Cognition 236 (2023) 105436

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Difficulty limits of visual mental imagery

Cristina R. Ceja *, Steven L. Franconeri

Northwestern University, United States of America

A R T I C L E I N F O

Keywords:
Mental imagery
Visual working memory
Capacity

A B S T R A C T

While past work has focused on the representational format of mental imagery, and the similarities of its operation and neural substrate to online perception, surprisingly little has tested the boundaries of the level of detail that mental imagery can generate. To answer this question, we take inspiration from the visual short-term memory literature, a related field which has found that memory capacity is affected by the number of items, whether they are unique, and whether and how they move. We test these factors of set size, color heterogeneity, and transformation in mental imagery through both subjective (Exp 1; Exp 2) and objective (Exp 2) measures – difficulty ratings and a change detection task, respectively – to determine the capacity limits of our mental imagery, and find that limits on mental imagery are similar to those for visual short-term memory. In Experiment 1, participants rated the difficulty of imagining 1–4 colored items as subjectively more difficult when there were more items, when the items had unique colors instead of an identical color, and when they scaled or rotated instead of merely linearly translating. Experiment 2 isolated these subjective difficulty ratings of rotation for uniquely colored items, and added a rotation distance manipulation (10° to 110°), again finding higher subjective difficulty for more items, and for when those items rotated farther; the objective measure showed a decrease in performance for more items, but not for rotational degree. Congruities between the subjective and objective results suggest similar costs, but some incongruities suggest that subjective reports can be overly optimistic, likely because they are biased by an illusion of detail.

1. Introduction

Imagine two circles, red and blue, horizontally arranged. You likely feel that you can picture both circles at once. Now imagine four circles (red, blue, green, and yellow) placed at the vertices of an invisible square. You may find it more difficult to picture all four circles simultaneously - perhaps only a subset is clearly 'visible' at a time. Now try to rotate that set of four circles, as a complete image, by 90° clockwise; if that operation felt easy, then you may hold an unusual talent (or at least a proclivity for self-deception). You likely found that you encountered a capacity limitation for these operations, such that this list of tasks became increasingly difficult as you attempted to imagine a greater number of items with a greater variety of colors – and especially when you needed to transform them across a simulated rotation.

Performing these tasks in your mind's eye requires mental imagery, the combination of an internal generation of a perceptual experience without its concurrent visual input (see Pearson, 2019 for review; Kosslyn, Thompson, & Ganis, 2006), as well as the ability to mentally transform that information by rotation (for review, see Zacks, 2008). Early work in the mental imagery literature primarily focused on testing whether functional constraints of mental imagery were similar to those found in online perception. People take longer to imagine scanning across a greater distance in a previously-memorized map (Kosslyn, Ball, & Reiser, 1978) or dot array (Borst & Kosslyn, 2010; Finke & Pinker, 1982), to imagine items farther in the periphery (Finke & Kosslyn, 1980), or need to imagine 'zooming in' on an item to verify that it has a specific property when it is typically small in the real world (e.g., imagining a mouse and determining whether it has claws) (Kosslyn, 1975). There are also well-documented constraints on mental transformation. People take longer to verify that two items have the same shape when their sizes are increasingly different, suggesting that they mentally 'scale' the item over time, as tested up to a 5:1 ratio (Bundesen & Larsen, 1975), and take longer to decide whether two items are the same when one is rotated to a progressively different degree than the other, suggesting that they must transform one of the items for comparison (Shepard & Metzler, 1971; Just & Carpenter, 1985; but see also

* Corresponding author at: Department of Psychology, Northwestern University, 2029 Sheridan Road, Cresap Hall 112, Evanston, IL 60208, United States of America.

E-mail address: crceja@u.northwestern.edu (C.R. Ceja).

https://doi.org/10.1016/j.cognition.2023.105436
Received 11 October 2021; Received in revised form 25 February 2023; Accepted 5 March 2023
Available online 10 March 2023

Stieff, 2007).

More recent work on mental imagery provides a deeper understanding of its internal process. There is increased evidence that imagery does indeed function in a similar manner to perception, albeit a weaker form (for review, see Pearson, Naselaris, Holmes, & Kosslyn, 2015; Pearson, 2019). Imagining a bright scene (i.e., a sunny sky) or dark scene (i.e., a dark room) can result in pupil constriction and dilation, respectively, comparable to those found in perception (Laeng & Sulutvedt, 2014). Similar to perception, prolonged imagining of a stimulus moving in one direction can induce a motion aftereffect where viewers perceive a subsequent perceptual probe to be moving in an opposite direction from the imagined movement (Winawer, Huk, & Boroditsky, 2010). Associative learning also occurs in the same manner for imagery as in perception – a perceptual Gabor stimuli will evoke an emotional response if a voluntary mental image of the same Gabor was previously conditioned with either pleasant or aversive stimuli (Lewis, O'Reilly, Khuu, & Pearson, 2013).

Yet there is currently very little work that has tested the limits of this internal mental imagery process. To our knowledge, only one study has explored the type of capacity limits that we seek for visual mental imagery. The study relied on a clever technique: assuming that perception and mental imagery share similar architectures, imagining an item in a particular location should strengthen perception of that item presented later (Keogh & Pearson, 2017). In this binocular rivalry test, different images were presented to participants at the same location across the two eyes, creating a bistable percept that could be biased by previously seen items at that location. Participants were primed with several placeholders (lines providing information on orientation, color, and location) and asked to imagine red horizontal and/or green vertical lines at these specific locations. Following this display, they were presented with a brief binocular rivalry display consisting of one red horizontal and one green vertical stimulus at one of the seven placeholder locations, and reported the dominant rivalry pattern at that location. The adaptation effect weakened in both strength and specificity when participants were asked to imagine more items, and its strength also weakened when those items were unique (e.g., a mixed collection of colors and orientations instead of all green verticals).

Relevant to the present experiments, this study also collected subjective reports of the quality of the imagined 'adaptation' displays. Participants again performed the same binocular rivalry task, but now also rated the vividness of the imagined displays on a scale from 1 ("least vivid") to 4 ("most vivid"). Subjective vividness linearly decreased when people were asked to imagine more placeholders. These vividness ratings were also related to the strength of the adaptation effect, such that there was more priming for items that were subjectively rated as more vivid. Other work has found a similar relationship between subjective vividness ratings for a battery of naturalistic mental imagery tasks (the Vividness of Visual Imagery Questionnaire; Marks, 1973) and binocular rivalry priming (Pearson, Rademaker, & Tong, 2011). Subjective ratings of vividness have also been found to become more predictive of binocular rivalry priming after a training period (Rademaker & Pearson, 2012).

