmole, Parra, Della Sala & Logie, 2008							
Exp. 1	Common shapes and colors	2,4 or 6	Yes	Change Detection	753 ms	906 ms	No
Exp. 2	Difficult-to-name shapes and colors	4	No	Change Detection	1000 ms	906 ms	No
Exp. 3	Common shapes and colors	6	Yes	Reconstruction task with location cue	753 ms	1 or 5 sec	No
Read, Rogers, & Wilson, 2016							
Exp. 1.	Common colored squares in certain locations.	4	No	Change Detection	500 / 2000 ms (simultaneous / sequential)	900 ms	No
Exp. 2.	Common shapes in colors, at certain locations.	3	No	Change Detection	500 ms (simultaneous)	900 ms	No
Rhodes, Parra & Logie, 2015	Common shapes and colors	3	Yes	Change Detection	900 or 2500 ms	1000 ms	No

Parra, Abrahams, Logie & Della Sala, 2009b							
Exp. 1	Objects constructed using object shapes, defined by a figure and ground area, filled with non-basic colors.	3	No	Change Detection (say same or different)	2000 ms	900 ms	No
Exp. 2	Same	4	No	Change Detection (say same or different)	1000 ms	900 ms	No
Pertzov, Heider, Liang, & Husain, 2015	Complex fractals	1 or 3	No	Reconstruction; select target and drag to its original location: misbinding errors calculated by rate of "swap errors"	1 or 3 sec	1 or 4 seconds	No
Brown et al., 2017							
Exp. 1	Common shapes and colors	3	Yes	Change Detection (one probe)	900 ms or 1500 ms	1000 ms	No
Exp. 2	Common shapes and colors	3	Yes	Change Detection (one probe)	500 ms/item (sequential presentation of 3 items)	1000 ms	No
Exp. 3	Common shapes and colors	3	Yes	Change Detection (one probe)	900ms	1000 ms (including suffix distractor)	Yes
Brown & Brockmole, 2010							
Exp. 1		3	Yes		900 ms	1000 ms	No

Exp. 2	Common colors and shapes Common colors and shapes	3	Yes	Change detection (one probe) Change detection (one probe)	1500 (ms)	1000 ms	Yes
Kinjo, 2010							
Exp. 1	Cards with common colors, shape, and numbers.	1,2 or 3 cards	No	Change detection	2000 ms	6000 ms	Yes (for 'binding' 2 and 3 features)
Exp. 2	Same	Same	No	Change detection	Self-paced encoding (maximum 2 minutes)	6000 ms	No (older adults could bind three features)
Killin, Abrahams, Parra, Della Sala, 2018	Uncommon shapes (same as in our study) and non-primary colors	3	No	Change Detection	2000 ms	900 ms	No
Cowan, Naveh-Benjamin, Kilb & Saults, 2006							
Exp. 1a	Common colored-squares	4, 6, 8, or 10	No	Change Detection	250 ms	1000 ms	Yes
Exp. 2a	Same	Same	No	Same as above, but separate block for item and binding trials	250 ms	1000 ms	Yes (small)
Brockmole & Logie, 2013	Either common shapes or animals in common shapes.	1, 2,3 or 4	No	Reconstruction of objects by clicking on a color patch, then shape,	2 seconds per item (items shown simultaneously, if four items; total encoding = 8 seconds)	Immediate recall	Yes

				and then location.			
Kessels, Hobbel, & Postma, 2007	Common objects presented in grid	7	No	Reconstruction of all items *Objects only*, *Positions only*, *Binding;* associate objects with the correct location.	3 seconds per item (sequential presentation)	Immediate recall	Yes
Peich, Husain, & Bays, 2013	Colored bars (continuous)	1 or 3	No	Reconstruction; Misbinding. Recreate targets' (continuous) color and orientation, probed by location. Feature-binding errors = incorrect report of color or orientation belonging to an item at un-probed location.	200 ms or 2000 ms	1000 ms (including 100 ms pattern mask)	Yes
Mitchell, Johnson, Raye, Mather & D'Esposito, 2000a Exp. 1	Common object drawing in different common colors	3	No	Change Detection (comparing object only, location only, or combination).	1000 ms per item (sequential presentation)	8000 ms	Yes

Mitchell & Cusack, 2018.	Locations and colors (continuous report using color-wheel)	1-4	No	Reconstruction (continuous). Mis-binding given by probability of responding for the non-target distribution.	250 ms	900 ms	Yes (modest)

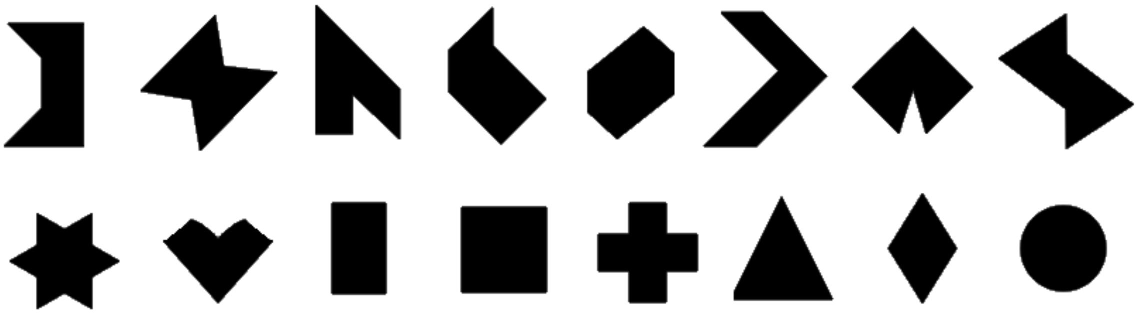


Figure A.1. The shapes used in all three experiments. Top; difficult-to-label to shapes, bottom; easy-to-label shapes.

Naming Data Analysis

After completing the memory task, participants filled out a questionnaire asking them to name the 16 shapes and 16 colors in the experiment (each shape and color appeared twice).[1] For example, participants saw a shape on the screen, with the text "Please name this shape (if you can't name it, please write 754)". The three-digit number differed for each stimulus, and was provided to ensure that not naming the item was not easier than typing in a name. The stimuli were the same in all three experiments, so we report the results for all three experiments combined.

