Article:
Atkinson, AL orcid.org/0000-0001-9536-6950, Baddeley, AD and Allen, RJ orcid.org/0000-0002-1887-3016 (2018) Remember some or remember all? Ageing and strategy effects in visual working memory. Quarterly Journal of Experimental Psychology, 71 (7). pp. 1561-1573. ISSN 1747-0218

https://doi.org/10.1080/17470218.2017.1341537

The Quarterly Journal of Experimental Psychology

ISSN: 1747-0218 (Print) 1747-0226 (Online) Journal homepage: http://www.tandfonline.com/loi/pqje20

Remember some or remember all? Ageing and strategy effects in visual working memory

Amy L. Atkinson, Alan D. Baddeley & Richard J. Allen

Publisher: Taylor & Francis & The Experimental Psychology Society
Journal: The Quarterly Journal of Experimental Psychology
DOI: 10.1080/17470218.2017.1341537

Remember some or remember all? Ageing and strategy effects in visual working memory

Amy L. Atkinson[a], Alan D. Baddeley[b], and Richard J. Allen[c]

a) Corresponding Author. School of Psychology, University of Leeds, Leeds, LS2 9JT. amy.atkinson@live.co.uk. +44(0)7907165249.
b) Department of Psychology, University of York, Heslington, York, Y010 5DD.
c) School of Psychology, University of Leeds, Leeds, LS2 9JT.

Author notes

The authors would like to thank Nelson Cowan, Thomas Aisthorpe and Ed Berry for useful discussion, and Emma Bonnick and Emily Earp for assistance with data collection. Experiment 1 formed part of a master’s degree at the University of York while Experiment 2 was part-supported by an Economic and Social Research Council PhD Discipline Studentship to Amy Atkinson. The authors declare no conflict of interest.

Abstract

Recent research (Bengson & Luck, 2015) has indicated that visual working memory capacity for unidimensional items might be boosted by focusing on all presented items, as opposed to a subset of them. However, it is not clear whether the same outcomes would be observed if more complex items were used which require feature binding, a potentially more demanding task. The current experiments therefore examined the effects of encoding strategy using multidimensional items in tasks that required feature binding. Effects were explored across a range of different age groups (Experiment 1) and task conditions (Experiment 2). In both experiments, participants performed significantly better when focusing on a subset of items, regardless of age or methodological variations, suggesting this is the optimal strategy to employ when several multidimensional items are presented and binding is required. Implications for task interpretation and visual working memory function are discussed.

Keywords: visual working memory; strategy/strategies; encoding; binding; ageing/aging

Remember some or remember all? Ageing and strategy effects in visual working memory

Visual Working Memory (VWM) allows incoming information to be temporarily stored and manipulated (Blacker, Curby, Klobusicky & Chein, 2014; Hartshorne, 2008; Luck & Vogel, 1997). It is considered essential for a myriad of human activities, including learning and navigation of the visual world (Blacker et al., 2014; Opitz, Schneiders, Krick, & Mecklinger, 2014), and is predictive of general cognitive ability and fluid intelligence (Fukuda, Vogel, Mayr, & Awh, 2010; Gold et al., 2010). However, capacity of VWM is constrained (Hartshorne, 2008), with young adults typically able to store only 3-4 items simultaneously (Luck & Vogel, 1997; Vogel & Machizawa, 2004). Moreover, large age-related declines are present (Brockmole, Parra, Della Sala, & Logie, 2008; Brockmole & Logie, 2013), with adults over 55 years old exhibiting a similar capacity to 8-9-year-old children (Brockmole & Logie, 2013). Research has therefore begun to explore how VWM capacity can be enhanced (e.g. Blacker et al., 2014).

One potential way to enhance working memory is through the use of encoding strategies, which refer to effortful, goal-directed processes intended to provide optimal encoding conditions (Dunning & Holmes, 2014). While relatively little research has explored strategy effects in VWM, one possible approach concerns the extent to which participants focus on some or all presented items. During the presentation phase of VWM tasks, individuals might focus on all presented stimuli, thus ensuring the item subsequently probed during retrieval was attended to. However, if the number of items exceeds capacity limits, VWM might be overloaded, rendering an individual unable to recall some, or even any items correctly (Gathercole, 2008). Moreover, even if loss of representations does not occur, focusing on all presented items may result in less precise representations (Donkin, Kary, Tahir, & Taylor, 2016; Pertzov, Avidan, & Zohary, 2009). Alternatively, individuals could focus on a subset of items once their capacity has been reached. This should lead to stronger or more precise representations, as individuals would have more time and resources to encode each item, and VWM capacity would be less likely to be exceeded (Donkin et al., 2016; Pertzov et al., 2009). However, under this approach, if an item is probed that is not in memory, participants would have to guess when responding (Donkin et al., 2016).

An assumption of several studies is that focusing on a subset of items is the optimum encoding strategy once an individual's capacity has been reached (Cusack, Lehmann, Veldsman, & Mitchell, 2009; Linke, Vincente-Grabovetsky, Mitchell, & Cusack, 2011). These studies have argued that individuals with a low VWM capacity (Linke et al., 2011) and low intelligence (Cusack et al., 2009), typically focus on all the items presented during encoding, even beyond capacity limits, which results in poor performance. In contrast, individuals with higher intelligence and a higher VWM capacity recognise that this strategy is maladaptive, and therefore focus on only a subset of items beyond capacity limits, which results in superior performance (Cusack et al., 2009; Linke et al., 2011).

However, research directly comparing such strategies has yielded evidence to suggest the contrary. Bengson and Luck (2015) presented young adults with a change detection task, in which participants were asked to indicate whether coloured square test arrays were identical or different to arrays displayed at encoding. Set size was randomly varied, with either four, six or eight coloured squares presented. Before the start of the task, participants were provided with explicit instructions on how to encode the information. They were either told: "Try to remember the entire array, no matter how many items are presented" (remember-all), "If you can't remember the entire array, focus on a subset and try to remember them well" (remember-subset), or "Do your best and try to get as many trials correct as possible" (do-your-best). Superior performance was observed in the remember-all condition, while remember-subset and do-your-best produced equivalent accuracy. This effect remained present at larger set sizes, suggesting that attempting to remember the entire array results in optimal performance, even after VWM capacity has been exceeded.

It should be noted, however, that the items used by Bengson and Luck (2015) were unidimensional, differing only by colour, with participants required to make a recognition memory judgment concerning only this feature dimension. In reality, visual objects are considerably more complex, comprising several features, such as colour, shape and depth, which must be accurately bound together (Ueno, Allen, Baddeley, Hitch, & Saito, 2011). This increased complexity may make it difficult for individuals to effectively encode all the visual information, potentially rendering a 'remember-all' strategy less than optimal. Moreover, conclusions drawn by Bengson and Luck (2015) were based on data from younger adults, who generally show optimal working memory abilities (Brockmole & Logie, 2013). Groups who typically experience VWM impairments, such as older adults (Brockmole et al., 2008; Brockmole & Logie, 2013), may show a different pattern of results as they may particularly struggle to effectively encode all the visual information. Indeed, Brown, Niven, Logie, Rhodes, and Allen (2017) recently suggested that older adults might be more likely to spontaneously implement a strategy of focusing on certain items from a presented set, to maintain performance levels in the face of limited cognitive resources. However, research has not explicitly examined the effects of encoding strategy instruction on VWM in older adults.

The current experiments explored the effects of encoding strategy in a range of different task conditions, using items which varied on two dimensions. Furthermore, in each of the different tasks, accurate responding required participants to encode both the colour and shape of items and successfully bind these features together. To explore whether strategy effects differ across age groups, younger and older adults were recruited in Experiment 1, with verbal cued recall adopted as the response task. Experiment 2 focused on young adult performance, and examined whether similar or differing strategy effects were observable using recognition tasks that re-presented either the whole display, as was the case with Bengson and Luck (2015), or a single probe at test.

Experiment 1

Experiment 1 explored the effects of strategy in younger adults (18-25 years) and older adults (55+ years) using a cued-recall VWM task which has previously been shown to detect effects of strategically prioritising single items in young adults (Hu, Allen, Baddeley & Hitch, 2016; Hu, Hitch, Baddeley, Zhang, & Allen, 2014). In each trial, participants were presented with several coloured shapes simultaneously. After a brief delay, they were cued with the outline of a shape and asked to recall the colour. A 2x2x3 design was implemented, manipulating instruction type (remember-subset or remember-all), age group (young adults or older adults) and set size (three, four or six items), with the primary dependent variable of VWM capacity. Participants subsequently completed a questionnaire, which assessed task difficulty, strategy adherence, how well they thought they performed, and how many shapes they focused on in the remember-subset condition.