An understanding of the limits in a related process, visual memory, could help provide these insights into the limits that might exist for visual mental imagery. The literature on visual memory has used a variety of demonstrations to generate constraints on the architecture of the visual system and test the (severely limited) capacity limits of memory. In memory, performance is found to degrade as people are asked to detect replacements of features (typically colors) between an encoding and test display, with a functional limit of around 3–4 items at most (Brady, Konkle, & Alvarez, 2011; Wilken & Ma, 2004; Zhang & Luck, 2008). Reproductions of a single color from memory is more precise than reproducing a set of unique colors (Zhang & Luck, 2008). When features are not replaced, but instead swapped among items or locations in a display, requiring the observer to remember not only a list of display features but the item or location where they were present, performance degrades to 2–3 stored color-item pairings (Alvarez & Thompson, 2009; Wheeler & Treisman, 2002; Xu & Franconeri, 2015). This limit drops to 1–2 color pairings when the items have moved since the encoding display, requiring a location remapping (Horowitz et al., 2007; Saiki, 2003; Saiki & Miyatsuji, 2009; Xu & Franconeri, 2015). Some evidence also suggests that recognizing a previously seen item display is disrupted by some types of motion more than others. When asked to confirm that a complex shape was part of a memorized set, responses were slower when the test image was rotated $>45^{\circ}$, but performance was robust for a shrinking manipulation until the image was only 20% of the original memorized size (Lam, Rensink, & Munzner, 2006). Performance in a visual memory task was unchanged when the display configuration was scaled at test (Jiang, Olson, & Chun, 2000) or suffered only slightly when the configuration translated horizontally or vertically (Hollingworth, 2007) from study to test, given that the items remained in the same spatial configuration.

These capacity limits serve as important constraints for models of the visual system. The functional memory limit of 3–4 items, and the interaction of that limit with the heterogeneity of their features, has generated debate over whether the visual system relies on a fixed set of 'memory slots' (Zhang & Luck, 2008), versus summary representations that consolidate information from multiple items (with that consolidation becoming rapidly less efficient after 3–4 items; Brady et al., 2011). The lower capacity for 'swap' displays reveals separable storage systems for what features are present in a display, versus a more limited system that binds those features to particular locations (Alvarez & Thompson, 2009; Franconeri, Alvarez, & Cavanagh, 2013). The even lower limit for keeping features bound to items as they move suggests a fundamentally spatiotopic (or even retinotopic) organization for visual memory (Golomb, Chun, & Mazer, 2008).

If mental imagery shows similar capacity hallmarks to visual working memory – degradation for more items, especially when features are unique, and when items move – it would allow for similar conclusions about the architecture that subserves visual mental imagery. Little work has tested this directly, but there is good reason to explain why similar factors of complexity might hinder representations in both mental imagery and in visual working memory, as these processes appear to be related in their neural correlates and representational formats. Both visual mental imagery and visual working memory are associated with activity in analogous brain regions in early visual cortex, frontal-parietal control regions, and occipital-temporal sensory regions (Albers, Kok, Toni, Dijkerman, & de Lange, 2013; Harrison & Tong, 2009; Kamitani & Tong, 2005; Serences, Ester, Vogel, & Awh, 2009; Slotnick, 2008; Slotnick, Thompson, & Kosslyn, 2012). Both processes also rely on representations sharing the same depictive format (Borst, Ganis, Thompson, & Kosslyn, 2012) - a pictorial representation that maintains how parts of the corresponding item are organized and spatially related to one another (Kosslyn, 1994). Visual mental imagery and visual working memory also have a similar influence on performance in certain tasks — for example, both processes can guide or bias attention in visual search tasks (Moriya, 2018; Olivers, Meijer, & Theeuwes, 2006; Soto, Heinke, Humphreys, & Blanco, 2005). Finally, performance across these two processes also appears to be correlated, such that viewers with greater mental imagery strength also have greater visual working memory precision and capacity (Keogh & Pearson, 2014). Relatedly, viewers often report using strategies involving mental imagery while performing visual working memory tasks (Berger & Gaunitz, 1979; Harrison & Tong, 2009), and those with greater mental imagery ability tend to perform better in these tasks, as compared to those with poorer mental imagery ability (Berger & Gaunitz, 1979; Keogh & Pearson, 2011).

However, it is important to note that, while visual mental imagery and visual working memory are related, there is also evidence that they are at least somewhat *distinct* processes. Specific visual interference techniques, such as dynamic visual noise or irrelevant visual input, have been shown to selectively impair recall performance when using a pegword mnemonic (which involves the generation of mental images to memorize a sequence of words or letters), but not when these stimuli were memorized and retained through rehearsal (Andrade, Kemps, Werniers, May, & Szmalec, 2002; Borst, Niven, & Logie, 2012; Quinn & McConnell, 1996; van der Meulen, Logie, & Della Sala, 2009); irrelevant visual input are, therefore, interpreted to interfere with the activation of visual information stored in the visual buffer that is being used to generate mental images (Kosslyn et al., 2006; Logie, 1995). Conversely, other interference techniques, such as retaining a spatial tapping pattern, have been shown to impair the retention of letters in visual short-term memory, but not visual mental images generated with a pegword mnemonic (van der Meulen et al., 2009). This double-dissociation appears to indicate that the underlying system that aids in mental imagery, the visual buffer, is not identical to the system involved with visual memory, referred to as the visual cache (Andrade et al., 2002; Logie, 1995, 2003; Pearson, 2001; Pearson, Logie, & Gilhooly, 1999).

Past findings have further dissociated visual mental imagery from visual working memory. Specific areas of the parietal control regions, occipital-temporal sensory regions, retrosplenial cortex, and middle temporal gyrus are more strongly activated in working memory tasks than in mental imagery tasks, suggesting at least some differences in the two processes (Slotnick et al., 2012). Another study used a battery of tasks to assess image generation and manipulation ability in both mental imagery and visual working memory, and found no significant associations between these processes in either children or adults (Bates & Farran, 2021). Those with poor mental imagery ability (Keogh & Pearson, 2011) or even aphantasia (Jacobs, et al., 2017; Keogh, Wicken, & Pearson, 2021) can still perform well in visual working memory tasks (albeit, often easy working memory tasks for those with aphantasia).

Past studies begin to fill in our understanding of whether mental imagery is subject to the same capacity constraints as visual short-term memory: subjective ratings showed that imagining more items was tougher, and the objective priming test showed that imagining more items, or items that were unique, were also tougher (Keogh & Pearson, 2017). Yet this work leaves important unanswered questions, even for subjective ratings. Subjectively, does heterogeneity feel like it hurts performance? How about when items move? What if they translate, which seems robust for visual memory, versus rotate, which is far more damaging for memory? Having the answers to these questions would allow similar inferences for mental imagery as they do in the memory literature, with respect to spatiotopy, feature binding, and robustness to transformation. These links would be even easier to make if subjective mental imagery tests used displays that were similar to those used in the working memory literature - small arrays of colored shapes. Most subjective measures of mental imagery have relied on more complex displays, like maps (Kosslyn et al., 1978) that make the number of items tougher to manipulate, or feature conjunctions (i.e., color and orientation; Keogh & Pearson, 2017) that can be difficult for the visual system to handle (Wolfe & Cave, 1999).