We found 'decisive' evidence that participants opted for "I cannot name this color" more often for colors classified as difficult-to-label (8.2%) than for easy-to-label colors (1.2%) ($1/B = 9.7 \times 10^{32}$). Overall, there was no strong evidence that younger and older adults differed in how often they were unable to name colors (younger 6.3%; older, 3.1%; $1/B = 0.66$). However, we found some evidence for an interaction ($1/B = 5.18$) suggesting that younger adults were *less* able to provide names for the difficult-to-label colors (unable for 10.7%; cf. older adults 5.8%). For easy-to-label colors, younger adults were unable to label 1.9% and older adults 0.04%.

Next, for colors where a name was provided, we found strong evidence that difficult-to-label colors ($M = 7.1$; $SD = 4.5$) were described with more characters than easy-to-label colors ($M = 5.2$, $SD = 2.2$), $1/B = 1.4 \times 10^{149}$. There was no clear evidence that younger and older adults differed in how many characters they used overall (younger: $M = 5.8$; $SD = 3.6$; older: $M = 6.6$; $SD = 3.7$), ($1/B = 0.62$), nor for an interaction ($1/B = 0.27$). For the difficult-to-label colors younger adults used $M = 6.7$, $SD = 4.6$ characters, and older adults $M = 7.5$, $SD = 4.3$. For easy-to-label colors, younger adults used $M = 4.9$, $SD = 1.7$ characters, older adults $M = 5.6$, $SD = 2.5$.

For shapes, we also found 'decisive' evidence that participants opted for "I cannot name this shape" more often for shapes classified difficult-to-label (41.7%) than for easy-to-label shapes (2.1%; $1/B = 6.3 \times 10^{325}$). Overall, there was no evidence that younger and older adults differed in how often they were unable to name shapes (younger 23.0%; older 20.8%; $1/B = 0.14$), but there was some evidence for an interaction ($1/B = 9.96$). Younger adults were unable to provide names for 44.5% of the difficult-to-label shapes, the older adults for 39.1%. For the easy-to-label shapes, younger adults were unable to label 2.5%, older adults 1.7%. A closer look at the data revealed that 20 participants were unable to label the heart, which was intended to be easy-to-label.

Next, we found 'decisive' evidence that difficult-to-label shapes ($M = 10.3$; $SD = 9.7$) were described with more characters than easy-to-label shapes ($M = 6.8$, $SD = 3.0$), $1/B = 3.2 \times 10^{6}$. There was no clear evidence that younger and older adults differed in the overall number of characters they used (younger; $M = 7.7$, $SD = 5.6$; older; $M = 8.4$; $SD = 7.2$), ($1/B = 0.20$), nor for an interaction ($1/B = 0.068$. For the difficult-to-label shapes younger adults used $M = 9.8$, $SD = 8.6$ characters, and older adults $M = 10.8$, $SD = 10.5$. For easy-to-label shapes, younger adults used $M = 6.5$, $SD = 2.4$ characters, and older adults $M = 7.0$, $SD = 3.4$.

References

Alala, B., Mwangi, W., & Okeyo, G. (2014). Image representation using RGB color space. *International Journal of Innovative Research and Development*, *3*(8).

Appendix B: Supplementary data and analyses for Chapter 3: Change-Detection

Table B.1

Experiment 4. Effects of the experimental factors on memory performance, proportion correct.

	B	Error	1/*B*
Age Group × Trial Type × Label-Ability	298.25	± 2.58%	3.33×10^{-3}
Trial Type × Label-Ability	1.37×10^{-4}	± 2.61%	7.29×10^{3}
Age Group × Trial Type	129.70	± 3.08%	7.71×10^{-3}
Label-Ability × Age Group	34.13	± 4.09%	0.029
Trial Type	5.28×10^{-118}	± 2.79%	2.0×10^{117}
Label-Ability	2.50×10^{-59}	± 4.54%	4.0×10^{58}
Age Group	13.35	± 3.69%	0.075

Note. 1/*B* provides evidence for retaining the parameter in the model when greater than 1.

Table B.2

Experiment 4. Effects of the experimental factors on memory performance, d'.

	B	Error	1/*B*
Age Group × Trial Type × Label-Ability	8.36	± 3.60%	0.12
Trial Type × Label-Ability	9.02×10^{-10}	± 4.48%	1.11×10^{9}
Age Group × Trial Type	8.55	± 3.40%	0.12
Label-Ability × Age Group	3.83	± 2.95%	0.26
Trial Type	1.67×10^{-52}	± 2.97%	5.99×10^{51}
Label-Ability	5.43×10^{-37}	± 2.66%	1.84×10^{36}
Age Group	3.30	± 3.36%	0.30

Note. 1/*B* provides evidence for retaining the parameter in the model when greater than 1.

Table B.3

Experiment 4. Means, SDs and Cohen’s d for memory performance, by experimental factors

		Mean	SD		Mean	SD	d
Age Group	Younger	.78	.42	Older	.76	.42	0.031
Trial Type	Single (col./shape)	.86/.78	.35/.41	Binding	.67	.47	-0.47/-0.25
Label-Ability	Easy-to-name	.83	.38	Difficult-to-Name	.71	.45	-0.27

Appendix C: Supplementary data and analyses for Chapter 4: Cognitive Aging and Verbal Labeling in Continuous Visual Memory

Table C.1

BF_{10}s for the effects of the experimental factors for all four responses (between-item model).

Predictor	Parameter		
	Probability memory (P^M)	**Probability continuous (P^O)**	**Continuous imprecision (σ^O)**
All responses			
Age Group	1.81	.106	10.98
Verbalization	6.14×10^{18}	9.34	3.25×10^{5}
Age Group × Verbalization	7.28	5.05	.10064

Table C.2

BF_{10}s for the effects of subset analyses of the verbalization manipulation, for all four responses (between-item model).

Predictor	**Parameter**		
	Probability memory (P^M)	**Probability continuous (P^O)**	**Continuous imprecision (σ^O)**
Silence vs. Suppression			
Age Group	.45	.37	5.46
Verbalization	3.19×10^{21}	60.49	6.06×10^{4}
Age Group × Verbalization	5.61	.025	.24
Silence vs. Labelling			
Age Group	8.35	.084	7.04
Verbalization	1.19×10^{11}	.042	1.33
Age Group × Verbalization	.013	17.02	.021
Suppression vs. Labelling			
Age Group	1.36	.089	13.20
Verbalization	3.01×10^{38}	350.58	4.68×10^{8}
Age Group × Verbalization	118.18	.55	.26

Table C.3

BF_{10}s for the effects of the experimental factors in the within item models.

