Memory & Cognition (2020) 48:1196–1213
https://doi.org/10.3758/s13421-020-01043-3

Cognitive aging and verbal labeling in continuous visual memory

Alicia Forsberg [1,2] · Wendy Johnson [1] · Robert H. Logie [1]

Published online: 29 May 2020
© The Author(s) 2020

Abstract

The decline of working memory (WM) is a common feature of general cognitive decline, and visual and verbal WM capacity appear to decline at different rates with age. Visual material may be remembered via verbal codes or visual traces, or both. Souza and Skóra, *Cognition*, *166*, 277–297 (2017) found that labeling boosted memory in younger adults by activating categorical visual long-term memory (LTM) knowledge. Here, we replicated this and tested whether it held in healthy older adults. We compared performance in silence, under instructed overt labeling (participants were asked to say color names out loud), and articulatory suppression (repeating irrelevant syllables to prevent labeling) in the delayed estimation paradigm. Overt labeling improved memory performance in both age groups. However, comparing the effect of overt labeling and suppression on the number of coarse, categorical representations in the two age groups suggested that older adults used verbal labels subvocally more than younger adults, when performing the task in silence. 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, continuous representations, but instead boosted memory via additional verbal (categorical) memory traces. These results challenged the assumption that visual memory paradigms measure the same cognitive ability in younger and older adults, and highlighted the importance of controlling differences in age-related strategic preferences in visual memory tasks.

Keywords Visual working memory · Cognitive aging · Delayed estimation · Verbal labeling · Memory precision

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; Gazzaley, Cooney, Rissman, & D'Esposito, 2005; Johnson, Logie, & Brockmole, 2010; Jost, Bryck, Vogel, & Mayr; 2010; Park et al., 2002; 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 multicomponent 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 (Gruber, 2001; Jonides et al., 1996; but see also D'Esposito & Postle, 2015). Furthermore, visual WM capacity appears to decline at a faster rate with age than verbal WM capacity (Johnson et al., 2010), supporting the notion that they do not rely on the same mechanisms. In this paper, we investigated 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.

Electronic supplementary material The online version of this article (https://doi.org/10.3758/s13421-020-01043-3) contains supplementary material, which is available to authorized users.

✉ Alicia Forsberg
aliciaforsberg@missouri.edu

[1] The University of Edinburgh, Edinburgh, UK

[2] Department of Psychological Sciences, The University of Missouri, University of Missouri-Columbia, 210 McAlester Hall, Columbia, MO 65211-2500, USA

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 and Schooler, 1991), 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, if you put down your drink at a party and an identical glass filled with something else appeared next to it, marking your own drink in your mind could be achieved using a verbal description ('the yellow one'), as well as a visual representation of the yellowness of the drink in it. The visual memory representation would help you grab your champagne instead of what would be someone else's orange juice. 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; Paivio, 1971; Saito, Logie, Morita, & Law, 2008). 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 subvocal rehearsal of verbal material is a separate mechanism that can support memory; however, the existence 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 labeling 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; Peich, Husain, & Bays, 2013; van den Berg, Shin, Chou, George, & Ma, 2012; Wilken & Ma, 2004; Zhang & Luck, 2008, 2009, 2011). 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 et al., 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. (2015) 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 subvocally 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 long-term memory (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.

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 would be in line with literature suggesting that older adults show more severe deficits for visuospatial than verbal material (see Jenkins, Myerson, Joerding, & Hale, 2000; see also Bopp & Verhaeghen, 2005; Johnson et al., 2010; Leonards, Ibanez, & Giannakopoulos, 2002; Logie & Maylor, 2009; Myerson, Hale, Rhee, & Jenkins, 1999; Park et al., 2002, for a meta-analysis regarding age-deficits in verbal memory). 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. However, in older people, verbal memory appeared related to a more general factor, 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 subprocesses 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 labeling of stimuli, and younger and older adults differ in the extent to which they rely on such labeling, 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 efficient verbal labeling and rehearsal and bound objects do not—while 'red, blue, green' may be feasible to verbalize during a typical memory retention interval, 'red-circle, blue-square, green-triangle' would likely be much more cumbersome (Brockmole & Logie, 2013; see Forsberg, Johnson, & Logie, 2019, for a summary of feature-binding paradigms and the role of verbal labeling, but see also Sense, Morey, Prince, Heathcote, & Morey, 2016). 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 labeling 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 manipulating verbal labeling in the delayed estimation paradigm. We conducted a conceptual replication of Souza and Skóra (2017, Experiment 4), including both younger and older participants. Like them, we used a mixture model to distinguish rough categorical ('verbal') representations from continuous ('visual') representations, based on knowledge of the specific shade (Hardman et al., 2017). This model relies on the assumption that correct responses in this color memory task belong to one of these two types of representations: continuous