If the benefits of the remember-all strategy are not limited to simple unidimensional items and apply across VWM more generally, we would expect to extend the findings reported by Bengson and Luck (2015) to paradigms using more complex objects that require feature binding. Alternatively, it is possible that participants may benefit from the remember-subset strategy under these circumstances, as the increased complexity and novel binding requirement may make it difficult for participants to effectively encode all the visual information presented. An interaction between instruction type and age group is also possible, as older adults may benefit from the remember-subset instruction more than younger adults a result of their poorer VWM abilities.

Method

Participants. Forty-four participants took part. Participants had no known learning difficulties, had normal or corrected-to-normal vision and no colour blindness, and were recruited through word of mouth from communities in and around the Universities of York and Leeds.

The younger adult sample comprised 20 participants aged between 18-25 years (Mean age (M) = 21.66, Standard Deviation (SD) = 1.95; 3 males, 17 females; M. years of education = 15.95, SD = 1.47;). Their mean standardised score on the Spot the Word (STW) task, used to estimate verbal IQ (Baddeley, Emslie, & Nimmo-Smith, 1992), was 10.95 (SD = 1.85).

Twenty-four older adults, aged 55-81 years, also completed the experiment (Mean age (M) = 65.94, SD = 9.19; 7 males, 17 females; M. years of education = 12.96, SD = 4.74). Two older adults were excluded for not following the instructions, one due to poor performance on the primary task (two SDs below the mean for the age group), and one due to a history of strokes. Final analysis was conducted on the data for 20 older adults (M. age = 66.29, SD = 7.76; M. years of education = 13.30, SD = 5.08; 5 males, 15 females). All older adults were healthy and community-dwelling, with no known dementia or cognitive impairment. The Mini Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) indicated that no participants showed signs of cognitive impairment (R = 25-29, M = 27.20, SD = 1.40). Their mean standardised score on the STW task was 9.25 (SD = 2.97). As is common in ageing studies, the younger adults had significantly more years of education than the older group, $t(22.15) = 2.24$, $p = .035$. The younger adults group also had a significantly higher standardised verbal IQ score, $t(31.80) = 2.17$, $p = .037$.

Materials. Visual working memory task. The visual memory task was run on a 13.3" MacBook Air using SuperCard (Version 4.5). Stimuli were drawn from a pool of eight shapes (circle, cross, triangle, arch, flag, star, diamond, chevron) and eight colours (red, yellow, green, blue, turquoise, black, purple, grey), and were randomly paired on each trial, selecting without replacement. Test cues consisted of unfilled outline shapes. All stimuli subtended a visual angle of 1.49º, based on an approximate viewing distance of 50cm.

Spot the Word (STW). The STW task (Version B) was used to assess verbal IQ (Baddeley et al., 1992). Participants were presented with 60 pairs of items, each comprising a real word and a pseudo-word (e.g. Kitchen - Harrick, Epicene - Floricity), and had to decide which item was real. Raw scores were then converted to standardised scores using age-corrected norms (Baddeley et al., 1992).

Questionnaire. A questionnaire was designed, comprising 5-point Likert scales, to assess how easy participants found the conditions and how well they thought they performed. Participants were also asked how many shapes they focused on in the remember-subset condition. The questionnaire also assessed strategy adherence, enabling us to exclude participants who did not follow the instructions correctly.

Design and procedure. The study implemented a 2x3x2 mixed design, with instruction type (remember-all or remember-subset) and set size (3, 4 or 6 shapes) as within-subject variables and age group (younger or older adults) as the between-subjects variable.

The VWM task consisted of two blocks of 90 trials: one for each instruction type condition, with order of blocks fully counterbalanced. Within each instruction condition, there were 30 trials for each set size (3, 4 and 6), randomly intermixed within each block. At the start of each instruction block, participants completed six practice trials, two at each set size.

Each condition commenced with the provision of written instructions. In the remember-all condition, participants were told “try to remember the colours of all of the shapes in the display, regardless of how many are presented”. In the remember-subset condition, they were told “if you can’t remember the colour of all the shapes in the display, focus on just some of them and try to remember them well”.

The experimental task is illustrated in Figure 1. Each trial began with presentation of a randomly generated two-digit number at screen centre for 1500ms. Participants repeated this number aloud for the duration of the trial, in order to prevent use of verbal coding (Baddeley, 1986). Next, a fixation cross appeared for 1000ms, followed by the display of coloured shapes (for 1000ms). Shapes were presented on a white background, at one of eight possible locations positioned at compass points around the screen centre. Following a 1000ms delay, a single shape outline appeared at the centre of the screen, with participants required to verbally state the original colour of this shape. Participants then pressed the space bar to move onto the next trial.

(Figure 1 about here)

Each instruction block was separated into three sections, each containing 30 trials, divided by short breaks. Participants were reminded of the instruction (remember-all or remember-subset) following the practice trials and at each break point. The STW task was completed between the instruction-type blocks of the VWM task, and the task questionnaire was administered at the end of the study.

Data analysis

The dependent variable was VWM capacity (K). This was calculated based on Cowan's (Chen & Cowan, 2013; Cowan, 2001) formula, adapted to fit the parameters of the current paradigm. Proportion correct (c) was related to capacity estimate (K), number of items in the memory array (N), and number of response options, that is, number of items in the experimental set (R), in the following formula:

$$c = \frac{K}{N} + \left(1 - \frac{K}{N}\right) \times \frac{1}{R}$$

This formula was transformed in order to convert proportion correct to capacity estimates for each participant in each condition:

$$K = \left(c - \frac{1}{R}\right) \times \frac{R \times N}{R - 1}$$

As appropriate, ANOVAs and t-tests were conducted. However, due to their reliance on null hypothesis significance testing (NHST), these techniques do not give an indication of the relative likelihood of the alternative hypothesis over the null hypothesis and do not allow an exploration of whether non-significant findings reflect equivalent performance across conditions or groups (Barchard, 2015; Mulder & Wagenmaker, 2016). Bayesian factor analysis was therefore also conducted using JASP (https://jasp-stats.org/). This compares the alternative hypothesis against the null hypothesis, thus allowing a comparison of the two, as well as an assessment of equivalence when non-significant differences are found (Jeffreys, 1961; Mulder & Wagenmaker, 2016). The results of these analyses are reported alongside the findings from the frequentist methods, and are interpreted using the guidelines set out by Jeffreys (1961). In each case, we first report the preferred model (i.e. with the highest Bayes Factor vs. the intercept-only null model) to emerge from the analysis. We then report the Bayes Factor for each component of interest in turn. These latter values were obtained by comparing a model containing each of these components, against a model with that component omitted.

Results

Accuracy. Mean capacity estimates (K) and standard error (SE) as a function of instruction type, set size and age group are displayed in Figure 2.

(Figure 2 about here)

A 2 (instruction type) x 3 (set size) x 2 (age group) mixed ANOVA yielded a significant effect of instruction type, $F(1, 38) = 9.60$, $MSE = 0.21$, $p = .004$, $\eta^2 p = .20$, with a higher VWM capacity emerging in the remember-subset condition ($M = 2.12$, $SE = 0.09$) relative to the remember-all condition ($M = 1.93$, $SE = 0.09$). A significant effect of age group was also found, with younger adults ($M = 2.37$, $SE = 0.12$) exhibiting a higher capacity than older adults ($M = 1.68$, $SE = 0.12$), $F(1, 38) = 15.96$, $MSE = 1.79$, $p < .001$, $\eta^2 p = .30$. A significant effect of set size was also found, $F(2, 76) = 17.48$, $MSE = 0.28$, $p < .001$, $\eta^2 p = .32$. Collapsing across instruction types and age groups, Bonferroni pairwise comparisons revealed significant differences between set sizes 3 ($M = 2.17$, $SE = 0.06$) and 6 ($M = 1.74$, $SE = 0.13$; $p<.001$) and set sizes 4 ($M = 2.17$, $SE = 0.10$) and 6 ($p<.001$), but not between set sizes 3 and 4 ($p = 1.00$). There was also a significant interaction between instruction type and set size, (Greenhouse-Geisser corrected $F(1.60, 60.63) = 3.45$, $MSE = 0.22$, $p = .048$, $\eta^2 p = .08$). Subsidiary analysis revealed no effect of instruction at set size 3 ($t(39) = .15$, $p = .883$), but an effect at set size 4 ($t(39) = 2.10$, $p = .043$) and 6 ($t(39) = 2.77$, $p= .009$). A significant interaction was also observed between set size and age group, $F(2, 76) = 5.07$, $MSE = 0.28$, $p = .009$, $\eta^2 p = .12$. Subsidiary analysis revealed that the effect of set size was not significant in the younger adults (Greenhouse-Geisser corrected $F(1.51, 28.66) = 2.54$, $MSE = 0.37$, $p = .109$, $\eta^2 p = .12$), but was in the older adults ($F(2, 38) = 19.78$, $MSE = 0.29$, $p < .001$, $\eta^2 p = .51$). No significant interaction was found between instruction type and age group, $F(1, 38) = 0.04$, $MSE = 0.21$, $p = .845$, $\eta^2 p = .001$, indicating that the remember-subset strategy enhanced capacity similarly across groups. No significant three-way interaction was found, (Greenhouse-Geisser corrected $F(1.60, 60.63) = 0.54$, $MSE = 0.22$, $p = .545$, $\eta^2 p = .01$.