Predictor	Parameter		
	Probability memory (P^M)	**Probability continuous (P^O)**	**Continuous imprecision (σ^O)**
First response only			
Age Group	.28	.90	.21
Verbalization	1.05×10^6	.27	2.59
Age Group × Verbalization	.22	2.83	4.97
All responses			
Age Group	5.35	.29	1.77
Verbalization	3.49×10^{20}	.84	4.64
Age Group × Verbalization	5.65	5.61×10^6	2.49

Table C.4

BF_{10}s for the effects of the experimental factors in the consistency check analysis (excluding participants who did the labeling condition prior to the silence condition).

Predictor	**Parameter**		
	Probability memory (P^M)	**Probability continuous (P^O)**	**Continuous imprecision (σ^O)**
First response only			
Age Group	.094	.15	.69
Verbalization	4.20×10^5	.033	.95
Age Group × Verbalization	.64	.38	.62
All responses			
Age Group	.13	.104	2.50
Verbalization	2.40×10^{16}	.40	1.90×10^3
Age Group × Verbalization	10.07	77.15	.176

Table C.5

BF_{10}s for the effects of subset analyses of the verbalization manipulation for the consistency check analysis (excluding participants who did the labeling condition prior to the silence condition).

Predictor	**Parameter**		
	Probability memory (P^M)	**Probability continuous (P^O)**	**Continuous imprecision (σ^O)**
First response only			
Silence vs. Suppression			
Age Group	.13	.18	.98
Verbalization	2.61×10^7	.21	.44
Age Group × Verbalization	.014	2.02	.22
Silence vs. Labeling			
Age Group	.090	.11	.64
Verbalization	1.58×10^2	.11	.70
Age Group × Verbalization	1.20	.044	.83
Suppression vs. Labeling			
Age Group	.12	.27	.66
Verbalization	1.24×10^{11}	.18	2.48
Age Group × Verbalization	5.62	.13	1.18

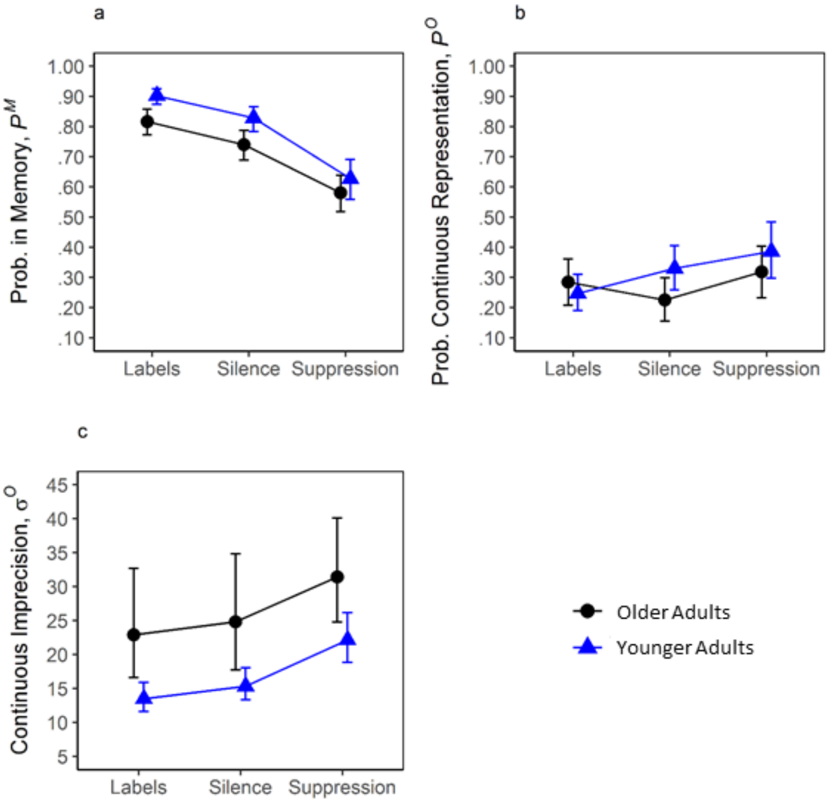


Figure C.1. Memory for all items (between-item model). Panel a. Group-level probability of having the probed item in memory. Panel b. The group-level probability that memory representation is continuous. Panel c. The imprecision of the group-level continuous memory representation.

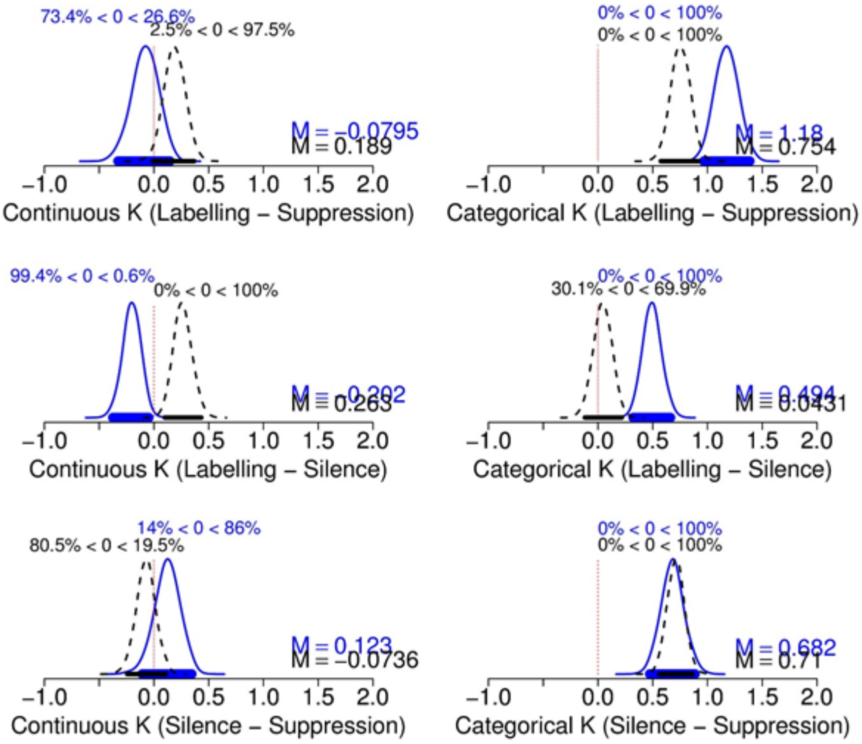


Figure C.2. Memory for all items (between-item model). Posterior differences in continuous and categorical K for specified comparisons. Mean values (M) larger than 0 for condition (A – B) indicates larger estimates in condition A than B. Each panel presents the percentages of the curves that are above and below 0 (null effect), the means (M), and the 95% credible intervals of the means (bars underneath each curve). Older adults in dotted black, younger in solid blue.

