

The relation between the capacities of imagination and visual memory in the short-term

Christopher Atkin^{1*}, Christina J. Howard¹, Thom Baguley¹, Joshua Baker² & Duncan Guest¹

¹NTU Psychology, Nottingham Trent University, Nottingham, UK

²Department of Psychology, University of Essex, Colchester, UK

* Correspondence to Dr Christopher Atkin; christopher.atkin02@ntu.ac.uk

Authors note: There are no conflicts of interest to disclose. Materials, code, and data are shared (Experiment 1 – 5) and Experiments 2 – 5 were pre-registered (please see individual Open Science Framework website links at the beginning of each Experiment).

Word count: 20,945

Abstract

Visual imagery and short-term memory utilise similar brain networks, but the extent to which they are related remains unclear. Here we explore whether the capacity of visual imagery (as yet unknown) is similar to the known capacity limits of visual working memory (VWM) and visual short-term memory (VSTM). Experiment 1 explored capacity limits in imagination, VWM and VSTM using a novel paradigm that for the first time provided estimates of capacity across these tasks. Imagination capacity was lower than that of VWM and VSTM. Experiments 2-4 eliminated alternative explanations of this capacity difference. Manipulating the time available to generate, update and maintain an image (imagination task) or encode, update and maintain an image (VWM task) did not influence performance in either task (Experiment 2). Manipulating the cue location and the size of the cued area had no specific influence on the imagination task (Experiment 3). Changing the test display (Experiment 4) showed that presenting all items at test (configural information) benefitted VSTM performance, presenting a single item benefitted VWM performance and manipulating test display had no impact on imagination performance. In Experiment 5, increasing object complexity eliminated the difference between VSTM and imagination capacity, however VWM capacity remained higher than that of imagery. For the first time, these experiments using analogous tasks demonstrate a difference in the observed capacities of VWM and imagery and provide the first measurable indication of the extent to which top-down (imagery), versus bottom-up activation of sensory systems (memory) supports the representation of perceptual stimuli.

Significance statement

To date, capacity as indicated by the number of items people can imagine at any time has been an unknown. This contrasts with research into visual short-term memory, where capacity limitations have long been researched. This article establishes that the observed capacity for imagination is lower than that of memory in the short-term. Given that imagery differs from memory by relying solely on top-down activation of the visual system, this indicates for the first time the extent of the contribution of sensory information to supporting the capacity of memory in the short-term.

Keywords: visual imagery, visual working memory, visual short-term memory, computational modelling, capacity estimates.

Introduction

Imagination is a fundamental aspect of cognition, enabling us to traverse time and space and think about concepts and contexts in different ways. As a central part of our everyday experience, it is important to understand the cognitive and neurological processes that support it. In the visual domain the extent to which visual memory, visual imagery and visual perception are distinct, overlapping or involve some common processes is a fundamental question about the nature of visual cognition. Despite this, somewhat surprisingly, the literatures on imagery and visual memory have evolved relatively independently and this issue has increasingly become a focus of recent research (Dijkstra et al., 2019).

In a recent review, Pearson (2019) proposed a model in which visual imagery and visual working memory (VWM) are seen as highly related. In support of this model is evidence from self-report measures where participants say they use mental imagery during retention to support performance in VWM tasks (e.g., Berger & Gaunitz, 1979) as well as evidence that has shown a relationship in the neural activation observed in imagery and VWM tasks (e.g., Albers et al., 2013). More recently Keogh and Pearson (2017) have shown that imagery has capacity limitations, with the precision of imagery decreasing with the number of to be imagined items, much in the way that capacity limitations are observed in VWM (Bays & Hussan, 2008). However, to date no one has attempted to measure the capacity of imagination and compare this to that of visual working memory. This is an important step in understanding the relationship between imagery and VWM. Should the observed capacities be similar, it would provide evidence of a significant overlap in the underlying processes supporting both functions. Should they differ, it would suggest that in imagery there are additional or different processes that impact how information is processed, maintained and utilised. Importantly, imagery is also posited to be involved in a wide range of cognitive processes, from prospective memory, to navigation, motor control and moral decisions (see Pearson, 2019 for a review). The extent to which imagery is capacity limited is important in understanding the extent to which it may support these cognitive tasks. The present work therefore develops a paradigm that for the first time allows estimation of the capacity of both imagination and VWM in terms of the number of items they can hold within this task in order to assess the extent to which their capacities are similar.

Unlike visual perception, visual imagery and visual memory can be conceptualised as visual representation in the absence of current corresponding visual input. The extent to which imagery is depictive and relies upon the same processes as visual perception (Kosslyn et al.,

2006; Kosslyn, 1994) or is based on propositional knowledge of what things would look like in an imagined situation (Pylyshyn, 2002) has been the focus of much debate. To address this issue, recent work has explored the extent of overlap in neural activation between perception and imagery at different levels of processing. Typically, these studies involve memorising a set of images that are then either presented again or participants are asked to imagine them. This has shown that there is some overlap in activation in the visual cortex during imagery and visual perception (Klein et al., 2004; Pearson, 2019; Pearson et al., 2015; Slotnick et al., 2005). For example, Klein et al. (2004) asked participants to view or imagine simple flashing stimuli that were either oriented horizontally or vertically. At lower visual areas they found evidence for some retinotopically organised activation for imagined stimuli, with activation patterns during imagination clearly differentiated for horizontal and vertical stimuli and this activation matching closely that seen in perception. Furthermore, the impact on low-level visual processing has also been found in behavioural studies using simple stimuli (Keogh & Pearson, 2017; Pearson et al., 2008). For example, Pearson et al. (2008) demonstrated that both seeing and imagining gratings with different colours and orientations could influence which stimulus was then seen in a binocular rivalry task.

Studies using more complex stimuli have also shown considerable overlap between imagery and perception in higher visual areas. Johnson and Johnson (2014) got participants to view or imagine previously memorised natural scenes and found that throughout scene selective brain areas, imagery partially recreated the same patterns of item-specific neural activity that were produced during perception. Notably, using multivariate pattern analysis (MVPA) a classifier was less able to distinguish between activation patterns from perception and imagery at higher levels such as the retrosplenial cortex than at lower visual areas. This suggests that greater abstraction of scene information is happening at higher levels which makes activation less distinguishable between perception and imagery (Johnson & Johnson, 2014). Similarly, Lee et al. (2012) examined activation patterns when participants viewed or imagined object images they had previously been trained on. They showed similarities in activation for objects between imagery and perception throughout the ventral processing stream. In particular they showed that activation patterns could be used to decode which object (from a set of 10) was being either perceived or imagined, albeit with weaker decoding power for imagery. Activation patterns from perception could also be used to discriminate which object was being imagined. There were, however, clear differences in activation patterns in imagery and perception, indicating differences in the type of processing in these two tasks. Importantly perception involves bottom-up input that is later modulated by top-down processes

whereas imagery has no bottom-up input and so early sensory regions receive top-down signals only making it likely that those signals are impoverished relative to those elicited by the actual object image. A recent review of the literature (Dijkstra et al., 2019) concluded that perception and imagery share a variety of neural mechanisms throughout the ventral stream from V1 onwards, with more overlap in higher levels. There is therefore substantial evidence that imagery simulates perception to some degree.

Given that imagery appears to involve creating visual representations that resemble perception, how does it relate to VWM? Visual imagery is similar to VWM in that both require active representation of, and manipulation of, visual information. VWM also involves the activation and manipulation of long-term memories (Cowan, 2019; Baddeley, 2000) and although imagery is similar in this regard, it is also associated with using long term memory to generate novel images (e.g. a pink elephant wearing a top hat standing on a sun lounger). Despite their clear conceptual similarities, imagery and VWM have largely been investigated as separate literatures, with a much firmer understanding of how they relate required (Tong, 2013). Behavioural studies exploring the relation between visual imagery and VWM have indicated clear similarities. For example, visual interference has been found to attenuate performance in both imagery and VWM tasks (Baddeley & Andrade, 2000; Darling et al., 2009; Dean et al., 2005; Dean et al., 2008; Keogh & Pearson, 2011, 2014; McConnell & Quinn, 2004; Sherwood & Pearson, 2010; Valenti & Galera, 2020; Vasques et al., 2016). However, irrelevant visual input has been shown to disrupt performance in imagery tasks but not during VWM tasks (Andrade et al, 2002, although see Borst et al, 2012a). As such it is not clear if imagery and VWM tap the same or separate underlying processes. It is recognised that imagery appears to be a more active process compared to short-term storage (Logie, 1995; Pearson, 2001). While both may require a representation to be held in some form of visual cache (Borst et al, 2012a), Andrade et al. (2002) suggested that imagery generation requires an active visual representation store, likened to an imagery buffer, whereas the retention of visual stimuli in VWM is maintained in a passive store which makes it less susceptible to irrelevant visual input. This may explain why irrelevant visual input interferes with image generation but does not disrupt image retention (Borst et al., 2012b).

In terms of neural activation, both visual imagery and VWM appear to involve lower and higher-level visual areas (Harrison & Tong, 2009; Kosslyn & Thompson, 2003; Kosslyn et al., 1995). For example, studies using transcranial magnetic stimulation (TMS) have found that disrupting early visual areas can impair performance in both imagery and VWM tasks

(Cattaneo et al., 2012; van de Ven & Sack, 2013). In an fMRI study Albers et al. (2013) used multivariate pattern analysis to measure and compare activity in the early visual areas (V1-V3) in visual perception, visual imagery and visual working memory tasks. In their task participants were presented with oriented gratings and either retained these or rotated them in their minds eye. Activation patterns were similar in imagination and VWM conditions. Moreover, a decoder trained on activation during perception could reliably predict which stimuli were being imagined or held in VWM, pointing towards some similarity in neural activity between these three processes. Evidence from behavioural tasks and tasks measuring neural activation therefore point to a strong relationship between imagery and VWM.

As outlined above, Pearson (2019) has argued that there is substantial evidence for a relationship between VWM and visual imagery, with an important line of evidence being that the tasks have similar capacity functions, that is, they appear similarly impacted by the increase in set-size. What is not clear however, is whether if VWM and visual imagery are underpinned by similar processes, they exhibit similar capacity limits. Many researchers have suggested a capacity for VWM and visual short-term memory of around 4 (± 1) items (Cowan, 2001; Luck & Vogel, 1997; Zhang & Luck, 2008). Others, however, have instead characterised VWM as a resource that can be flexibly distributed across items (see Bays, Catalao & Husain, 2009; Bays & Husain, 2008) leading to more precise representation of items when fewer are held in VWM with a decrease in fidelity as set-size increases. Relatedly the same appears to be true for perceptual representations of attended objects (see Alvarez & Franconeri, 2005; Howard & Holcombe, 2008) in the absence of the requirement to maintain representations in memory. As discussed by Brady et al. (2011), models of working memory capacity must be able to explain capacity both in terms of how many objects appear to be stored and understanding the content of such memory representations, such as the number, quality and characteristics of encoded features, and how these features may be hierarchically stored in object representations (Brady et al., 2011).

In comparison to the visual memory literature, there has been relatively little work examining capacity limits in imagination. Keogh and Pearson (2017) investigated visual imagery capacity limitations by measuring the impact of imagery on a subsequent binocular rivalry task. Representations in imagination biased perception in the subsequent rivalry task, with the degree of biased dominance serving as a marker of sensory strength of the imagined object. Priming strength from imagery decreased as the the number of objects to be imagined increased and as they increased in complexity. This demonstrates that imagination is limited

in capacity, and the observed reduction in precision of imagery as load increased has clear parallels to findings with visual memory (Bays & Husain, 2008). Whilst providing important evidence of capacity limitations in imagery, Keogh and Pearson's (2017) task did not probe what was actually being imagined. As it measured the impact imagery had on a subsequent task this does not enable direct estimation of the observed capacity of visual imagery.

Recently, Ceja and Franconeri (2023) examined capacity limits in a visual imagery task in which objects that were initially presented then had to be mentally translated in space. Set-size was manipulated as well as whether the objects were heterogeneous or homogeneous in colour and three types of spatial translation were assessed. They found subjective difficulty of the task increased with larger set-size and when the task required representing objects with different unique colours as well as varying for different types of spatial transformations. Set-size effects were also observed using an objective measure of change detection, but there were only minimal effects of varying spatial transformation (in this case, rotation) on performance. They concluded that capacity limitations in visual imagery appear sensitive to factors similar to those that impact VWM. Balaban and Ullman (2024) have also investigated the capacity for imagining moving items in a task involving imagining the continued motion of animated objects. Their data were best fit by a model in which only motion of a single item could be imagined at any one time, suggesting a very low capacity. Importantly, both the Balaban and Ullman (2024) and the Ceja and Franconeri (2023) study looked at capacity limitations in terms of imagery being used to translate a previously shown image. As such imagination was used not to generate a new image but to use the minds-eye to alter an existing one. Whilst showing our ability to do this is limited in capacity, it is important to note that these tasks may involve a significant number of additional or different processes compared to imagining static objects.

The extant literature on capacity limitations in imagery therefore show that imagery is a capacity limited system that shows some commonalities with VWM. However, to date no work has yet tried to provide a direct measurement of the capacity of imagery within a given task that can be compared to the capacity of VWM in an analogous task. Here, we present a paradigm that enables direct measurement of capacity in imagery and short-term visual memory and thus allows for direct comparison of capacity limitations between imagery and visual memory tasks. Note we use the term short-term visual memory simply because our paradigm explores this in several ways that we distinguish below. Whilst previous approaches to studying the similarities between imagery, perception and short-term visual memory have tended to present stimuli that have to be memorised (Johnson & Johnson, 2014; Lee et al.,

2012) or manipulated (Albers et al., 2013) an issue with this is that what is being measured during imagery is influenced both by sensory and mnemonic processing. Critically, the shared influence of low-level sensory stimulation from imagery and memory representations of specific images may have contaminated attempts to assess similarities between perception, attention and imagery. Similarly in the investigations of imagery capacity limitations, Balahan and Ullman (2024) and Ceja and Franconeri (2023) also present images and then ask for them to be manipulated in imagination. Even the Keogh and Pearson (2017) task, which went to lengths to avoid such contamination, used placeholders to cue imagery that provided stimulus specific information (colour and orientation). In the paradigm presented here, a purer imagery task is used; one where neither perceptual input nor its subsequent memory trace can influence performance. This allows for accurate estimates of imagery capacity. Moreover, short-term visual memory versions of the task allow for direct comparisons of observed capacity estimates between visual imagery and short-term visual memory, providing one of the strongest tests of the relation between these cognitive processes to date.

In our paradigm participants are shown a grid within which either a subset of squares (the memoranda) visibly turned black during encoding (for memory tasks), or in which cues outside of the grid indicated which squares to imagine turning black (for the imagery task). Critically, although the representations in imagination and in memory were identical (black squares within a grid), they were never actually seen in the imagery task. Manipulating the number of “black” squares allowed for capacity estimation through application of the WoMMBAT (Working Memory Modelling using Bayesian Analysis Techniques; Morey & Morey, 2011) model. Two different forms of visual memory task were used to provide more informative comparisons with the imagination task, one in which the memoranda were built up sequentially (e.g., each display showing one block to be memorised), and one in which the memoranda were shown simultaneously (e.g., one display showing all blocks to be memorised). If visual imagery and short-term visual memory rely on similar underlying processes and resources, then capacity estimates in the short-term visual memory tasks should be similar and strongly correlated with that of the imagination tasks.

Experiment 1

The experiment draws from previous paradigms such as change detection (Luck & Vogel, 1997) and cumulative imagery (Avons & Sestieri, 2005). An imagery task was used and two short-term visual memory tasks. These tasks all used a 6x6 matrix grid. In one of the visual memory tasks a number of these squares turned black very briefly before the matrix

disappeared. We refer to this as the visual short term memory task (VSTM) because no manipulation of the information in memory is required. In the other visual memory task, the squares turned black sequentially. This is referred to as the VWM task because the information in memory has to be continually updated. We recognise that the terms VWM and VSTM are sometimes used interchangeably in the literature. However, explicit recognition that one of these tasks is relatively passive in nature (VSTM) and one requires more active updating (VWM) is a useful and in our view captures a core feature of the distinction. In the imagination task two arrow cues (one on the column, one on the row) indicated that the participant had to imagine the grid square at the intersection of the arrow cues as black. The to-be-imagined array was built up sequentially. At test, a matrix was displayed with black squares filled in and participants had to indicate whether a cued square was in the same, or different location to that being represented. Set size was manipulated in all the tasks. In the VWM and the imagery tasks increasing the set size increased the number of steps in the sequence.

Method

Transparency and openness

We report how we determined our sample size, all manipulations, and all measures in the study. Data and research materials are available on the Open Science Framework: [OSF | The same or different? Capacity limitations in visual imagery versus visual perception of simple structural objects](https://osf.io/fpwb9/) (Atkin et al., 2025), which serves as the central repository for all five experiments. This project contains links to individual components for Experiments 1–5, with pre-registrations for Experiments 2–5 accessible via these links. The data and research materials for Experiment 1 are available at: <https://osf.io/fpwb9/>.

Participants

A convenience sample of forty-two participants (29 women), aged between 19 and 34 years ($M = 20.7$, $SD = 3.84$) and with normal or corrected to normal vision took part. This was a generous sample in comparison with those of previous similar behavioural studies investigating visual imagery and VWM ($N < 36$; e.g., Keogh & Pearson, 2011, 2014, 2017). Moreover, an initial pilot experiment using this paradigm indicated that this sample was sufficiently powered to yield an effect of set size in the imagery task. A priori power analysis was conducted using MorePower 6.0.4 (Campbell & Thompson, 2012) for a one-way ANOVA

design with five-levels (set-size 3, 4, 5, 6, 7) using an alpha α of .05 and an F ratio from the pilot study ($F = 19.65$). Results indicated that a total sample of 42 participants in the current experiment to have a .99% power. Each participant received a total of £15 for completing the study. For this and all subsequent experiments data collection was approved by Nottingham Trent University Business, Law and Social Sciences Research Ethics Committee and consent was obtained from each participant. Data were collected in 2018.

Apparatus and Materials

All experiments were run on a Lenovo ThinkCentre M79 10J7 using a 24" monitor (60-Hz refresh rate) with a 1360x768-pixel resolution. The imagery, VWM and the VSTM tasks were programmed using Psychopy 1.84.2 (Peirce, 2007). Colour is expressed as deviations from mid grey in Psychopy which takes values between 1 and -1. Forwards and Backwards Corsi Block Tapping Tasks were run using Millisecond Inquisit Lab 5 software. A viewing distance was set at 57cm for each of the tasks. An online questionnaire was administered before the laboratory tasks using Qualtrics to measure imagery vividness (Vividness of Visual Imagery Questionnaire and Vividness of Object and Spatial Imagery). An online questionnaire asking participants what cognitive strategy they used during the imagination and VWM tasks were administered using Qualtrics after completion of each of the tasks. Previous evidence has stated that cognitive strategies should be accounted for when analysing data as they have been found to influence performance during tasks (Pearson & Keogh, 2019).

Vividness of Visual Imagery Questionnaire (VVIQ)

The VVIQ consists of 16 items which investigates the vividness of an individual's experience of visual imagery (Marks, 1973). It is a self-report inventory which consists of four subsections that each require participants to generate a mental image of specific scenes and situations (e.g., 'think of some relative or friend whom you frequently see - but who is not with you at present - and consider carefully the picture that comes before your mind's eye'). Subsequently, 4 items relating to each mental image are investigated (e.g., 'the exact contour of face, head, shoulders and body'). Each item is based on 5-point rating scale (1= 'Perfectly clear and as vivid as normal vision to me' to 5 = 'No image at all, you only "know" that you are thinking of the object'). The total score ranges from 16-80 for each participant with low

scores indicating vivid imagery. The VVIQ has been widely reviewed and has been found to have very good internal consistency and reliability: Cronbach's $\alpha = .88-.95$ (Campos & Pérez-Fabello, 2009) with Mark's (1973) original study finding Cronbach's $\alpha = .89$ and good construct and criterion validity. In the current experiment all items produce high internal reliability with Cronbach's $\alpha = .94$.

Vividness of Object and Spatial Imagery (VOSI)

Vividness of object and spatial imagery was assessed using a 28-item scale (Blazhenkova, 2016). This measure was introduced as it measures both object and spatial imagery. The self-report inventory comprises two factors that assess object and spatial imagery. Each factor is assigned 14 items. Object imagery probes an individual's ability to create colourful, vivid and specified objects (e.g., 'Appearance of a candle fire'), whereas spatial imagery items investigate spatial relations and transformation of imagery (e.g., 'motion of the planets on a model of the solar system'). Each item is based on 5-point rating scale (1 = 'No image at all, you only "know" that you are thinking of the object' to 5 = 'Perfectly clear and vivid as normal vision'). Total scores range from 14-70 for each factor with high scores indicating vivid object imagery or vivid spatial imagery. A grand total for 28-items range from 28-140 with high scores indicating vivid object and spatial imagery. Each factor has been found to have convergent and discriminative validity and good internal reliability: object imagery Cronbach's $\alpha = .88$ and spatial imagery Cronbach's $\alpha = .89$ (Blazhenkova, 2016). From the current experiment, all factors produce high internal reliability with object imagery Cronbach's $\alpha = .90$ and spatial imagery Cronbach's $\alpha = .86$.

Strategy questions (imagination)

A series of questions were developed to investigate the strategies used by participants for the imagination task. In total there were seven items. 6 items were rated on a 5-point Likert-style scale that ranges from 1 'fully disagree' to 5 'fully agree' and contained questions such as 'After the cues were presented, I used them as I was asked to'. There was no total score, each item represented a separate score. 1 item requires participants to 'describe the strategy used to complete the task?'