A 2 x 3 x 2 mixed Bayesian ANOVA was also conducted to assess the relative strength of evidence for each main effect and interaction. The strongest model included main effects of instruction type, set size and age group and an interaction between set size and age group (BF > 1000 versus the intercept only model, ± 2.34%). The inclusion of a main effect of set size (BF > 1000), group (BF = 87.47) and the interaction between set size and age group (BF = 14.13) was strongly favoured, whilst the inclusion of an instruction type effect was moderately favoured (BF = 7.40). The analysis did not support the inclusion of an interaction between instruction type and age group (BF = 0.20), instruction type and set size (BF = 0.67) or instruction type, age group and set size (BF = 0.21).

Questionnaire. Task difficulty and judgements of performance ratings are displayed in Table 1, as a function of instruction type and age group. A 2 (instruction) x 2 (age group) mixed ANOVA on the task difficulty ratings revealed significant effects of instruction type, $F(1, 38) = 29.18$, $MSE = 0.27$, $p < .001$, $\eta^2 p = .43$, with participants reporting the remember-subset condition ($M = 2.33$, $SE = 0.12$) was easier than the remember-all condition ($M = 1.70$, $SE = 0.10$;. A significant effect of age group was also found, $F(1, 38) = 4.33$, $MSE = 0.65$, $p = .044$, $\eta^2 p = .10$, with young adults ($M = 2.20$, $SE = 0.13$) reporting the task was easier than the older adults ($M = 1.83$, $SE = 0.13$). There was also a significant interaction between instruction type and age group, $F(1, 38) = 5.65$, $MSE = 0.27$, $p = .023$, $\eta^2 p = .13$. Subsidiary analysis revealed that both the younger adults ($t(19) = 5.60$, $p < .001$) and the older adults ($t(19) = 2.10$, $p = .049$) reported finding the remember-subset condition easier, though the reported difference was larger in the younger adults. A 2 x 2 Bayesian mixed ANOVA revealed that the strongest model included a main effect of instruction type and age group, and an interaction between instruction type and age group (BF > 1000, versus the intercept only model, ± 4.27%). The inclusion of a main effect of instruction was strongly supported (BF > 1000), whilst the inclusion of a main effect of age group (BF = 2.72) and the interaction between instruction type and age group was anecdotally supported (BF = 1.71).

A 2 (instruction) x 2 (age group) mixed ANOVA on judgements of performance ratings revealed significant effects of instruction type, $F(1, 38) = 4.87$, $MSE = 0.50$, $p = .033$,

η^2p = .11, with participants judging that they performed better in the remember-subset condition (M = 2.43, SE = 0.11) compared to the remember-all condition (M = 2.08, SE = 0.12). A significant effect of age group was also found, F(1, 38) = 13.82, MSE = 0.52, p = .001, η^2p = .27, with young adults (M = 2.55, SE = 0.11) predicting better performance than older adults (M = 1.95, SE = 0.11). A significant interaction between instruction type and age group also emerged, F(1, 38) = 4.87, MSE = .50, p = .033, η^2p = .11. Subsidiary analysis demonstrated that the younger adults believed they performed better in the remember-subset condition (t(19) = 2.90, p = .009), whereas the older adults judged that their performance did not differ (t(19) = 0.00, p = 1.00). A 2 x 2 Bayesian mixed ANOVA revealed that the strongest model included a main effect of instruction type and age group, and an interaction between instruction type and age group (BF = 98.36, versus the intercept only model, ± 1.47%). The inclusion of a main effect of age group was strongly supported (BF = 16.28), whilst the inclusion of an interaction between instruction type and age group was moderately supported (BF = 3.05). The inclusion of a main effect of instruction type was anecdotally supported (BF = 2.28).

(Table 1 about here)

The questionnaire also assessed how many shapes participants focused on in the remember-subset condition. As one younger adult failed to answer this question, data from 19 younger adults and 20 older adults were analysed. No significant difference was found between younger (M = 3.03, SE = 0.14) and older (M = 2.82, SE = 0.16) adults, t(37) = 0.98, p = .332, BF = .46.

Discussion

This experiment provides clear evidence that strategy use at encoding affects performance on VWM tasks. In line with the assumptions of previous studies (Cusack et al., 2009; Linke et al., 2011), using a remember-subset strategy led to higher capacity estimates. This effect was only found at set sizes 4 and 6, though this is somewhat unsurprising given that participants were told to only apply the strategy in trials they found difficult, and three items is typically within capacity limits (Luck & Vogel, 1997; Hartshorne, 2008; Vogel & Machizawa, 2004).

No significant interaction emerged between instruction type and age, indicating that both groups benefitted from the remember-subset instruction to similar extents. While we would note that the younger adults had a significantly higher standardised verbal IQ, this might, if anything, have increased the likelihood of observing an interaction as the groups differed more than was originally anticipated. The absence of such an effect therefore supports the conclusion that strategy effects were unaffected by participant characteristics, and that, at least in the current paradigm, the superiority of the remember-subset instruction is reliable across age groups and different VWM capacities.

With regards to the questionnaire, both groups reported finding the remember-subset condition easier. In line with performance outcomes, the younger adults also judged that they performed better the remember-subset condition. In contrast, the older adults thought that their performance did not differ in the instruction conditions, despite exhibiting higher accuracy when told to focus on a subset of items. This suggests that the older adults were less aware of how the strategies affected their performance, in line with several previous studies suggesting that meta-memory decreases with age (Bruce, Coyne, & Botwinick, 1982; Bunnell, Baken, & Richards-Ward, 1999; Crumley, Stetler, & Horhota, 2014, but see Dunlosky & Hertzog, 1997; Halamish, McGillivray, & Castel, 2011; Rabinowitz, Ackerman, Craik, & Hinchley, 1982).

The apparent superiority of a remember-subset strategy contrasts with the conclusions drawn by Bengson and Luck (2015), who suggested that focusing on the entire array leads to optimal performance using unidimensional stimuli. This suggests that the effectiveness of encoding strategies may depend on the complexity of items and whether binding is required. When items are unidimensional, it may be easier and more effective to encode all the to-be-remembered visual information that is presented. In contrast, when items are more complex and require binding between features, focusing on a subset of items appears to be the optimal strategy.

However, there were several other differences between the current experiment and the Bengson and Luck (2015) study, which may potentially explain the differences in results found. Firstly, the display presented at retrieval differed between studies, with participants in the Bengson and Luck study presented with the entire array, whilst participants in the current study were presented with a single probe in the centre of the screen. Bengson and Luck suggested participants may have exhibited a higher VWM capacity in the remember-all condition as this encouraged formation of a holistic representation of the entire array, which could then be used to detect changes to the overall scene. In contrast, our single probe cued recall paradigm would prevent participants from using an ensemble representation to enhance performance. In this case, applying limited resources and/or capacity to a subset of items may become the more productive encoding strategy.

Another factor which differed between the studies was the retrieval method used. In the current experiment, participants were required to actively recall the colour of the shape, whereas in Bengson and Luck (2015), participants could rely on recognition to detect whether a change had occurred. Recall is thought to be more demanding than recognition (Craik & McDowd, 1997), which may have made it more difficult for participants to successfully remember all the items presented in the current study.