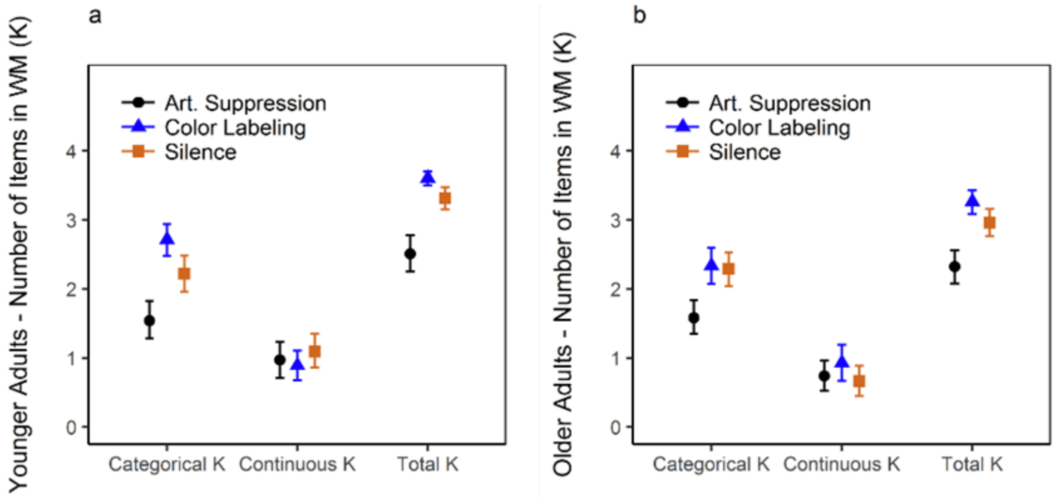


Figure. C.3. Memory for all items (between-item model). Categorical, Continuous and Total K, by age group and verbalization condition.

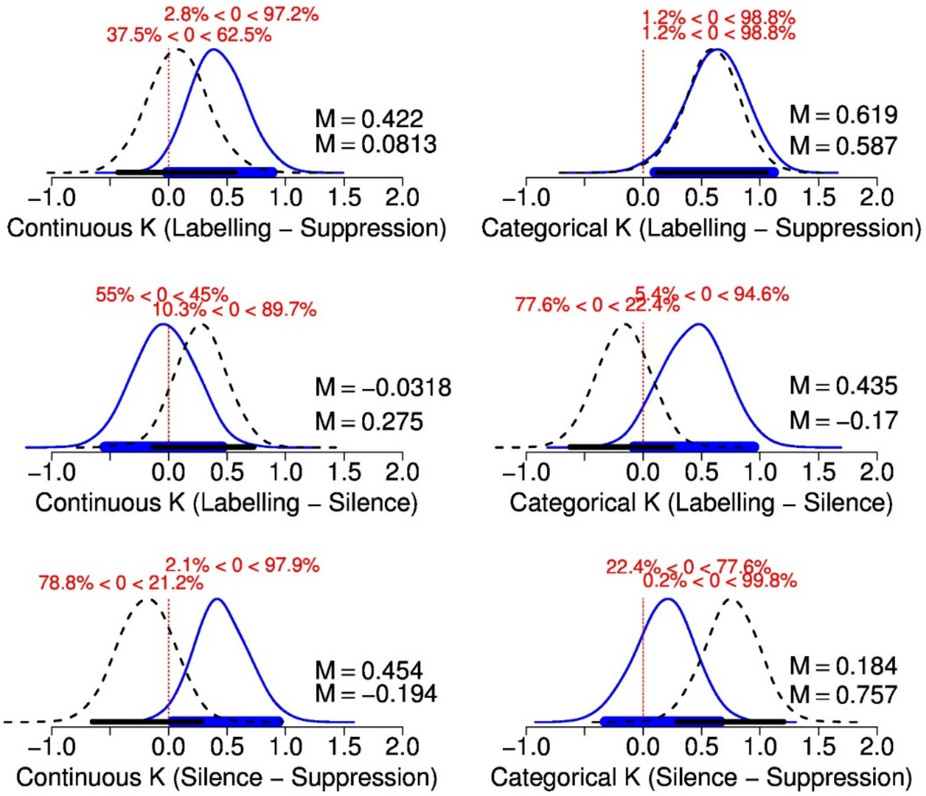


Figure C.4. Memory for the first item only (between-item model), in the consistency check analysis (excluding participants who did the labeling condition prior to the silence condition). Posterior differences in continuous and categorical K for specified comparisons. Mean values (M) larger than 0 for condition (A – B) indicates larger estimates in condition A than B. Each panel presents the percentages of the curves that are above and below 0 (null effect), the means (M), and the 95% credible intervals of the means (bars underneath each curve). Older adults in dotted black, younger in solid blue.