Strategy question (VWM)

The strategy question investigates the individual's strategy during the working memory task. The single question requires participants to 'describe the strategy used to complete the task'.

Design and procedure

Participants first completed the VOSI and VVIQ. The remaining tasks were completed with task order counterbalanced across participants. Once the imagination or VWM task was complete; questions were administered for each of these tasks regarding strategies used. It took approximately 180 minutes to complete the study.

Forward and Backward Corsi Block Tapping Task

These were used as alternative measures of VWM capacity (Vandierendonck et al., 2004). Participants were required to reproduce a block sequence in the correct order that continually increased by one in length if the correct sequence was reproduced at least once out of two attempts. Participants began the task attempting to reproduce a 2-block sequence with 9 being the maximum block length tested. In the Forward version of the task each trial began with a display (3cm x 3cm) of 9 blue blocks for 1000ms. The first block of the sequence changed color to yellow for 1000ms before returning to blue and the next block of the sequence being shown. After the full sequence on that trial was shown participants had to reproduce the sequence by clicking on the relevant squares with a mouse. If they made a mistake, they could press a reset button that allowed them to restart their response sequence. After each attempted sequence feedback (Correct/Incorrect) was provided centrally. The backward version of the task was identical with the exception that participants were asked to recall the sequence backwards. The tasks measure both the longest sequence remembered (Corsi span, CS) and the total number of correct sequences (Corsi total, CT).

VWM task

Each trial began with the word 'Ready' presented centrally for 500ms followed by a light grey (0.5, 0.5, 0.5) fixation dot, presented on the left [X,Y (-337.5, 0)] or right [X,Y (337.5, 0)] of the display on a white background for 500ms. This was then replaced at the same location by a light grey (0.5, 0.5, 0.5) 6 x 6 grid (6.72cm x 6.72cm) with line width 2.6mm and individual square size of 1.12cm x 1.12cm). Black blocks (-.2, -.2, -.2) sequentially populated

random locations on the grid every 2000ms (Figure 1). Participants were asked to fixate the centre of the grid. The length of the sequence was determined by set size. Immediately following the final presentation, a dynamic visual white noise mask (6.72cm x 6.72cm) was presented at the array location for 1000ms.

Subsequently, the test array was presented on the opposite side of the display and comprised the same grid containing black blocks either in the previously cued locations ('same' trials) or with one of these blocks moved by two squares either horizontally, vertically or diagonally ('different' trials), with the constraint that no such moved block could be directly adjacent to any other black blocks. A post-cue queried a randomly-selected item (a white outline around a block) from one of the presented black blocks on 'same' trials or the block that had a changed location on 'different' trials. Participants responded either 'same' (Z key) or 'different' (M key) as quickly and as accurately as possible. They then self-reported vividness (low '1' to high '4'). Upon response feedback (Correct/Incorrect) was presented centrally for 500ms.

After a 10-trial practice block, participants completed 180 trials over 3 blocks. Within each block, each set size (3, 4, 5, 6, 7) was presented 12 times (six same and six different trials) in a random order. Participants could take a short break after each block.

Imagination task

The imagination task was identical to the VWM task, except that no squares turned black. Instead, participants sequentially built up an image of a set of black blocks in their mind, with arrow cues presented at the top of a column and the left side of a row indicating that a block should be imagined at the intersection of that column and row (Figure 1). Arrows changed position every 2000ms with the sequence dependent on set size. Immediately following the final presentation of arrow cues, cues disappeared and a blank grid was presented for 1000ms. The test array was shown as per the VWM task, with vividness ratings also given (with 10 practice trials and 180 trials structured in the same way). Unlike VWM/VSTM no visual mask was used as these are used to prevent visual persistence which can inflate estimates of performance. No such visual persistence should occur without sensory input. The task was also designed such that in the imagery condition the grid in which participants were imagining remained visible throughout. This ensured that participants were still holding the image in their

mind when the test display appeared on the other side of the screen, and thus that the test display probed what was currently being imagined.

VSTM task

The VSTM task was identical to the VWM task except that the black blocks were presented simultaneously for 100ms followed by the mask. The test display and questions were identical. Set size differed to the other tasks as piloting indicated the task was easier and so a greater range was used (3, 5, 7, 9). As such there were 144 trials across 3 blocks, with each block presenting each set size 12 times (six same and six different trials) in a random order. Participants were able to take a short break after each block.

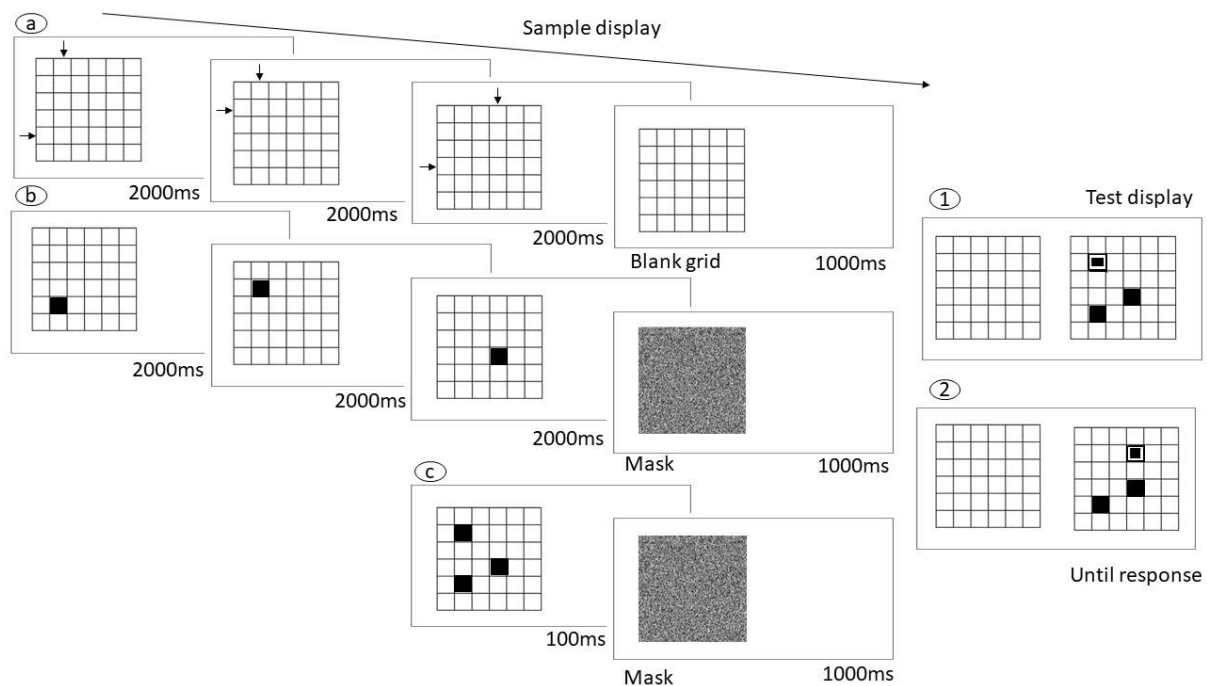


Figure 1. Stimuli and sequence of events in visual tasks with set-size 3 items. (a): *Imagination task* - participants were asked to create a black square in their minds eye at the intersection between the two arrow cues. (b): *VWM task* - black squares were presented in a sequential manner. (c): *VSTM task* - all black squares were presented simultaneously. (1) same trial (2) change trial.

Results

Statistical analysis

Data were analysed in accordance with the respective pre-registrations. ANOVA and t test analysis were performed using SPSS (Version 29.0; IBM Corp., 2022). The main dependent variables (d' and vividness) are analysed using within subjects ANOVAs and paired samples t tests (Bonferroni corrections), with Alpha level set at .05. In Signal Detection Theory, d' measures sensitivity and is calculated as $d' = z(p(\text{hits})) - z(p(\text{false alarms}))$. Extreme values (1 or 0) were adjusted to 0.99 for hits and 0.01 for false alarms. This correction for extreme values are consistent with the recommendations and ensure that all probabilities fall within a usable range for the z-transformation (Stanislaw & Todorov, 1999). Effect sizes for ANOVAs are reported using generalised eta squared. Eta squared was chosen as the measure of effect as it is comparable across different research designs (Baguley, 2009). Effect sizes for t tests are reported using Cohen's d. All other analysis that was pre-registered have been included in the Supplementary Materials. However, please note proportion correct and response time have not been correlated with questionnaire measures, as d' is the central measure of performance, and there is no theoretical reason to correlate response time with the questionnaire measures. In addition, Bayes factors (BF_{10}) were calculated using JASP software (Love et al., 2015) to measure evidence for the null versus alternative hypotheses (see Jarosz & Wiley, 2015) for ANOVAs and t tests. The Bayes factor provides an odds ratio for these hypotheses, where values less than 1 favour the null hypothesis and values greater than 1 favour the alternative hypothesis. Previous research has categorized the strength of Bayes factors with labels such as "substantial" (3), "strong" (10), and "decisive" (100) (Wetzels et al., 2011). WoMMBAT (Morey & Morey, 2011) was used to model capacity of imagination and visual memory (VWM and VSTM). All Bayes analyses were completed at the request of the journal and were not pre-registered.

Data were trimmed to exclude responses that were too fast (<300ms) or too slow (>6000ms). 300ms is considered a reasonable threshold in behavioural studies to exclude excessively fast responses (e.g., Welford, 1980), which may suggest anticipation rather than deliberate decision-making. The main analysis included all participant data. Nine participants reported using alternative cognitive strategies (e.g., 'I gave each black block a specific number to encode its location'). Statistical analysis was re-conducted with these participants removed. Since no evidence of a substantive difference was found between the data with all participants and with the nine participants removed, for transparency the analysis with all participants data is reported.

Data on d' and vividness ratings are shown in Figure 2. In all tasks increasing set size worsened d' and vividness. Across tasks, d' and vividness appeared worst in the imagery task, followed by the VWM task, with d' and vividness appearing best in the VSTM task. For the VWM and imagination tasks, a 2 (task: VWM, imagination) x 5 (set-size: 3, 4, 5, 6, 7) repeated measures analysis of variance (ANOVA) was carried out on accuracy (d') and vividness ratings in two separate ANOVAs. For VWM, imagination and VSTM tasks, a 3 (task: VWM, imagination, VSTM) x 3 (set-size: 3, 5, 7) repeated measures ANOVA was carried out on accuracy (d') and vividness ratings in 2 separate ANOVAs.

The 2 (task: VWM, imagination) x 5 (set-size: 3, 4, 5, 6, 7) repeated-measures ANOVA on d' yielded an effect of task, $F(1, 41) = 57.68, p < .001, \eta_g^2 = .26, BF_{10} = 3.133 \times 10^{+6}$, with VWM performance better than imagination performance. There was an effect of set-size, $F(4, 164) = 57.24, p < .001, \eta_g^2 = .37, BF_{10} = \infty$, signifying that as set-size increased, performance decreased. No interaction effect was found, $F(4, 164) = .86, p = .487, \eta_g^2 < .01, BF_{10} = .231$. The same ANOVA on vividness ratings yielded an effect of task, $F(1, 41) = 49.18, p < .001, \eta_g^2 = .39, BF_{10} = 1.195 \times 10^{+6}$, with VWM vividness being better than imagination. An effect of set-size, $F(1.75, 71.88) = 106.84, p < .001, \eta_g^2 = .44, BF_{10} = 1.035 \times 10^{+14}$, signifying that as set-size increases, vividness scores decreased. There was an interaction effect between task and set-size, $F(2.42, 99.02) = 3.07, p = .018, \eta_g^2 = .01, BF_{10} = 6.54$. Post hoc analysis found that VWM vividness decreased between set-size three and four ($p = .001, d = .70, BF_{10} = .441.51$), four and five ($p < .001, d = 1.14, BF_{10} = 2.576 \times 10^{+6}$), five and six ($p < .001, d = .82, BF_{10} = 4516.24$) and no effect between six and seven ($p = .052, d = .46, BF_{10} = 7.034$). Imagination vividness decreased between set-size three and four ($p < .001, d = 1.26, BF_{10} = 2.531 \times 10^{+7}$), four and five ($p < .001, d = .83, BF_{10} = 5128.95$), six and seven ($p < .001, d = .79, BF_{10} = 2808.254$) and no effect between five and six ($p = .181, d = .38, BF_{10} = .2.422$).

To compare across all three tasks a 3 (task: VWM, imagination, VSTM) x 3 (set-size: 3, 5, 7) repeated-measures ANOVA on d' yielded an effect of task, $F(2, 82) = 46.08, p < .001, \eta_g^2 = .31, BF_{10} = 1.471 \times 10^{+11}$. Post hoc analysis found that VSTM performance was significantly better than imagination performance, ($p < .001, d = .1.33, BF_{10} = 5.332 \times 10^{+18}$),

VWM performance was significantly better than imagination performance, ($p < .001$, $d = .1.13$, $BF_{10} = 1.104 \times 10^{+12}$), and VSTM performance was significantly better than VWM performance, ($p = .041$, $d = .40$, $BF_{10} = 12.95$). There was an effect of set-size, $F(2, 82) = 142.33$, $p < .001$, $\eta_g^2 = .46$, $BF_{10} = 1.053 \times 10^{+14}$, signifying that as set-size increased, performance decreased, and no interaction effect, $F < 1$, $BF_{10} = .089$. The same ANOVA on vividness ratings vividness yielded an effect of task, $F(1.34, 56.86) = 35.74$ $p < .001$, $\eta_g^2 = .34$, $BF_{10} = 1.279 \times 10^{+9}$. Post hoc analysis found that VSTM vividness was significantly better than imagination vividness ($p < .001$, $d = .1.06$, $BF_{10} = 1.342 \times 10^{+16}$), VWM vividness was significantly better than imagination vividness, ($p < .001$, $d = .1.05$, $BF_{10} = 1.951 \times 10^{+15}$), and there were no significant difference between VSTM and VWM vividness, ($p = .266$, $d = .27$, $BF_{10} = 2.58$). There was an effect of set-size, $F(1, 41) = 136.7$, $p < .001$, $\eta_g^2 = .42$, $BF_{10} = \infty$, signifying that as set-size increases, vividness scores decreased. There was no interaction effect between task and set-size, $F(4, 164) = 2.35$, $p = .057$, $\eta_g^2 = .01$, $BF_{10} = 1.89$.

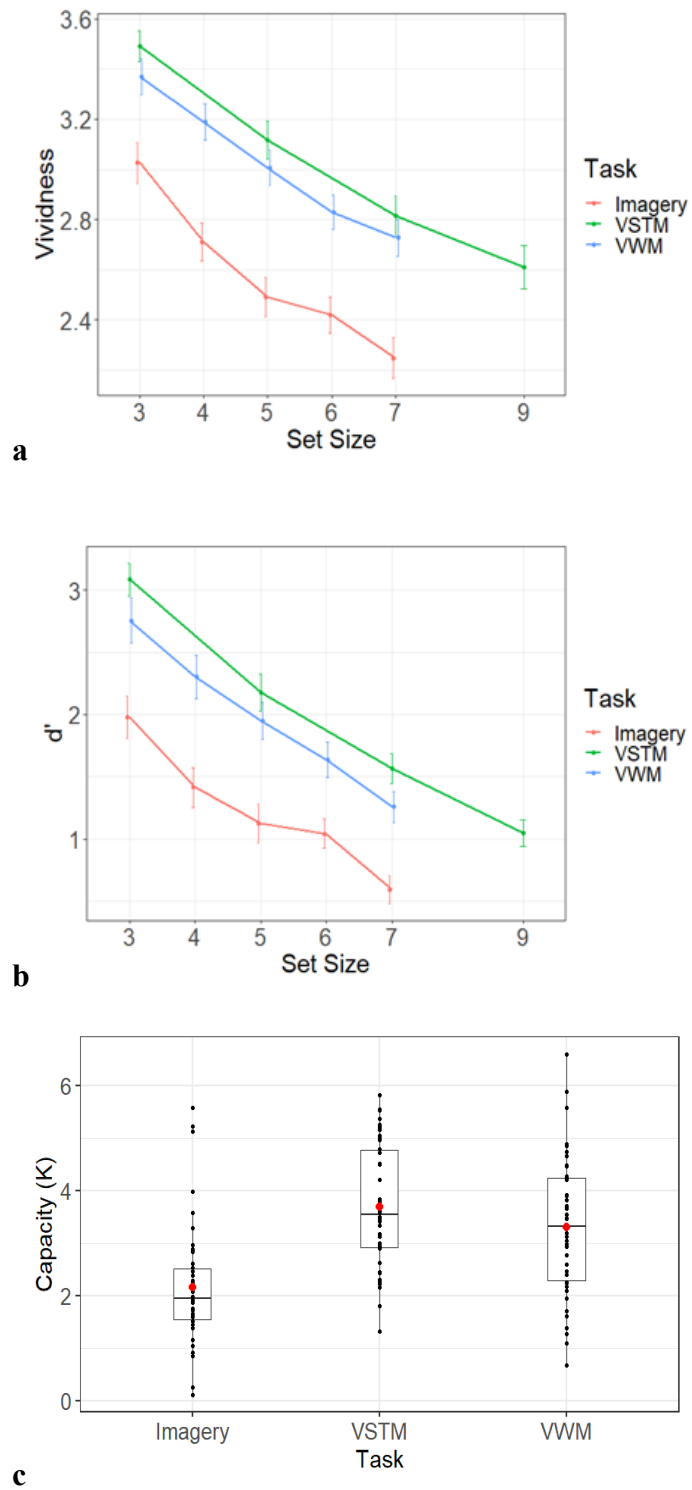


Figure 2. d' (a) and vividness (b) as a function of set-size for the imagery, VWM and VSTM tasks (error bars indicate standard error). Observed capacity estimates (k) for participants in the imagery, VSTM and VWM tasks are shown in (c) (the mean is indicated by the large red dot).

Capacity Estimates

WoMMBAT (Morey & Morey, 2011) was used to model capacity of imagination and visual memory (VWM and VSTM). The term *observed capacity* refers to the capacity demonstrated within these tasks, while *system capacity* denotes the theoretical maximum capacity of the underlying system. Importantly, differences in observed capacity do not necessarily imply differences in system capacity. The model is a hierarchical Bayesian extension of a non-hierarchical model developed for single-cue change detection and produces capacity estimates (k) for each participant (Rouder et al., 2008). Two separate models were run on data from the VWM, VSTM and imagery tasks simultaneously in order to estimate capacity. The first model contained participant as a factor and the second model contained both participant and task as factors. The WoMMBAT fit for each model was determined by the *Deviance Information Criterion* (DIC) (lower values indicate better fit). The modelling revealed that the participant by task interaction model provided the lowest DIC (20583.70) compared to participant as a factor (21274.80), indicating that task is important in order to understand differences in k (see the Supplemental Materials for full details on modelling procedure and statistics). This model yielded a mean observed capacity (k) of 3.69 items in the VSTM task, 3.31 items in the VWM task and 2.16 items in the imagery task. Estimates of k for each participant for each task are shown in Figure 2c. Strikingly, there was no overlap between the inter-quartile ranges of capacity estimates between imagery and VSTM, with only a small overlap between the inter-quartile ranges of imagery and VWM. A one-way repeated-measures ANOVA was conducted to compare the effect of task (imagination, VWM, VSTM) on k . There was a significant effect of task, $F(2, 82) = 42.05, p < .001, \eta_g^2 = .51, BF_{10} = 2.407 \times 10^{+10}$. Post hoc analysis found that VWM k was significantly higher than imagination k ($p < .001, d = 1.13, BF_{10} = 2.360 \times 10^{+6}$) and VSTM k was higher than imagination k ($p < .001, d = 1.24, BF_{10} = 2.054 \times 10^{+7}$), with no significant difference between VWM and VSTM k ($p = .102, d = .34, BF_{10} = 1.43$). As in these tasks the whole display was presented at test even though a single item was cued, modelling was repeated using the WoMMBAT whole display change detection model (Morey & Morey, 2011), with very similar output (k estimates of 3.95, 3.65 and 2.69 for VSTM, VWM and imagery respectively).

Correlations (corrected for multiple testing) between measures are shown in Table 1 (means and standard deviations of the questionnaire measures can be found in the

supplementary materials). Estimates of k were highly positively correlated between tasks, with the strongest correlation between imagery and VWM, and the weakest between imagery and VSTM. The same pattern was true of d' . This is expected given that both imagery and VWM require storage and manipulation of information and that previous research has shown similar neural activation between the tasks (Albers et al., 2013; Dijkstra et al., 2017). Self-report measures of vividness at the trial level and subjective ability to construct vivid imagery (VOSI and VVIQ) did not correlate with estimates of performance (d') or capacity (k , Corsi Total or Corsi Span). From the Corsi task estimates the Corsi total correlated with imagery performance and k , but not the memory equivalents.

Table 1. Correlations between k , d' , and vividness for each task (Imagination (Image), VWM and VSTM), Forward and backwards Corsi span (CS) and Corsi total (CT), VVIQ and VOSI (spatial, object and total).