The present work relies on displays similar to those used in the visual working memory literature to test whether subjective reports (Exp. 1, Exp. 2) and objective measures (Exp. 2) of performance within a participant's mental imagery reveal the same capacity-limiting factors found in previous memory studies. These experiments systematically manipulate the number of imagined items (Exp. 1, Exp. 2), whether they are unique or identical colors (Exp. 1), how they move (linear translation, scaling, and rotation; Exp. 1), and the degree of rotational transformation (Exp. 2), to measure the impact of each factor and its potential interactions on capacity. While these experiments cannot directly measure subjective capacity in a way that suggests a certain number of items, they *can* show that the same capacity-limiting factors – and their interactions – affect subjective ratings in the same way that they affect objective measures of performance. Although the numerical rating that a participant gives for subjective difficulty might be subjective, the way that these numbers change across conditions is informative.

Mental imagery might be affected by the same factors that limit capacity for visual working memory. Online perception can offload much of the task of internally representing the external world to the external world itself, and mental imagery could be more impoverished than one might think (just like visual memory) when that external signal is missing. It is also possible that mental imagery is not affected by all the same factors as visual working memory. Internally-generated representations could function like sustained perceptual input, with mental imagery acting as an internal reference to maintain access to detail (Mohr, Linder, Dennis, & Sireteanu, 2011; Pearson et al., 2015; Tartaglia, Bamert, Mast, & Herzog, 2009). If so, then increasing the number of objects, or their heterogeneity, in an imagined display might not impact mental imagery performance as much as it does for visual working memory, because those statistics do not change.

To preview our results, the objective and subjective measures of Experiments 1 and 2 suggest capacity limitations that are broadly similar to those found in visual working memory tasks. Subjective difficulty reports were greater when imagining larger set sizes, rotation and scaling transformations (compared to translation), different features, and across greater degrees of rotation. Accuracy in objective change detection performance was lower with an increase in set size, although not with an increase in the degree of transformation. Additionally, performance was significantly lower when colored items swapped locations, mirroring how visual working memory capacity is negatively impacted when one must bind both feature and location information (Alvarez & Thompson, 2009; Xu & Franconeri, 2015).

2. Experiment 1

In Experiment 1, we examined subjective reports of difficulty in mentally updating the locations of uniquely or identically colored items varying in set size across simple transformations (linear translation, rotation, or scaling; see Fig. 1).

Note that there are other potential methods for testing subjective performance in mental imagery tasks (e.g., the vividness ratings found in Keogh & Pearson, 2017 and Marks, 1973). In this work, however, we relied on a subjective difficulty rating where participants were to interpret "difficulty" in terms of a few factors, including ease of transformation, whether they could maintain the items and their relationships to one another in their mind's eye during the trial, and how clear the image was during this transformation. This definition of difficulty aligns in many respects with how vividness is defined in previous work on imagery (i.e., how vivid and clear the image is in your mind) to provide more confidence that the findings are linked or due to imagery.

2.1. Materials and methods

2.1.1. Participants

Thirty participants (18–19 years old) completed the experiment. Sample sizes for all experiments were determined a priori to mirror conventional sample sizes in the mental imagery literature (e.g., Keogh & Pearson, 2011; Pearson et al., 2011). All participants had normal or corrected-to-normal vision, were given course credit for participation, and gave informed written consent.

2.1.2. Stimuli and apparatus

The experiment was controlled by a MacOS computer running Qualtrics. The displays were presented on a 23-in. LCD monitor with a 60-Hz refresh rate and 1440 × 900-pixel resolution, and viewed at an approximate distance of 56 cm, with approximately 28 pixels per degree visual angle (dva). The displays shown during a given trial were static and illustrated colored circles performing one of three types of transformations: either linear translation (a movement of 9.1 dva in length), 90° rotation, or scaling.

For the *linear translation* condition, participants were shown a static display with filled, colored circle(s) in the top half of the display, each