An understanding of whether these factors are important in determining strategy effects would provide further insight into the conditions in which each strategy is likely to be useful. This was examined in Experiment 2.

Experiment 2

Experiment 2 explored whether task factors, namely the retrieval method used (recall vs. recognition) and the display presented at retrieval (whole display vs single probe), which subsequently affects an individuals' ability to use an ensemble representation, are important in determining strategy effects. As in Experiment 1, participants were either told to remember all items or to focus on a subset. To explore whether retrieval method is an important factor in determining the directionality of strategy effects, a change detection recognition paradigm was used in Experiment 2, rather than the cued-recall method used in Experiment 1. If recognition supports a remember-all strategy (as in Bengson & Luck, 2015) while a more demanding recall process is better suited to remember-subset, then we might expect participants to perform better in the remember-all condition in this experiment.

In addition, display at retrieval was manipulated, with participants either responding based on a single probe (as in Experiment 1), or the whole array (as in Bengson & Luck, 2015). This allowed us to explore whether the ability to use an ensemble representation is likely to be an important factor, as this configural information would be useless in the single probe condition, but may be useful when the entire array is displayed. Utilisation of an ensemble representation would be more likely to emerge when the whole display is encoded, rather than a subset of the presented items. Therefore, if the ability to use an ensemble representation is indeed important in determining strategy effects, we would expect an interaction between strategy and test display, whereby participants exhibit better performance in the remember-subset condition when presented with a single probe, but better performance in the remember-all condition when presented with the whole array. Conversely, if neither the ability to use an ensemble representation nor the retrieval method are important factors, we might again expect participants to perform better in the remember-subset condition. Such a finding would replicate and extend the results from Experiment 1, providing further evidence that focusing on a subset of items is the optimal strategy to use when stimuli are more complex and feature binding is required.

Given that no significant interaction was found between instruction type and age group in the previous experiment, Experiment 2 focused on young adult participants. Moreover, given that no significant differences between instruction types were found at set size 3, Experiment 2 only assessed performance at set sizes 4 and 6.

Method

Participants. Twenty participants aged between 18-30 years took part (M. age = 20.50, SD = 3.17; 20 females; M. years of education = 15.10, SD = 1.74). Participants had no known learning difficulties, normal or corrected-to-normal vision and no colour blindness.

Materials. Visual working memory task. The visual memory task was run on a 13.3" MacBook Air using SuperCard (Version 4.5). Stimuli used were the same as Experiment 1. Test screens consisted of either one coloured shape displayed in the centre of the screen or the whole array, dependent upon the condition. In 50% of trials in each block, the test array was identical to the presentation array, requiring a 'same' response. In the other 50% of trials, the association between colour and shape of two of the items was switched, requiring a 'different' response.

Questionnaire. A similar questionnaire was administered to participants as in Experiment 1. This assessed task difficulty, adherence to strategies, how well people thought they performed and the number of shapes focused on in the remember-subset conditions. This questionnaire was separated into four sections, with one for each combination of instruction type and display conditions.

Design and Procedure. The study implemented a 2x2x2 mixed design, with instruction type (remember-all or remember-subset), display (single probe or whole display) and set size (4 or 6 shapes) as within-subject variables. The dependent variables were accuracy on the VWM task and reaction time.

The VWM task consisted of four blocks of 60 trials: one for each combination of the instruction type and retrieval display conditions (All-Whole, All-Single, Subset-Whole, Subset-Single). Order of the instruction conditions and the order of display blocks within the instruction conditions was counterbalanced. Within each block, there were 30 trials for each set size (4 and 6), which were randomly intermixed. At the start of each instruction block, participants completed four practice trials, two at each set size.

As with Experiment 1, each condition commenced with the provision of written instructions. The strategy instructions (i.e. remember all or remember subset) were identical to those given in Experiment 1. The experimental task is illustrated in Figure 1. Each trial began with the presentation of a randomly generated two-digit number (which participants were required to continuously repeat) at the centre of the screen for 2000ms. Next, a fixation cross appeared for 1000ms, followed by the display of coloured shapes (for 1000ms). As in Experiment 1, shapes were presented on a white background, at one of eight possible locations positioned at compass points around the screen centre. Following a 1000ms delay, a testing screen appeared, and participants had to indicate whether a change had occurred. In the single probe condition, one item was presented at the centre of the screen. In the whole display condition, the entire array was presented. Participants responded using a keyboard, pressing ‘z’ if the item(s) was/were the same, and pressing ‘/’ if the item(s) was/were different. Each instruction block was separated into two sections, each containing 30 trials, divided by a short break.

At the end of the experiment, participants completed the questionnaire. Participants also completed the STW task between the instruction blocks, to aid comparisons with Experiment 1.

Data analysis

Accuracy was determined by a corrected recognition score, calculated by subtracting false alarms (whereby participants responded ‘different’, but the correct answer was ‘same’) from hits (whereby participants responded ‘different’ and the correct answer was ‘different’). This outcome measure was selected, rather than a capacity measure, as the formulae for calculating VWM capacity in change detection tasks are either inappropriate for tasks involving binding (Cowan, 2001; Cowan, Blume, & Saults, 2012; Pashler, 1988), or can only currently be calculated for a limited number of set sizes (Cowan et al., 2012). Reaction time (RT) was also measured, with values above or below 2.5 standard deviations from the mean removed. As with Experiment 1, Bayes factor analysis was conducted using JASP and interpreted using the guidelines set out by Jeffreys (1961).

Results

Accuracy. Mean corrected rejection score (and SE) as a function of instruction type and display are displayed in Figure 3. A 2 (instruction type) x 2 (display) x 2 (set size) repeated measures ANOVA yielded a significant effect of instruction type, F(1, 19) = 5.58, MSE = 0.06, p = .029, $\eta^2 p$ =.23, with participants exhibiting a higher VWM capacity in the remember-subset condition (M = 0.39, SE = 0.03) relative to the remember-all condition (M = 0.30, SE = 0.04). A significant effect of display was also found, F(1, 19) = 12.00, MSE = 0.05, p = .003, $\eta^2 p$ = .39, with participants exhibiting a higher VWM capacity in the whole display condition (M = 0.40, SE = 0.03) relative to the single probe condition (M = .28, SE = .03). A significant effect of set size also emerged, F(1, 19) = 35.33, MSE = 0.03, p <.001, $\eta^2 p$

= .65, with participants performing better in the set size 4 condition (M = 0.42, SE = 0.03) than the set size 6 condition (M = .26, SE = 0.03). No significant interaction between instruction type and display was found, $F(1, 19) = 0.66$, MSE = 0.03, $p = .426$, $\eta^2 p = .03$. There were also no other significant interactions ($F \leq .44$, $p \geq .514$).

A 2 x 2 x 2 Bayesian repeated measures ANOVA was also conducted. The strongest model favoured the inclusion of main effects of instruction type, display and set size (BF > 1000, versus the intercept only model, ± 1.53%). The inclusion of main effects of instruction type (BF = 10.97), display (BF > 1000) and set size (BF = 419.65) were all strongly favoured. The model did not favour the inclusion of interactions between instruction type and display (BF = 0.32), instruction type and set size (BF = 0.26), set size and display (BF = 0.25) or instruction type, display and set size (BF = 0.34).

(Figure 3 about here)

RT. Mean RT (and SE) as a function of instruction type, display and set size are displayed in Table 2. A 2 (instruction type) x 2 (display) x 2 (set size) repeated measures ANOVA yielded a significant effect of set size, $F(1, 19) = 12.50$, MSE = 11221.04, $p = .002$, $\eta^2 p = .40$, with participants responding faster in the set size 4 condition (M = 909.65, SE = 37.89) than the set size 6 condition (M = 968.88, SE = 45.65). There were no significant main effects of instruction type (Remember-all M = 924.40, SE = 42.99; Remember-subset M = 954.13, SE = 46.24; $F(1, 19) = 0.73$, MSE = 48492.62, $p = .404$, $\eta^2 p = .04$) or display (Whole display M = 926.45, SE = 40.69; Single probe M = 952.08, SE = 44.24; $F(1, 19) = 1.41$, MSE = 18696.69, $p = .251$, $\eta^2 p = .07$), and no significant interactions ($F \leq .63$, $p \geq .436$).