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	Image k	-																
2	VWM k	.68*	-															
3	VSTM k	.47	.61*	-														
4	Image d'	.99*	.68*	.49	-													
5	VWM d'	.68*	.98*	.61*	.69*	-												
6	VSTM d'	.49	.62*	.97*	.51	.63*	-											
7	Image vividness	.28	.23	.08	.29	.26	.10	-										
8	VWM vividness	.49	.56*	.37	.49	.54*	.40	.57*	-									
9	VSTM vividness	.00	.11	.39	-.03	.07	.36	.33	.57*	-								
10	Forward (CS)	.46	.23	.18	.46	.25	.18	.07	.20	.03	-							
11	Forward (CT)	.53*	.32	.25	.54*	.31	.26	.09	.23	.05	.96*	-						
12	Backwards (CS)	.31	.20	.19	.35	.24	.17	.14	-.01	-.17	.31	.29	-					
13	Backwards (CT)	.49	.42	.38	.52*	.42	.37	.08	.14	-.04	.46	.50	.84*	-				
14	VVIQ	.17	.13	.06	.13	.10	.10	-.29	-.19	-.23	.01	-.02	.11	.06	-			
15	VOSI (object)	.08	.05	-.02	.08	.09	-.02	.21	.25	.25	-.07	-.05	-.28	-.19	-.59*	-		
16	VOSI (spatial)	.22	.13	.07	.21	.13	.03	.33	.36	.36	-.04	.00	-.17	-.05	-.54*	.77*	-	
17	VOSI (total)	.16	.10	.03	.15	.12	.00	.29	.33	.33	-.06	-.03	-.24	-.12	-.60*	.94*	.94*	

Note. $N = 42$, two-tailed test; * $p < .0003$ (Bonferroni correction).

Discussion

The paradigm in Experiment 1 is the first we know of that allows direct comparison of capacity limitations across imagery, VWM and VSTM tasks. Observed capacity estimates for the VWM and VSTM tasks were similar to those expected from the literature (~3-4 items), providing confidence in the validity of the paradigm (Cowan, 2001; Luck & Vogel, 1997). Quite striking was that the data appear to show the capacity estimate observed for imagery was reduced relative to VWM and VSTM. Intriguingly this suggests that imagery and visual memory differ in important respects, whilst the strong correlation between capacity measures indicates reliance on a set of similar processes. However, before drawing out the theoretical import of these findings it is necessary to consider whether differences in observed capacity were an artefact of methodology. There are several methodological issues that deserve exploration. One potential issue is that the amount of time available to generate and perceive objects in the visual imagery and VWM tasks was constrained to 2000ms. It could be that this simply does not enable time to create a stable representation in imagination. Critically, it could be that the fixed duration disproportionately impacted visual imagery as the imagery process might take longer than memory encoding. Indeed, previous research has suggested that imagery requires greater reliance on recruitment of top-down processes for the generation and maintenance of internally generated images (Dijkstra et al., 2017; Pearson, 2019; Schlegel et al., 2013). If duration was a limiting factor, then providing more time to generate and maintain an image should improve performance. This is explored in Experiment 2.

Second, the use of the arrow cues in the imagery task may have drawn attention away from the areas where the black squares were being imagined, interfering with their representation, or alternatively led to less spatial precision in imagery generation as the cueing method created the possibility of incorrect identification of the intersection. This is explored in Experiment 3. Finally, it may be that the presentation of the test display somehow favored the memory tasks because it contained the configuration of the black squares, and this configural information may have been processed more easily within the memory tasks. This is explored in Experiment 4.

Experiment 2

Experiment 2 aims to understand if extended presentation time improves imagination performance so that it is comparable to VWM performance. It may be that the observed capacity estimate differences between imagination and VWM in Experiment 1 were due to imagination requiring extra time to generate and maintain a visual image. This experiment was preregistered on the Open Science Framework. The preregistration, along with associated data and research materials, are available at: <https://osf.io/xk6zs/>.

Method

Participants

Sample size planning was conducted using one-sided $\alpha = .05$ and was based on the 42 participants from Experiment 1 using the most relevant parameters of interest (set-size 4 - imagery and VWM d'), effect size .82 (Cohen's d), and $r = .48$ (paired correlation). A priori power analysis indicated a sample size of 28 to have a 99% power to detect such an effect. Twenty-eight participants (25 women) took part in the experiment aged between 19 and 30 years ($M = 19.73$, $SD = 1.18$) and had normal or corrected to normal vision. The study took approximately 70 minutes to complete. Each participant received 7 research credits for taking part as a Psychology student through the Psychology Sona System. Data were collected in 2019.

Apparatus and materials

This was the same as Experiment 1. However, the Forwards and Backwards Corsi Block Tapping tasks were not included and neither was the VSTM task.

Design and procedure

Participants first completed the set of questionnaires as described in Experiment 1. Upon completion, the order of the imagination and VWM tasks was counterbalanced across participants using a Balanced Latin Square Design. Once the imagination task or VWM task was complete; questions were administered for each of these tasks which investigated the strategies used to complete the task. A repeated measures design was used.

The design and procedure were the same as Experiment 1, with the following exceptions. Only set-size 4 was used and the arrow cues (imagination task) or filled blocks (VWM task) were presented for either 2 seconds or 4 seconds. For each task the practice block contained 8 trials: four trials were allocated to each of the conditions (duration: either 2 seconds or 4 seconds) and of the four trials two was allocated to the “same” test array presentation and two to the “different” test array presentation. Each of the two trials were either presented to the left or right of the display. For each task three equally balanced main blocks comprised 72 trials. Each duration condition therefore comprised 36 trials, 18 “same” trials and 18 “different” trials of which 9 were presented to the left and 9 presented to the right of the display. Trial order was randomised.

Results

Data were trimmed and excluded as in Experiment 1. Figure 3 shows d' and vividness for both tasks across the 2s and 4s conditions. As expected, d' and vividness were worse for imagination in comparison to VWM. Increasing presentation duration appeared to have no impact on performance. A 2 (task: VWM, imagination) x 2 (duration: 2 second, 4 second) repeated measures analysis of variance (ANOVA) was carried out on performance (d') and vividness ratings in 2 separate ANOVAs. The analysis was completed on data from all participants and excluding three participants who reported using alternative cognitive strategies (e.g., ‘I gave each black block a specific number to encode its location’). Since no evidence of a difference was found between these analyses the analysis with all participants in is reported. The analysis on d' yielded a significant effect of task only, $F(1, 27) = 131.83, p < .001, \eta_g^2 = .60, BF_{10} = 3.800 \times 10^{+8}$, with no significant effect of presentation duration, $F < 1, BF_{10} = .277$, and no significant interaction, $F(1, 27) = 1.47, p = .236, \eta_g^2 < .01, BF_{10} = .527$. The same ANOVA on vividness ratings yielded a significant effect of task, $F(1, 27) = 76.78, p < .001, \eta_g^2 = .66, BF_{10} = 6.160 \times 10^{+6}$. There was no significant effect of presentation duration, $F < 1, BF_{10} = .758$, but there was a significant interaction effect, $F(1, 27) = 6.10, p = .020, \eta_g^2 = .03, BF_{10} = 2.78$. Post hoc analysis showed no significant effect of presentation duration on imagination, $t(28) = 1.26, p = .219, d = .24, BF_{10} = .408$, whereas in the VWM task increasing presentation duration significantly decreased vividness, $t(28) = 2.30, p = .029, d = .43, BF_{10} = .1.89$.

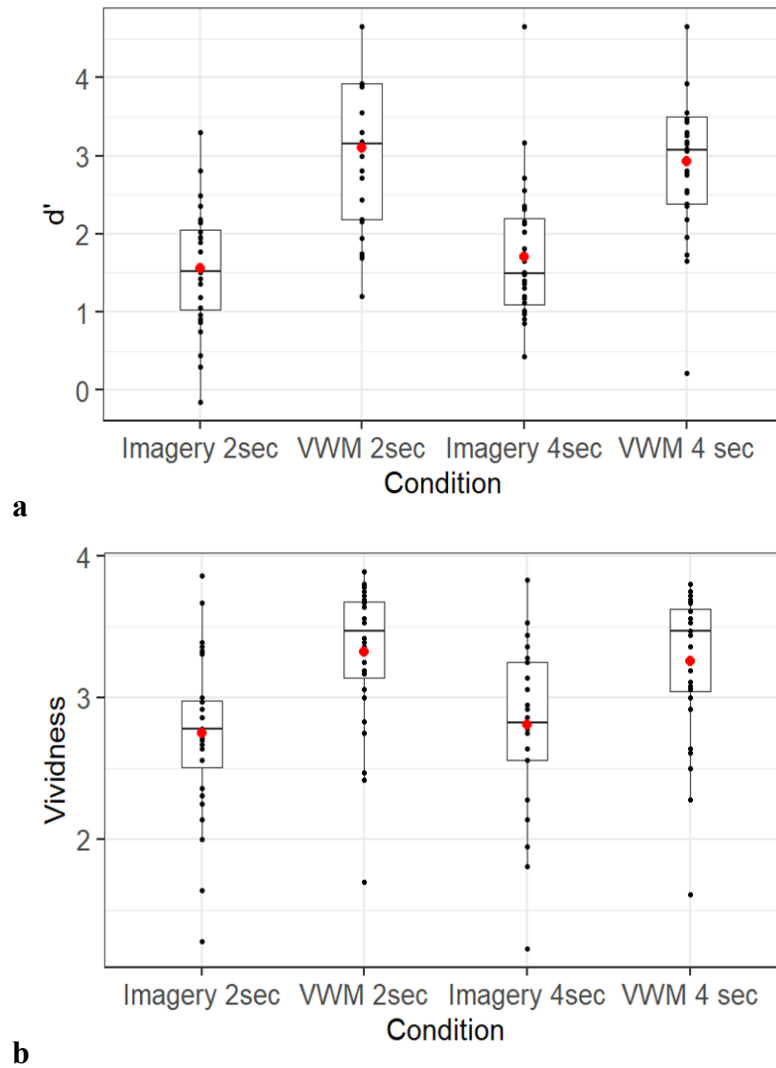


Figure 3. d' (a) and vividness (b) for presentation time (2 and 4 seconds) for each task (imagery and VWM).

Correlations (corrected for multiple correlations) between measures are shown in Table 2 (means and standard deviations of the questionnaire measures can be found in the supplementary materials). Strong positive correlations were found between imagination and VWM d' and vividness. This is consistent with imagination and VWM sharing similar processes and potentially sharing neural networks (Albers et al., 2013; Dijkstra et al., 2017) and corroborates the findings from Experiment 1. There were no significant correlations between imagination measures (d' and vividness) and questionnaire measures and no

significant correlations between VWM measures (d' and vividness) and questionnaire measures.

Table 2. Correlations among imagination (Image) and VWM d' and vividness (combined 2 and 4 second conditions for each task) and questionnaires: VVIQ, VOSI and subscales

		1	2	3	4	5	6	7	8
1	Image d'	-							
2	VWM d'	.70*	-						
3	Image vividness	.34	.24	-					
4	VWM vividness	.40	.46	.86*	-				
5	VVIQ	.06	.12	-.09	-.03	-			
6	VOSI (object)	-.07	-.03	-.03	-.11	-.82*	-		
7	VOSI (spatial)	.00	-.09	.10	.06	-.69*	.68*	-	
8	VOSI (total)	-.04	-.06	.04	-.02	-.82*	.90*	.93*	-

Note. $N=28$, two-tailed test; * $p < .001$. (Bonferroni correction).

Discussion

Experiment 2 shows that presentation duration had no effect on performance in the imagination task. Thus, in Experiment 1, performance differences and the observed capacity differences between VWM and imagery was not simply due to more time being needed to generate and maintain items in the imagination task. This is perhaps unsurprising as Keogh and Pearson (2017) also found that increasing imagery generation time from 6–12s did not improve priming in their binocular rivalry task. Moreover, evidence suggests that imagery generation occurs within 1s (Pearson et al., 2008) and so 2s should have been ample time to generate a robust image.

Although Experiment 2 rules out a role of presentation duration, there were other aspects of Experiment 1 that may have selectively harmed performance in the imagination condition. The paradigm was purposefully created so that nothing was presented at places where images needed to be created. Rather a pair of arrow cues indicated the intersection where participants had to imagine a black block. However, this cuing methodology may either divert attention away from the to-be-imagined location and/or increase positional noise in terms of the imagination being created at the correct location. Experiment 3 was designed to eliminate

these explanations of the performance differences between imagination and VWM in Experiment 1.

Experiment 3

The aim of this experiment was to explore the extent to which the arrow cues in the imagery task may have drawn attention away from the areas where the black squares were being imagined, interfering with their representation, or alternatively led to less spatial precision in imagery generation as the cueing method created the possibility of incorrect identification of the intersection. This experiment was preregistered on the Open Science Framework. The preregistration, along with associated data and research materials, are available at: <https://osf.io/pbqxx/>.

Method

Participants

Sample size planning was conducted using one-sided $\alpha = .05$ and was based on the 42 participants from Experiment 1 using the most relevant parameters of interest (set-size 5 - imagery and VWM d'), effect size .79 (Cohen's d), and $r = .52$ (paired correlation). A priori power analysis indicated 32 participants in the current experiment to have a 99% power to detect such an effect. Thirty-two participants (22 women) took part in the experiment aged between 19 and 31 years ($M = 22.3$, $SD = 4.4$) and had normal or corrected to normal vision. The study took approximately 180 minutes to complete. Each participant received 18 research credits for taking part as a Psychology student through the Psychology Sona System. Data were collected in 2019.

Apparatus and materials

This was the same as Experiment 1.

Design and procedure

Participants first completed the set of questionnaires as described in Experiment 1. Upon completion, participants completed 11 different experimental conditions (see Figure 4). Conditions were counterbalanced using a Balanced Latin Squared Design. A repeated measures

design was used. Once all the imagination or VWM tasks were complete; questions were administered for each of the tasks which investigated the strategies used to complete the task.

The design and procedure were the same as the set size 5 condition in Experiment 1 aside from the following differences. The 8 (A–H) conditions (Figure 4 Panels A-H) comprised a 2 (imagination task or VWM task) x 2 (global grid size large or small) x 2 (4 x 4 or 6 x 6 grid locations) design. The grid used in Experiment 1-2 was used for comparison and consisted of the large global grid size with 6 x 6 locations (Figure 4 panel B and panel F). If the difference observed between imagination and VWM performance in Experiment 1 was due to the arrow cues directing attention away from the to be imagined object, then decreasing the grid size, and thus decreasing the absolute distance of the cue from the to-be-imagined object, should improve performance in the imagination condition. Otherwise any impact of reducing grid size would be expected to be similar across the VWM and imagination tasks. In terms of the number of grid locations, then reducing this should improve performance in both tasks by reducing uncertainty. However, if the use of a cue increased the noise in where to generate the imagination in previous Experiments, then reducing location uncertainty should preferentially benefit imagination. This should particularly be the case when the global grid size is small.

Conditions I and J were imagination conditions only. In these conditions the arrow cues stayed in the same place they were in the standard Panel B condition used in Experiments 1-2, but the global grid size was reduced. As such conditions I and J were identical to conditions A and C, except that arrow cues were further away from the grid. If using an arrow cue harmed performance in the imagination task of Experiment 1 because it increased the noise in perception of where to imagine the object, then increasing the distance from the arrow cue to the grid should harm performance in conditions I and J relative to conditions A and C.

Condition K was a VWM condition, using the standard size VWM grid (Panel F) but with the addition of arrow cues. We included this condition to see whether the presence of arrow cues diverted attention away from the to-be-memorised object and thus reduced performance relative to the condition in which the grid size and grid locations were the same but for which there were no additional arrow cues (condition F).

In terms of the stimuli, the large global grid size with 6 x 6 locations was as previously described in Experiment 1. The large global grid size and 4 x 4 locations consisted of a grid

that measured 6.72 cm x 6.72 cm and consisted of 16 squares with each square measuring 1.68 cm x 1.68 cm. The small global grid size and 6 x 6 locations consisted of a grid that measured 4.48 cm x 4.48 cm and consisted of 36 squares with each square measuring .75 cm x .75 cm. The small global grid size and 4 x 4 locations consisted of a grid that measured 4.48 cm x 4.48 cm and consisted of 16 squares with each square measuring 1.12 cm x 1.12 cm. The arrow cues in conditions A-D were placed at the top of a column and the left side of a row indicating that a block should be imagined at the intersection of that column and row. The stimuli in the VWM conditions E-H were the same as the imagination conditions A-D but instead of the arrow cue, a block was filled black sequentially as per Experiment 1. The imagination conditions I and J were the same conditions A and C other than that the arrow cues were placed as if a large global grid size was presented (as in conditions B and D). Finally, condition K was the same as condition F but with the addition of arrow cues (used in the same way as in condition B).

Each condition had a practice block and a main block. The practice block contained 4 trials: 2 were allocated to the “same” test array presentation and 2 to the “different” test array presentation. 1 of the “same” trials and 1 of the “different” trials were presented to the left of the display the other “same” and “different” trials were presented to the right of the display. The main block for each condition contained 36 trials. 18 “same” trials and 18 “different” trials of which 9 of each were presented to the left and 9 of each were presented to the right of the display. All trials were randomised within a block.

Results

Data trimming and exclusion criteria were the same as in Experiment 1.

Figure 5 shows d' and vividness for imagination and VWM for Conditions A-H. Manipulating global grid size appeared to have a minimal effect on performance in both imagination and VWM, whereas reducing grid locations appeared to improve performance in both imagination and VWM. A 2 (task: imagination, VWM) x 2 (number of locations: 6 x 6, 4 x 4) x 2 (global grid size: large, small) repeated measures ANOVA was conducted on d' and vividness. For d' There was a significant effect of task with VWM performance producing better scores than imagination, $F(1, 31) = 39.87, p < .001, \eta_g^2 = .22, BF_{10} = 18124.72$, and a significant effect of number of locations with fewer locations producing better scores, $F(1, 31) = 86.4, p < .001, \eta_g^2 = .22, BF_{10} = 2.596 \times 10^{+7}$. There was no significant effect of global grid size, $F < 1, BF_{10} = .186$, and no significant interaction effects; task and number of locations, $F < 1, BF_{10} = .702$, task and global grid size, $F(1, 31) = 3.2, p = .083, \eta_g^2 < .01, BF_{10} = .379$, number of locations and global grid size, $F(1, 31) = 2.9, p = .099, \eta_g^2 < .01, BF_{10} = .362$, and no significant three way interaction effect, $F(1, 31) = 2.7, p = .111, \eta_g^2 < .01, BF_{10} = .328$. The same ANOVA on vividness ratings yielded a significant effect of task with VWM performance producing greater vividness, $F(1, 31) = 64.21, p < .001, \eta_g^2 = .26, BF_{10} = 1.624 \times 10^{+6}$, and a significant effect of number of locations with fewer locations producing better vividness, $F(1, 31) = 26.01, p < .001, \eta_g^2 = .17, BF_{10} = .788.22$. There was no significant effect of global grid size, $F < 1, BF_{10} = .167$, and no significant interaction effects; task and number of locations, $F(1, 31) = 1.67, p = .205, \eta_g^2 < .01, BF_{10} = 1.198$, task and global grid size, $F(1, 31) = 1.88, p = .180, \eta_g^2 < .01, BF_{10} = .302$, number of locations and global grid size, $F < 1, BF_{10} = .151$, and no significant three-way interaction effect, $F < 1, BF_{10} = .043$.