Function used to generate color values

```
def LAB2RGB(L, a, b, radius): # draws a circle in CIELab colour space with specified centre (L, a, b) and raduis then converts to RGB, trimming nonsense values

    colours = []
    # create CIELab colours
    for ang in range(1, 361):
        theta = ang * pi / 180.000 # converts angle to radian
        A = a + radius*numpy.cos(theta)
        B = b + radius*numpy.sin(theta)

        # Lab to XYZ
        var_Y = (L + 16) / 115.000
        var_X = A / 500.000 + var_Y
        var_Z = var_Y - B / 200.000

        # filter X, Y, Z with threshold 0.008856
        if  var_Y**3 > 0.008856: var_Y = var_Y**3
        else: var_Y = ( var_Y - 16 / 116.000 ) / 7.787
        if var_X**3 > 0.008856: var_X = var_X**3
        else: var_X = ( var_X - 16 / 116.000 ) / 7.787
        if var_Z**3 > 0.008856: var_Z = var_Z**3
        else: var_Z = ( var_Z - 16 / 116.000 ) / 7.787

        # reference points
```

```
ref_X =  95.047
ref_Y = 100.000
ref_Z = 108.883

X = ref_X * var_X / 100.000
Y = ref_Y * var_Y / 100.000
Z = ref_Z * var_Z / 100.000

# covert XYZ to RGB
var_R = X * 3.2406 + Y * -1.5372 + Z * -0.4986
var_G = X * -0.9689 + Y * 1.8758 + Z * 0.0415
var_B = X * 0.0557 + Y * -0.2040 + Z * 1.0570

# gamma correction to IEC 61966-2-1 standard
if var_R > 0.0031308: var_R = 1.055 * ( var_R ** ( 1 / 2.400 ) ) - 0.055
else: var_R = 12.92 * var_R
if var_G > 0.0031308: var_G = 1.055 * ( var_G ** ( 1 / 2.400 ) ) - 0.055
else: var_G = 12.92 * var_G
if var_B > 0.0031308: var_B = 1.055 * ( var_B ** ( 1 / 2.400 ) ) - 0.055
else: var_B = 12.92 * var_B

# trim
if (var_R*255) > 255: R = 255
elif (var_R*255) < 0: R = 0
else: R = round(var_R*255)

if (var_G*255) > 255: G = 255
```

```
        elif (var_G*255) < 0: G = 0
        else: G = round(var_G*255)

        if (var_B*255) > 255: B = 255
        elif (var_B*255) < 0: B = 0
        else: B = round(var_B*255)

        colours.append([R,G,B])
    return numpy.array(colours)



# To call this; specify L, a, b and radius.
colours = LAB2RGB(L = 50, a = 20, b = 20, radius = 60)
```

Output from the All-Item analyses

Here, we present and discuss the output of the model including memory for all four items. For all four items, the between-item model also had a smaller WAIC than the within-item model (Δ = -545.6). Therefore, we present and discuss parameter estimates from this model below, see Table S3 for output from the within-item models.

Memory performance: Parameter estimates

For all items, there was 'anecdotal' evidence for a main effect of age on the probability of remembering colors (P^M), as well as a 'decisive' verbalization effect ($BF_{10} = 6.14 \times 10^{18}$) and an interaction between the two ($BF_{10} = 7.28$), suggesting that verbalization affected performance differently in the two age groups. Specifically, the younger adults were comparatively more impaired by suppression (see Fig. S3; see also suppression vs. labelling analysis, Age Group × Verbalization $BF_{10} = 118.18$, Table S2). There was no main effect of age on the probability of continuous representations, but a 'substantial' main effect of verbalization ($BF_{10} = 9.34$), and evidence for an interaction with age ($BF_{10} = 5.05$). For precision, we observed 'strong' evidence for an age effect ($BF_{10} = 10.98$) and a 'decisive' verbalization effect ($BF_{10} = 3.25 \times 10^5$), but no Age Group × Verbalization interaction, suggesting that verbalization instructions had similar effects on continuous memory precision in both age groups.

Preventing Labelling (Silence vs. Suppression)

When modelling performance for all four items in the arrays together, suppression also reduced the probability of remembering (P^M; $BF_{10} = 3.19 \times 10^{21}$). Also, there was some evidence for an interaction with age, such that the younger adults were comparatively more impaired by suppression ($BF_{10} = 5.61$).

Suppression increased the probability of continuous representations (P^O; BF_{10} = 60.49) in both age groups (see Fig. S1). Hence, when responding to all items, suppression reduced categorical responding in both age groups, not just for the older adults (in contrast to the traditional, first-presented items analysis above). Categorical memory representations in older adults' reduced credibly under suppression (*M* = - 0.71 items), and in younger adults (*M* = - 0.68 items), but there were no credible differences in continuous K in either age group (see Fig. S2). There was also ‘decisive’ evidence that suppression resulted in reduced precision ($BF_{10} = 6.06 \times 10^4$), and this was not observed differently in the age groups.

Enforcing Labelling (Silence vs. Labelling)

Overt labelling also improved memory (P^M) for both age groups for all items ($BF_{10} = 1.19 \times 10^{11}$). There was no evidence of a main effect of labelling on the probability of continuous responding (P^O), but substantial evidence for an interaction between labelling and age group (BF_{10} = 17.02), indicating that enforced labelling made young adults less likely to have continuous representations but older adults slightly more, compared to silence (see Fig. S1). Compared to spontaneous performance in silence, younger adults' categorical memory representations increased credibly when instructed to label (*M* = 0.49 items), while their continuous K credibly decreased (*M* = - 0.20 items). In contrast, older adults' categorical memory capacity under instructed labelling did not differ credibly from their performance in silence (see Fig. S2), while their continuous capacity increased credibly (*M* = 0.26). There was ‘anecdotal' evidence that instructed labelling increased precision (BF_{10} = 1.33) to equal extents in the age groups.

The Labelling Benefit (Labelling vs. Suppression)

All items. Including the subsequent three items in the analysis produced a different pattern. While overt labelling improved memory (P^M) in both age groups compared to suppression ($BF_{10} = 3.01 \times 10^{38}$), it also increased the probability of categorical responding ($1 - P^O$, $BF_{10} = 350.58$), seemingly not to different extents in the age groups (Age Group × Verbalization; $BF_{10} = 0.55$). Similar to first-presented items, labelling led to credible increases in categorical K compared to suppression for participants of both age groups (younger *M* = 1.18, older *M* = 0.75 items). Surprisingly, the older adults’ continuous K benefitted as well (*M* = 0.19 items), but the younger adults’ did not (see Fig S2). This suggests that labelling benefitted memory performance differently as participants responded to subsequent items. We observed decisive evidence that suppression decreased precision (σ^O) compared to overt labelling ($BF_{10} = 4.68 \times 10^8$), to similar extents in both age groups (Age Group × Verbalization; $BF_{10} = .26$).