A 2 x 2 x 2 Bayesian repeated measures ANOVA revealed that the strongest model included set size (BF = 10.07, versus the intercept only model, ± 1.87%). The inclusion of set size was strongly favoured (BF = 13.42). The inclusion of main effects of instruction type (BF = 0.60) and display (BF = 0.47), and the interactions between instruction type and display (BF = 0.19), instruction type and set size (BF = 0.20), set size and display (BF = 0.22), and instruction type, display and set size (BF = 0.33) was not favoured.

(Table 2 about here)

Questionnaire. Self-reported task difficulty and judgements of performance ratings are displayed in Table 3. A 2 (instruction type) x 2 (display) repeated measures ANOVA on the task difficulty ratings revealed a significant effect of instruction type, F(1, 19) = 4.42, MSE = 2.06, p = .049, η^2p = .19, with participants reporting the remember-subset condition (M = 2.80, SE = 0.17) was easier than the remember-all condition (M = 2.13, SE = 0.22). There was no significant effect of display (Whole display M = 2.60, SE = 0.17; Single probe M = 2.33, SE = 0.15; F(1, 19) = 1.72, MSE = 0.88, p = .206, η^2p = .08). There was also no significant interaction between instruction type and display (F(1, 19) = 0.66, MSE = 0.47, p = .425, η^2p = .03). A 2 x 2 Bayesian repeated-measures ANOVA revealed that the strongest model included a main effect of instruction type (BF = 10.92, versus the intercept only model, ± 1.08%). Inclusion of a main effect of instruction type was strongly supported (BF = 12.20). Inclusion of a main effect of display (BF = 0.45) and an interaction between instruction type and display (BF = 0.34) was not supported.

A 2 (instruction type) x 2 (display) repeated measures ANOVA on judgements of performance ratings revealed a significant effect of instruction type (F(1, 19) = 8.05, MSE = 0.97, p = .011, η^2p = .30), with participants suggesting that they performed better in the remember-subset condition (M = 2.88, SE = 0.12) relative to the remember-all condition (M = 2.25, SE = 0.19). No significant effect of display emerged (Whole display M = 2.58, SE = 0.13; Single probe M = 2.55, SE = 0.15; $F(1, 19) = 0.02$, MSE = 0.64, $p = .891$, $\eta^2 p = .001$). There was also no significant interaction between instruction type and display ($F(1, 19) = 1.51$, MSE = 0.21, $p = .234$, $\eta^2 p = .07$). A 2 x 2 Bayesian repeated-measures ANOVA was also conducted. This revealed that the strongest model included a main effect of instruction type (BF = 72.86, versus the intercept only model, ± 1.16%). Inclusion of a main effect of instruction type was strongly supported (BF = 72.56), whilst inclusion of a main effect of display (BF = 0.23) and an interaction between instruction type and display (BF = 0.38) was not supported.

With regard to how many shapes participants focused on in the remember-subset conditions, there was no significant difference between the whole display (M = 2.67, SE = 0.09) and single probe (M = 2.62, SE = 0.09) conditions, $t(19) = 0.46$, $p = .649$, BF = .26).

(Table 3 about here)

Discussion

Participants performed significantly better when they focused on a subset of items, compared to when they attempted to remember all the visual information presented. This was found despite no significant differences in RT between instruction type conditions, suggesting this effect is not due to a speed-accuracy trade off. This extends the instruction effect found in Experiment 1, providing further evidence that focusing on a subset of items is the optimal encoding strategy when objects are multidimensional and binding is required. It is also in line with the assumptions of several previous studies (Cusack et al., 2009; Linke et al., 2011), which have suggested that trying to remember a subset of items is likely to result in better performance beyond capacity limits.

Superiority of the remember-subset instruction was found despite a change in the retrieval method from cued recall (Experiment 1) to recognition (Experiment 2), suggesting that variations in the retrieval method used between Experiment 1 and Bengson and Luck (2015) cannot explain the differences in findings between these studies. Moreover, no significant interaction was found between instruction type and display, indicating that, at least in the current paradigm, the ability to use an ensemble representation is not an important factor in determining which strategy is most beneficial.

A significant effect of display was found, however, with participants exhibiting superior memory performance in the whole display condition. This differs from findings reported by Wheeler and Treisman (2002), who found that participants exhibited a higher VWM capacity when presented with single probe in trials where binding was required, but is in line with those reported by Johnson, Hollingworth, and Luck (2008), who also observed superior performance in whole display conditions.

Findings from the questionnaire are in line with Experiment 1, with participants reporting that they found the remember-subset condition easier. Participants also correctly judged that they performed better in the remember-subset condition. In contrast, however, participants suggested that their performance did not differ between display conditions, despite exhibiting higher accuracy when the whole display was presented. This suggests that individuals are aware of how strategies affect performance, though they may have less insight into the effects of other task factors (Koriat, Bjork, Sheffer, & Bar, 2004; Kornell & Bjork, 2009).

General Discussion

Previous research has suggested that participants can exhibit a higher VWM capacity by focusing on all items presented, as opposed to just a subset, even after capacity has been exceeded (Bengson & Luck, 2015). However, these findings were limited to memory for unidimensional stimuli (coloured squares), making it difficult to ascertain whether similar results would be found if more complex items were used that require feature binding. It is plausible to predict that a different pattern of results might be found under these circumstances, as the increased complexity of items and the novel feature binding element may make it more difficult for individuals to effectively encode all the visual information in a display. The present experiments therefore examined the effects of encoding strategies in VWM tasks using multidimensional items, varying in both shape and colour. To respond accurately, participants had to bind these features together to form object-based representations of the items. We examined the effectiveness of remember-all vs. remember-subset encoding strategies on this ability across different age groups and task conditions.

In Experiment 1, the effects of strategy were explored in younger and older adults using a cued-recall task, which has previously been shown to be sensitive to the effects of other forms of encoding strategy (Hu et al., 2014, 2016). In Experiment 2, a change detection recognition paradigm was employed to examine whether retrieval method is an important factor in determining effects of strategy. The display presented at retrieval was also manipulated in Experiment 2, to explore whether the ability to use an ensemble representation affects the relationship between strategy use and performance on VWM tasks. In both experiments, participants performed significantly better when explicitly told to focus on a subset of items when they found the task difficult, compared to when they were told to focus on all the to-be-remembered items. This was found across age groups, retrieval methods and display types.

These findings are in line with the assumptions of several previous studies, which have suggested that trying to remember a subset of items is likely to lead to optimal performance once capacity has been exceeded (Cusack et al., 2009; Linke et al., 2011). Although these outcomes differ from the conclusions drawn by Bengson and Luck (2015), we do not consider these findings to be inconsistent, as Bengson and Luck explored effects of strategy on tasks requiring memory for simple, unidimensional items. As such, it is possible the effect of strategy depends on the complexity of items and whether binding is required.

There are, however, other differences between the current experiments and the study conducted by Bengson and Luck (2015), which may provide alternative explanations for the differences in findings. In the current study, presentation time was substantially longer (1000ms vs 100ms in the Bengson and Luck design), though, if anything, this might have made it easier to effectively encode all the visual information, as it would have increased viewing time per item (Donkin et al., 2016). Alternatively, the differences in findings may reflect use of articulatory suppression in our studies, a technique not employed by Bengson and Luck (2015). However, this is also unlikely, given the brevity of their presentation duration, together with evidence that articulatory suppression does not significantly influence performance in VWM tasks (Morey & Cowan, 2005; Hardman & Cowan, 2015).

The remember-subset advantage observed in the current study can be likened to the outcomes reported by Hu et al. (2014, 2016), who found that memory for an item from within a short sequence was enhanced if participants were instructed to prioritise that item over others during encoding. These endogenously-driven prioritisation effects appear to be executive-dependent (Hu et al., 2016), and may reflect active storage in an accessible and privileged state within the focus of attention, potentially within the episodic buffer (Hu et al., 2014). The remember-subset advantage observed in the current study represents a different form of internally-motivated item selection, though it may similarly reflect attended items entering a privileged and more accessible state. However, it is possible the subset advantage observed here does not rely on executive resources to the same extent as the prioritisation effect found by Hu et al. Supporting this, the older adults in Experiment 1 benefitted from the remember subset strategy to a similar extent as the younger adults, despite this group typically exhibiting reduced executive function (Kirova, Bays, & Lagalwar, 2015).