representations, which are visual in nature (as they are based on a precise representation of what the specific color looked like), and categorical representations, which are not necessarily visual (as they tend to cluster around prototypical color values, and could theoretically be maintained without a visual representation (e.g., only remembering the verbal label 'red'). Hardman et al. (2017) found that when they included a mechanism for remembering rough categories (e.g., 'purple'), the model outperformed the continuous representation-only model. However, we did not assume that continuous representations as classified by the model correspond *perfectly* to visual memory traces, and categorical to verbal. Indeed, the link between verbal labeling and the two types of representations is an empirical question, which we explore with the explicit labeling manipulations below.

We investigated the following three questions:

(1) Do older adults spontaneously use verbal labels more than younger adults 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 subvocally label, the number of continuous versus categorical 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/or rehearsal of labels impair older adults' memory performance more than it impairs younger adults' performance? If older adults' 'visual' memory performance (as measured by the visual delayed estimation task) depends on verbal labeling (i.e., a verbal strategy) to compensate for poor visual memory, their memory, overall, should be more impaired while verbal labeling is disrupted by articulatory suppression, compared with younger adults.
(3) Do older adults benefit from labels in the same way as younger adults? This can be tested by comparing instructed labeling with suppression (i.e., disrupted labeling) in the 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, instructed labeling would result in fewer continuous representations, but more categorical representations, which would be consistent with the verbal recoding hypothesis (see Schooler & Engstler-Schooler, 1990). Alternatively, labeling might be beneficial because it adds a verbal (categorical) representation to the visual WM trace (the dual-trace [visual + verbal] hypothesis). Then, the number of continuous ('visual') representations would be the same, but there would be additional, categorical ('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. The method and analyses were preregistered via the Open Science Framework (osf.io/m64px).

We examined whether 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 performance under instructed overt labeling, articulatory suppression to prevent labeling, and in silence (similar to Souza & Skóra, 2017) to test the following three hypotheses.

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

H_2: *Do younger and older adults differ in the probability that the representation in memory is continuous as opposed to categorical*? If participants in one age group 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 other age group. Also, if participants in different age groups benefit differently from verbal labels, labeling (compared with suppression) may result in increases in different types of representations (i.e., differential gains in continuous and/or categorical representations).

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 differential favoring of categorical representations over continuous may contribute to differences in precision between younger and 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 (i.e., continuous 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,[1] we recruited 32 younger adults and 33 older adults, and excluded and replaced two younger and three older adults for not completing all trials, per our preregistered exclusion criteria. The final sample consisted of 30 younger adults (18–27 years old, M = 22.0, SD = 2.7, nine males), and 30 older adults (62–78 years, M = 68.6, SD = 4.9, 10 males). 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 cutoff 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 preregistered (osf.io/m64px).

Materials and procedure

The study had a mixed design, with age group (younger or older) as a between-subjects factor, and a within-subjects 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.

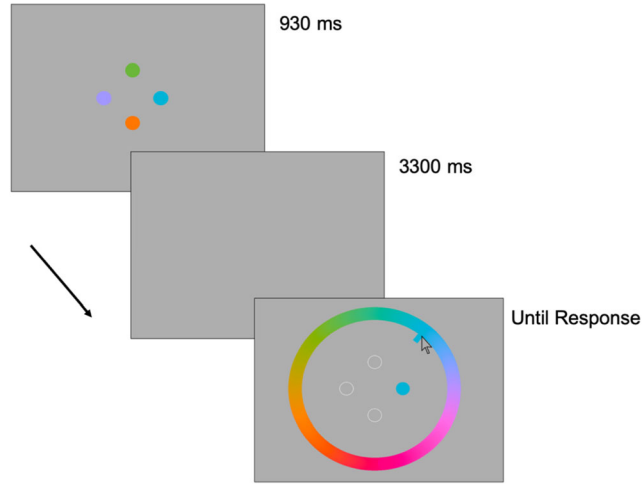

Fig. 1 Outline of a typical trial. This is a visual representation where items are not drawn to scale. The function used to generate the color values are available in the Supplementary Materials. (Color figure online)

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 Fig. 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).