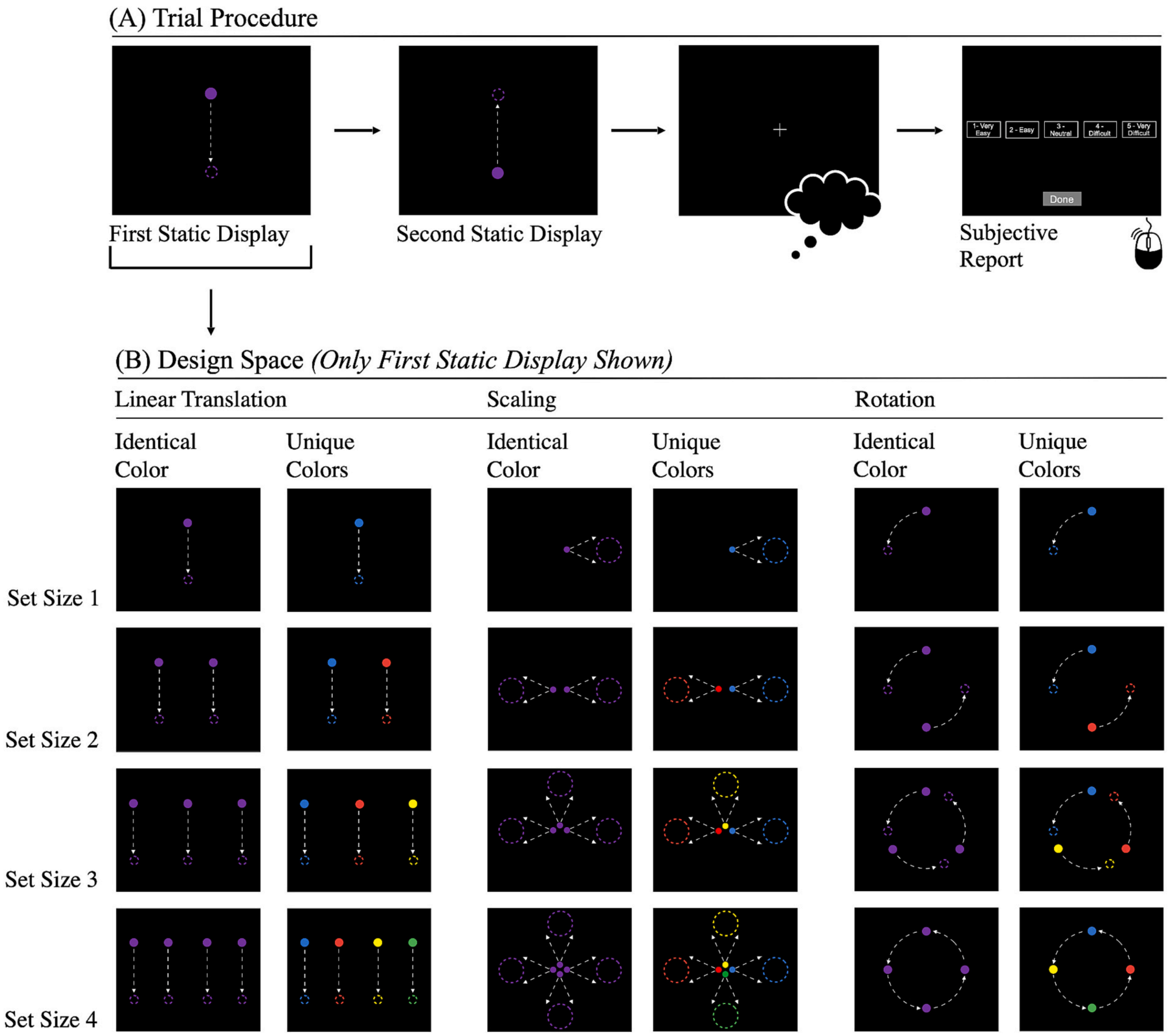

Fig. 1. Procedure and design space for Experiment 1 (displays are not to scale). (A) Trial Procedure: Participants were shown two static displays (in this example, a first static image indicating a downward movement (*First Static Display*) followed by a second static image indicating an upward movement (*Second Static Display*)). Participants were then shown a black screen and asked to imagine this movement. Finally, participants rated the difficulty of performing this transformation on a 1–5 scale ("Very Easy" to "Very Difficult"; *Subjective Report*). (B) Design Space: All displays varied in set size (*see rows*) or color heterogeneity (*see columns*); only set size 1 displays described here in the caption. *Linear translation condition:* Participants first viewed a static display with a filled circle at the top of the screen and a dotted line pointing downwards to an outlined circle at the bottom of the screen (downward translation; shown here). This was followed by a second static display with a filled circle at the bottom of the screen and a dotted arrow pointing upwards to an outlined circle at the top of the screen (upward translation; not shown here). Participants were then shown a blank screen (not shown here) and were asked to imagine the circle(s) moving downward together, and then back up to their starting point. *Scaling transformation condition:* Participants first viewed a static display that had a circle near the center of the screen with two dotted arrows diverging outwards toward a larger, outlined circle on the side of the screen (scaling up; shown here). This was followed by a second static display with a larger filled circle on the side of the screen and two dotted arrows converging inward toward an outlined circle near the center of the screen (scaling down; not shown here). Participants were then shown a blank screen (not shown here) and were asked to imagine the circle expanding as it moved outward toward the edge of the screen, and then shrinking back to its original size as it moved back toward the center. *Rotation transformation condition:* Participants first viewed a static display with a filled circle and a curved, dotted line pointing counter-clockwise toward an outlined circle (counter-clockwise rotation; shown here). This was followed by a second static display with the previously outlined circle now filled and a dotted arrow pointed clockwise to an outlined circle in the same location as the previously filled circle (clockwise rotation; not shown here). Participants were then shown a blank screen (not shown here) and were asked to imagine the circle rotating counterclockwise and then clockwise 90°.

with a dotted arrow pointing to an unfilled, dotted circle in the bottom half of the screen, indicating a downward motion (see Fig. 1). This was then followed by a static display with filled, colored circle(s) appearing in the bottom half of the screen and a dotted arrow for each pointing to an unfilled, dotted circle in the top half of the screen, indicating an upward motion (note: this is not displayed in the design space in Fig. 1). For the *90° rotation condition,* participants were shown a static display with filled, colored circle(s), each with a dotted arrow curving 90° counterclockwise to an unfilled, dotted circle (see Fig. 1). This was then followed by a static display with filled, colored circle(s) at the position of the previously unfilled, dotted circle(s) (and vice versa for the previously filled, colored circle(s)). Each dotted arrow now curved clockwise 90° (note: this is not displayed in Fig. 1). For the *scaling* condition, participants were shown a static display with filled, colored circle(s) placed toward the center of the display, each with two parallel, dotted arrows pointing to the sides of the screen that diverged as they reached the sides. Each pair of arrow heads pointed to a larger, unfilled, dotted circle. This indicated a scaling up motion (see Fig. 1). This was then followed by a static display with larger, filled, colored circle(s), each with a pair of dotted arrows converging toward the center of the screen to a smaller, unfilled, dotted circle, indicating a scaling down motion (note: this is not displayed in Fig. 1).

For all transformation conditions, the number of items ranged from 1 to 4 items on any given display, with each circle subtended 1.7 dva in width (with the exception of the larger circle in the scaling condition). These items could all be the same color (purple (RGB: 112, 48, 160)) or different colors (blue (RGB: 68, 114, 196), red (RGB: 255, 0, 0), green (RGB: 0, 176, 80), or yellow (RGB: 255, 255, 0)).

2.1.3. Procedure

rials began with each static display shown for 500 ms (e.g., view a static display indicating a downward motion for 500 ms, followed by a static display indicating an upward motion for 500 ms). Participants then viewed a black screen for 15 s, where they were instructed to imagine the colored circle(s) moving along the path indicated by the dotted arrows. Participants were then asked to rate the level of difficulty in imagining the transformation (subjective report task) on a 1–5 scale from "Very Easy" to "Very Difficult" (additional survey questions that are not central to the primary results reported here are available in the Supplementary Materials). Once the final response was completed for the survey, the next trial was initiated.

Each participant was tested across 6 blocks of 4 trials, for a total of 24 trials. Participants were given a self-initiated break between each block. Blocks were separated by transformation type (linear translation, 90° rotation, or scaling) and color type (identical-color or unique-colors), resulting in 6 distinct blocks (translation/identical-color, translation/unique-colors, rotation/identical-color, rotation/unique-colors, scaling/identical-color, scaling/unique-colors) that were randomized across the experiment.

2.2. Analysis and results

Difficulty ratings were predicted using a linear mixed effects model, with set size, the type of transformation, and whether the colors were unique or identical as the fixed effects, and subjects as the random effect. The model object was then subjected to a Wald Chi-Squared test to check for main effects for each of the factors. To test for differences between the levels of each factor, post-hoc analyses were conducted on the model using least squares means with Tukey adjustments.