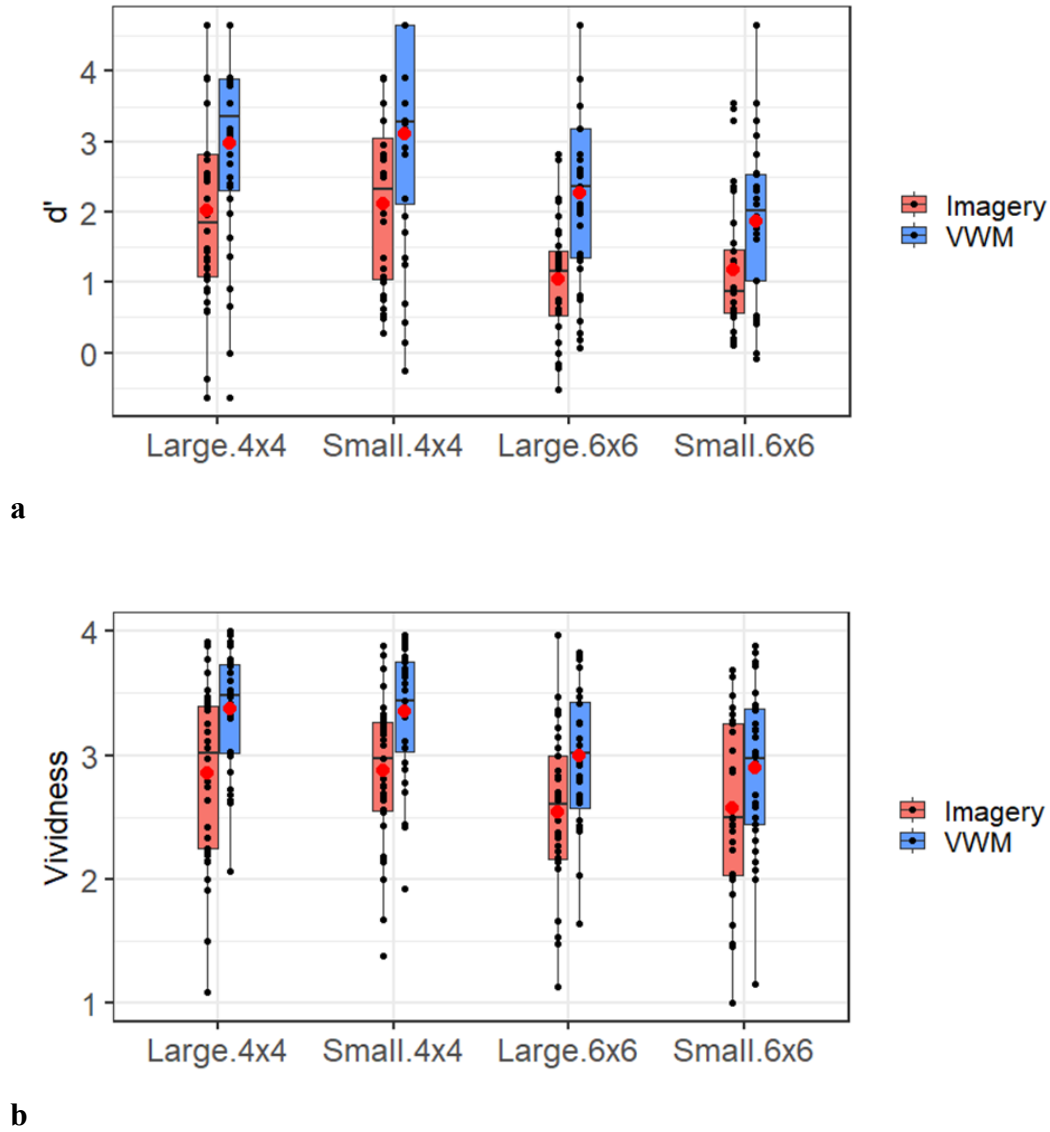
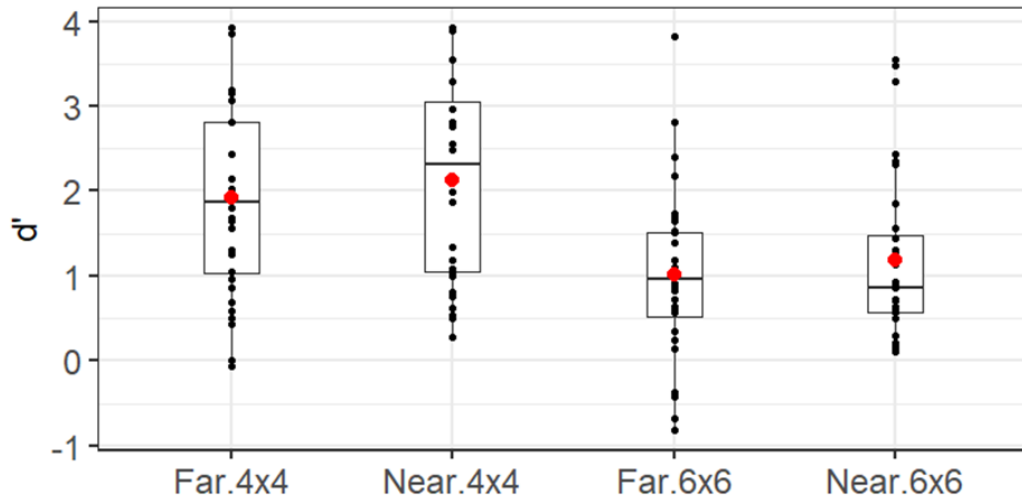


Figure 5. d' (a) and vividness (b) for the imagery task (conditions A-D) and the VWM task (conditions E-H) in which number of locations (4 x 4, 6 x 6) and global grid size (small, large) were manipulated.

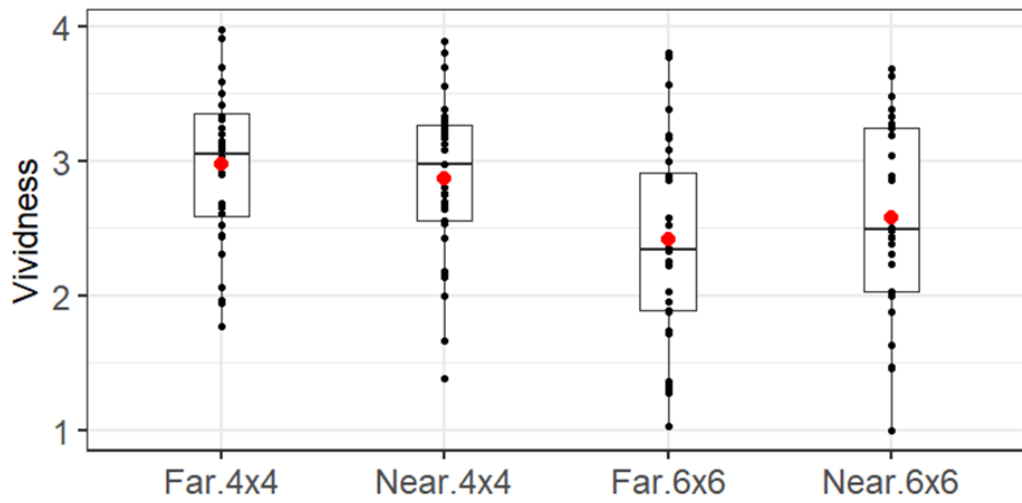
Figure 6 shows d' and vividness for the imagination conditions where the arrow cue was far from the grid (conditions I and J) and for convenience replots the data from Conditions A and B where the arrow cue was near to the grid for ease of comparison. There appears to be little systematic impact of the cue distance on either d' or vividness. A 2 (cue location; far, near) x 2 (number of locations: 4 x 4, 6 x 6) repeated measures ANOVA was conducted on d' . There was a significant effect of number of locations, with improved performance when there were fewer grid locations, $F(1, 31) = 43.05, p < .001, \eta_g^2 = .37, BF_{10} = 40181.83$. There was no

main effect of cue location, $F(1, 31) = 2.87, p = .101, h_g^2 = .02, BF_{10} = .444$, and no significant interaction effect, $F < 1, BF_{10} = .346$. The same ANOVA on vividness ratings yielded a significant effect of number of locations with 4 x 4 number of locations producing better vividness scores than 6 x 6 number of locations, $F(1, 31) = 34.06, p < .001, \eta_g^2 = .33, BF_{10} = 9922.94$. There was no significant effect of cue location, $F < 1, BF_{10} = .818$. There was a significant interaction effect between the number of locations and cue location, $F(1, 31) = 5.23, p = .028, \eta_g^2 = .05, BF_{10} = 3.157$. Post hoc analysis found that the far cue with 4 x 4 locations produced significantly better vividness scores than the far cue with 6 x 6 locations, $t(31) = 6.36, p < .001, d = .83, BF_{10} = 37287.89$, and the near cue with 4 x 4 locations produced significantly better vividness scores than the near cue with 6 x 6 locations, $t(31) = 3.01, p = .005, d = .82, BF_{10} = 7.71$. There was a significant effect of cue distance with the near cue with 6 x 6 locations producing better vividness scores than the far cue with 6 x 6 locations, $t(31) = 2.05, p = .049, d = .36, BF_{10} = 1.18$. Finally, there was no significant difference between near cue and 4 x 4 locations and far cue with 4 x 4 locations, $t(31) = 1.27, p = .214, d = .22, BF_{10} = .393$.

Figure 7 shows d' and vividness from the VWM conditions F and K which differed in whether the VWM task contained arrow cues or not. There appeared to be no real impact of the inclusion of the arrow cue. A paired t test showed no impact of the cue on d' $t(31) = 1.26, p = .218, d = .22, BF_{10} = .387$, or on vividness, $t(31) = 1.49, p = .146, d = .26, BF_{10} = .515$

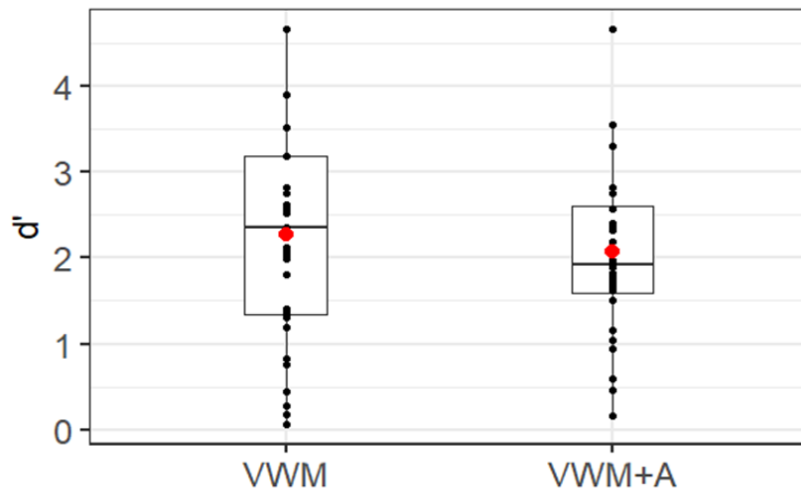


a

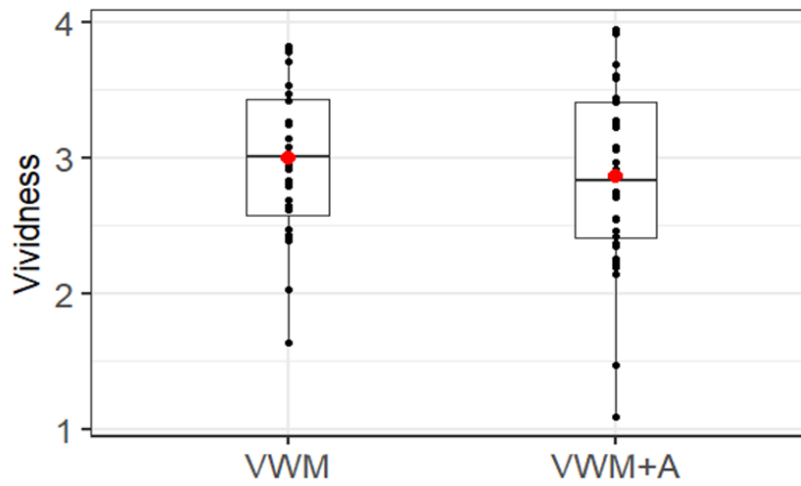


b

Figure 6. d' (a) and vividness (b) for the imagination task conditions A, B, I and J which differed in the number of grid locations (4 x 4, 6 x 6) and cue location (far, near).



a



b

Figure 7. d' (a) and vividness (b) for the VWM task (conditions F and K) which differed in whether the VWM task contained arrow cues (VWM+A) or not (VWM).

Table 3 shows correlations between imagination d' and VWM d' with questionnaire responses (means and standard deviations of the questionnaire measures can be found in the supplementary materials). Consistent with previous experiments, there was a strong positive correlation between imagery and VWM for both d' and imagery. There were no significant correlations between imagination measures (d' and vividness) and questionnaire measures and no significant correlations between VWM measures (d' and vividness) and questionnaire measures.

Table 3. Correlations among imagination (Image) and VWM d' and vividness (combined: number of locations: 6x6, 4x4 and global grid size: large, small for each task) and questionnaires: VVIQ, VOSI and subscales

		1	2	3	4	5	6	7	8
1	Image d'	-							
2	VWM d'	.69*	-						
3	Image vividness	.44	.21	-					
4	VWM vividness	.47	.46	.85*	-				
5	VVIQ	-.23	-.22	.02	-.14	-			
6	VOSI (object)	.23	.48	.03	.26	-.62*	-		
7	VOSI (spatial)	-.01	.01	.02	.15	-.57*	.57*	-	
8	VOSI (total)	.10	.23	.03	.22	-.66*	.84*	.92*	-

Note. $N=32$, two-tailed test; $p < .001$ (Bonferroni correction).

Discussion

Experiment 3 explored whether the use of an arrow cue in the imagination task in Experiment 1 could have led to detrimental performance in that task, thus accounting for performance and the observed capacity differences between the VWM and imagination tasks. Several hypotheses were explored. First was whether the cue diverted attention away from the to-be-imagined object. Manipulating the size of the grid, and thus the absolute distance from the arrow from the point at which an imagination was to occur, had no impact on imagination. In addition, including the arrow cue in a VWM task to see if it captured attention and reduced performance showed that this was not the case. As such there is little indication that the arrow cue selectively impacted performance in the imagination task by diverting attention away from the area in which participants had to build up an image.

The second consideration was that the arrow cue, because it is more distal from the location of the to-be-imagined stimulus, might lead to increased noise about where to generate the imagination, leading to worse performance in the imagination task relative to the VWM task where there was much less location uncertainty. If this were the case, then reducing the number of locations in the grid should enhance performance in both tasks, but more so in the imagination task. However, whilst reducing the number of grid locations did enhance performance, it did not do so selectively for imagination. The same hypothesis was also explored in the imagination task by manipulating the distance of the arrow cue from the grid. Increasing this distance had no impact on imagination performance. There was a significant decrease in vividness when the arrow cue was further away from the grid, but only for the 6 x 6 display with Bayesian analysis indicating inconclusive evidence in favour of the alternative hypothesis.

Taken together, this suggests very little evidence that the arrow cue disrupted performance in the imagination task to such an extent that it could explain the large performance differences observed in Experiment 1. It is worth noting that other research has found an impact of cue location on imagery. Borst and Kosslyn (2010) developed a spatial imagery task that used an arrow cue to determine if imagined dots were in a specific location. They found that increasing the distance between the arrow cue and test location increased errors. However, their task required participants to study a pattern of dots and then draw them on paper and so there was a large memory load. In the imagination task presented here memory load was minimal as participants simply had to build up an image that was still live at test and did not require generating the image from memory.

The reduced performance in imagination relative to VWM in Experiment 1 therefore cannot be explained in terms of needing more time to generate images to update memory. Nor does it appear to be because the arrow cues interfered with the act of generating an image. A final possibility is that there was something about the test display that disadvantaged the imagination task. Experiment 1 used identical test displays for the imagery, VWM and VSTM tasks (e.g., participants were shown all black blocks with one block being highlighted white). It could be that this type of display benefited memory performance as the stimuli were presented during build-up, which could enable better configural processing because then there would be greater similarity between the appearance of the test display and of the stimuli presented during build up, and so presenting items in their configuration at test might enhance

memory performance relative to imagination. In addition, the presentation of the black blocks in the test array may interfere with imagination processes more than memory task because these blocks were not seen during imagination, whereas they were seen during build-up for VWM and VSTM. Experiment 4 therefore manipulated the type of test display. At test participants were either presented with the whole configuration with the decision made about a single item within this (this is as per Experiments 1-3 and referred to as the post-cue display), shown the whole display and asked if it was the same or different, or shown a display with a single item and asked whether this was the same or different.

Experiment 4

Experiment 4 aimed to assess if the presentation of the test display somehow favored the memory tasks because it contained the configuration of the black squares, and this configural information may have been processed more easily within the memory tasks compared to the imagination task. This experiment was preregistered on the Open Science Framework. The preregistration, along with associated data and research materials, are available at: <https://osf.io/gamdy/>.

Method

Participants

Sample size planning was conducted using one-sided $\alpha = .05$ and was based on the 42 participants from Experiment 1 using the most relevant effect of interest (set-size 5 - VWM and VSTM d'), effect size .24 (Cohen's d), and $r = .41$ (paired correlation). A priori power analysis indicated 28 participants in the current experiment to have a 96% power to detect such an effect. Twenty-eight participants (19 women) took part in the experiment aged between 19 and 30 years ($M = 21.6$, $SD = 4.6$) and had normal or corrected to normal vision. The study took approximately 140 minutes to complete. Each participant received 14 research credits for taking part as a Psychology student through the Psychology Sona System. Data were collected in 2020.

Apparatus and materials

This was the same as Experiment 1.

Design and procedure

Participants completed the set of questionnaires as described in Experiment 1 followed by completing the imagination and visual memory tasks, the order of which were counterbalanced across participants using a Balanced Latin Square Design. Once the imagination or VWM task was complete; questions were administered for each of these tasks which investigated the strategies used to complete the task.

The design and procedure for the imagination, VWM and VSTM tasks were the same as Experiment 1 (see Figure 8), with the following exceptions. Only set-size 5 was used. In addition, there were three types of test array: all five black blocks (whole display), all five black blocks with one block highlighted white (post-cue display) or a single black block (single cue display). Participants were asked to judge whether all of these blocks (whole display) or the cued block was originally cued or not (see Figure 8: panel A). The practice block contained 12 trials: 4 trials were allocated to each of the conditions (test array: whole display, post-cue or single-cue) and of the 4 trials 2 were allocated to the “same” test array presentation and 2 to the “different” test array presentation. 2 of the 4 trials were presented to the left of the display the other 2 were presented to the right of the display. 3 equally balanced main blocks comprised a total of 108 trials. Each test array therefore comprised 36 trials, 18 “same” trials and 18 “different” trials of which 9 were presented to the left and 9 presented to the right of the display. All trials were randomised.

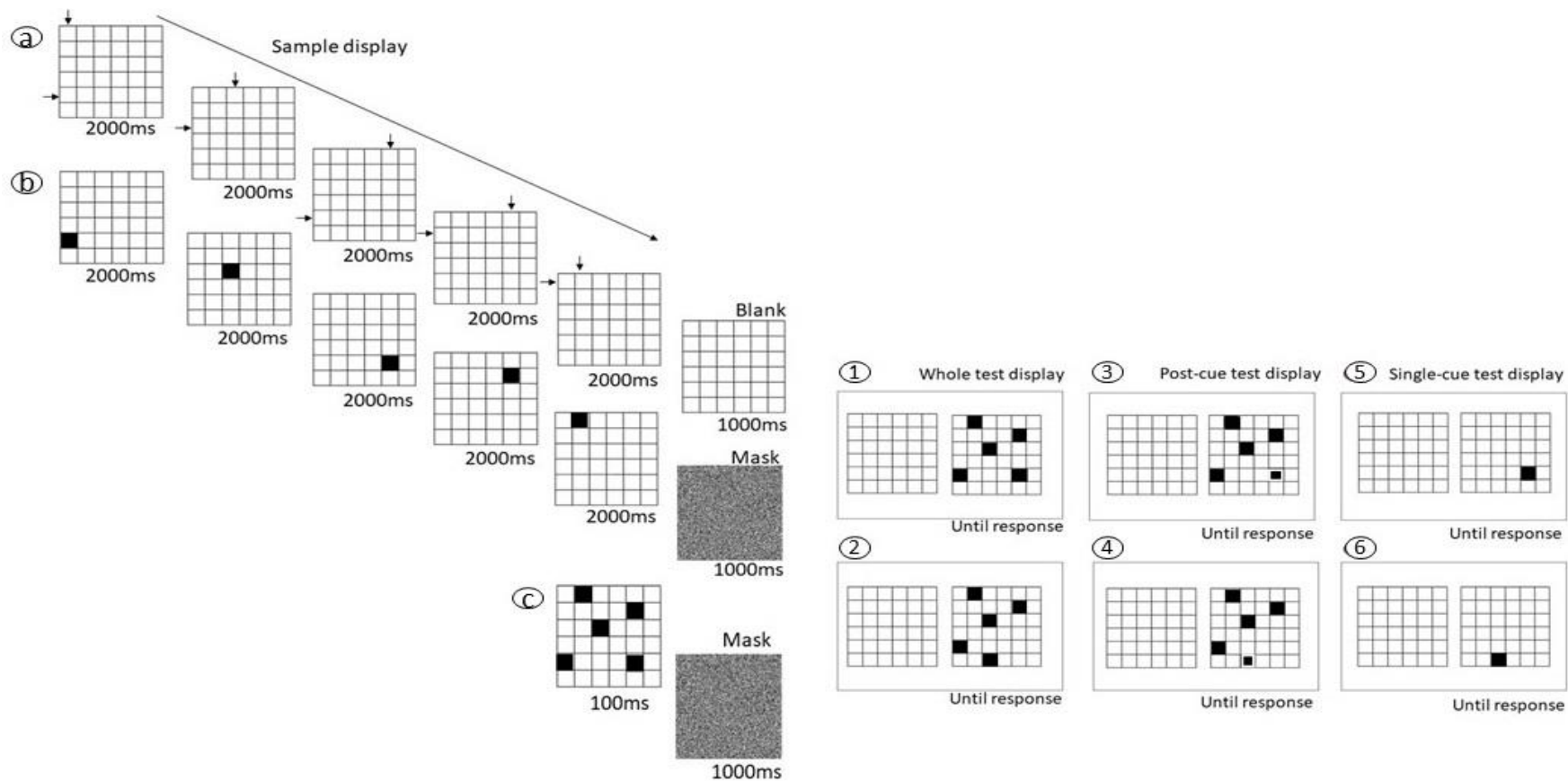


Figure 8. Stimuli and sequence of events on visual tasks. Panel A: Imagination task -participants were asked to create a black square in their minds eye at the intersection between the two cue pointers. Panel B: VWM task—Instead of creating a black square in the mind’s eye the black square was perceptually presented in a sequential manner. Panel C: VSTM task—all black squares were presented simultaneously. 1 and 2 are whole display change detection tests, 3 and 4 are post-cue change detection tests and 5 and 6 are single-cue change detection tests. 1, 3 and 5) same trial 2, 4 and 6) change trial. An error in 2, 4 and 6 indicates participants could not successfully imagine / remember all five locations accurately.

Results

Data were trimmed as in Experiment 1. Two participants reported using alternative cognitive strategies (e.g., ‘I gave each black block a specific number to encode its location’). Statistical analysis was re-conducted with these participants removed. Since no evidence of a substantive difference was found between the data with all participants and with the two participants removed, only the analysis with all participants data is reported.