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,[2] centered in the color space with L = 50, a = 20, b = 20, radius = 60, and then converted to RGB values, trimming nonsense values (see Supplementary Materials). Next, they saw a grey interstimulus interval (ISI) screen of 3,300 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 on the shade in the color-wheel they recalled having been in that

[1] We based this sample size on Experiment 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 because each participant performed three different conditions, with a set size of four 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).

[2] 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.

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 (Version 1.82.01; Peirce, 2007) and displayed on a 22-in. LCD monitor, with a diagonal of 20.6-in., and a screen resolution of 1,680 × 1,050. 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?' We 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.

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) versus 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 Fig. 2b). Alternatively, responses could be informed by continuous, more fine-grained representations of the specific hue, varying linearly with the studied colors (see Fig. 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 Fig. 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 Fig. 2d), or in accordance with certain categories (categorical guessing, see Fig. 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 2h shows all responses from an imaginary participant. This imaginary participant has five color categories. The heights of the distributions in Fig. 2f show the likelihood that different response angles would be chosen for a specific study angle, for each of the four response types. Figure 2e illustrates all the responses classified into the four different categories. Figure 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 Fig. 2a. 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 Fig. 2c, and uniform (continuous) guessing in Fig. 2d. 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; $1 - P^O$).

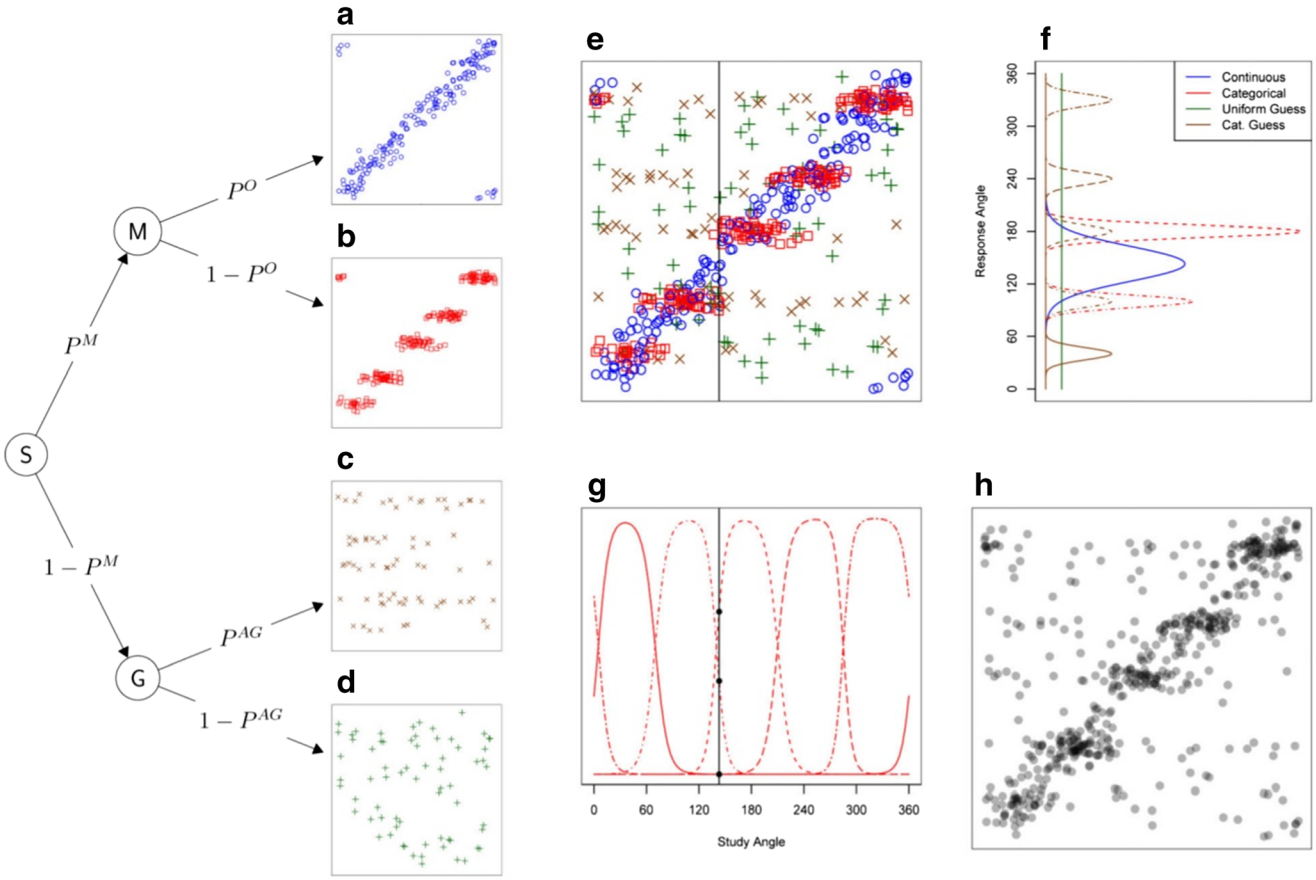

Fig. 2 Multinomial process tree for the model and related plots for Hardman et al.'s (2017) categorical-continuous model. For all scatterplots, the *x*-axis represents the studied color hue and the *y*-axis the response hue. **a** Continuous memory: responses vary linearly with the studied hue. The width of the diagonal line indicates continuous imprecision. **b** Categorical memory: for a range of studied hues, the same categorical response is provided. The width of the categorical bands reflects categorical imprecision. **c** Categorical guessing: guessing is distributed over categories. **d** Random guessing. **e** show the points in panels **a–d** combined. **f** Shows response densities for the four different response types for a single study angle (indicated by vertical line in **e**). **g** Shows the function giving the probability that a given study angle will be assigned to the given category. **h** Illustrates the same points as in **e**, but without information about the response type. Reprinted from "Categorical Working Memory Representations Are Used in Delayed Estimation of Continuous Colors," in K. O. Hardman, E. Vergauwe, and T. J. Ricker, 2017, *Journal of Experimental Psychology: Human Perception and Performance*, *43*(1), 30. (Color figure online)

Thus,