See Fig. 2 for results. A significant main effect of set size was found, $X^2(3) = 180.80$, $p < 0.001$, $\eta_p^2 = 0.213$. Difficulty ratings significantly increased as set size increased from 1 to 3 ($M_{\text{Set Size 1}} = 1.32$, $SE_{\text{Set Size 1}} = 0.05$, 95% $CI_{\text{Set Size 1}} = [1.09, 1.56]$; $M_{\text{Set Size 2}} = 1.96$, $SE_{\text{Set Size 2}} = 0.08$, 95% $CI_{\text{Set Size 2}} = [1.72, 2.19]$; $M_{\text{Set Size 3}} = 2.27$, $SE_{\text{Set Size 3}} = 0.09$, 95% $CI_{\text{Set Size 3}} = [2.03, 2.50]$), all $ps \leq 0.001$. However, there was no significant difference in difficulty ratings between set size 3 and set size 4 ($M = 2.32$, $SE = 0.09$, 95% CI = [2.09, 2.56]), $p = 0.910$. There was also a significant main effect for the type of transformation, $X^2(2) = 78.26$, $p < 0.001$, $\eta_p^2 = 0.105$. Difficulty ratings were significantly lower for linear translation ($M = 1.61$, $SE = 0.06$, 95% CI = [1.38, 1.84]) than for the scaling transformation ($M = 2.07$, $SE = 0.07$, 95% CI = [1.84, 2.30]), $p < 0.001$. There was no significant difference between scaling and rotation ($M = 2.23$, $SE = 0.08$, 95% CI = [2.00, 2.45]), $p = 0.074$. We also found a significant main effect of whether the colors were unique or identical, $X^2(1) = 33.95$, $p < 0.001$, $\eta_p^2 = 0.048$, with difficulty ratings significantly greater for unique colors ($M = 2.14$, $SE = 0.06$, 95% CI = [1.92, 2.36]) than for identical colors ($M = 1.79$, $SE = 0.05$, 95% CI = [1.57, 2.02]), $p < 0.001$.

Additionally, there was a significant interaction between type of transformation and set size, $X^2(6) = 21.25$, $p = 0.002$, $\eta_p^2 = 0.031$, indicating that participants rated larger set sizes as more difficult when the translation type was rotated or scaled than when it was linearly translated (see Fig. 2). There was also a significant interaction between set size and whether the colors were unique or identical, $X^2(3) = 11.62$, $p = 0.009$, $\eta_p^2 = 0.017$, indicating that larger set sizes were rated more difficult when the colors were unique compared to when they were all identical. There was no interaction between transformation type and whether the colors were unique or.

identical, $X^2(2) = 0.548$, $p = 0.761$, $\eta_p^2 = 0.001$. There was also no significant three-way interaction between set size, whether the colors were unique or identical, and type of transformation, $X^2(6) = 12.17$, $p = 0.058$, $\eta_p^2 = 0.018$.

In summary, subjective difficulty ratings were greater for larger set sizes, for the scaling and rotation transformations (compared to linear translation), and for items with unique colors (compared to those with identical colors). Furthermore, larger set sizes resulted in even higher difficulty ratings when those items rotated or scaled rather than linearly translated, and when the colors were unique rather than all identical.

Although the participants were naïve to previously established capacity-limiting factors of working memory, their subjective difficulty reports paralleled findings from past objective measures of visual working memory capacity, with set size, motion-transformation, and color heterogeneity impacting performance.

3. Experiment 2

Experiment 1 suggests that many of the same capacity-limiting factors that impact visual working memory – set size, motion-transformation, and color heterogeneity – also impact subjective difficulty in a mental imagery task. Experiment 2 again relies on this subjective difficulty rating from Experiment 1, but also adds a replication of those capacity limits in a modified visual memory paradigm that contains objective measures of performance accuracy, allowing a comparison between mental imagery difficulty ratings and objective accuracy in the types of memory change detection tasks that are used to measure memory capacity.

In this objective paradigm, participants perform a change detection task where they are instructed to imagine a specific configuration of colored circles rotating to a specific degree, and respond whether a subsequently shown display matches the rotated configuration of circles

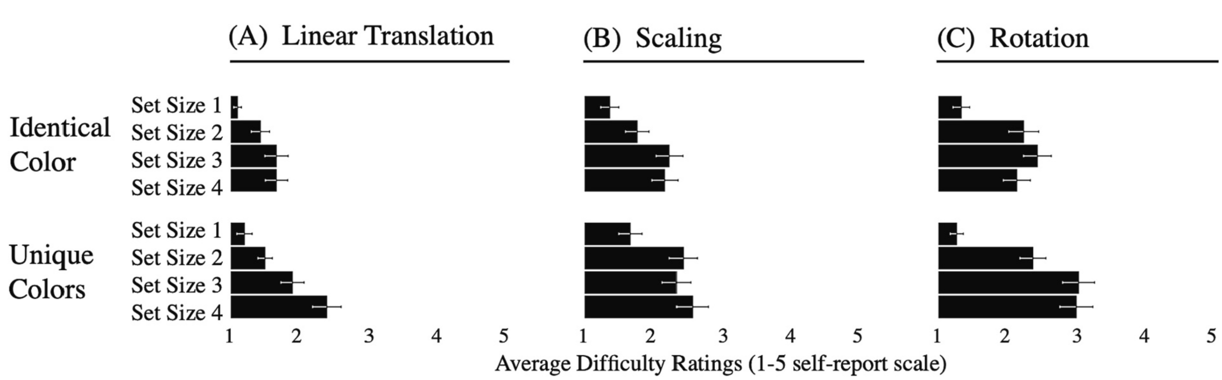

Fig. 2. Results for Experiment 1. Average subjective difficulty ratings for linear translation trials (A), scaling trials (B), and rotation trials (C) for set sizes 1–4 and identical/unique colors. Error bars indicate standard error.

they were instructed to imagine. These change detection responses are then either correct or incorrect (e.g., incorrect if the subsequently shown display matched what participants should have imagined, but participants reported that the subsequently shown display did *not* match what they imagined). While this task is inspired from the memory literature, imagining an image transformation and determining whether your imagined display matches the subsequently shown display should involve mental imagery in order to perform this image regeneration and manipulation (i.e., imagining the rotation of the circles).

Apart from the addition of an objective change detection task to the subjective difficulty reports, Experiment 2 mirrors the design of Experiment 1, but with the following exceptions for both subjective and objective conditions. Rather than investigating a variety of different transformation types, Experiment 2 focuses solely on rotation, as this type of transformation is one of the most commonly used transformation types in previous mental imagery work. Given that Experiment 2 focuses on rotation, the degree of rotation is also manipulated to determine whether there is a threshold for mental rotation before performance is negatively impacted, similarly to how previous work in visual working memory has found the threshold to be 45° before performance is negatively impacted (Lam et al., 2006). Additionally, items in Experiment 2 are now all different colors to allow for an objective change detection measure.

3.1. Materials and methods

3.1.1. Participants

Twenty participants (18–34 years old) completed the experiment. All participants had normal or corrected-to-normal vision, were compensated $10 for participation, and gave informed written consent.

3.1.2. Stimuli and apparatus

The displays were controlled by a MacOS computer running MATLAB. The displays were presented on a 23-in. LCD monitor with a 60-Hz refresh rate and 1440 × 900-pixel resolution, and viewed at an approximate distance of 56 cm, with approximately 28 pixels per dva.

The moving displays always consisted of four circles (1.8 dva each in width) connected to one another by a larger, white circle (14.4 dva in width; see Fig. 3). All circles in the moving display were outlined in white, but not filled in. The four circles could be positioned in a " + " (circles positioned at 0°, 90°, 180°, and 270°) or " × " (circles positioned at 45°, 135°, 225°, and 315°) orientation, depending on the subsequent orientation of the static display. The moving display could then rotate 10°, 60°, or 110° clockwise or counterclockwise as a whole.