Alternatively, the remember-subset advantage may have emerged due to intrinsic drawbacks that accompany the remember-all strategy. Focusing on all items would mean participants only had a short viewing time for each item, thus potentially resulting in the generation of imprecise representations (Donkin et al., 2016). Furthermore, attempting to remember all items may have resulted in an overload of VWM on some trials (Gathercole, 2008), making it difficult for participants to effectively maintain or retrieve the visual information effectively. Given that the cognitive mechanisms underlying such effects are currently unclear, it would be beneficial for future work to explore this further.

Nevertheless, these findings have important outcomes, adding to a growing body of literature suggesting that strategy use is an important factor in determining VWM capacity (Bengson & Luck, 2015; Logie, 2011; Morrison, Rosenbaum, Fair, & Chein, 2016). Researchers investigating working memory should be aware of these strategy effects as differences in spontaneous use between participants or across conditions may confound results (Bengson & Luck, 2015). In order to minimise such effects of strategy, Bengson and Luck (2015) suggest researchers should provide neutral task instructions. However, neutral instructions would allow participants to decide which strategy to employ, which may then vary between participants or across conditions (Donkin et al., 2016). Instead, one suggestion might be that researchers should provide specific task instructions in order to reduce or control strategy effects.

Younger adults correctly judged that they performed better in the remember-subset condition in both experiments, though they did not successfully judge how differences in the display presented at retrieval affected performance (Experiment 2). Taken together, this suggests that young adults are able to assess the effectiveness of strategies, but may be less able to judge how task factors affect performance (Koriat et al., 2004; Kornell & Bjork, 2009). This might reflect the level of control one has over these factors, as individuals can generally adjust encoding strategies but not task features. In contrast, older adults were not able to judge how the strategies affected their performance. This is in line with several previous studies suggesting that subjective measures of memory become less accurate with age and are poor in older adults (Bruce et al., 1982; Bunnell et al., 1999; Crumley et al., 2014, but see Dunlosky & Hertzog, 1997; Halamish et al., 2011; Rabinowitz et al., 1982). This may also explain why older adults use encoding strategies less frequently (Devolder & Pressley, 1992; Zacks, 2011); if they experience a lack of internal feedback regarding the effectiveness of strategies, they may be less likely to apply them.

In summary, the experiments presented here provide evidence that strategy use yields small but reliable effects on VWM capacity. In tasks using items comprised of multiple features that must be accurately bound together, focusing on a subset of items consistently results in better performance than trying to remember all the visual information, regardless of age group, retrieval type and test display.

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), 298-313. doi: 10.1037/0096-3445.135.2.298

Baddeley, A. D. (1986). Working memory. Oxford: Oxford University Press.

Baddeley A., Emslie H and Nimmo-Smith I. (1992). The Speed and Capacity of Language - Processing Test. Bury St Edmunds: Thames Valley Test Company.

Barchard, K. A. (2015). Null Hypothesis Significance Testing Does Not Show Equivalence. Analyses of Social Issues and Public Policy, 15(1), 418-421. doi: 10.1111/asap.12095

Bengson, J. J., & Luck, S. J. (2016). Effects of strategy on visual working memory capacity. Psychonomic Bulletin & Review, 23(1), 265-270. doi: 10.3758/s13423-015-0891-7

Blacker, K. J., Curby, K. M., Klobusicky, E., & Chein, J. M. (2014). Effects of action video game training on visual working memory. Journal of Experimental Psychology: Human Perception and Performance, 40(5), 1992-2004. doi: 10.1037/a0037556

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). doi: 10.3389/fpsyg.2013.00012

Brockmole, J. R., Parra, M. A., Della Sala, S., & Logie, R. H. (2008). Do binding deficits account for age-related decline in visual working memory? Psychonomic Bulletin & Review, 15(3), 543-547. doi: 10.3758/PBR.15.3.543

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

Bruce, P. R., Coyne, A. C., & Botwinick, J. (1982). Adult age differences in metamemory. *Journal of Gerontology*, *37*(3), 354-357. doi: 10.1093/geronj/37.3.354

Bunnell, J. K., Baken, D. M., & Richards-Ward, L. A. (1999). The effect of age on metamemory for working memory. *New Zealand Journal of Psychology*, *28*(1), 23-29.

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

Chen, Z., & Cowan, N. (2013). Working memory inefficiency: Minimal information is utilized in visual recognition tasks. Journal of Experimental Psychology: Learning, Memory, and Cognition, 39(5), 1449-1462. doi: 10.1037/a0031790

Craik, F. I., & McDowd, J. M. (1987). Age differences in recall and recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13(3), 474-479. Doi: 10.1037/0278-7393.13.3.474

Crumley, J. J., Stetler, C. A., & Horhota, M. (2014). Examining the relationship between subjective and objective memory performance in older adults: A meta-analysis. *Psychology and Aging, 29*(2), 250-265. doi: 10.1037/a0035908

Cusack, R., Lehmann, M., Veldsman, M., & Mitchell, D. J. (2009). Encoding strategy and not visual working memory capacity correlates with intelligence. Psychonomic Bulletin & Review, 16(4), 641-647. doi: 10.3758/PBR.16.4.641

Devolder, P. A., & Pressley, M. (1992). Causal attributions and strategy use in relation to memory performance differences in younger and older adults. *Applied Cognitive Psychology*, *6*(7), 629-642. doi: 1 0.1002/acp.2350060706

Donkin, C., Kary, A., Tahir, F., & Taylor, R. (2016). Resources masquerading as slots: Flexible allocation of visual working memory. Cognitive Psychology, 85, 30-42. Doi: 10.1016/j.cogpsych.2016.01.002

Dunning, D. L., & Holmes, J. (2014). Does working memory training promote the use of strategies on untrained working memory tasks? Memory & Cognition, 42(6), 854-862. doi: 10.3758/s13421-014-0410-5

Dunlosky, J., & Hertzog, C. (1997). Older and younger adults use a functionally identical algorithm to select items for restudy during multitrial learning. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *52*(4), P178-P186. doi: 10.1093/geronb/52B.4.P178

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research, 12(3), 189-198.

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

Gathercole, S. E. (2008). Working memory in the classroom. Psychologist, 21, 382-385.

Gold, J. M., Hahn, B., Zhang, W. W., Robinson, B. M., Kappenman, E. S., Beck, V. M., & Luck, S. J. (2010). Reduced capacity but spared precision and maintenance of working memory representations in schizophrenia. Archives of General Psychiatry, 67(6), 570-577. doi: 10.1001/archgenpsychiatry.2010.65

Hardman, K. O., & Cowan, N. (2015). Remembering complex objects in visual working memory: Do capacity limits restrict objects or features? Journal of Experimental Psychology: Learning, Memory, and Cognition, 41(2), 325-347. doi: 10.1037/xlm0000031

Halamish, V., McGillivray, S., & Castel, A. D. (2011). Monitoring one's own forgetting in younger and older adults. *Psychology and Aging*, *26*(3), 631-635. doi: 10.1037/a0022852

Hartshorne, J. K. (2008). Visual working memory capacity and proactive interference. PLoS ONE, 3(7): e2716. doi: 10.1371/journal.pone.0002716

Hu, Y., Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2016). Executive control of stimulus-driven and goal-directed attention in visual working memory. *Attention, Perception, & Psychophysics*, 1-12. doi: 10.3758/s13414-016-1106-7

Hu, Y., Hitch, G. J., Baddeley, A. D., Zhang, M., & Allen, R. J. (2014). Executive and perceptual attention play different roles in visual working memory: Evidence from suffix and strategy effects. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1665-1678. doi: 10.1037/a0037163

Jeffreys, H. (1961). Theory of probability (3rd edition). New York: Oxford University Press.