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

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

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. The (young adult) suppression condition was used as a cornerstone in the modeling. 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 cutoff points; therefore, some subjective interpretation is inevitable when describing the result. Typically, BF = 1 is considered 'no evidence', BF between 1 and 3 is considered 'anecdotal' (Wetzels & Wagenmakers, 2012) or 'not worth more than a bare mention' (Jeffreys, 1961), and BF greater than 3 is considered 'substantial'[3], between 10 and 30 'strong', 30 and 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 preregistered 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%); $\chi^2(1, N = 60) = 0.34$, $p = .56$. We also conducted some exploratory analyses to test whether there was a differential effect of suppression and/or overt labeling, compared with performance in silence, between self-reported labelers and nonlabelers. We found no clear evidence that participants who reported using labeling in the silence condition differed from those who did not report having labeled (see Supplementary Materials for further details about these analyses).

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 preregistration, 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 the Supplementary Materials.

We fit all models with three parallel chains of 10,000 iterations each, with a burn-in of 2,000 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 rerunning 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,[4] 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 ($\Delta = -190.7$), indicating a better fit to the data. Therefore, we only discuss the results of this model (see Supplementary Materials for output from the within-item models).

[3] The common meaning of "substantial" has changed over time (R. D. Morey, 2015).

[4] While the experimenter did not monitor or record which specific labels were used, the occurrence of labeling was monitored throughout the session through constant experimenter presence. All participants were able to say four color names in the experimental time frame, as established during the practice session.

Memory performance: Parameter estimates

Figure 3 shows the group-level probability that the first-presented items were in memory (Fig. 3a), the group-level probability that they were stored continuously (Fig. 3b), and the group-level imprecision of continuous representations (Fig. 3c), by age group and verbalization condition. See Table 1 for BF_{10}s for the factors and interactions, and Table 2 for BF_{10}s for all subset comparisons. 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, older adults did not appear comparatively more impaired under suppression, as suppression impaired the younger adults' performance comparatively more, when compared with labeling (see Fig. 3; also confirmed by subset analysis contrasting suppression with labeling; Age Group × Verbalization BF_{10} = 29.06; see Table 2). Alternatively, this difference can also be interpreted as younger adults being comparatively more able to benefit from overt labeling, than older adults. Or, both these effects could coexist. 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 versus Continuous K

We also calculated estimated Categorical versus 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 continuous ('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