The Labelling Benefit despite Interference and Delay

In the paper, we focused on the effect of labelling in the ‘traditional' analysis: memory for the first-presented item only. This is standard practice because memory for subsequent items is "tainted" by interference from previous responding. Arguably, however, in real-world contexts maintaining representations despite interference is common, and interchangeable use of visual/verbal representations is likely involved in this process. For instance, some suggest that perceptual memories are lost via 'sudden death' rather than gradual decay (Zhang & Luck, 2009), whereas memories with verbal labels are more robust (Donkin, Nosofsky, Gold, & Shiffrin, 2015). For all four items, labelling (compared to suppression) increased categorical representations but not continuous representations in the younger adults. This differed from Souza and Skóra’s (2017) results; they found categorical and continuous benefits of labelling for one item and all items alike. However, our longer presentation time (930 compared to 250 ms; used to ensure older adults would be able to perceive and label colors) may have induced these differences by making our task easier. Visual traces were likely much stronger, even under suppression and response interference, for our participants.

Surprisingly, for all four items, in older adults both categorical and continuous representations increased with overt labelling compared to suppression. Compared to silence, labelling only boosted continuous representations. This supports the idea that older adults maximized categorical representations by sub-vocally rehearsing in silence but that saying labels out loud – thus producing auditory traces – protected their visual (i.e. continuous) representations from the response interference associated with responding to all four items, or the delay imposed as they provided responses one by one. Indeed, the memory decay with time from the initial presentation to giving the final response was likely exacerbated in older adults due to slower processing speed (Brown, Brockmole, Gow, & Deary, 2012). Also, the appearance of the color-wheel as a response device likely interferes with the original visual trace (Donkin, et al., 2015). As participants look around the wheel and provide their responses, this interference might affect older adults differently if they are more susceptible to distraction (supported by, e.g., Gazzaley, Cooney, Rissman, & D'Esposito, 2005), and overt labelling might have helped reduce the decline of perceptual traces despite such interference. These results suggested that labels can play different roles when providing multiple memory responses in participants of different age groups.

References

Brown, L. A., Brockmole, J. R., Gow, A. J., & Deary, I. J. (2012). Processing speed and visuospatial executive function predict visual working memory ability in older adults. *Experimental aging research*, *38*(1), 1-19.

Donkin, C., Nosofsky, R., Gold, J., & Shiffrin, R. (2015). Verbal labeling, gradual decay, and sudden death in visual short-term memory. *Psychonomic Bulletin & Review*, *22*(1), 170-178.

Gazzaley, A., Cooney, J. W., Rissman, J., & D'esposito, M. (2005). Top-down suppression deficit underlies working memory impairment in normal aging. *Nature neuroscience*, *8*(10), 1298.

Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. *Cognition*, *166*, 277-297.

Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological science*, *20*(4), 423-428.

Appendix D: Supplementary data and analyses for Chapter 5: Strategic Mediation in Working Memory Training in Younger and Older Adults

Table D.1

Participant demographics by age and strategy group (all participants).

	Younger Adults			Older Adults		
	Control	Strategy	*p*	Control	Strategy	*p*
N	29	31		30	30	
Age	23.0 (3.96)	22.0 (2.98)	.29	70.3 (5.69)	68.3 (5.11)	.144
Gender F/M	21/8	23/8	1.0	20/10	17/13	.60
Education	16.2 (2.81)	15.7 (2.46)	.55	15.5 (3.43)	16.2 (3.22)	.409
Pre-training N-back composite	0.59 (4.94)	-0.23 (5.97)	.577	1.04 (4.88)	-1.33 (5.54)	.096

Note. Values in parentheses are standard deviations. *P*-values were calculated from *t*-tests for continuous variables and χ^2 test for gender. The N-back composite score were the summed values of the z-transformations of the average and maximum level accuracy in the adaptive digit N-back task, and d-prime values and RTs for correct responses in the letter and colour N-back tasks.

Table D.2

Mean values (standard deviations) for the pre-post measures per group at pre- and post-test, for younger adults (all participants).

	Control Group (N = 29)				**Strategy (N = 31)**			
	Pre	**Post**	***r***	***d***	**Pre**	**Post**	***r***	***d***
Trained Digit N-back								
Maximum level	4.28 (1.71)	5.52 (2.16)	0.66	0.76	3.83 (1.49)	6.23 (1.98)	0.47	1.31
Average level	2.72 (0.91)	3.41 (1.02)	0.70	0.91	2.52 (0.94)	3.84 (1.09)	0.53	1.33
Task-specific near transfer								
Letter 2-back (d-prime)	2.25 (0.94)	2.48 (0.96)	0.71	0.32	2.12 (1.07)	2.73 (1.09)	0.47	0.55
Letter 3-back (d-prime)	1.19 (0.76)	2.00 (1.15)	0.55	0.83	1.09 (1.03)	2.39 (1.15)	0.4	1.08
Colour 2-back (d-prime)	2.03 (0.78)	2.54 (0.93)	0.39	0.54	1.92 (1.12)	2.73 (1.13)	0.58	0.79
Colour 3-back (d-prime)	0.90 (0.82)	1.69 (1.22)	0.52	0.75	1.02 (0.57)	2.21 (1.18)	0.43	1.11
Letter 2-back RT (ms)	803.85 (108.04)	686.08 (127.44)	0.51	-1.00	779.64 (141.24)	647.69 (149.98)	0.47	-0.88
Letter 3-back RT (ms)	802.58 (120.04)	676.99 (100.28)	0.36	-1.00	774.63 (213.40)	633.67 (135.03)	0.43	-0.71
Colour 2-back RT (ms)	811.47 (119.21)	696.87 (115.34)	0.26	-0.80	819.43 (122.33)	662.84 (148.07)	0.37	-1.02
Colour 3-back RT (ms)	857.86 (129.50)	721.56 (106.42)	0.24	-0.93	830.50 (227.99)	674.33 (152.25)	0.35	-0.69
Task-general near transfer								
Selective updating of digits	32.38 (8.14)	33.00 (7.08)	0.78	0.12	33.42 (9.76)	35.90 (8.07)	0.65	0.33
Digit span (correct items)	34.52 (10.00)	34.10 (8.83)	0.73	-0.06	33.68 (9.35)	35.52 (8.97)	0.38	0.18
Digit span (maximum	6.79 (2.06)	7.28 (1.53)	0.7	0.33	7.16 (1.93)	7.42 (2.05)	0.24	0.11
Running memory	25.31 (4.49)	26.28 (5.32)	0.49	0.19	23.94 (5.20)	26.32 (5.93)	0.57	0.46

Note. Exclusions to specific analyses apply.