Figure 9 shows that overall performance (d') tended to be worse for imagery, slightly better for VWM and better again for VSTM. However, it is clear that this was dependent on the test display used. The whole and single post-cue conditions showed similar levels of d' within each task, whereas in the single-cue condition, d' was similar for imagery and VSTM, but markedly better in VWM. A 3 (task: imagination, VWM, VSTM) x 3 (test array: whole display, post-cue, single-cue) repeated-measures ANOVA was conducted on d' and yielded an effect of task, $F(2, 54) = 22.06, p < .001, \eta_g^2 = .24, BF_{10} = 4.307 \times 10^{+8}$. Post hoc analysis found that VWM performance was significantly better than imagination performance ($p < .001, d = .88, BF_{10} = 619.84$), VSTM performance was significantly better than imagination performance ($p < .001, d = .1.03, BF_{10} = 4236.39$), and there was no significant difference between VWM and VSTM performance ($p = .355, d = .30, BF_{10} = .633$). There was no effect of test array, $F(2, 54) = 1.03, p = .365, \eta_g^2 = .01, BF_{10} = .964$, but there was a significant interaction between task and test array, $F(4, 108) = 8.76, p < .001, \eta_g^2 = .12, BF_{10} = 23558.82$. Post hoc analysis for the imagination task found no significant pairwise differences between the single-cue test and the whole display test ($p = .093, d = .43, BF_{10} = 1.80$), single-cue test and post-cue test ($p = .999, d = .14, BF_{10} = .262$) or whole display and post-cue test ($p = .557, d = .26, BF_{10} = .458$). For the VWM task, the single-cue test produced better performance than the whole display test ($p = .021, d = .55, BF_{10} = 6.24$) and the post-cue test ($p = .030, d = .52, BF_{10} = 4.60$), whereas no significant difference was found between whole test display and the post-cue test ($p = .954, d = .19, BF_{10} = .321$). For the VSTM task the whole test display produced better performance than the single-cue test ($p = .002, d = .66, BF_{10} = 22.25$), the post-cue test array produced better performance than single-cue test ($p = .001, d = .78, BF_{10} = 88.74$), and no significant differences were found between whole test display and post-cue test ($p = .999, d = .04, BF_{10} = .204$). For each test array condition, post hoc comparisons were made

between each pair of tasks. All of these were significant ($p < .024$, $BF_{10} > 5.86$), apart from the single-cue condition between imagery and VSTM ($p = .426$, $d = .29$, $BF_{10} = .554$).

Data on vividness ratings (Figure 9) showed lower vividness for imagery in comparison to VWM and VSTM. However, only VSTM vividness appeared influenced by test array, with lower vividness ratings for the single-cue than either the whole display or post-cue. A 3 (Task: imagination, VWM, VSTM) $\times$ 3 (test array: whole, post-cue, single-cue) repeated measures ANOVA was conducted on mean imagery vividness. There was a main effect of task, $F(2, 54) = 25.99$, $p < .001$, $\eta_g^2 = .38$, $BF_{10} = 2.672 \times 10^{+12}$. Post hoc analysis found that VWM vividness ratings were significantly more vivid than imagination vividness ratings ($p < .001$, $d = 1.18$, $BF_{10} = 4.806 \times 10^{+11}$), VSTM vividness ratings were significantly better than imagination vividness ratings ($p < .001$, $d = 1.10$, $BF_{10} = 4.009 \times 10^{+9}$), and there was no significant difference between VSTM and VWM vividness ratings ($p = .820$, $d = .21$, $BF_{10} = .363$). There was a significant effect of test array, $F(2, 54) = 14.69$, $p < .001$, $\eta_g^2 = .09$, $BF_{10} = 1.735 \times 10^{+9}$. Post hoc analysis demonstrated that the whole test display vividness ratings were significantly better than the post-cue vividness ratings ($p = .028$, $d = .53$, $BF_{10} = 3.122$) and the single-cue vividness ratings ($p < .001$, $d = .85$, $BF_{10} = 2679.49$) and the post-cue test array vividness ratings were significantly better than the single-cue vividness ratings ($p = .003$, $d = .61$, $BF_{10} = 43.40$). There was a significant interaction between task and test array, $F(4, 108) = 13.28$, $p < .001$, $\eta_g^2 = .09$, $BF_{10} = 1.462 \times 10^{+7}$. Post hoc analysis found no significant differences for imagination between any of the test conditions ($p > .170$, $BF_{10} = .208 - 1.14$). Similarly, for the VWM task, no significant differences were found for vividness scores between any of the test conditions ($p > .594$, $BF_{10} < .438$). For the VSTM task, the whole test display produced better vividness ratings than the single-cue test ($p < .001$, $d = .104$, $BF_{10} = 2730.12$), the post-cue test produced better vividness ratings than the single-cue test ($p < .001$, $d = .95$, $BF_{10} = 802.28$), and no significant differences were found between whole test display and post-cue test ($p = .395$, $d = .29$, $BF_{10} = .585$). All post hoc analysis comparing task vividness between task for the different test conditions were significant ($p < .05$, $BF_{10} > 3.79$), apart from the post-cue condition between VWM and VSTM ($p = .091$, $d = .43$, $BF_{10} = 1.83$).

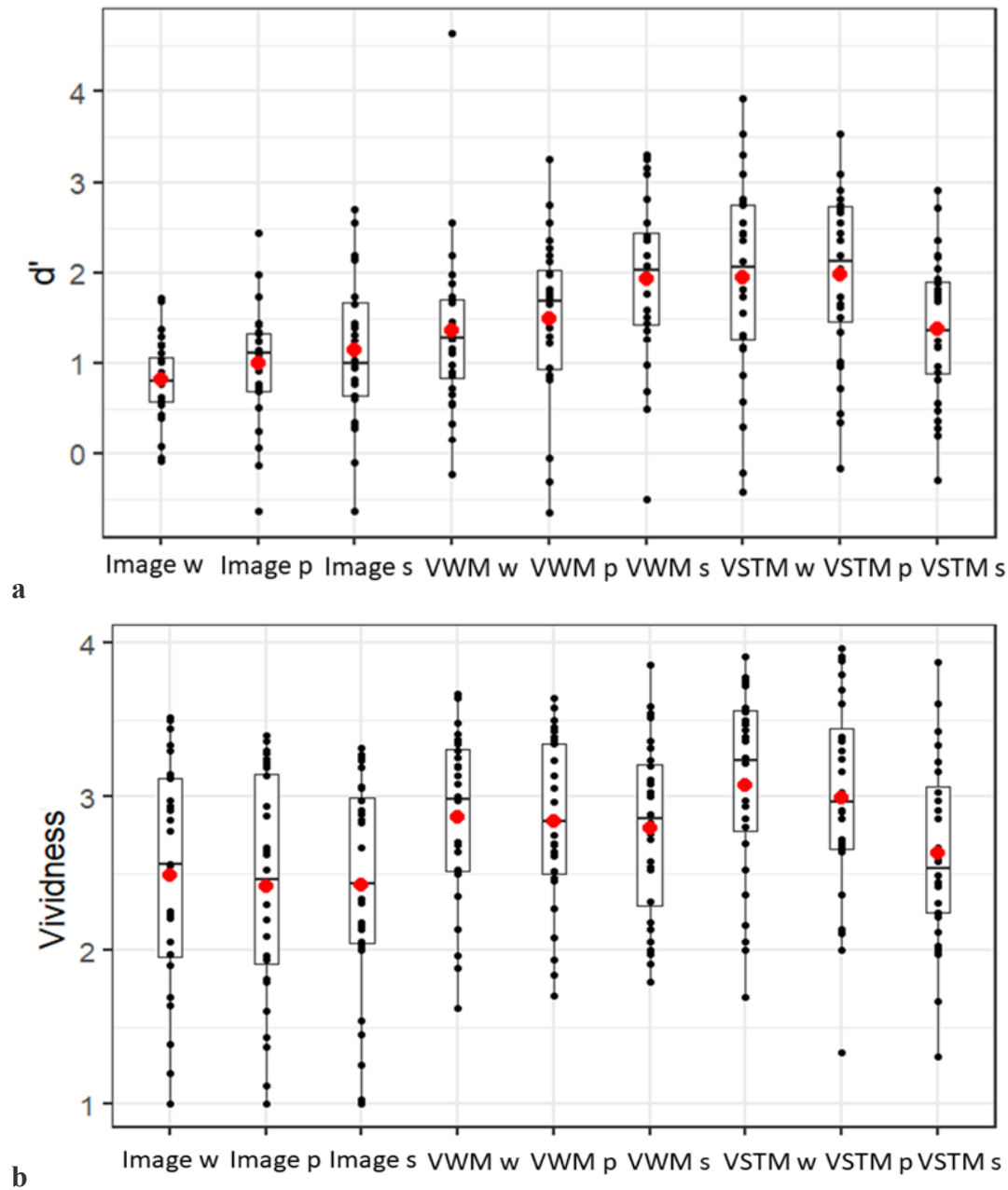


Figure 9. d' (a) and vividness (b) for each task (imagination (image), VWM and VSTM) with each test array presentation (whole display (w), post-cue (p) and single-cue(s)).

Means and standard deviations of the questionnaire measures can be found in the supplementary materials. Table 4 shows strong correlations between imagination, VWM, VSTM performance and vividness, with the exception of imagery and VWM performance and imagery and VSTM performance which did not significantly correlate. The lack of a significant correlation between imagery and VWM performance was surprising given the strong significant correlations from Experiment 1 – 3 between imagery and VWM performance. However, it is worth noting that the Bonferroni correction is very conservative, which

increases the likelihood of Type II errors. Imagery and VWM vividness also positively correlated with the VOSI. Given no significant relationship was found between self-report measures of vividness and the VOSI in Experiments 1 – 3 it is difficult to interpret the meaning of this finding.

Table 4. Correlations among imagery, VWM and VSTM d' and vividness (combined: whole display, post-cue and single-cue for each task) and questionnaires: VVIQ, VOSI and subscales.

		1	2	3	4	5	6	7	8	9	10
1	Image d'	-									
2	VWM d'	.52	-								
3	VSTM d'	.47	.74*	-							
4	Image vividness	-.07	-.39	-.25	-						
5	VWM vividness	.09	-.14	-.13	.88*	-					
6	VSTM vividness	.03	-.10	.09	.81*	.85*	-				
7	VVIQ	.22	.26	.13	-.53	-.52	-.47	-			
8	VOSI (object)	-.18	-.27	-.18	.61*	.53	.41	-.84*	-		
9	VOSI (spatial)	-.25	-.04	-.23	.54	.60*	.47	-.67*	.77*	-	
10	VOSI (total)	-.23	-.17	-.22	.62*	.60*	.47	-.81*	.95*	.93*	-

Note. $N=28$, two-tailed test; * $p < .001$ (Bonferroni correction).

Discussion

Experiment 4 explored the extent to which the type of test display influenced performance across imagery, VWM and VSTM tasks, to see whether this could account for the large performance differences across tasks seen in Experiment 1. In particular, as all the items were shown at test in Experiment 1 even though only one was cued, this might benefit the VWM and VSTM task in which greater configural processing could have occurred due to the blocks being shown during each trial. In contrast, presenting many black blocks at test may have disrupted imagination processes because no black blocks were presented during imagination generation, they were only imagined. However, overall imagery performance (d') and vividness remained lower than that of VWM and VSTM with VWM and VSTM producing similar performance and vividness. This pattern is similar to that seen in Experiment 1. The type of test array used did influence performance, however this differed in each task. For VWM performance was better in the single cue condition, with comparable performance in the other two types of test display. In contrast, VSTM performance was considerably poorer in the single-cue condition than the other two test displays which had comparable levels of performance. Unlike both VWM and VSTM, test display had no impact on imagery

performance. This suggests that configural information influences imagery, VSTM and VWM differently.

The differing impact of test array across VSTM, VWM and imagination indicate that these tasks are not underpinned by identical processes. That VSTM performance was better when configural information was presented at test (whole display and post-cue conditions) supports the relational information hypothesis (Alvarez, 2011) and the findings of Jiang et al. (2000) that VSTM is supported by global spatial relations. Indeed, when relational information at decision was removed then VSTM performance was similar to imagery performance in the single cue condition, suggesting that relational information was supporting VSTM performance in Experiment 1. In contrast, the marked performance advantage in the single cue condition for the VWM task suggests that in this task configural information was used to a lesser extent, such that presenting configural information at test actually interfered with decision making. This may likely be because the stimulus was built up sequentially, with no display having multiple black blocks in it until test.

Alternatively, there may simply be a benefit when the manner of encoding (sequential or simultaneous) matches the test used (single item or all items presented). For example, if participants did not integrate all the steps of the sequence when presented with sequential information, then it might be that performance is better for a single probe test display, as this better matches how the information was processed (a single step). Although possible, previous studies have shown a simultaneous presentation advantage for both whole display and single-cue test arrays (Lecerf & de Ribaupierre, 2005). Furthermore, in this experiment, imagery was not similarly affected with the test display having little impact on imagery performance. However, it is possible that imagery performance did not benefit from any particular test display simply because the test display included black squares whereas the stimulus build-up in imagery did not have this sensory input. It should be noted however that we used a Corsi task in Experiment 1 (see also Experiment 5). The Corsi task is an explicit measure of sequence memory, and if participants were remembering a sequence rather than an integrated whole, VWM performance/capacity should have correlated strongly with the Corsi measures. There was very little evidence for strong relationships between these measures.

Taken together these findings suggest that the type of test display used is not responsible for the performance differences between imagination and VWM in Experiment 1. Experiments 2-4 therefore rule out a number of alternative explanations of why imagery performance (and

thus observed capacity) was worse than that of VWM in Experiment 1. Worse imagery performance was not due to imagery generation taking longer as increasing the duration of each sequential step had no impact on imagery performance (Experiment 2). Neither was there any evidence that the arrow cue methodology disadvantaged the imagery task by capturing attention or being a less precise indication of where to generate an image (Experiment 3). Finally, neither was it due to the test display inadvertently supporting VWM performance or harming imagery performance. As such it appears that the capacity observed in the imagery task was reduced relative to the capacity observed for the VWM task.

It is important to recognise however, that in Experiment 1, we modelled capacity in terms of the number of objects, the assumption being that the paradigm was measuring how many objects could be imagined. However, a potential issue is that the task was more a spatial imagery task than an object imagery task (Kosslyn, 1994). We do not believe that this was the case for several reasons. First the imagery task required sequential generation of an image of an object at a given location and this should have led to objects being represented as separate objects. Secondly, manipulating the type of test in Experiment 4 showed no evidence of configural processing in imagery (unlike for the VSTM and VWM tasks). Lastly, the test in Experiment 1 was about the location of a single object, not about the configuration as a whole. Nevertheless, to address this in Experiment 5 we modified the task to focus on object features.

Experiment 5

Experiment 5 aimed to assess how varying an object feature (colour) influenced capacity estimates observed in the memory and imagination tasks. Rather than black squares, squares of varying colours were presented/imagined in the visual memory and imagery tasks respectively. As such Experiment 5 increased the complexity of the objects to be presented/imagined. At test a cue was presented at one of the presented/imagined locations and the task was to indicate whether the object at that location was the same colour as the test cue. Importantly, as this task requires binding of colour and location information it is a direct test of object imagery. This experiment was preregistered on the Open Science Framework. The preregistration, along with associated data and research materials, are available at: <https://osf.io/e34q7/>.

Method

Participants

Sample size planning was conducted using one-sided α .05 and was based on the 42 participants from Experiment 1 using the smallest effect size observed previously (set-size 5 - VWM and VSTM d'), effect size .24 (Cohen's d), and $r = .41$ (paired correlation). A priori power analysis indicated 27 participants in the current experiment to have a 98% power to detect such an effect. Twenty-seven participants (17 women) took part in the experiment aged between 18 and 39 years ($M = 27.19$, $SD = 5.72$) and had normal or corrected to normal vision. The study took approximately 150 minutes to complete. Each participant received 15 research credits for taking part as a Psychology student through the Psychology Sona System. Data were collected in 2020.

Apparatus and materials

This was the same as Experiment 1 (including the Forwards and Backwards Corsi Block Tapping tasks).

Design and procedure

Participants first completed the set of questionnaires as described in Experiment 1. Upon completion, the imagination, VWM task, VSTM task, forward Corsi block task and backward Corsi block task were counterbalanced across participants using a Balanced Latin Square Design. A repeated measures design was used. Once the imagination task or VWM task was complete; questions were administered for each of these tasks which investigated the strategies used to complete the task.

VWM task

The design and procedure were the same as Experiment 1 with the following exceptions. Participants had to remember a set of coloured blocks which were sequentially presented every two seconds within a grid (Figure 10: panel B). The number of blocks (set-size) was manipulated (3, 4, 5 and 6). There were 6 possible colours: red (1, -1, -1), green (-1, 1, -1), blue (-1, -1, 1), yellow (-1, 1, 1), purple (0.5, 0, 0.5) and cyan (0, 1, 1). Repetition of colour was allowed. However, within a trial, the same colour could not be repeated more than twice and only one colour could be repeated (the reason for this is discussed below in the design and procedure of the imagination task). At test, only a single-cue was presented in the test array. Using a test where the full test array was presented would have enabled participants to

strategically count which colours were presented in the test array. Using a single cue at test therefore eliminated this possibility. Participants were asked to judge whether this block was originally viewed or not. On half the trials the test display was the originally cued block (“same” trial), and in the other half differed such that the colour of the presented block was randomly selected from other colours previously used for objects in that trial (“different” trials). This prevented participants making a decision based on a colour they knew they had not seen in that trial. In all cases, the cued location was a location that an object had been presented in that trial.

The practice block contained 8 trials: two trials were allocated to each of the set-size conditions (3, 4, 5, 6) and of the two trials, one trial was allocated to the “same” test array presentation and two to the “different” test array presentation. Half of the trials were presented from the left of the display and the other half was presented from the right of the display, equally balanced between the “same” and “different” test array presentations.

Participants then completed 144 trials across 3 equally balanced main blocks. Within each block, each set-size (3, 4, 5, 6) was presented 12 times, there were 6 same trials and 6 different trials, all presented in a random order. Participants were able to take a short break after completing each block.

Imagination task

The design and procedure were the same as Experiment 1, with the same alterations as presented in the above design section for VWM with the following exceptions. The to-be imagined coloured blocks were cued with a black (-.2, -.2, -.2) letter on the column and a black (-.2, -.2, -.2) arrow on the row of the grid indicating participants to imagine a coloured block at the intersection of these cues. The colour to be imagined was set by the letter presented on the column (R = red, G = green, B = blue, Y = yellow, P = purple and C = cyan). The cues therefore designated a location and a colour to be imagined (Figure 10: panel A). Repetition of one colour per trial was allowed because there are only a limited amount of highly discernible colours that can be allocated to specific letters. For example, B = blue, therefore B for black could not be used as one of the cued colours. In addition, some letter cues are similar in structure and so were avoided. For example, O could represent the colour orange, however the structure of the letter O is similar to that of G.

VSTM task

The design and procedure were the same as Experiment 1, with the same alterations as presented in the above design section for VWM with the following exception. Coloured blocks were presented simultaneously during the sample array for 100ms (Figure 10: panel C).

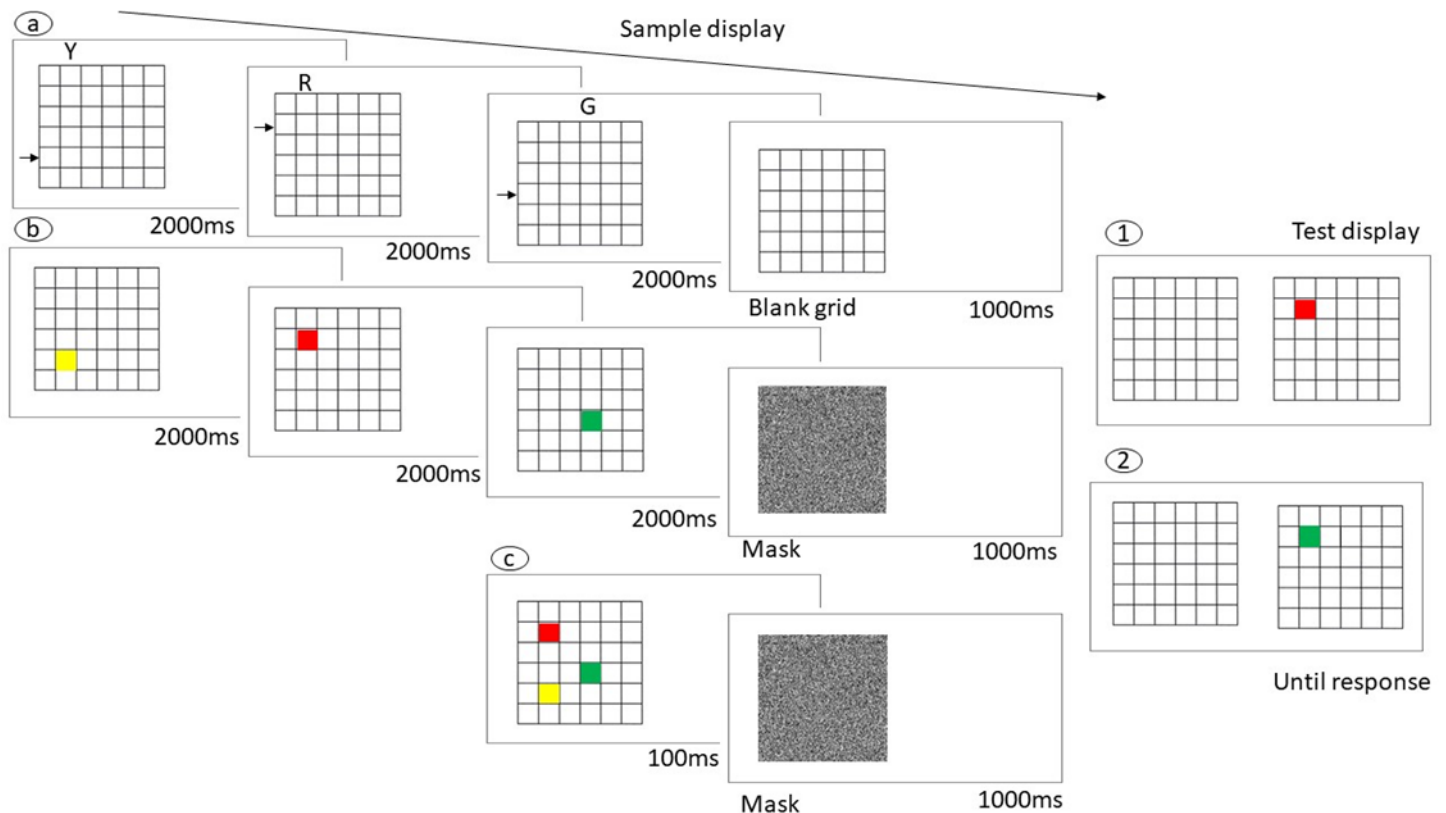


Figure 10. Stimuli and sequence of events in visual tasks with set-size 3 items. Panel A: Imagination task - participants were asked to create coloured blocks in their minds eye at the intersection between the letter cue and the arrow cue (e.g., at the start of the sequence participants should imagine a yellow block at the intersection of the Y and the arrow). Panel B: VWM task - coloured blocks were visibly presented on screen in a sequential manner. Panel C: VSTM task - all coloured blocks were presented simultaneously. 1) same trial 2) change trial.