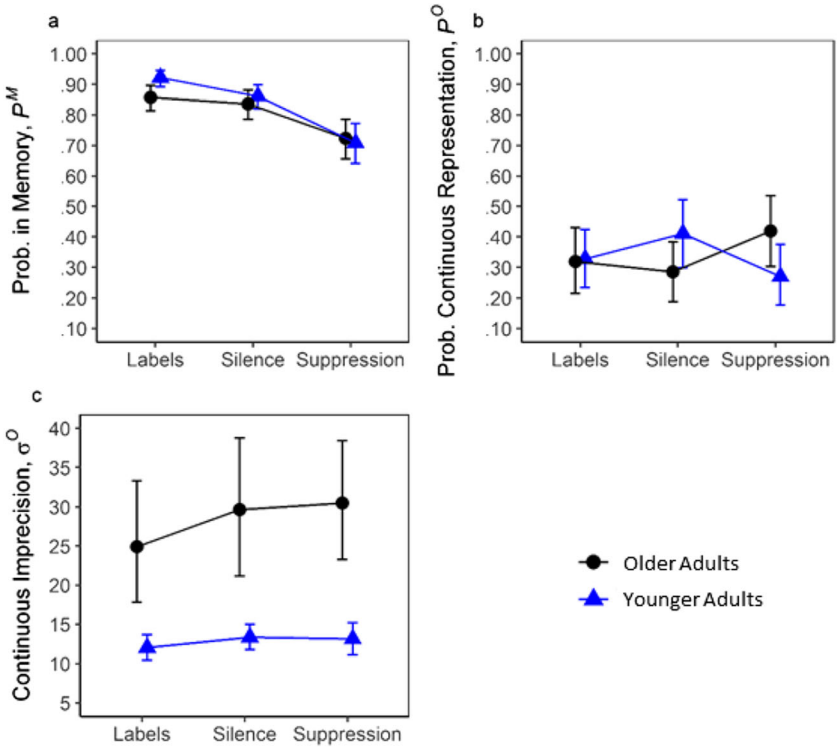

Fig. 3 Memory for the first item only. **a** The group-level probability of having the probed item in memory. **b** The group-level probability that memory representation is continuous. **c** The imprecision of the group-level continuous memory representation. Error bars depict 95% credible intervals of the parameters

Table 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

second, categorical and verbal trace, resulting in increased categorical and Total K, but no change in Continuous K. Posterior differences in Categorical and Continuous K by verbalization condition and age group are presented in Fig. 4. To test differences in labeling in the age groups compared with 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 subvocal labeling when performing the task in silence we quantified the labeling benefit by comparing labeling versus suppression (see Fig. 5 for estimates of total, Categorical and Continuous K).

Preventing silent 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 2). This was true for both age groups, with no evidence of an 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, the proportion of continuous representations under suppression relative to silence decreased for younger adults but increased for older adults (see Fig. 3a). Although this was the case for the proportion of continuous representations, when considering the absolute capacity, Continuous K did not change credibly under suppression in older adults ($M = +0.26$ items). In younger adults, however, it was credibly reduced under suppression ($M = -0.65$; see Fig. 4a). Categorical memory representations in older adults were credibly reduced by suppression ($M = -0.71$ items), but not in younger adults ($M = -0.04$ items; see Fig. 4b). 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 ($BF_{10} = .15$).

Enforcing labeling (silence vs. labeling)

Overt labeling improved memory (P^M) compared with performing the task in silence (BF_{10} = 42.03), in both

Table 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	210.36	.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