Table D.3

Mean values (standard deviations) for the pre-post measures per group at pre- and post-test, for older adults (all participants).

	Control Group (N = 30)				**Strategy (N = 30)**			
	Pre	**Post**	***r***	***d***	**Pre**	**Post**	***r***	***d***
Trained Digit N-back								
Maximum level	3.10 (0.92)	3.83 (1.29)	0.51	0.64	2.77 (0.68)	3.63 (1.22)	0.39	0.76
Average level	1.94 (0.55)	2.55 (0.76)	0.58	0.98	1.90 (0.41)	2.40 (0.74)	0.55	0.81
Task-specific near transfer								
Letter 2-back (d-prime)	1.85 (0.79)	2.31 (0.86)	0.63	0.64	1.65 (0.89)	1.98 (0.86)	0.38	0.34
Letter 3-back (d-prime)	0.76 (0.48)	1.28 (0.88)	0.45	0.65	0.70 (0.59)	1.28 (0.88)	0.48	0.74
Colour 2-back (d-prime)	1.81 (0.75)	2.09 (0.83)	0.53	0.36	1.36 (0.81)	1.92 (0.96)	0.43	0.58
Colour 3-back (d-prime)	0.77 (0.58)	0.94 (0.76)	0.16	0.19	0.51 (0.50)	0.86 (0.76)	0.42	0.48
Letter 2-back RT (ms)	1017.30 (165.77)	869.92 (178.94)	0.82	-1.42	1000.64 (179.96)	918.08 (151.49)	0.51	-0.5
Letter 3-back RT (ms)	1002.24 (174.61)	936.59 (167.59)	0.70	-0.49	996.27 (170.11)	932.69 (161.45)	0.67	-0.47
Colour 2-back RT (ms)	1013.51 (166.92)	909.24 (160.31)	0.64	-0.75	1040.48 (161.13)	948.80 (145.40)	0.8	-0.94
Colour 3-back RT (ms)	1071.56 (160.84)	959.52 (199.16)	0.54	-0.64	1023.64 (177.20)	995.01 (144.42)	0.45	-0.17
Task-general near transfer								
Selective updating of digits	24.63 (11.48)	30.43 (11.33)	0.75	0.72	24.50 (13.14)	26.33 (13.16)	0.83	0.24
Digit span (correct items)	33.23 (8.24)	34.37 (7.91)	0.64	0.17	32.17 (8.73)	33.63 (8.19)	0.63	0.2
Digit span (maximum span)	6.93 (1.36)	7.23 (1.36)	0.18	0.17	6.67 (2.04)	6.83 (1.72)	0.54	0.09
Running memory	24.33 (4.33)	23.80 (5.29)	0.51	-0.11	22.67 (5.14)	23.83 (4.61)	0.59	0.26

Note. Exclusions to specific analyses apply.

Table D.4

ANCOVA results for the trained task and for the transfer measures. (all participants).

		F	***p***	***d / ηp²***
Trained Digit N-back				
Maximum level	Strategy	4.17	.065	0.37
	Age	**20.67**	**<.001**	**0.87**
	Interaction	3.58	.073	0.03
Average level	Strategy	2.85	.094	0.30
	Age	**17.86**	**<.001**	**0.81**
	Interaction	**6.74**	**.021**	**0.06**
Task-specific near transfer				
Letter 2-back (d-prime)	Strategy	0.13	.785	0.07
	Age	2.23	.268	0.28
	Interaction	3.4	.194	0.03
Letter 3-back (d-prime)	Strategy	2.15	.268	0.27
	Age	**14.38**	**.001**	**0.71**
	Interaction	1.47	.365	0.01
Colour 2-back (d-prime)	Strategy	1.00	.466	0.19
	Age	**6.54**	**.041**	**0.49**
	Interaction	0.29	.679	<.001
Colour 3-back (d-prime)	Strategy	2.27	.268	0.29
	Age	**22.42**	**<.001**	**0.90**
	Interaction	0.96	.466	0.01
Letter 2-back (RT in ms)	Strategy	0.58	.597	0.14
	Age	**11.61**	**.004**	**-0.68**
	Interaction	3.28	.194	0.03
Letter 3-back (RT in ms)	Strategy	0.52	.598	-0.13
	Age	**49.98**	**<.001**	**-1.39**
	Interaction	0.45	.607	<.001
Colour 2-back (RT in ms)	Strategy	0.09	.785	-0.06
	Age	**21.83**	**<.001**	**-0.96**
	Interaction	1.94	.285	0.02
Colour 3-back (RT in ms)	Strategy	0.07	.785	0.05
	Age	**45.07**	**<.001**	**-1.34**

	Interaction	2.82	.23	0.03
Task-general near transfer				
Selective updating of digits	Strategy	0.59	.799	-0.14
	Age	<.001	.976	0.01
	Interaction	6.51	.145	0.05
Digit span (correct items)	Strategy	0.47	.799	0.13
	Age	<.001	.976	0.01
	Interaction	0.64	.799	0.01
Digit span (maximum span)	Strategy	0.28	.799	-0.10
	Age	0.77	.799	0.16
	Interaction	0.28	.799	<.001
Running memory	Strategy	1.29	.799	0.21
	Age	4.82	.181	0.40
	Interaction	0.01	.976	<.001

Note. To correct for multiple comparisons, Benjamini-Hochberg adjusted p-values were applied for group comparisons on each pre-post outcome measure.

Table D.5

Classification scheme for strategy types based on participants' self-reports at post-test, used by independent raters.