Johnson, J. S., Hollingworth, A., & Luck, S. J. (2008). The role of attention in the maintenance of feature bindings in visual short-term memory. Journal of Experimental Psychology: Human Perception and Performance, 34(1), 41-55. doi: 10.1037/0096-1523.34.1.41

Kirova, A. M., Bays, R. B., & Lagalwar, S. (2015). Working memory and executive function decline across normal aging, mild cognitive impairment, and Alzheimer's disease. BioMed Research International, 2015: 748212. doi: 10.1155/2015/748212

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: the role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*(4), 643-656. doi: 10.1037/0096-3445.133.4.643

Kornell, N., & Bjork, R. A. (2009). A stability bias in human memory: overestimating remembering and underestimating learning. *Journal of Experimental Psychology: General*, *138*(4), 449-468. doi: 10.1037/a0017350

Linke, A. C., Vicente-Grabovetsky, A., Mitchell, D. J., & Cusack, R. (2011). Encoding strategy accounts for individual differences in change detection measures of

VSTM. *Neuropsychologia*, 49(6), 1476-1486. doi: 10.1016/j.neuropsychologia.2010.11.034

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. Nature, 390(6657), 279-281. doi:10.1038/36846

Logie, R. H. (2011). The functional organization and capacity limits of working memory. Current Directions in Psychological Science, 20(4), 240-245. doi: 10.1177/0963721411415340

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), 703-713. doi: 10.1037/0278-7393.31.4.703

Morrison, A. B., Rosenbaum, G. M., Fair, D., & Chein, J. M. (2016). Variation in strategy use across measures of verbal working memory. *Memory & Cognition*, 1-15. doi: 10.3758/s13421-016-0608-9

Mulder, J., & Wagenmakers, E. J. (2016). Editors' introduction to the special issue "Bayes factors for testing hypotheses in psychological research: Practical relevance and new developments". *Journal of Mathematical Psychology*, *72*, 1-5. doi: 10.1016/j.jmp.2016.01.002.

Opitz, B., Schneiders, J. A., Krick, C. M., & Mecklinger, A. (2014). Selective transfer of visual working memory training on Chinese character learning. Neuropsychologia, 53, 1-11. doi: 10.1016/j.neuropsychologia.2013.10.017

Pertzov, Y., Avidan, G., & Zohary, E. (2009). Accumulation of visual information across multiple fixations. Journal of Vision, 9(2). doi:10.1167/9.10.2

Rabinowitz, J. C., Ackerman, B. P., Craik, F. I., & Hinchley, J. L. (1982). Aging and metamemory: The roles of relatedness and imagery. *Journal of Gerontology*, *37*(6), 688-695. doi: 10.1093/geronj/37.6.688

Sense, F., Morey, C. C., Prince, M., Heathcote, A., & Morey, R. D. (2016). Opportunity for verbalization does not improve visual change detection performance: A state-trace analysis. Behavior Research Methods, 1-10. doi:10.3758/s13428-016-0741-1

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

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. Journal of Experimental Psychology: General, 131(1), 48-64. doi: 10.1037/0096-3445.131.1.48

Zacks, R. T. (1982). Encoding Strategies Used by Young and Elderly Adults in a Keeping Track Task. *Journal of Gerontology*, *37*(2), 203-211. doi: 10.1093/geronj/37.2.203

Figure captions

Figure 1. The experimental paradigm used (with a 4-item trial as an illustrative example). In Experiment 1, participants were presented with an outline of a shape at retrieval and asked to recall the colour. In Experiment 2, participants had to indicate whether a change in colour-shape combination had occurred. Figure not to scale.

Figure 2. Mean VWM capacity (K) and SE in Experiment 1, as a function of instruction type and set size for the younger adults (A) and older adults (B).

Figure 3. Mean corrected recognition score and SE in Experiment 2, as a function of instruction type and display at set size 4 (A) and set size 6 (B).

Table 1

Table 1: Mean (and SE) task difficulty ratings (1 = very difficult, 5 = very easy) and judgements of performance (1 = very poor, 5 = very good) in Experiment 1, as a function of instruction type and age group.

	Difficulty – All	Difficulty Subset	Performance – All	Performance - Subset
Younger	1.75 (0.14)	2.65 (0.18)	2.20 (0.16)	2.90 (0.18)
Older	1.65 (0.13)	2.00 (0.15)	1.95 (0.17)	1.95 (0.14)

Table 2

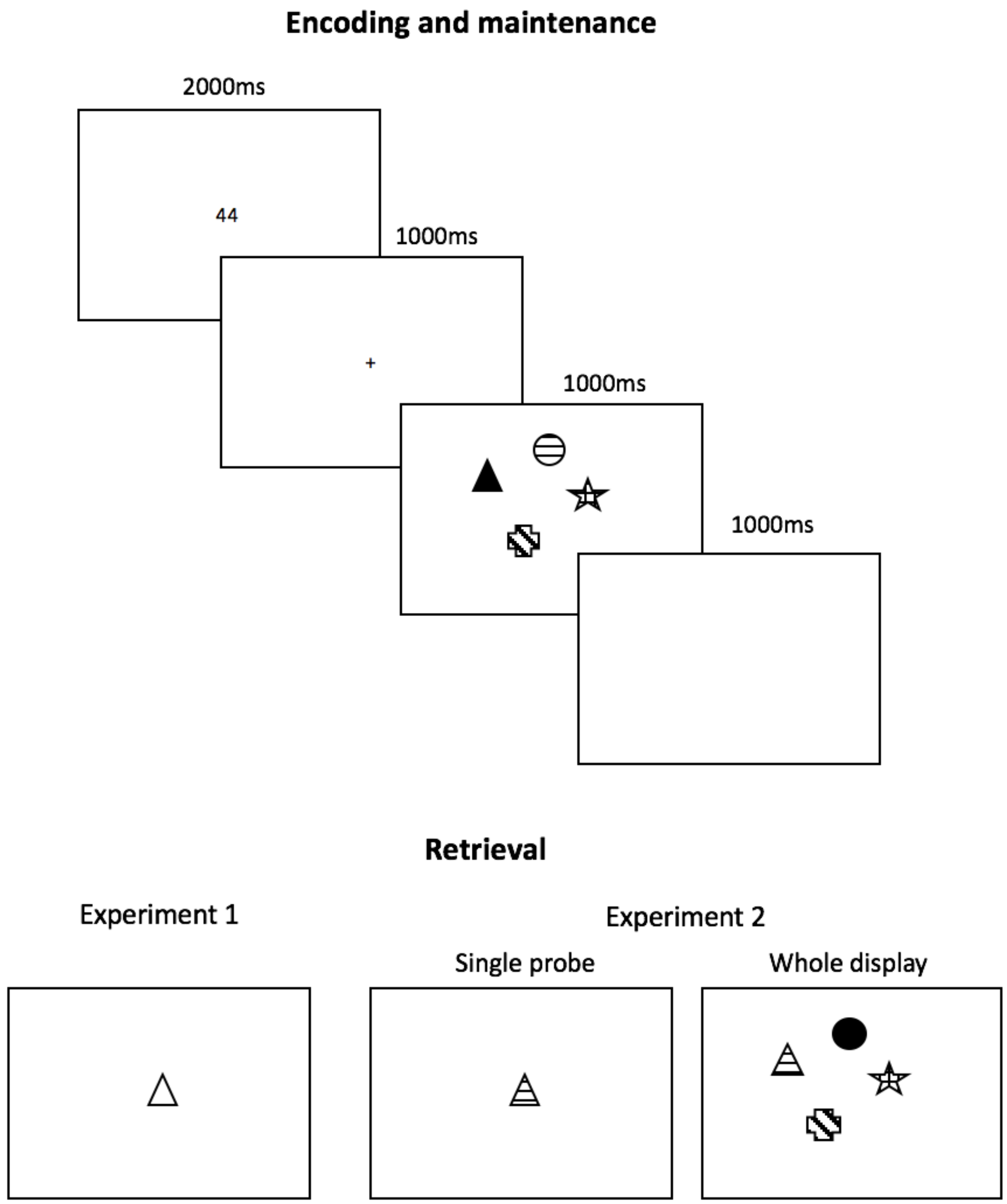
Table 2: Mean reaction time (RT) and SE in Experiment 2, as a function of instruction type, display and set size.