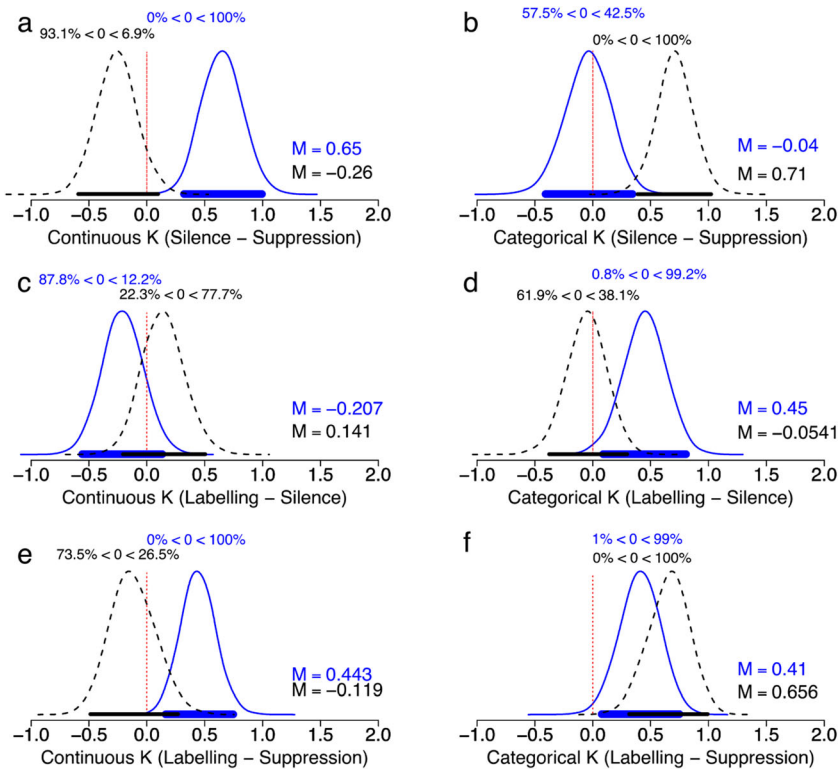

Fig. 4 Mean values (M) larger than zero for condition (**a** - **b**, e.g., Silence - Suppression) indicates larger estimates in Condition **a** than **b**).

age groups (there was no evidence for an interaction with age; BF_{10} = .14). It did not influence the probability of having a continuous (as opposed to categorical) representation of the first item (P^O), regardless of age group (see Table 2). Labeling did not produce a credible change in Continuous K in either age group (see Fig. 4c). However, compared with performance in silence, younger adults' categorical memory representations increased credibly when instructed to label (M = +0.45 items; see Fig. 4d). In contrast, older adults' categorical memory capacity under instructed labeling did not differ credibly from their performance in silence (see Fig. 4d). Furthermore, there was some evidence that overt labeling increased precision (i.e., reduced imprecision; σ^O) for the first item (BF_{10} = 3.02), but not to different extents in the age groups.

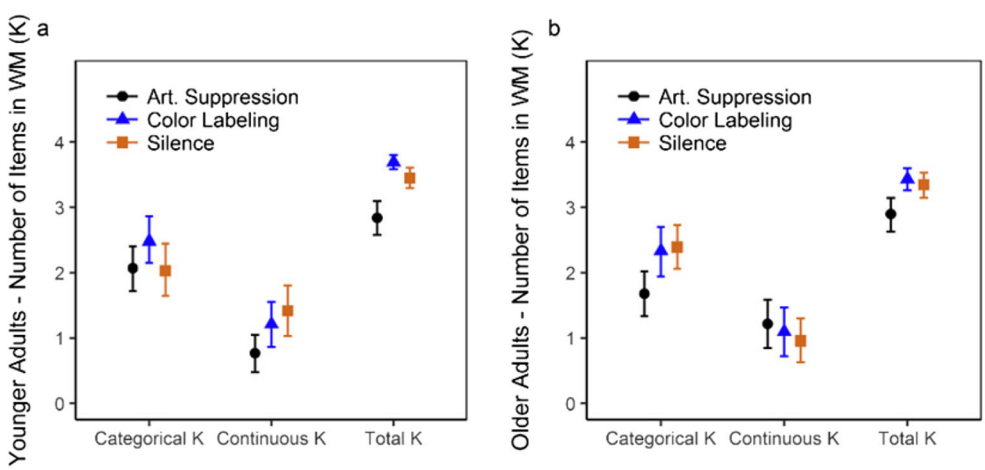

Fig. 5 Memory for the first item only. Categorical, continuous, and total K, by age group and verbalization condition. Error bars depict 95% credible intervals

The labeling benefit (labeling vs. suppression)