Scoring	**Example**
Rehearsal (1)	"I repeated the digits silently in my mind" "I repeated a list of letters in my mind" "Repeating out loud the letters"
Grouping (2)	"I created groups of 3 digits" "I grouped the letters in pairs" "When the sequence was long enough I remembered the items in groups of 4" "I remembered the digits in groups"
Updating (3)	"I created a group of digits in my mind and dropped the last digit when a new digit appeared" "Replaced the first letter with the latest letter that appeared on the screen." "I tried to replace each color one at a time as a new color came up"
Grouping and comparison (4)	"I split the digits into different series, and compared those to each other" "Held each sequence of digits in mind, removed the first and added a new digit to the end. After that, I checked if it was the same as the one just dropped."
Semantics (5)	"I created words from the letters (e.g., C-R-S = Corn – Rose – Sand)" "I converted the digits to melodies, e.g., 1356 = DO-MI-SO-LA"
Phonology (6)	"I made up lists based on initial parts of the digit names such as se-fi-ni (7-5-9)" "I tried to make syllables out of the letters"
Imagery (7)	"I tried to associate each digit with some image in my mind" "I tried to visualize the letters as snakes"
Visualization (8)	"I visualized the numbers" "I tried to visualize the letter sequence in my mind"
Familiarity (9)	"I chose the letters that felt most familiar" "I recalled the digits that were familiar"
Guessing (10)	"I just used intuition" "I started somewhere in the middle of the sequence, and did not memorize the first digits in the sequence at all"
Other strategies (11)	"I made up a song based on the letters" "Yes" "I tried to keep all the digits in my mind" "This task was difficult"
No strategy use (12)	"I pressed the N-key if the current white box was the same as the white box presented before it, or the M-key if it was not the same"

Table D.6.

The frequency of occurrence of reported strategy types by the control participants

	N-back Digit		*N-back Letter*		*N-back Colour*	
	Younger(%)	Older(%)	Younger(%)	Older(%)	Younger(%)	Older(%)
Rehearsal	3.45	31.03	6.90	27.59	3.45	24.14
Grouping	13.79	10.34	10.34	13.79	10.34	3.45
Updating	0.00	0.00	0.00	0.00	3.45	0.00
Grouping and comparison	10.34	0.00	13.79	0.00	3.45	0.00
Semantics	0.00	0.00	3.45	0.00	0.00	0.00
Phonology	0.00	0.00	0.00	0.00	0.00	0.00
Imagery	0.00	0.00	0.00	0.00	0.00	0.00
Visualization	3.45	3.45	0.00	0.00	0.00	3.45
Familiarity	0.00	0.00	0.00	0.00	0.00	0.00
Guessing	0.00	0.00	0.00	0.00	0.00	0.00
Other strategies	27.59	17.24	13.79	6.90	17.24	13.79
No strategy use	41.38	37.93	51.72	51.72	62.07	55.17

Table D.7.

Level of Detail rating instructions for strategies. The examples were created as an aid for the independent raters.

CODING CRITERIA: LEVEL OF DETAIL

Scoring	Description	Example responses
No strategy use (0)	*Ticks the "No" –alternative and has no comments in the text box.* *OBS! If the participant has commented anything on his/her strategy use of in the comment box, yet ticked "No", make your evaluation based on the comment. In other words, if the comment shows that the participant has used some kind of strategy, he/she can get points (depending on the content of the comment).*	
Yes-alternative + non-specific description of strategy (1 point)	*There is not a clear strategy in the comment. The comment may be short or long, but the description of the strategy is non-specific. The participant may also repeat the task instructions provided at the beginning of each pre- and posttest task (i.e., there is nothing new in the comment except for the instructions that already has been provided).* - **Examples:** • Only writes that he/she tried to remember the stimuli in his/her head/mind/read aloud but does not describe how. • Writes that he/she has acted on intuition • Reports that he/she did not use any consistent strategy • Reports that he/she used different strategies depending on the situation.	• "I memorized the digits in my mind" • "I used intuition" • "I did not use any strategy consistently" • "I used different strategies depending on the situation" • "I tried to keep the items in my head" • "I tried to remember the digits in my head" • "I read the colors aloud" • "I said the numbers aloud"
Yes-alternative + general strategy description (2 points)	*The comment clearly indicates that the participant has used a strategy (i.e., he/she writes the way in which the strategy was used, not only that "I tried to remember / keep the items in my head / read aloud") but the description of the strategy remains at a general level.* - **Examples** • Writes that he/she repeated the stimuli series in his/her mind. • Writes that he/she read the items in series in his/her mind. • Writes that he/she used e.g., rhythm, grouping, fingers to memorize the items.	• "I repeated the letters in my mind in two or three groups" • "I memorized the digits in pairs, such as 52-48" • "Grouping the digits in separate series"

Yes-alternative + two or more details of the used strategy in addition to the general strategy description (3 points)	*The comment clearly indicates that the participant have used a strategy, and describes related details of the strategy.* - **Example:** • Writes that he/she split the stimuli series into separate chunks and compared them to each other.	• "I tried to keep 2 numbers back in my mind while saying 1 number back out loud to try to keep in my mind the 2 differing numbers.", • "Memorized the four letters as one sequence and replaced each letter as a new one came up" "I split the digits into different series, and compared those to each other"

Motivation and Expectations (also including non-compliant participants)

Below, we report analyses conducted including all participants (i.e., also including those strategy-group participants who reported not complying with the strategy instruction).

Alertness, motivation, and expectations. Participants' expected improvement during training assessed right before they started the training, to assess if one group receiving a strategy would make a difference. There was no difference in expectation between strategy and control groups in younger (t(55.09) = 0.08, p = .94) or older adults t(56.38) = 0.56, p = .58). We also compared expected improvement on each task between pre-test to post-test, after the strategy group had received the strategy instruction. There were no significant differences on any of the tasks, in either age group (all p-values ≥ .13). Similarly, there we no differences in self-reported alertness or motivation in either age groups just after they completed the training session (all p-values ≥ .20).

Training Improvement. Two independent samples t-tests found no overall effect of Strategy Instruction in the younger (t(58) = 1.08, p = .28, d = 0.28; strategy group M = 4.37 digits, control group M = 3.93 digits), or in the older adults (t(58) = -0.39, p = .69, d = 0.10, strategy group M = 2.48 digits, control group: M = 2.58 digits) on average performance across all 20 trials.

Pre-test Composite Scores: Compliant vs Non-Compliant Strategy group participants

Two exploratory independent samples t-tests found no differences in pre-test composite score between compliant and non-compliant participants in the younger ($t(26) = -0.002$, $p = .999$, $d = -0.001$), or in the older adults ($t(25) = -0.161$, $p = .873$, $d = -0.064$).