Results

Data were trimmed as per Experiment 1. Five participants reported using alternative cognitive strategies (e.g., ‘I gave each black block a specific number and verbally named it colour to encode its location’). Statistical analysis was re-conducted with these participants

removed. Since no evidence of a difference was found between the data with all participants and with the five participants removed, only the analysis with all participants data is reported.

Imagination, VWM and VSTM results

Figure 11 shows d' (a) and vividness (b) for all tasks as a function of set-size. In all tasks d' and vividness decreased as set-size increased. VWM performance was better than that of imagination and VSTM, and VSTM performance was visually better than imagination. A 3 (task: Imagination, VWM, VSTM) $\times$ 4 (set-size: 3, 4, 5, 6) repeated-measures ANOVA was conducted on d' and yielded an effect of task, $F(2, 52) = 3.66, p = .032, \eta_g^2 = .06, BF_{10} = 1.31$. Post hoc analysis found that VWM performance was significantly better than imagination performance ($p = .008, d = .64, BF_{10} = 2009.05$) (note that whilst the Bayes Factor for the main effect of task was inconclusive, the Bayes Factor on the post hoc analysis shows decisive support for a performance difference). There was no significant difference between VSTM performance and imagination performance, ($p = .562, d = .26, BF_{10} = .903$) and no significant differences between VWM and VSTM performance ($p = .735, d = .23, BF_{10} = .709$). Note that these analysis suggest it is hard to interpret the VSTM performance which sat between the imagination and VWM performance, but was significantly different from neither, with little clear evidence for the null. There was an effect of set-size, $F(3, 78) = 80.11, p < .001, \eta_g^2 = .39, BF_{10} = \infty$, signifying that as set-size increased, performance decreased, and no interaction effect, $F < 1, BF_{10} = .06$. The same ANOVA on vividness ratings yielded a non-significant effect of task, $F(1.58, 41.18) = 1.51, p = .231, \eta_g^2 = .04, BF_{10} = .043$, an effect of set-size, $F(1.25, 32.41) = 65.34, p < .001, \eta_g^2 = .31, BF_{10} = .5459 \times 10^{+14}$ signifying that as set-size increases vividness scores decrease, and no interaction effect, $F < 1, BF_{10} = .035$.

Estimating capacity

The same single-cue WoMMBAT model (Morey & Morey, 2011) was used as in Experiment 1. An estimate of capacity for VWM, imagination and VSTM was generated using the same parameters, the same modelling procedure and the same two models (participant variance model and participant by task) as Experiment 1. The first model contained participant as a factor and the second model contained both participant and task as factors. The modelling revealed that the participant by task interaction model provided the lowest DIC (11187.85) compared to participant as a factor (11391.20), indicating that task is important in order to understand differences in k (see the Supplemental Materials for full details on modelling

procedure and statistics). This model yielded a mean observed capacity (k) of 2.79 items in the VWM task, 2.57 items in the VSTM task and 2.31 items in the imagery task. Estimates of k for each participant for each task are shown in Figure 11 (c). Interestingly, there was significant overlap between the inter-quartile ranges of k estimates between the three tasks. A one-way repeated-measures ANOVA was conducted to compare the effect of task (imagination, VWM, VSTM) on k . There was a significant effect of task, $F(2, 52) = 3.41, p = .041, \eta_g^2 = .12, BF_{10} = 1.42$. Post hoc analysis found that VWM k was significantly higher than imagination k ($p = .006, d = .66, BF_{10} = .18.98$), there was no evidence of a difference between VSTM k and imagination k ($p = .450, d = .29, BF_{10} = .541$), and no evidence of a difference between VWM and VSTM k ($p = .999, d = .18, BF_{10} = .309$). Note that whilst there was not clear support for the main effect of task from the Bayesian analysis, there was very clear support for a difference between the capacities for VWM and the imagination. However, the support for the null effect between VSTM and imagination capacities was inconclusive.

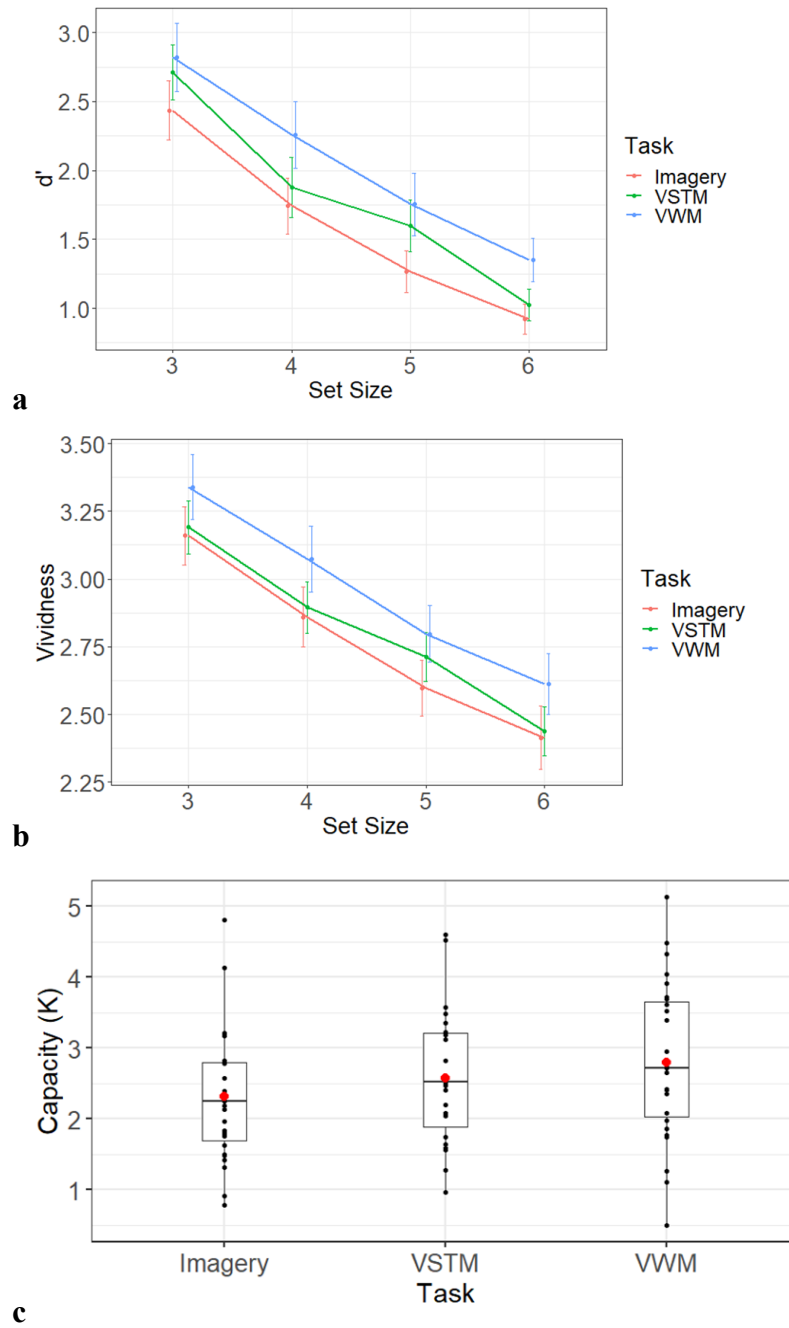


Figure 11. d' (a) and vividness (b) as a function of set-size for the imagery, VWM and VSTM tasks (error bars indicate standard error). Capacity estimates (k) for participants in the imagery, VSTM and VWM tasks (c). The boxplot means are represented by the red dot.

Means and standard deviations of the questionnaire measures can be seen in the supplementary materials. Strong positive correlations were found between between imagery and VWM (Table 5). This is consistent with imagery and VWM sharing similar processes and

corroborates the findings from Experiment 1 - 3. Surprisingly, despite research suggesting VWM and VSTM use a similar short-term maintenance component (for a review see, Aben et al., 2012) and Experiments 1 and 4 showing a strong relation between performance in these tasks, there was no significant correlation found between VWM and VSTM nor between imagery and VSTM. The forward Corsi span only significantly correlated with VWM d' .

Table 5. Correlations between k , d' , and vividness for each task (Imagination (Image), VWM and VSTM), Forward and backwards corsi span (CS) and corsi total (CT), VVIQ and VOSI (spatial, object and total).

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Image k	-															
2	VWM k	.76*	-														
3	VSTM k	.49	.35	-													
4	Image d'	.98*	.71*	.49	-												
5	VWM d'	.75*	.98*	.35	.71*	-											
6	VSTM d'	.46	.35	.99*	.46	.34	-										
7	Image vividness	.20	.13	-.16	.25	.19	-.21	-									
8	VWM vividness	.33	.48	-.16	.35	.54	-.19	.67*	-								
9	VSTM vividness	.04	-.07	.57*	.04	-.10	.58*	.07	-.04	-							
10	Forward (CS)	.37	.52	.18	.35	.55*	.17	.43	.29	.01	-						
11	Forward (CT)	.38	.48	.25	.36	.49	.24	.45	.26	.12	.95*	-					
12	Backwards (CS)	.38	.37	.51	.37	.34	.49	-.01	-.07	.14	.31	.41	-				
13	Backwards (CT)	.42	.37	.50	.41	.36	.48	-.07	-.08	.11	.36	.47	.84*	-			
14	VVIQ	.08	.19	.10	.01	.11	.14	-.38	-.27	-.21	-.01	.03	.16	.16	-		
15	VOSI (object)	-.23	-.20	-.11	-.15	-.11	-.15	.42	.33	.22	-.03	-.07	-.28	-.27	-.79*	-	
16	VOSI (spatial)	-.13	-.17	-.15	-.08	-.06	-.19	.42	.16	.10	.19	.10	-.30	-.18	-.50	.56*	-
17	VOSI (total)	-.20	-.21	-.15	-.13	-.09	-.19	.48	.27	.17	.10	.03	-.33	-.25	-.72*	.86*	.91*

Note. $N = 27$, two-tailed test; * $p < .0003$ (Bonferroni corrections).

Comparisons with Experiment 1

Given that Experiment 5 focused additionally on object features (colour) as well as location, whereas Experiment 1 focused on object location only, analyses were run to compare

d' , vividness and observed capacity estimates between these tasks. All participant data was used in all analyses. A 2 (experiment: objects, colours) x 4 (set-size: 3, 4, 5, 6) mixed ANOVA was used in the analysis on imagery and VWM. A 2 (experiment: object, colours) x 2 (set-size: 3, 5) mixed ANOVA was used in the analysis on VSTM. Finally, a 2 x (experiment: object, colour) x 3 (task: imagination, VWM, VSTM) mixed ANOVA was used in the analysis on imagery, VWM and VSTM k .

A 2 x (experiment: object, colour) x 4 (set-size: 3, 4, 5, 6) mixed ANOVA was used in the analysis on imagery d' with the experiment as a between subjects independent variable yielded no effect of experiment, $F(1, 67) = 1.05, p = .309, \eta_g^2 = .01, BF_{10} = .577$. There was an effect of set-size $F(3, 201) = 50.12, p < .001, \eta_g^2 = .19, BF_{10} = .3.753 \times 10^{+14}$, signifying that as set-size increased, performance decreased. There was also a significant interaction effect between experiment and set-size, $F(3, 201) = 2.76, p = .043, \eta_g^2 = .01, BF_{10} = 1.073$. Post hoc analysis found that performance decreased in both Experiment 1 and 5 between set-size three and four (both p 's $< .001$, both $BF_{10} < .441.51$). However, in Experiment 1, performance did not decrease significantly between set-size 4 and 5 ($p = .141, BF_{10} = 1.97$), whereas performance did decrease significantly in Experiment 5 between set-size 4 and 5 ($p = .018, BF_{10} = 7.72$). Importantly, the direct comparisons between Experiment 1 and Experiment 5, for each of the set-sizes: 3, 4, 5 and 6 did not reach significance (all p 's $> .100$, all $BF_{10} < .817$). A 2 x (experiment: object, colour) x 4 (set-size: 3, 4, 5, 6) mixed ANOVA was used in the analysis on VWM d' with the experiment as a between subjects independent variable yielded no effect of experiment, $F < 1, BF_{10} = .277$. There was an effect of set-size $F(3, 201) = 47.76, p < .001, \eta_g^2 = .16, BF_{10} = 3.002 \times 10^{+15}$, signifying that as set-size increased, performance decreased, and no interaction effect, $F < 1, BF_{10} = .118$. A 2 x (experiment: object, colour) x 4 (set-size: 3, 4, 5, 6) mixed ANOVA was used in the analysis on VSTM d' with the experiment as a between subjects independent variable yielded an effect of experiment, $F(1, 67) = 5.28, p = .025, \eta_g^2 = .06, BF_{10} = 2.48$. There was an effect of set-size $F(3, 201) = 87.06, p < .001, \eta_g^2 = .21, BF_{10} = 7.849 \times 10^{+10}$, signifying that as set-size increased, performance decreased and no interaction effect, $F < 1, BF_{10} = .993$.

A 2 x (experiment: object, colour) x 4 (set-size: 3, 4, 5, 6) mixed ANOVA was used in the analysis on imagery vividness with the experiment as a between subjects independent variable yielded no effect of experiment, $F < 1, BF_{10} = .418$. There was an effect of set-size $F(1.65, 110.25) = 110.17, p < .001, \eta_g^2 = .19, BF_{10} = 2.310 \times 10^{+14}$, signifying that as set-size

increased, vividness decreased, and no significant interaction effect, $F(3, 201) = 1.56, p = .218, \eta_g^2 < .01, BF_{10} = .335$. A 2 x (experiment: object, colour) x 4 (set-size: 3, 4, 5, 6) mixed ANOVA was used in the analysis on VWM vividness with the experiment as a between subjects independent variable yielded no effect of experiment, $F(1, 67) = 1.46, p = .232, \eta_g^2 = .02, BF_{10} = .800$. There was an effect of set-size $F(1.89, 126.45) = 101.72, p < .001, \eta_g^2 = .17, BF_{10} = \infty$, signifying that as set-size increased, vividness decreased, and no interaction effect, $F(1.89, 126.45) = 2.65, p = .078, \eta_g^2 < .01, BF_{10} = 1.213$. A 2 x (experiment: object, colour) x 4 (set-size: 3, 4, 5, 6) mixed ANOVA was used in the analysis on VSTM vividness with the experiment as a between subjects independent variable yielded an effect of experiment, $F(1, 67) = 11.03, p = .001, \eta_g^2 = .12, BF_{10} = 20.54$. There was an effect of set-size $F(1, 67) = 99.73, p < .001, \eta_g^2 = .15, BF_{10} = 4.126 \times 10^{+11}$, signifying that as set-size increased, vividness decreased and no interaction effect, $F(1, 67) = 1.50, p = .225, \eta_g^2 < .01, BF_{10} = 1.726$.

Estimating capacity

A 2 x (experiment: object, colour) x 3 (task: imagination, VWM, VSTM) mixed ANOVA on observed capacity (k) with the experiment as a between subjects independent variable yielded an effect of task, $F(2, 134) = 28.58, p < .001, \eta_g^2 = .10, BF_{10} = 2.090 \times 10^{+12}$, with k being significantly lower in imagery than VWM ($p < .001, BF_{10} = 1.058 \times 10^{+8}$) and VSTM ($p < .001, BF_{10} = 2.772 \times 10^{+6}$), with no evidence of a difference between VWM and VSTM ($p = .999, BF_{10} = .222$). There was an effect of experiment $F(1, 67) = 4.35, p = .041, \eta_g^2 = .04, BF_{10} = 1369.52$, with Experiment 1 producing higher overall k than Experiment 5. There was also a significant interaction effect between task and experiment, $F(2, 134) = 11.63, p < .001, \eta_g^2 = .05, BF_{10} = 3131.06$. Post hoc analysis found significant differences between experiments for VSTM k ($p < .001, d = 1.15, BF_{10} = 248.29$), with k greater in Experiment 1, with no significant differences found between experiments for imagery k ($p = .589, d = .24, BF_{10} = .286$) or VWM k ($p = .097, d = .46, BF_{10} = .835$) (see Figure 12).

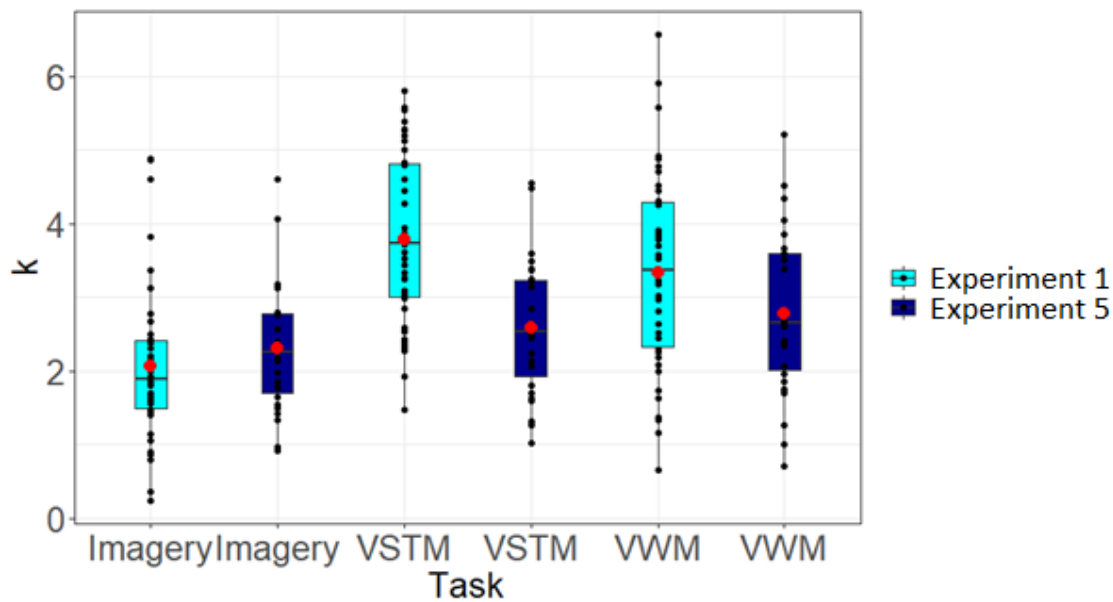


Figure 12. Capacity estimates (k) for each participant in the object experiment (Experiment 1) and colour experiment (Experiment 5) for imagination (Imagery), VSTM and VWM tasks. Large red dots show the mean observed capacity.

Discussion

This experiment was designed to determine if imagery and visual memory (VWM and VSTM) observed capacity is influenced by increasing object complexity. Testing the fidelity of memory or imagery of a feature (colour) of an object at a particular location also created a more stringent test of the capacity of object imagery. Relative to Experiment 1, the major difference was that observed capacity of VSTM declined significantly, such that it did not differ with that for imagery. This contrasts with Experiment 1 where VSTM observed capacity was significantly greater than that of imagery. In Experiment 5, a single cue needed to be used to prevent decisions being based on whether or not there was a repeated colour at test in “different” trials. As such it may well be that removing all the stimuli from the display clearly harmed VSTM performance and observed capacity in Experiment 5 due to the loss of configural information, and that it was not the increased complexity of the objects that harmed performance. This finding is consistent with Experiment 4, in which a single cue at test rendered performance in the VSTM task similar to that of imagery. Alternatively, given that the focus of attention appears crucial when binding information in memory (Awh & Jonides, 2001; Fougny & Marois, 2006; Treisman & Zhang, 2006; Wheeler & Treisman, 2002; Woodman & Luck, 2004), it is possible that the short simultaneous presentation of objects in

the VSTM task limited the extent to which binding of location and colour could be performed (Yamamoto & Shelton, 2009). However, given the results of Experiment 4, it seems more likely that the VSTM task was impacted by the lack of configural information. Despite VSTM observed capacity decreasing in Experiment 5, it was neither significantly different from VWM nor imagination. Although the absence of a significant difference in observed capacity between VSTM and VWM means that firm conclusions cannot be drawn, it is of note that the design of the task in Experiment 5 eliminated a clear difference in apparent observed capacity between imagination and VSTM. One possibility is that it was the requirement to bind colours and locations that was responsible for this elimination of capacity differences. Another possibility is that the single post-cue was responsible for this elimination of capacity differences, similar to the results seen in Experiment 4 for the single post-cue conditions for imagination and VSTM. In other words, the capacity of VSTM may resemble that of imagination when the objects are more complex or when configural processing is made more difficult.