The static displays still included the four circles connected by a larger, white circle in a " + " or " × " orientation; but now, 2–4 of these four circles could be filled in with a color (any random combination of red (RGB: 228, 26, 28), blue (RGB: 55, 126, 184), orange (RGB: 255, 127, 0), green (RGB: 77, 175, 74), yellow (RGB: 255, 255, 51), or purple (RGB: 152, 78, 163)). These RGB values were chosen from the '6-class qualitative' ColorBrewer color set (http://colorbrewer2.org/), and were approximately perceptually equiluminant.

All degrees mentioned are in the context of a 360° circle with 0° at the topmost position on the circle. For set size 2 displays, the two colored circles were always presented across from one another on opposite sides of the larger circle, in four possible orientations (colored circles positioned at 0° and 180° (shown in Fig. 3); positioned at 45° and 225°; positioned at 90° and 270°; or positioned at 135° and 315°). For set size 3, the three colored circles were presented in a triangle configuration in four possible orientations (colored circles positioned at 0°, 90°, and 270° (shown in Fig. 3); positioned at 45°, 135°, and 315°; positioned at 90°, 180°, and 270°; or positioned at 135°, 225°, and 315°). For set size 4, the four colored circles could be presented in two possible orientations (colored circles positioned at 0°, 90°, 180°, and 270° (shown in Fig. 3), or positioned at 45°, 135°, 225°, and 315°).

3.1.3. Procedure

Trials began with a 500 ms central fixation cross, where participants were instructed to keep their eyes during the trial. Participants then viewed a string of four random consonants (e.g., "L P J X") for 2 s and were instructed to repeat this string aloud until the response screen. This verbal suppression encouraged participants to visually represent the colors of these circles and not simply encode their respective verbal labels ("red", "blue", "green", etc). A central fixation cross was again

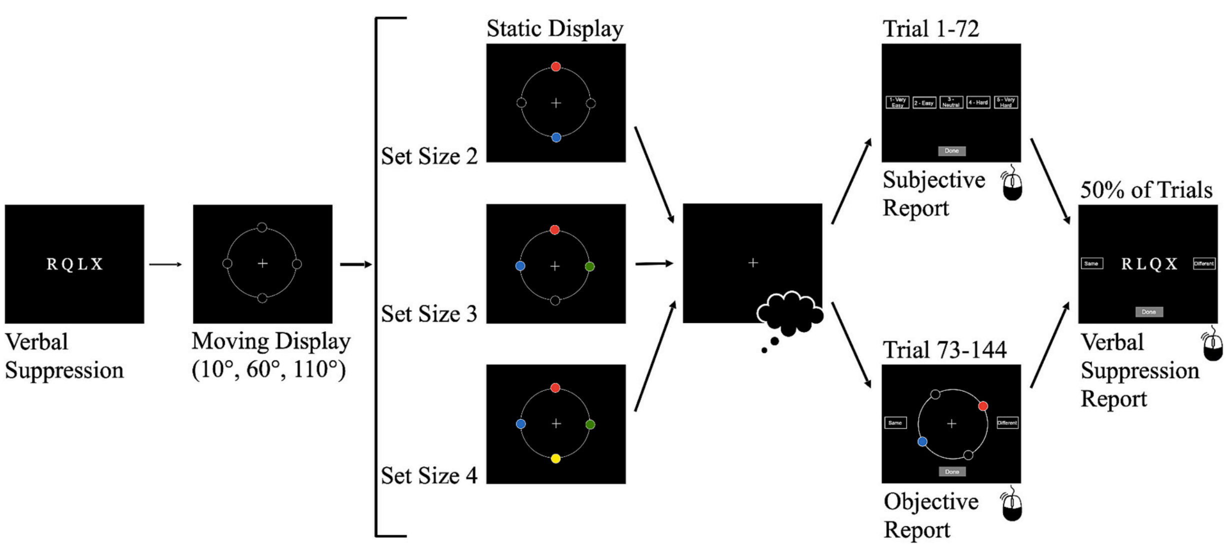

Fig. 3. Procedure for Experiment 2 (displays are not to scale). Participants were shown a string of letters to repeat aloud during the trial (*Verbal Suppression*). They then saw a moving display of unfilled circles (shown here in a " + " orientation) rotating either 10°, 60°, or 110° clockwise or counterclockwise. This was then followed by a static display of unique-colored circles (again shown in a " + " orientation) ranging in set size from 2 to 4. Participants then viewed a black screen with a fixation cross, where they were instructed to imagine the colored circles actually rotating the same degree as the unfilled, dotted circles. Finally, participants performed the response tasks: For trials 1–72, participants indicated the difficulty of performing this transformation on a 1–5 scale ("Very Easy" to "Very Difficult"; subjective report). For trials 73–144, participants performed a change detection task for this transformed display (objective report). On 50% of all trials, either response was followed by a verbal suppression report, where participants indicated whether a string of letters matched the string that they repeated aloud during the trial.

displayed for 500 ms, followed by the moving display. In the moving display, there was a fixation cross and four unfilled, white circles (located in a " + " or " × " orientation) connected by an unfilled, white circle. This configuration could rotate 10°, 60°, or 110° clockwise or counterclockwise. This moving display was followed by a 1 s fixation cross and then the static display for 1.5 s. In the static display, the fixation cross and 4 unfilled, solid white circles connected by the solid white circle were again displayed; however, now 2–4 of the unfilled, solid white circles could be filled in with a color (i.e., red, orange, yellow, blue, green, purple). This display was followed by a central fixation cross for 3.5 s, during which participants were instructed to imagine the static display rotating the same degree of distance as the moving display. Participants then viewed the response screen for their respective condition (either subjective or objective).

In the subjective condition, participants reported the perceived difficulty of imagining the static display rotating the same distance as the moving display (subjective report). Participants were shown a screen with a scale from 1 to 5, with the options of "1 - Very Easy", "2 - Easy", "3 - Neutral", "4 - Hard", or 5 - "Very Hard." To enter their subjective ratings, participants clicked the corresponding option with the mouse, followed by clicking the "Done" box to advance.

In the objective condition, participants were shown a second display of colored circles and performed a change detection task (i.e., does this second display match the transformed display you imagined during the trial?; referred to as the objective report). Half of all trials showed a second display that was an identical display to the one imagined during the trial (no change trial; i.e., if asked to imagine a display with 2 colored and 2 uncolored circles rotated 10° clockwise, this second display would show the exact same configuration rotated 10° clockwise). The remaining half of trials showed a second foil display that differed from the display imagined during the trial in one of the following ways: two colors present in the display could swap positions (color swap foil), a novel color could replace another color in the display (novel color foil), or the configuration could be rotated the wrong distance (incorrect rotation foil; all options were equally likely to occur). Participants were explicitly informed during the instruction period that any of these foil displays could occur. To select their responses, participants used the mouse to click a "Same" box to the left of the display (indicating that the second display matched the imagined display) or a "Different" box to the right of the display (indicating a difference between the second display and the imagined display), and then a "Done" box once they were satisfied with their response. This initiated the next trial or the verbal suppression task (see below).