Overt labeling improved memory (P^M) in both age groups compared with suppression ($BF_{10} = 3.11 \times 10^{11}$). There was no evidence of any effect of labeling on the probability of continuous (as opposed to categorical) responding (P^O). However, the absolute 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; see Fig. 4f). 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; see Fig. 4e). 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' Continuous K increased slightly under suppression ($M = +0.12$ items). However, this increase was not within the 95% credible interval (see Fig. 4e), which would be required to support the verbal recoding hypothesis (i.e., the number of verbal representations increases at the expense of the visual representation). Instead, the older adults' gain fit best with the dual-trace hypothesis; labeling is beneficial because it adds a verbal (categorical) representation to the visual WM trace, without changing the number of visual representations. Furthermore, there was some evidence that suppression reduced precision (i.e., increased imprecision 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 versus silence comparison for continuous responding (P^O). However, Categorical versus Continuous K comparisons were similar to the original results (see Supplementary Materials), 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 comparatively less impaired 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 than younger adults when performing the task in silence? (2) Depend more on verbal labels for visual memory performance than younger adults? (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 extent to which they spontaneously applied subvocal labeling 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 manipulating subvocal labeling (silence vs. suppression) affected the age groups' probabilities to respond continuously differently was 'decisive'. Specifically, compared with when performing the task 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 subvocally in silence. In contrast, older adults' categorical memory capacity in silence did not differ credibly from their performance under instructed labeling (see Fig. 4d) but was poorer under suppression (see Fig. 4b)—consistent with spontaneous subvocal 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 with 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 subvocal labeling, 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 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 extent 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 tested whether older adults' memory (P^M) would be comparatively more impaired during concurrent suppression compared with the two other conditions (i.e., when labeling was prevented). This was not supported. Instead, we observed strong evidence that suppression impaired the younger adults' performance comparatively more when contrasting it with labeling. This Age × Labeling interaction left it unclear whether younger adults were able to benefit more from labeling, or were comparatively more disrupted by suppression. Either way, these results contradicted the idea that older adults depend on a less impaired 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. Thus, interestingly, older adults' increased tendency to use verbal labels when performing the task silently (as discussed above) did not appear to correspond to a *need* to use such labels to maintain good overall memory performance, compared with younger adults.

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 by the CatContModel (i.e., the imprecision parameter analyzed in this paper should not be influenced by categorical responding)—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 instructed overt labeling with suppression (prevented labeling) 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., preexisting visual representations; Brady, Konkle, & Alvarez, 2011) can boost visual (continuous) WM (Alvarez & Cavanagh, 2004; Buttle & Raymond, 2003; Curby & Gauthier, 2007; Curby, Glazek, & Gauthierm, 2009; Olsson & Poom, 2005; 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 Fig. 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). If so, this effect may not be noticeable in studies using categorical stimuli. It is also likely that our experimental conditions reflect different degrees of labeling (overt labeling: nearly always, silence: sometimes, suppression: less often). If the effect of labeling on Categorical and Continuous K differs in magnitude, and silence is not a true middle ground (which is unlikely), this might also explain this pattern of results. 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 'cool' 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 with 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 CatContModel (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 CatContModel 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 versus continuous representations in a way that would not be possible using a set of fixed, categorical to-be-remembered items. Moreover, including a perceptual matching task to ensure that younger and older adults did not differ in memory precision due to for instance greater motor noise in older adults may have been informative. However, Loaiza and Souza (2018) found that while there was an age difference in perceptual matching, it was much smaller than that observed in the WM task, suggesting that it did not fully account for the age differences in WM precision memory (see also, Souza, 2016; Loaiza & Souza, 2019). Also, age-related perceptual and/or motor differences are unlikely to vary between our experimental conditions. Next, there were instances where participants coughed, sneezed, or otherwise lapsed attention during a trial, and may not have said all four colors on that specific trial (or, indeed, maintained suppression perfectly). This would have been a very small minority of trials. However, recording sounds would have allowed us to exclude such trials, and ensure that the number of such instances did not vary between participants in the two age groups. Finally, while the experimenter monitored that color labels were said aloud by all participants in the labeling condition, we did not monitor which label was applied to which color. Thus, it is possible that participants may have used unrelated, random color names simply to comply with the instruction. However, the large beneficial effect of labeling, in both age groups, indicates that this was not the norm, or that assigning any kind of label can be beneficial.

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 predetermined, 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 simply 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 suppression prevented verbal labeling, which suggested that older adults' visual memory performance did not *depend* more on verbal labeling.

However, older adults appeared to rely more on coarse, categorical representations than younger adults when doing the task in silence. Furthermore, these representations appeared to be supported by subvocal labeling, since they were specifically reduced under suppression. Finally, while we replicated Souza and Skóra's (2017) 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. 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.

Acknowledgements The work was supported by a PhD studentship from The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/L501530/1). Funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and Medical Research Council (MRC) is gratefully acknowledged.

Open practices statements This experiment was preregistered (osf.io/m64px). Note the minor amendment reducing the number of trials (osf.io/pjtc6). The data and analysis code are available at (osf.io/bzd2n).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

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.

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.

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: Oxford University Press.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29.

Baddeley, A. D., & Hitch, G. J. (2018). The phonological loop as a buffer store: An update. *Cortex, 112,* 91–106. doi:https://doi.org/10.1016/j.cortex.2018.05.015

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* (p. 28–61). New York, NT: Cambridge University Press. doi:https://doi.org/10.1017/CBO9781139174909.005

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.