Estimates of observed capacity for VWM did not appear to decrease significantly in Experiment 5 relative to Experiment 1, although the Bayes Factor was inconclusive. This suggests that object complexity has little impact on VWM observed capacity. These findings are aligned with previous research suggesting that VWM capacity is not influenced by feature complexity (Luck & Vogel, 1997; Luria & Vogel, 2011), although research has suggested that encoding multiple dimensions of an object reduces performance in VWM tasks (Shin & Ma, 2017; Wang et al., 2017; Wheeler & Triesman, 2002). In terms of imagination, object complexity did not appear to impact observed capacity, and critically, observed capacity for imagination was significantly lower than VWM capacity. We discuss the theoretical implications of this below.

General Discussion

Prior research suggests that imagery and VWM are related (Harrison & Tong, 2009; Keogh & Pearson, 2011, 2014; Kosslyn & Thompson, 2003) and show some similarity in neural activation patterns in the visual system (Albers et al., 2013). Whilst it is known that imagery is capacity limited (Keogh & Pearson, 2017, Ceja & Franconer, 2023; Balaban & Ullman, 2024), the capacity of imagery in terms of the number of imagined objects has never previously been measured. In five experiments we present a paradigm that enables measurement of capacity in imagery, VWM and VSTM tasks using directly analogous tasks

and explore whether performance differences between these tasks were due to methodological factors including; time to build up an image (Experiment 2), the cue used in the imagery task capturing attention or leading to a less precise location of the generated image (Experiment 3) or whether the inclusion of all objects at test selectively enhanced/impaired visual memory/imagery respectively (Experiment 4). The results of Experiments 2-4 established that performance differences between tasks in Experiment 1 did not appear to be caused these features of the design, Experiment 5 extended the findings of Experiment 1 by exploring whether similar findings were found when the complexity of the stimuli were increased. In Experiment 5 the blocks presented/imagined were coloured and the test required correct binding of colour and location information. The central finding was that the observed capacity of VWM remained above that of imagery. Taken together these results show that across several tasks, the capacity observed for imagination is consistently lower than that of VWM.

Observed capacity estimates revealed similar capacities in VSTM and VWM (~3 items) to previous estimates (Cowan, 2001; Luck & Vogel, 1997). In comparison, the observed capacity for imagery was lower (~2 items). It is important to note that imagery observed capacity was significantly below that of VWM capacity in both Experiment 1 and 5, but only lower than of VSTM in Experiment 1. We return to this finding below, but do not believe it detracts from the main finding that the capacity we observe in the imagery tasks appears lower than that of VWM tasks. Unlike previous tasks that have explored the relationship between imagery and VWM (Albers et al., 2013), the measurement of performance was not influenced by sensory or mnemonic contamination from the to-be-imagined stimuli. Importantly, the methodology also enabled modelling of capacity in a set of tasks that were directly analogous, for the first time providing a rigorous assessment of how capacity varies between imagery and short-term visual memory tasks.

The lower observed capacity of imagery relative to VWM indicates the importance of the sensory system in supporting memory maintenance and rehearsal (Cattaneo et al., 2009; Postle, 2016). Although the construction and maintenance of imagery can activate sensory systems (Albers et al., 2013; Lee et al., 2012) this appears to be insufficient to enable robust representations of multiple items to be created and maintained at the same level as when there is initial sensory input. The notion that retrieval invokes the sensory processes involved at encoding has been repeatedly suggested (Barsalou, 2008) with evidence for similar cortical activations when encoding and retrieving the same information (Nyberg et al., 2000) and

similar processing dynamics (Kent et al., 2014). In Experiments 1 and 5 the percentage of VWM capacity that remains when taking account of imagery capacity provides an estimation of the extent to which VWM is supported by sensory processes, given that the major difference between these tasks was the sensory input. This suggests that initial sensory input supports around 17-35% of VWM capacity in these tasks. Whilst prominent VWM theories mention the role of sensory input (e.g., Baddeley & Hitch, 1974; Baddeley, 2000; Barrouillet et al., 2004; Cowan, 2016) this is relatively minimally described. For example, Cowan et al.'s (2020) Embedded Process Model discusses activation of a brief sensory store and describes how incoming information can be held in the focus of attention over a period of time. Here we show the importance of that initial sensory store, for the first time showing the extent to which this supports VWM capacity.

Why is imagery capacity lower than that of VWM?

It is important to consider the different explanations as to why the observed capacity of imagery appears lower than that of VWM. Note that throughout we have also used the term observed capacity to denote that this was the capacity we observed in these tasks. We differentiate this from the theoretical capacity of the system and discuss this further below. In our tasks both imagery and VWM require the maintenance of a representation but differ in the source of the information on which that representation is based, with the source being self-generated (imagery) or visual input (VWM). It is possible that the observed capacity differences arise because of this distinction and how these sources of information are processed and representations formed or in the resources required to maintain representations derived from different information. It may also be that observed capacity differences arise because imagery and VWM rely on different systems which differ in their capacity. We explore each of these in turn.

A reduced imagery capacity relative to VWM capacity may be due to the differences in activation levels throughout the visual cortex. Early visual cortex has been shown to be less active during imagery than during visual perception, with imagery and perception showing more similar activation to one another moving up the processing stream (Pearson, 2019; Pearson et al., 2015; also see Bartolomeo et al., 2020). Similarly, overall activation of early visual areas during imagery remains relatively low compared to that of VWM (Albers, et al, 2012). Nevertheless, the pattern of neural activity show similarities between imagery, visual perception and VWM (Albers et al., 2013; Cichy et al., 2012; Lee et al., 2012). For example, a

multivariate pattern classifier (MVPC) was able to decode low level features in areas V1 and V2 for both imagery and VWM stimuli, suggesting that they share a similar pattern of activation in the early occipital areas (Albers et al., 2013). As such the visual system has been conceptualised as a ‘dynamic blackboard’ which can be used to form visual representations from both low-level visual input and higher-level top-down input. In the case of imagery, this top-down input cascades backwards through the visual system, using prior knowledge to activate sensory and spatial representations to support generation of the image (Albers et al., 2013; Pearson, 2019). As imagery is reliant on top-down activation only, reduced activation of early visual areas relative to VWM in which top-down and bottom-up processes determine activation may make it difficult to generate and maintain these images. Alternatively, it may be that greater reliance on recruitment of top-down processes for the generation and maintenance of internally generated images (Schacter et al., 2007; Schlegel et al., 2013) may not reduce activation in early visual areas per se but still produce less precise activation patterns (Albers et al., 2013; Dijkstra et al., 2017). In either case, the lower capacity observed in imagery may reflect a reduced ability to process and store these less stable representations in a way that is immune from interference.

It may also be that the reduced observed capacity in imagery is a result of the greater top-down processing required in comparison to stimulus driven tasks (Dijkstra et al., 2017), such as the recruitment of attentional processes (Mayer et al., 2007; Zanto et al., 2011). Dijkstra et al. (2017), for example, investigated imagery and visual perception using dynamic causal modelling and found differences in connectivity whereby stronger BOLD coupling activations from top-down connections were found when imagining objects relative to perceiving objects. This could be due to the extra attentional resources needed to generate and maintain imagined experiences in the absence of visual input (Dijkstra et al., 2017). Lower observed capacity in imagery could therefore reflect the increased requirement of top-down attentional processes to generate and maintain images when these same resources underpin system capacity. Partially speaking against this as an explanation here is that allowing more time for a limited capacity attentional system to support imagery by increasing the time available for image generation had no effect on performance (Experiment 2). It remains a possibility however that more limited capacity attentional resources are required for ongoing maintenance of images rather than maintenance of VWM representations.

The above explanations are theoretically consistent with the notion that the capacity limits of imagery arise within the same general architecture that is used to support VWM. In this conceptualisation, the overall system capacity is the same, but in imagery the additional

processing requirements when activation is top-down only makes it harder to maintain and store information within this system. This could be due to the resources that support a limited capacity storage system also being used more to generate and maintain imagery (e.g., attention), or because the type of activation produced in image generation is harder to maintain and store within this limited capacity system. If imagery and VWM were to utilise the same general architecture, this would imply that the capacity of this system is the same for both tasks. Whilst this may be the case, it seems that, at least in the tasks presented here, during imagery the ability to utilise the full system capacity is impaired such that in reality the ability to store and maintain object representations is reduced for imagery compared with VWM. In other words, the system capacity is stable, but the observed capacity differs. It is of course, difficult to tell apart such a perspective from one in which the difference in observed capacities is due to the two tasks involving differing capacity limited systems. The only way of doing so is to show that, under certain circumstances imagery capacity can be increased to be comparable with VWM capacity. In Experiments 2-4 we demonstrate that the reduced capacity observed for imagery was not an artefact of different aspects of the methodology used that selectively impaired the imagery task.

Another possibility then is that imagery capacity is lower than that of VWM because they rely on separate systems with different capacities. For example, whilst a visual cache (Logie 1995, 2003, 2011) is responsible for short-term storage of representations of perceived stimuli, Pearson (2001) suggests that imagery relies on a separate visual buffer that images are generated into. This buffer decays quickly unless the image is preserved, which requires additional processing demands. The buffer is seen as a more active system in which visual representations can be manipulated and is contrasted with the more passive visual cache (Andrade et al., 2002; Borst et al, 2012a; Logie, 1995; 2003; 2011; Quinn & McConnell, 2006; Van der Meulen et al., 2009). It is possible that this visual buffer may have a more limited capacity than the visual cache. If that were the case, then even if the visual buffer generated images that were then transferred to the same visual cache as that used in VWM, the result would be a lower overall capacity for imagery.

If imagery and VWM are underpinned by similar mechanisms, then we might expect to see strong relationships between measures of performance. Whilst our correlational data shows a strong relationship between imagery and VWM in terms of performance and capacity, there are some aspects that could be interpreted to suggest differences. For example, scores in the Corsi task, normally taken to be a measure of VWM capacity, appeared to correlate more strongly and reliably with imagery capacity than it did with visual memory capacity in

Experiment 1 (this was not the case in Experiment 5 where no systematic patterns of correlations were observed between these measures). Importantly, the Corsi task requires reconstruction of the sequence in which blocks are highlighted in the mind's eye. In the VWM task presented here, the sequence was not task relevant even though it might have been encoded as part of a memory strategy. Reconstruction of a dynamic event may well involve processes additional to that required for maintenance and storage of non-dynamic information, and these may also relate to processes required in imagery, such as image generation. Indeed, activation patterns in imagery become more similar to simple perceptual tasks the further up the visual hierarchy where more top-down processing is being integrated into a representation (Pearson, 2019), suggesting there may be different processes underpinning imagery and sensory processing at lower levels of processing.

Taken together, our findings do not provide direct evidence that imagery and VWM are underpinned by the same, or different systems. They do, however, offer compelling evidence that, despite imagery and VWM being strongly related, the observed imagery capacity is lower than that of VWM. As such our findings appear broadly consistent with the recent Reverse Hierarchical Model of imagery (Dijkstra, 2024; Pearson, 2019; also see Ahissar & Hochstein, 2004). This model sees VWM, and imagery as strongly related and was developed to account for data on the brain networks involved in imagery. It suggests that imagery involves an interconnected network that spans from the frontal cortex to the occipital regions, which works in reverse order to that of perceptually driven images. In the first stage, the frontal cortex begins the process of generating an image. The second stage involves activation of the medial temporal lobe, which is associated with memory encoding and memory storage. Finally, the model posits that neural activation then propagates to the early visual cortex and also the parietal cortex if the representation has spatial properties. As noted above, the additional processing, maintenance and storage demands of this system may underpin the reduced capacity of imagery that we observed. The model also suggests that the overlap between imagery and perception is greater at higher level brain areas, because that is where the trigger to generate the image comes from (frontal and medial temporal lobe) and so it makes sense that activation patterns there are stronger and more perception like. Viewed in this way, the results here appear to demonstrate just how important early visual areas are in supporting the creation and maintenance of multiple items in imagery.

Imagery capacity as fixed slots?

There has been much debate in the literature as to how capacity limitations arise in visual memory (for review, see Oberauer et al., 2016). Two of the main theories to account for the capacity of VWM are slot-based models (e.g., Awh et al., 2007; Cowan, 2001; Luck & Vogel, 1997) and resource-based models (e.g., Alvarez & Cavanagh, 2004; Bays & Husain, 2008; Franconeri et al., 2013; Wilken & Ma, 2004; see van den Berg et al., 2014 for comparison between models). The purpose of the work reported here was to directly compare modelled capacity estimates for imagination and different short-term visual memory tasks. Doing so required creating a new paradigm in which performance limitations in imagery and short-term visual memory could be assessed in comparable tasks. The modelling approach used (Morey & Morey, 2011) was selected because it was the approach that we believe best fit this new paradigm. Whilst this approach produces estimates of both VWM and imagery capacity in terms of the number of items, we do not claim that capacity limitations are due to the number of slots or due to a general limited capacity resource. Rather, the findings simply indicate that imagery has a lower observed capacity in comparison to VWM. Indeed, our own work in other visual tasks views attention as a limited capacity resource that can be split amongst multiple objects flexibly (Howard & Holcombe, 2008) or used to flexibly weight visual information in decision (Guest & Lamberts, 2010, 2011). Recently, tasks showing that imagery is capacity limited have used procedures in which imagery is measured using continuous measures, and these may be more appropriate to testing between slot-based or resource-based models of imagery (Keogh & Pearson, 2017; Balahan & Ullman, 2024).

VSTM and imagery

Across all experiments, performance was significantly better in the VWM task than the equivalent imagery task. This was not the case for VSTM however. VSTM performance was only better than the equivalent imagery performance when the test displays included all the blocks initially presented. Where this was not the case and memory was probed by a single cue (Experiment 4 and Experiment 5), VSTM performance was equivalent to that of imagination. This pattern of results was quite striking and suggests that in the VSTM task (where all the blocks are presented simultaneously for a brief time before being masked) configural information was used to support memory such that when this configural information was absent at test, performance worsened. This has some important considerations for memory research insofar as studies need to be careful as to the extent their design taps into memory for configurations rather than individual objects. Of additional interest is that performance for

VSTM fell to levels equivalent to that observed in imagination. As noted previously, attention may be critical for binding information to location in memory (Triesman & Zhang, 2006) and the brief presentation of stimuli in the VSTM task may have made this more difficult and made participants more reliant on configural information. That performance was similar for VSTM and imagery when no configural information could be used at test raises the possibility that attention also mediated the ability to bind images to locations, and this process may be more resource intensive when activation levels in early visual areas are diminished.

Object and Spatial Imagery

Although imagery is often thought of as a single construct, there is evidence to suggest a distinction between object imagery (imagining an object) and spatial imagery (focusing on the spatial relations between things). For example, Kosslyn (1994) suggested they have a different neural basis, with object imagery involving early visual areas (e.g., V1) and spatial imagery involving posterior parietal cortex. There is also evidence that aphantasics who have poor object imagery have unaffected spatial imagery (Dawes et al., 2020), and that these different types of imagery relate differently to memory (e.g., event memory) (Sheldon et al., 2017). However, pure spatial and object imagery tasks are rare; typically imagery tasks involve both spatial and object imagery (Kosslyn et al., 1997), and various studies have found that both object and spatial imagery tasks lead to activation in ventral and dorsal pathways (Ishai et al., 2000; Kosslyn et al., 1997; Lambert et al., 2002; Mazard et al., 2004). It is therefore difficult to separate spatial and object imagery, as often tasks that draw primarily on object processing require some spatial processing and vice versa. Similarly, in our experiments it is likely that both spatial and object-based processes are at play, in both memory and imagery. As the purpose of the studies was to assess the number of items that could be imagined and relate this to the capacity of short-term visual memory, it is important to consider the extent to which performance in the tasks was driven by spatial, rather than object imagery.

First, it is important to note the similarity of our paradigm to tasks that have been used to measure capacity in terms of the number of items in short-term visual memory (Morey et al., 2011). Using a similar paradigm and thus extending it to imagery is therefore appropriate in terms of trying to establish commonalities between the capacities of short-term visual memory and imagery. The imagery task used also bears significant resemblance to established visual tasks. For example, some visual imagery tasks require participants to visualise simultaneously

presented squares within a matrix. Probed items are then presented in a blank matrix at test and participants decide if the probed items fall within the pattern that was visualised (Kosslyn et al., 1990). Other variations of this task have been developed such as the Cumulative Imagery Task (Avon & Sestieri, 2005). In this task participants are instructed to build up a pattern in their minds using black squares that are sequentially presented within a matrix. At test, a pattern is briefly shown, and participants are asked to decide if the pattern is the same or different to what they were asked to construct. As such our paradigm also resembles paradigms that have been used to study imagery. That is not to say these tasks do not draw on both spatial and object level imagery, merely that the task here is relatable to established imagery tasks. More recently, other methods to measure imagery have used priming of the visual system to explore the impact on a secondary binocular rivalry task (e.g., Chang et al., 2013; Keogh & Pearson, 2011, 2014; Pearson et al., 2008; Pearson et al., 2011; Sherwood & Pearson, 2010). These have provided unique insights, but also probably tap into both object and spatial processes, as the focus on low-level priming of the visual system makes distinguishing between spatial and object imagery non-trivial.

In the experiments reported here, the stimuli are built up sequentially (imagery and VWM) and the test is on a single object. As such there is good reason to suppose that the stimuli were imagined as a set of objects. Evidence from the experiments supports this. Notably, In Experiment 4 there were no differences between the types of test display on imagery performance, suggesting that configural processing, something more associated with spatial imagery, was relatively limited. Importantly, the VSTM task did show a clear reduction in performance when moving from using the post-cue or whole display at test to a single stimulus, and a clear explanation for this was the importance of spatial information at presentation when the black squares were presented simultaneously. Most importantly however is that Experiment 5 showed lower observed capacity for imagery compared to VWM and in this experiment the task was to identify whether the colour of the test item was correct. As such, this was a test of feature information of an object at a location. Therefore, we believe the current set of experiments provides insights predominantly about object-based imagery, whilst recognising that, as we asked participants to imagine objects at specific locations, spatial imagery will have been drawn upon to support performance to some extent.

The Role of Memory

One issue with a lot of the literature exploring the relation between memory and imagery, is the extent to which the studies test the systems underpinning imagery versus those underpinning memory. For example, in studies exploring the relations between these tasks at the neural level, participants need to be trained on a subset of images, such as a particular type of lamp (e.g., Lee et al., 2012). This means that when imagining that lamp, they are imagining the same lamp as that either being perceived or remembered in a short-term memory task. Making the stimulus (the lamp) the same in all tasks allows for a comparison of the neural signatures being evoked. However, it also means that the extent that something is imagined versus remembered is difficult to establish. In this paradigm, we attempted to create a situation in which the image is novel, and is built up, and then is probed whilst it is still being imagined. Thus, there was no mask to disrupt the imagery task. As such, we can be confident that what was being probed was the image that participants were imagining there and then. Of course, it may be that participants stopped holding that image in the mind eye, and instead committed it to memory, but there was no incentive to do so. Moreover, if they had done this, and we still saw a reduction in capacity, this would still suggest that the ability to create a robust and accurate image of multiple objects and transfer that to memory is reduced relative to creating a robust memory when the to be remembered items are presented. An interesting avenue for further research could be to explore this issue by trying to disrupt imagery. For example, whilst we used a mask in the VWM and VSTM tasks to minimise visual persistence, no mask was used in the imagery task as there was no visual stimulus to mask. Use of a mask may force participants to transfer an image into VWM.

Another issue to consider is what capacity limits in memory and imagery mean. In Experiment 1, the test explores whether participants have the correct memory by testing whether they can identify whether a test item was presented at that location or not. An error suggests that memory is fallible, and thus errors can be used to deduce the capacity of short-term memory. However, capacity here means capacity for storing and retrieving information accurately. Similarly, an error in an imagery task shows that the image being held by the participant is inaccurate. It does not follow however that this means someone cannot imagine a set number of items per se, simply that they cannot imagine them accurately (e.g., at their locations). Indeed, prior work has explored this in relation to memory for features not bound at locations (e.g., Luck & Vogel, 1997). In contrast, other work has investigated features independently of the feature bound to a location (e.g., Baddeley et al., 2011; Schneegans et al. (2021). Nevertheless, in Experiment 5, the test was not about location, but about whether the

item imagined at that location was a given colour. It is harder to argue in that situation that this is not testing the ability of participants to imagine a number of objects simultaneously. It is possible that object images could be free-floating, and thus tend to be imagined without reference to a spatial context. If this were the case, then any test that requires an image to be bound to a location will under report the number of items that can be imagined. This is an interesting avenue for further research. At the very least, the data from these Experiments suggests that the number of objects we can imagine at particular locations is inferior to the number of objects we can remember at particular locations.