On 50% of trials in the subjective condition and in the objective condition (or 36 trials in each condition), participants performed a verbal suppression task. In the verbal suppression task, participants were shown a string of letters in the center of the screen and were instructed to respond whether the string was the same or different to the string of letters they had been repeating aloud during the trial. Half of all verbal suppression response trials showed a string of letters identical to those repeated aloud during the trial. The remaining half of trials showed a different string: either any two letters swapped order in the string (this occurred on half of all different string trials), or a novel letter (not previously appearing in the string) replaced a previous letter in the string (this occurred on the remaining half of all different string trials). To select their responses, participants used the mouse to click a "Same" box to the left of the string of letters (indicating that the two strings were identical) or a "Different" box to the right of the string of letters (indicating that the two strings differed), and then a "Done" box once they were satisfied with their response. Feedback was provided after each response, with correct responses followed by a "Correct!" display for 1 s, while incorrect responses were followed by an "Incorrect!" display and a 5 s penalty before the next trial.

Each participant was tested in 8 blocks (with 4 blocks each for the subjective condition and the objective condition) of 18 trials, for a total of 144 trials. There was a 15 s break between each block. Participants always performed the subjective condition blocks first, followed by the objective condition blocks. In this way, participants' perceived accuracy on the objective task could not influence their ratings of perceived difficulty of mental rotation. Trials were randomized between blocks within each condition, and all variables were balanced throughout the experiment (i.e., all combinations of set size, degree of rotation, direction of rotation, and orientation were equally likely to occur during the subjective condition as in the objective condition).

3.2. Analysis and results

For the verbal suppression task in the subjective and objective conditions, participants with an average accuracy performance lower than 2 standard deviations below the mean (subjective condition: $M = 98.3\%$, $SD = 2.3\%$; objective condition: $M = 96.5\%$, $SD = 3.5\%$) were excluded from further analysis ($N = 20–2 = 18$). This stricter cut-off (as compared to a standard cut-off of 50–60% for a two-alternative forced choice task) was chosen to exclude any participants that were not performing the verbal suppression task accurately on the majority of trials. For the change detection task in the objective condition, an additional exclusion criteria of overall change detection accuracy performance lower than 60% (to exclude participants performing near chance) resulted in the exclusion of 2 additional participants, for a total of 16 participants. There were no additional exclusion criteria for difficulty ratings in the subjective condition.

For the subjective condition, difficulty ratings were predicted using a linear mixed effects model, with set size and the degree of rotation as fixed effects, and subjects as the random effect. The model object was then subjected to a Wald Chi-Squared test to check for main effects for each of the factors. All additional post-hoc analyses were conducted using least squares means with Tukey adjustments.

See Fig. 4 for results. Consistent with Experiment 1 (which tested set sizes 1–4 for rotational transformations), there was a significant main effect of set size, $X^2(2) = 89.70$, $p < 0.001$, $\eta_p^2 = 0.074$. Difficulty ratings significantly increased with set size ($M_{\text{Set Size 2}} = 2.27$, $SE_{\text{Set Size 2}} = 0.06$, 95% $CI_{\text{Set Size 2}} = [1.98, 2.56]$; $M_{\text{Set Size 3}} = 2.71$, $SE_{\text{Set Size 3}} = 0.06$, 95% $CI_{\text{Set Size 3}} = [2.43, 3.00]$; $M_{\text{Set Size 4}} = 2.93$, $SE_{\text{Set Size 4}} = 0.06$, 95% $CI_{\text{Set Size 4}} = [2.65, 3.22]$), all $ps \leq 0.006$. We also found a significant main effect for the degree of rotation, $X^2(2) = 250.45$, $p < 0.001$, $\eta_p^2 = 0.182$. Difficulty ratings significantly increased with rotational degree ($M_{10^\circ} = 2.00$, $SE_{10^\circ} = 0.05$, 95% $CI_{10^\circ} = [1.71, 2.29]$; $M_{60^\circ} = 2.85$, $SE_{60^\circ} = 0.06$, 95% $CI_{60^\circ} = [2.56, 3.13]$; $M_{110^\circ} = 3.07$, $SE_{110^\circ} = 0.06$, 95% $CI_{110^\circ} = [2.78, 3.36]$), all $ps \leq 0.006$. There was no significant interaction between set size and the degree of rotation, $X^2(4) = 5.12$, $p = 0.275$, $\eta_p^2 = 0.005$.

For the objective condition, change detection accuracy was predicted using a mixed effects model, with set size and the degree of rotation as the fixed effects, and subjects as the random effect. The model object was then subjected to a Wald Chi-Squared test to determine main effects for each of the factors. Additional post-hoc analyses were once again conducted using least squares means with Tukey adjustments.

In the objective change detection task, performance paralleled the subjective reports of difficulty with a main effect of set size, $X^2(2) = 14.36$, $p < 0.001$, $\eta_p^2 = 0.014$. Average accuracy in the change detection task for set size 2 was 87.24% ($SE = 1.70\%$, 95% CI = [83.61%, 91.86%]), compared to 81.25% for set size 3 ($SE = 1.99\%$, 95% CI = [76.58%, 87.08%]), $p = 0.051$. While there was no significant difference in change detection accuracy between set size 3 and 4 ($M = 76.82\%$, $SE = 2.16\%$, 95% CI = [71.57%, 83.46%]), $p = 0.282$, there was a significant difference between set size 2 and 4, $p < 0.001$. There was no significant main effect of the degree of rotation on change detection accuracy ($M_{10^\circ} = 82.81\%$, $SE_{10^\circ} = 1.93\%$, 95% $CI_{10^\circ} = [78.60\%, 88.52\%]$; $M_{60^\circ} = 83.85\%$, $SE_{60^\circ} = 1.88\%$, 95% $CI_{60^\circ} = [80.12\%, 89.62\%]$; $M_{110^\circ} = 78.65\%$, $SE_{110^\circ} = 2.09\%$, 95% $CI_{110^\circ} = [73.91\%, 85.22\%]$), $X^2(2) = 4.07$, $p = 0.131$, $\eta_p^2 = 0.004$. Overall, these objective findings indicate lower performance as more items needed to