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.

Baltes, P. B., & Baltes, M. M. (1990). Psychological perspectives on successful aging: The model of selective optimization with compensation. *Successful aging: Perspectives From the Behavioral Sciences*, *1*(1), 1–34.

Bayarri, M. J., & Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science, 19*(1), 58–80.

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.

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.

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., 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.

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.

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.

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.

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.

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

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. (2005). Working memory capacity limits in a theoretical context. In C. Izawa & N. Ohta (Eds.), *Human learning and memory: Advances in theory and application. The 4th Tsukuba international conference on memory* (pp. 155-175). Mahwah, NJ: Erlbaum.

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

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.

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, *66*, 115–142.

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.

Dickey, J. M. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics, 42*(1), 204–223.

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

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.

Dvorine, I. (1963). Quantitative classification of the color-blind. *The Journal of General Psychology, 68*(2), 255–265.

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

Forsberg, A., Johnson, W., & Logie, R. H. (2019). Aging and feature-binding in visual working memory: The role of verbal rehearsal. *Psychology and Aging*, *34*(7), 933–953.

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, & 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.

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

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.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika, 82*(4), 711–732.

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

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.

Hardman, K. O. (2017). CatContModel: for delayed estimation tasks (Version 0.8.0) [Computer software]. Retrieved from https://github.com/hardmanko/CatContModel/releases/tag/v0.8.0

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.

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.

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.

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.

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

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

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, 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.

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.

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

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

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.

Loaiza, V. M., & Souza, A. S. (2018). Is refreshing in working memory impaired in older age? Evidence from the retro-cue paradigm. *Annals of the New York Academy of Sciences, 1424*(1), 175–189.

Loaiza, V. M., & Souza, A. S. (2019). An age-related deficit in preserving the benefits of attention in working memory. *Psychology and Aging, 34*(2), 282.

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. (1995). *Visuo-spatial working memory*. Hove, UK: Erlbaum.

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

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., 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.

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

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.

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. (2015). 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/

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.

Oberauer, K. (2013). The focus of attention in working memory—from metaphors to mechanisms. Frontiers in Human Neuroscience, 7, 673.

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, Rinehart & Winston.

Park, D. C., & Reuter-Lorenz, P. (2009). The adaptive brain: aging and neurocognitive scaffolding. *Annual Review of Psychology*, *60*, 173–196.

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. 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.

Pashler, H. (1988). Familiarity and visual change detection. *Perception & psychophysics*, *44*(4), 369–378.

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.

Perfect, T. J., & Maylor, E. A. (2000). *Rejecting the dull hypothesis: The relation between method and theory in cognitive aging research*. Oxford, England: Oxford University Press.

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.

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.

Ramaty, A., & Luria, R. (2018). Visual working memory cannot trade quantity for quality. *Frontiers in Psychology*, *9*.

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., & 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. In R. Cabeza, L. Nyberg, & D. Park (Eds.), *Cognitive neuroscience of aging: Linking cognitive and cerebral aging* (pp. 186–217). Oxford, England: Oxford University Press.

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.

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.

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.

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.

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.

Shulman, H. G. (1971). Similarity effects in short-term memory. *Psychological Bulletin*, *75*(6), 399.

Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science*, *7*(5), 301–305.

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. (2016). No age deficits in the ability to use attention to improve visual working memory. *Psychology and Aging, 31*(5), 456.

Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. *Cognition*, *166*, 277–297.

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.

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 Ravenzwaaij, D., Cassey, P., & Brown, S. D. (2018). A simple introduction to Markov Chain Monte-Carlo sampling. *Psychonomic Bulletin & Review*, *25*(1), 143–154.

Vergauwe, E., Barrouillet, P., & Camos, V. (2010). Do mental processes share a domain-general resource?. *Psychological Science*, *21*(3), 384–390.

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.

Watson, J. B. (1924). The place of kinaesthetic, visceral and laryngeal organization in thinking. *Psychological Review*, *31*(5), 339.

Wetzels, R., & Wagenmakers, E. J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review, 19*(6), 1057–1064.

Whorf, B. L. (1956). Language, thought, and reality: Selected writings of Benjamin Lee Whorf (Ed. J. B. Carroll). Cambridge, MA: MIT Press; trad. It. 1970. *Linguaggio, pensiero e realtà*.

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, 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.

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.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.