Conclusions

In this paper we have developed a novel paradigm that enables capacity to be measured using comparable tasks for imagination, VWM and VSTM. We believe that this paradigm offers one of the cleanest and most direct methods for measuring and comparing the observed capacities of imagination and visual memory. As such, these studies for the first time indicate what the observed capacity of imagination is, and how this compares to short-term visual memory. In so doing, it demonstrates the contribution of sensory systems in supporting memory. However, this study uses a specific instantiation of a visual paradigm, and future research should investigate the capacity of imagination and visual memory using other task operationalisations (e.g., Farah, 1985, 1989) to see if the capacity findings replicate. Furthermore, while our findings are based on relatively homogenous samples of UK undergraduates, we would expect the mechanism underlying imagery to generalize to other healthy adults which could be confirmed with replications across cultures. The paradigm used also involved simple objects (e.g., coloured blocks) and it may be the case that more complex images are easier to retain in imagery. Brady et al. (2016) found an increase in VWM capacity when participants were presented with real world object relative to simple stimuli (although see Li et al., 2020) and this effect may be even greater in imagery.

Whilst we have demonstrated that in these tasks the capacity we observed for imagery appears lower than that of VWM, the exact mechanisms underpinning this are unclear, and much more work is needed to understand this. Nevertheless, we believe the findings are an important step in understanding the relations between imagery and short-term visual memory. Einstein is reported as saying “Logic will get you from A to B. Imagination will take you

everywhere”. The data suggest that imagination might not take you much further than imagining A and B.

References

- Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & De Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology*, 23(15), 1427–1431. <https://doi.org/10.1016/j.cub.2013.05.065>
- Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, 8(10), 457–464. <https://doi.org/10.1016/j.tics.2004.08.011>
- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, 15(3), 122–131. <https://doi.org/10.1016/j.tics.2011.01.003>
- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, 15(2), 106–111. <https://doi.org/10.1111/j.0963-7214.2004.01502006.x>
- Alvarez, G. A., & Franconeri, S. L. (2005). How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, 5(8), 524–534. <https://doi.org/10.1167/5.8.7>
- Andrade, J., Kemps, E., Werniers, Y., May, J., & Szmalec, A. (2002). Insensitivity of visual short-term memory to irrelevant visual information. *Quarterly Journal of Experimental Psychology*, 55A(3), 753–774. <https://doi.org/10.1080/02724980143000541>
- Atkin, C., Howard, J. C., Baguley, T., Baker, J., Guest., D. (2025, June 9). The same or different? Capacity limitations in visual imagery versus visual perception of simple structural objects. Retrieved from osf.io/kumsv
- Avons, S. E., & Sestieri, C. (2005). Dynamic visual noise: No interference with visual short-term memory or the construction of visual images. *European Journal of Cognitive Psychology*, 17(3), 405–424. <https://doi.org/10.1080/09541440440000104>
- Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in cognitive sciences*, 5(3), 119–126.
- Balaban, H., & Ullman, S. (2024). The capacity limits of tracking in the imagination. *Preprint*. <https://cris.openu.ac.il/en/publications/the-capacity-limits-of-tracking-in-the-imagination>
- Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423. <https://doi.org/10.1016/j.apm.2016.02.027>
- Baddeley, A., Allen, R. J., & Hitch, G. J. (2011). Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia*, 49(6), 1393–1400. <https://doi.org/10.1016/j.neuropsychologia.2010.12.042>
- Baddeley, A., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General*, 129(1), 126–145. <https://doi.org/10.1037/0096-3445.129.1.126>
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Baguley, T. (2009). Standardized or simple effect size: What should be reported?. *British journal of psychology*, 100(3), 603–617.

- Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, 133(1), 83–100. <https://doi.org/10.1037/0096-3445.133.1.83>
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*. <https://doi.org/10.1146/annurev.psych.59.103006.093639>
- Bartolomeo, P., Hajhajate, D., Liu, J., & Spagna, A. (2020). Assessing the causal role of early visual areas in visual mental imagery. *Nature Reviews Neuroscience*, 21(9), 517–517. <https://doi.org/10.1038/s41583-020-0349-4>
- Bays, P. M., & Husain, M. (2008). Europe PMC Funders Group Dynamic Shifts of Limited Working Memory Resources in Human Vision, 321(5890), 851–854. <https://doi.org/10.1126/science.1158023>
- Berger, G. H., & Gaunitz, S. C. (1979). Self-rated imagery and encoding strategies in visual memory. *British journal of psychology* (London, England : 1953), 70(1), 21–24. <https://doi.org/10.1111/j.2044-8295.1979.tb02137.x>
- Blazhenkova, O. (2016). Vividness of object and spatial imagery. *Perceptual and Motor Skills*, 123(2), 474–490. <https://doi.org/10.1177/0031512516666026>
- Borst, G., Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2012a). Representations in mental imagery and working memory: evidence from different types of visual masks. *Memory & cognition*, 40(2), 204–217. <https://doi.org/10.3758/s13421-011-0143-7>
- Borst, G., & Kosslyn, S. M. (2010). Individual differences in spatial mental imagery. *Quarterly Journal of Experimental Psychology*, 63(10), 2031–2050. <https://doi.org/10.1080/17470211003802459>
- Borst, G., Niven, E. & Logie, R.H. (2012b) Visual mental image generation does not overlap with visual short-term memory: A dual-task interference study. *Memory & Cognition* 40, 360–372. <https://doi.org/10.3758/s13421-011-0151-7>
- 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.
- Brady, T. F., Störmer, V. S., & Alvarez, G. A. (2016). Working memory is not fixed-capacity: More active storage capacity for real-world objects than for simple stimuli. *Proceedings of the National Academy of Sciences*, 113(27), 7459-7464.
- Campbell, J. I. D., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods*, 44(4), 1255–1265. <https://doi.org/10.3758/s13428-012-0186-0>
- Campos, A., & Pérez-Fabello, M. J. (2009). Psychometric quality of a revised version of the Vividness of Visual Imagery Questionnaire. *Perceptual and Motor Skills*, 108(1), 798–802. <https://doi.org/10.2466/pms.108.3.798-802>
- Cattaneo, Z., Bona, S., & Silvanto, J. (2012). Cross-adaptation combined with TMS reveals a functional overlap between vision and imagery in the early visual cortex. *NeuroImage*, 59(3), 3015–3020. <https://doi.org/10.1016/j.neuroimage.2011.10.022>
- Cattaneo, Z., Vecchi, T., Pascual-Leone, A., & Silvanto, J. (2009). Contrasting early visual cortical activation states causally involved in visual imagery and short-term memory. *European Journal of Neuroscience*. <https://doi.org/10.1111/j.1460-9568.2009.06911.x>
- Ceja, C. R., & Franconeri, S. L. (2023). Difficulty limits of visual mental imagery. *Cognition*, 236, 105436 <https://doi.org/10.1016/j.cognition.2023.105436>
- Chang, S., Lewis, D. E., & Pearson, J. (2013). The functional effects of color perception and color imagery. *Journal of Vision*, 13(10), 4–4. <https://doi.org/10.1167/13.10.4>

- Cichy RM, Heinzle J, Haynes JD (2012) Imagery and perception share cortical representations of content and location. *Cerebral Cortex* 22(2): 372-380
- Cowan, N. (2001). The magical number 4 in short-term memory : A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–185.
- Cowan, N. (2016). *Working memory capacity: Classic edition*. Routledge.
- Cowan, N. (2019). Short-term memory based on activated long-term memory: A review in response to Norris (2017). *Psychological Bulletin*, 145(8), 822-847.
- Cowan, N., Morey, C. C., & Naveh-Benjamin, M. (2021). An embedded-processes approach to working memory: How is it distinct from other approaches, and to what ends? In R. H. Logie, V. Camos, & N. Cowan (Eds.), *Working memory: State of the science* (44–84). Oxford University Press. <https://doi.org/10.1093/oso/9780198842286.003.0003>
- Darling, S., Della Sala, S., & Logie, R. H. (2009). Capacity limits of visual short-term memory are best explained by output interference rather than storage limitations. *Memory*, 17(8), 779–789. <https://doi.org/10.1080/09658210903186971>
- Dean, G. M., Dewhurst, S. A., Morris, P. E., & Whittaker, A. (2008). Dynamic visual noise: Further evidence for interference with visual imagery. *Memory*, 16(9), 810–818. <https://doi.org/10.1080/09658210802365404>
- Dean, G. M., Dewhurst, S. A., & Whittaker, A. (2005). Dynamic visual noise disrupts visual imagery. *Acta Psychologica*, 119(3), 289–301. <https://doi.org/10.1016/j.actpsy.2005.02.002>
- Dijkstra, N. (2024). Nuancing the heterarchical theory of visual mental imagery. *Physics of Life Reviews*, 49, 10-11.
- Dijkstra, N., Zeidman, P., Ondobaka, S., van Gerven, M. A. J., & Friston, K. (2017). Distinct Top-down and Bottom-up Brain Connectivity During Visual Perception and Imagery. *Scientific Reports*. <https://doi.org/10.1038/s41598-017-05888-8>
- Dijkstra, N, Bosch, S. E., & van Gerven, M. A. J. (2019). Shared Neural Mechanisms of Visual Perception and Imagery. *Trends in Cognitive Sciences*, 23(5), 423–434. <https://doi.org/10.1016/j.tics.2019.02.004>
- Farah, M. J. (1985). Psychophysical evidence for a shared representational medium for mental images and percepts. *Journal of Experimental Psychology: General*, 114(1), 91.
- Farah, M. J. (1989). Mechanisms of imagery-perception interaction. *Journal of Experimental Psychology: Human Perception and Performance*, 15(2), 203.
- Fougnie, D., & Marois, R. (2006). Distinct capacity limits for attention and working memory: Evidence from attentive tracking and visual working memory paradigms. *Psychological Science*, 17(6), 526-534. <https://doi.org/10.1111/j.1467-9280.2006.01736.x>
- Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources: Competitive content maps for attention and memory. *Trends in Cognitive Sciences*, 17(3), 134–141. <https://doi.org/10.1016/j.tics.2013.01.010>
- Guest, D., & Lamberts, K. (2010). The prioritization of perceptual processing in categorization. *Attention, perception & psychophysics*, 72(4), 1079–1096. <https://doi.org/10.3758/APP.72.4.1079>
- Guest, D., & Lamberts, K. (2011). The time course of similarity effects in visual search. *Journal of experimental psychology. Human perception and performance*, 37(6), 1667–1688. <https://doi.org/10.1037/a0025640>
- Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, 458(7238), 632–635. <https://doi.org/10.1038/nature07832>
- Howard, C. J., & Holcombe, A. O. (2008). Tracking the changing features of multiple objects: Progressively poorer perceptual precision and progressively greater perceptual lag. *Vision Research*, 48(9), 1164–1180. <https://doi.org/10.1016/j.visres.2008.01.023>

- IBM Corp. (2022). *IBM SPSS Statistics for Windows, Version 29.0*. Armonk, NY: IBM Corp
- Ishai, A., Ungerleider, L. G., & Haxby, J. V. (2000). Distributed neural systems for the generation of visual images. *Neuron*, 28(4), 979–990. [https://doi.org/10.1016/S0896-6273\(00\)00168-9](https://doi.org/10.1016/S0896-6273(00)00168-9)
- Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, 7, 2–9. <http://dx.doi.org/10.7771/1932-6246.1167>
- Johnson, M. R., & Johnson, M. K. (2014). Decoding individual natural scene representations during perception and imagery. *Frontiers in Human Neuroscience*, 8(1 FEB), 1–14. <https://doi.org/10.3389/fnhum.2014.00059>
- Kent, C., Guest, D., Adelman, J. S., & Lamberts, K. (2014). Stochastic accumulation of feature information in perception and memory. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2014.00412>
- Keogh, R. (2017). Different strokes for different folks: A behavioural and neural investigation of mental imagery and individual differences in cognitive strategies in visual working memory capacity, (November).
- Keogh, R., & Pearson, J. (2011). Mental imagery and visual working memory. *PLoS ONE*, 6(12). <https://doi.org/10.1371/journal.pone.0029221>
- Keogh, R., & Pearson, J. (2014). The sensory strength of voluntary visual imagery predicts visual working memory capacity. *Journal of Vision*, 14(12:7), 1–13. <https://doi.org/10.1167/14.12.7.doi>
- Keogh, R., & Pearson, J. (2017). The perceptual and phenomenal capacity of mental imagery. *Cognition*, 162, 124–132. <https://doi.org/10.1016/j.cognition.2017.02.004>
- Klein, I., Dubois, J., Mangin, J. F., Kherif, F., Flandin, G., Poline, J. B., ... Le Bihan, D. (2004). Retinotopic organization of visual mental images as revealed by functional magnetic resonance imaging. *Cognitive Brain Research*, 22(1), 26–31. <https://doi.org/10.1016/j.cogbrainres.2004.07.006>
- Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. Oxford: Oxford University Press.
- Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. MIT press.
- Kosslyn, Stephen M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin*, 129(5), 723–746. <https://doi.org/10.1037/0033-2909.129.5.723>
- Kosslyn, Stephen M., Thompson, W. L., Klm, I. J., & Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature*, 378(6556), 496–498. <https://doi.org/10.1038/378496a0>
- Lambert, S., Sampaio, E., Scheiber, C., & Mauss, Y. (2002). Neural substrates of animal mental imagery: Calcarine sulcus and dorsal pathway involvement—An fMRI study. *Brain Research*, 924(2), 176–183. [https://doi.org/10.1016/S0006-8993\(01\)03215-2](https://doi.org/10.1016/S0006-8993(01)03215-2)
- Lecerf, T., & de Ribaupierre, A. (2005). The structure of visuospatial memory in adulthood. *Swiss Journal of Psychology*, 64(4), 259–272. <https://doi.org/10.1024/1421-0185.64.4.259>
- Lee, S. H., Kravitz, D. J., & Baker, C. I. (2012). Disentangling visual imagery and perception of real-world objects. *NeuroImage*, 59(4), 4064–4073. <https://doi.org/10.1016/j.neuroimage.2011.10.055>
- Li, X., Xiong, Z., Theeuwes, J., & Wang, B. (2020). Visual memory benefits from prolonged encoding time regardless of stimulus type. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(10), 1998–2005. <https://doi.org/10.1037/xlm0000863>

- 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. (1995). Visuo-spatial working memory. Hove, England: Erlbaum.
- Logie, R.H. (2003). Spatial and visual working memory: A mental workspace. In D. Irwin & B. Ross (Eds.), *Cognitive vision: The psychology of learning and motivation* (Vol. 42, pp. 37–78). London, England: Elsevier.
- Love, J., Selker, R., Marsman, M., Jamil, T., Verhagen, A. J., Ly, A.,...Wagenmakers, E.-J. (2015). JASP (Version 0.6.5). <https://jasp-stats.org>
- Luck, S. J., & Vogel, E. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279–281. <https://doi.org/10.1038/36846>
- Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, 64(1), 17–24. <https://doi.org/10.1111/j.2044-8295.1973.tb01322.x>
- Mayer, J. S., Bittner, R. A., Nikolić, D., Bledowski, C., Goebel, R., & Linden, D. E. J. (2007). Common neural substrates for visual working memory and attention. *NeuroImage*, 36(2), 441–453. <https://doi.org/10.1016/j.neuroimage.2007.03.007>
- Mazard, A., Tzourio-Mazoyer, N., Crivello, F., Mazoyer, B., Mellet, E., & Denis, M. (2004). A PET meta-analysis of object and spatial mental imagery. *European Journal of Cognitive Psychology*, 16(5), 673–695. <https://doi.org/10.1080/09541440340000484>
- McConnell, J., & Quinn, J. G. (2004). Interference in visual working memory. *Quarterly Journal of Experimental Psychology*, 57A(8), 1323–1336. <https://doi.org/10.1080/02724980343000793>
- Morey, R. D., & Morey, C. C. (2011). WoMMBAT: A user interface for hierarchical Bayesian estimation of working memory capacity. *Behavior Research Methods*, 43(4), 1044–1065. <https://doi.org/10.3758/s13428-011-0114-8>
- Nyberg, L., Habib, R., McIntosh, A. R., & Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proceedings of the National Academy of Sciences of the United States of America*, 97(20), 11120–11124. <https://doi.org/10.1073/pnas.97.20.11120>
- Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, 142(7), 758–799. <https://doi.org/10.1037/bul0000046>
- Pearson, D. G. (2001). Imagery and the visuo-spatial sketchpad. In J. Andrade (Ed.), *Working memory in perspective* (pp. 33–59). Psychology Press
- Pearson, J. (2019). The human imagination: the cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience*, 20(10), 624–634. <https://doi.org/10.1038/s41583-019-0202-9>
- Pearson, J., Clifford, C. W. G., & Tong, F. (2008). The Functional Impact of Mental Imagery on Conscious Perception. *Current Biology*, 982–986. <https://doi.org/10.1016/j.cub.2008.05.048>
- Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental Imagery: Functional Mechanisms and Clinical Applications. *Trends in Cognitive Sciences*. <https://doi.org/10.1016/j.tics.2015.08.003>
- Pearson, J., Rademaker, R. L., & Tong, F. (2011). Evaluating the mind’s eye: The metacognition of visual imagery. *Psychological Science*, 22(12), 1535–1542. <https://doi.org/10.1177/0956797611417134>
- Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1–2), 8–13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Postle, B. R. (2016). How Does the Brain Keep Information “in Mind”? *Current Directions in Psychological Science*, 25(3), 151–156. <https://doi.org/10.1177/0963721416643063>
- Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, 25(2), 25–30. <https://doi.org/10.1016/j.zgesun.2007.12.005>

- Quinn, J. G., & McConnell, J. (2006). The interval for interference in conscious visual imagery. *Memory*, 14(2), 241–252. <https://doi.org/10.1080/09658210500210019>
- Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.0711295105>
- Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature Reviews Neuroscience*, 8(9), 657–661. <https://doi.org/10.1038/nrn2213>
- Schlegel, A., Kohler, P. J., Fogelson, S. V., Alexander, P., Konuthula, D., & Tse, P. U. (2013). Network structure and dynamics of the mental workspace. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.1311149110>.
- Schneegans, S., Harrison, W. J., & Bays, P. M. (2021). Location-independent feature binding in visual working memory for sequentially presented objects. *Attention, Perception, & Psychophysics*, 83(6), 2377–2393. <https://doi.org/10.3758/s13414-021-02245-w>
- Slotnick, S. D., Thompson, W. L., & Kosslyn, S. M. (2005). Visual mental imagery induces retinotopically organized activation of early visual areas. *Cerebral Cortex*, 15(10), 1570–1583. <https://doi.org/10.1093/cercor/bhi035>
- Sheldon, S., Amaral, R., & Levine, B. (2017). Individual differences in visual imagery determine how event information is remembered. *Memory*, 25 (3), 360–369. <https://doi.org/10.1080/09658211.2016.1178777>
- Sherwood, M. R., & Pearson, J. (2010). Closing the mind's eye: Incoming luminance signals disrupt visual imagery. *PLoS ONE*, 5(12), e15217. <https://doi.org/10.1371/journal.pone.0015217>.
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1), 137–149. <https://doi.org/10.3758/BF03207704>
- Tong, F. (2013). Imagery and visual working memory: one and the same? *Trends in Cognitive Sciences*, 26(5), 489–490. <https://doi.org/10.1016/j.tics.2013.08.005>
- Treisman, A., & Zhang, W. (2006). Location and binding in visual working memory. *Memory & Cognition*, 34(8), 1704–1719. <https://doi.org/10.3758/BF03195932>
- Valenti, R., & Galera, C. (2020). Examining the effects of external stimuli on visual imagery performance. *Journal of Cognitive Psychology*, 32(5-6), 520–530. <https://doi.org/10.1080/20445911.2020.1785632>
- Van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, 121(1), 124–149. <https://doi.org/10.1037/a0035234>
- van de Ven, V., & Sack, A. T. (2013). Transcranial magnetic stimulation of visual cortex in memory: Cortical state, interference and reactivation of visual content in memory. *Behavioural Brain Research*, 236, 67–77. <https://doi.org/10.1016/j.bbr.2012.08.001>
- Van Der Meulen, M., Logie, R. H., & Sala, S. D. (2009). Selective interference with image retention and generation: Evidence for the workspace model. *Quarterly Journal of Experimental Psychology*, 62(8), 1568–1580.
- Vandierendonck, A., Kemps, E., Fastame, M. C., & Szmalec, A. (2004). Working memory components of the Corsi blocks task. *British Journal of Psychology*, 95(1), 57–79. <https://doi.org/10.1348/000712604322779460>
- Vasques, R., Yanez, R., & Santos, J. (2016). Visual imagery and external distractions: Effects

- on working memory tasks. *Cognitive Processing*, 17(3), 213–224.
<https://doi.org/10.1007/s10339-016-0753-2>
- Welford, A. T. (1980). The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics*, 23(1), 5–31.
<https://doi.org/10.1080/00140138008924759>
- Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, 131(1), 48–64. <https://doi.org/10.1037/0096-3445.131.1.48>
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4(12), 1120–1135. <https://doi.org/10.1167/4.12.11>
- Woodman, G. F., & Luck, S. J. (2004). Visual search is slowed when visuospatial working memory is occupied. *Psychonomic Bulletin & Review*, 11(2), 269–274.
<https://doi.org/10.3758/BF03196569>
- Zanto, T. P., Rubens, M. T., Bollinger, J., & Gazzaley, A. (2011). Top-down modulation of visual feature processing: The role of the inferior frontal junction. *NeuroImage*, 53(2), 736–745. <https://doi.org/10.1016/j.neuroimage.2010.06.012>
- Zhang, W., & Luck, S. J. (2008). Discrete Fixed-Resolution Representations in Visual Working Memory Weiwei. *Nature*, 453(7192), 233–235.
<https://doi.org/10.1038/nature06860>.Discrete