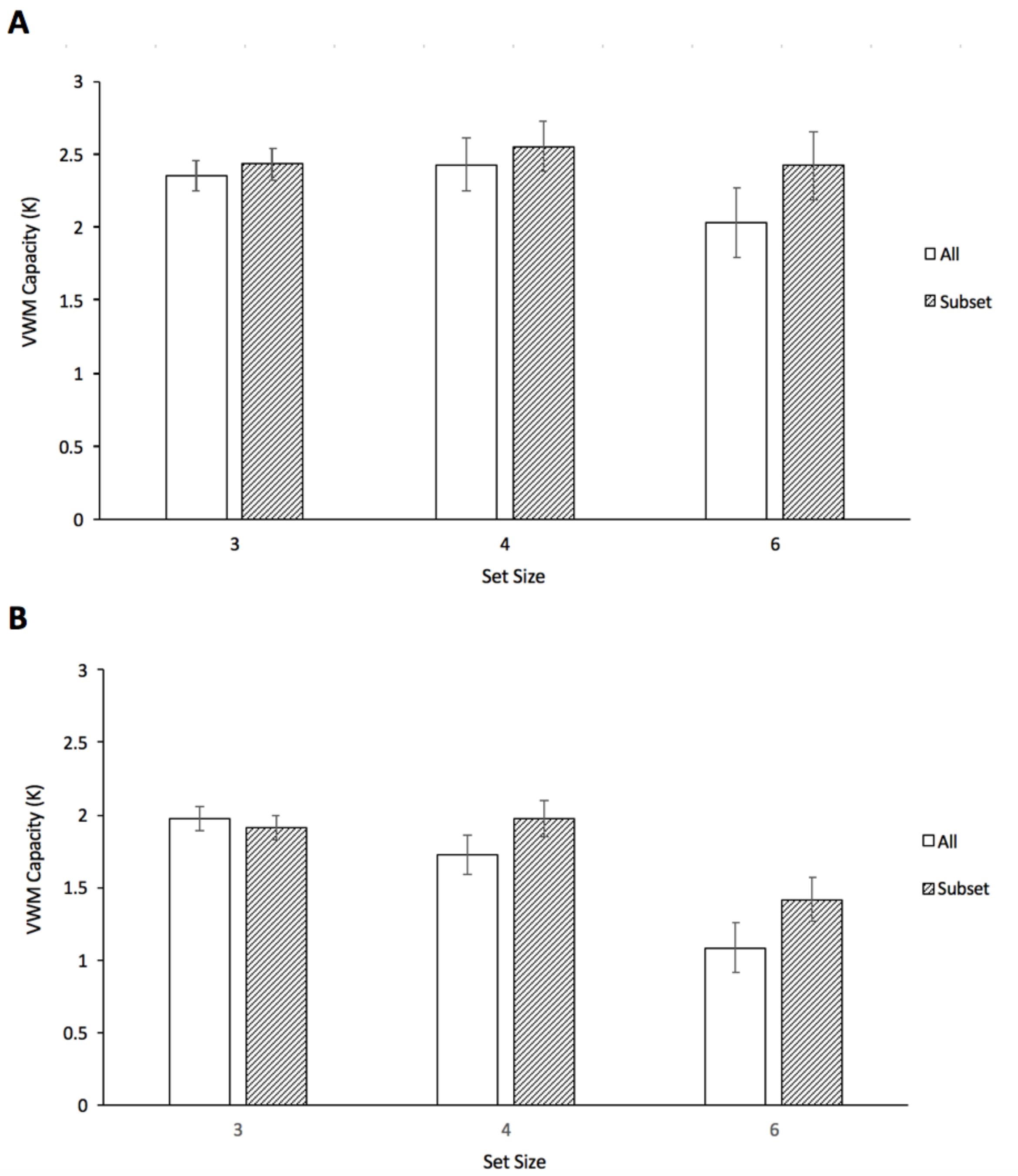
	All-Whole	All-Single	Subset-Whole	Subset-Single
Set size 4	872.15 (41.70)	918.50 (42.68)	909.75 (46.91)	938.20 (43.88)
Set size 6	948.40 (47.11)	958.55 (56.55)	975.50 (49.46)	993.05 (59.16)

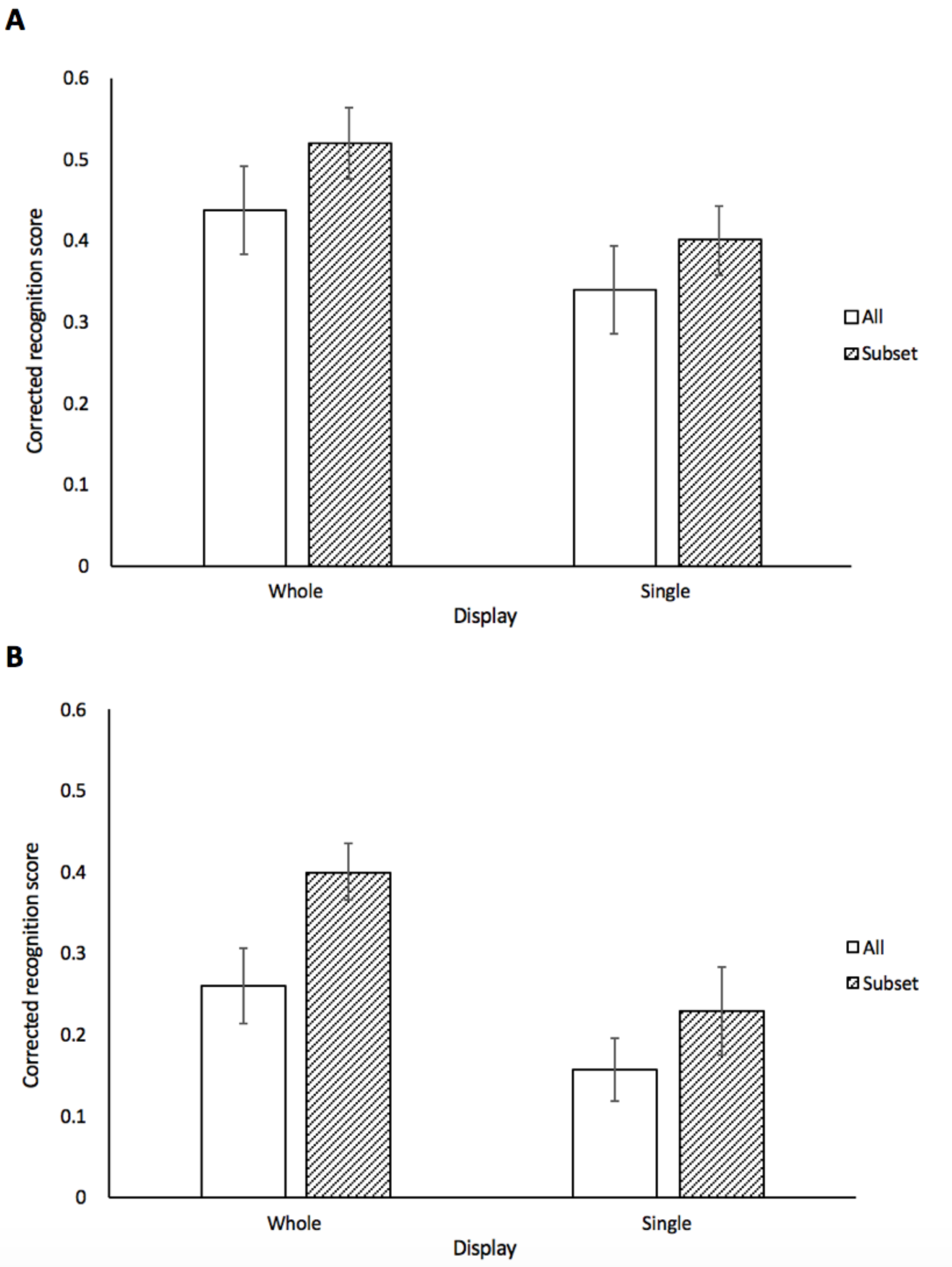
Table 3

Table 3: Mean (and SE) ratings of task difficulty (1 = very difficult, 5 = very easy) and judgements of performance (1 = very poor, 5 = very good) in Experiment 2, as a function of instruction type and display.

	Difficulty - All	Difficulty - Subset	Performance - All	Performance - Subset
Whole	2.20 (0.29)	3.00 (0.19)	2.20 (0.21)	2.95 (0.14)
Single	2.05 (0.22)	2.60 (0.23)	2.30 (0.21)	2.80 (0.19)


Encoding and maintenance
2000ms
44
1000ms
+
1000ms
1000ms
Retrieval
Experiment 1
Experiment 2
Single probe
Whole display


A
3
2.5
2
1.5
1
0.5
0
VWM Capacity (K)
3
4
6
Set Size
All
Subset
B
3
2.5
2
1.5
1
0.5
0
VWM Capacity (K)
3
4
6
Set Size
All
Subset


A
Corrected recognition score
0.6
0.5
0.4
0.3
0.2
0.1
0
Whole
Single
Display
All
Subset
B
Corrected recognition score
0.6
0.5
0.4
0.3
0.2
0.1
0
Whole
Single
Display
All
Subset