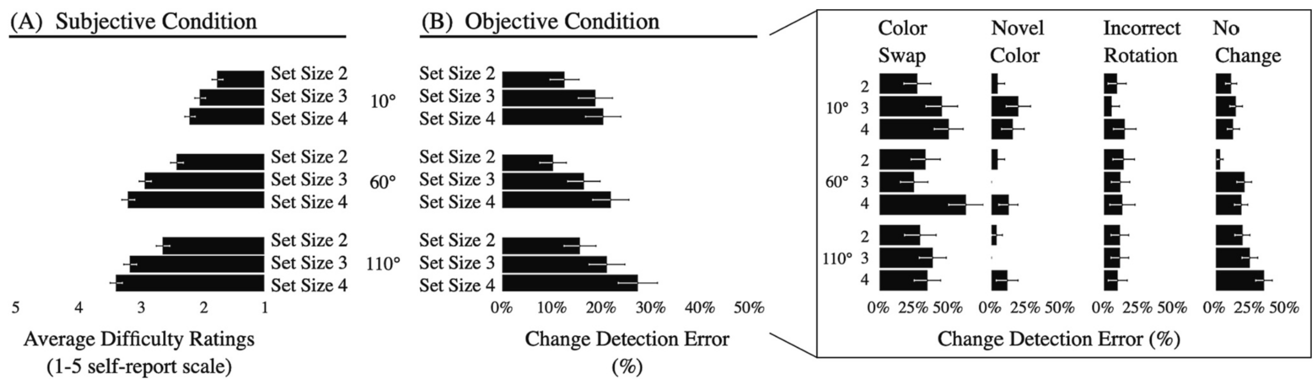

Fig. 4. Results for Experiment 2. (A) Average subjective difficulty ratings for the subjective condition. (B) Average change detection error for the objective condition. Inset displays show average change detection error for the objective condition broken down by change detection trial type (color swaps, novel colors, incorrect rotations, or no change). Error bars indicate standard error.

be maintained without exogenous input — however, a greater degree of rotation did not significantly affect performance. Additionally, there was no significant interaction between set size and degree of rotation, $X^2(4) = 0.71$, $p = 0.951$, $\eta_p^2 = 0.001$.

A second model with the type of change detection trial added as an additional independent variable was then run. In order to make the model viable, only color swaps, incorrect rotations, and no change trials were analyzed, as novel color trials had such high accuracy (M = 92.19%, SE = 1.94%) that there was no variability for the model to fit for some of the condition's levels (see Fig. 4; change detection error was 0% in novel color trials for set size 3 in both the 60° and 110° rotations). This second model was compared to a version of the first model in which novel color trials were also excluded. The second model was a better fit to the data, $X^2(18) = 73.25$, $p < 0.001$, suggesting that the type of change trials impacted performance. In this model, there was a significant main effect of set size, $X^2(2) = 7.97$, $p = 0.019$, and of change detection trial type, $X^2(2) = 42.54$, $p < 0.001$. There was also a significant interaction between degree of rotation and type of change detection trial, $X^2(4) = 9.96$, $p = 0.041$, indicating that the effect of rotational degree on change detection accuracy differed based on the change detection trial type.

Subsequent pairwise tests of the change detection trial types also only analyzed least squares means for color swaps, incorrect rotations, and no change trials, due to the low variability in the novel color trials. These tests showed that accuracy in the change detection task was not significantly different between incorrect rotation (M = 88.54%, SE = 2.30%, 95% CI = [83.92%, 94.22%]) and no change trials (M = 82.64%, SE = 1.58%, 95% CI = [80.48%, 90.29%]), $p = 0.319$. Change detection accuracy, however, was significantly greater for incorrect rotation and for no change trials compared to color swap trials (M = 61.98%, SE = 3.51%, 95% CI = [52.56%, 72.70%]); all $ps < 0.001$.

With this model, there was again no main effect of the degree of rotation, $X^2(2) = 4.74$, $p = 0.093$, and there were no significant interactions between degree of rotation and set size, $X^2(4) = 0.88$, $p = 0.928$, between set size and type of change detection trial, $X^2(4) = 3.91$, $p = 0.419$, or between degree of rotation, set size, and change detection trial type, $X^2(8) = 10.18$, $p = 0.252$.

Given the within-subject design, we were tempted to compute a correlation between subjective ratings and objective performance. However, we hesitate to examine these correlations with sample sizes that are designed to measure condition differences, but not sufficiently powered to measure individual differences. In particular, the vast majority of change trial errors came from swap trials, and the present design only contained 12 such trials per subject.

4. General discussion

In this work, it is important to note that mental imagery cannot be entirely dissociated from visual memory — participants must use working memory to remember how many items to imagine, the color of the items, and the movement of the items in order to imagine these transformations. While previous work had removed the need for working memory in order to test "pure" mental imagery capacity (Keogh & Pearson, 2017), the current work allows us to investigate mental imagery capacity as it commonly is used in a variety of tasks (e.g., when needing to decide if two simultaneously presented item sets are identical when one of the items is rotated (Meyerhoff, Jardine, Stieff, Hegarty, & Franconeri, 2021).

With this consideration in mind, the present study finds that visual mental imagery is limited in both subjective and objective tasks by many of the same factors as those found in previous working memory studies. In Experiment 1, participants rated it more difficult to imagine more items, items that rotated or scaled as opposed to linearly translated (especially when there were more items), and uniquely colored items. The higher difficulty ratings for imagining more objects could be due to similar forms of spatially-modulated crowding in mental imagery that are invoked for capacity limitations in visual memory and online visual tasks (Ahmad et al., 2017; Franconeri et al., 2013; Whitney & Levi, 2011). The movement deficits are compatible with findings from objective visual memory paradigms, with the possible exception of relatively high subjective difficulty for scaling motion, which has shown minimal to no negative effects on objective task performance in the visual memory literature (Jiang et al., 2000; Lam et al., 2006). One possible explanation for this difference in results might be that the image manipulation required here did not convey a sense of scale invariance, as the change in image size was not shown in tandem with a change in viewing distance (e.g., the objects got larger, rather than smaller, the farther they moved from fixation). In past memory work, in contrast, a robust sense of such scale invariance may have led to better objective performance against scaling.

In Experiment 2, items were always uniquely colored and always rotated, and participants also performed an objective change detection test. Participants rated it more difficult to imagine more items (from 2 to 3 to 4), and items that rotated farther (from 10° to 60° to 110°). Objective performance mirrored the trend of worse performance with more items, but surprisingly showed no additional cost when those items rotated beyond 10°. While these subjective reports cannot provide numeric capacity estimates for mental imagery, this finding suggests that set size is a capacity-limiting factor in visual mental imagery, the same way that it limits objective visual memory performance.

Why does the angle of rotation impact subjective ratings, but not objective performance? One explanation for this incongruity is that maintaining unique item identities across a rotation (detecting 'swaps') was simply too hard in the objective task, so participants focused primarily on detecting other types of changes. Indeed, participants easily noticed both new colors in the test display and violations of the prescribed rotation angle, but showed massive error rates when colors swapped locations between the encoding and test displays. This difficulty for detecting swaps is a theme in past visual memory research, where capacity for linking colors to particular objects drops to 1 or 2 items when these items with unique colors have moved since the encoding display (Horowitz et al., 2007; Saiki, 2003; Saiki & Miyatsuji, 2009; Xu & Franconeri, 2015). The present data show even *worse* performance in swap detection than this past work – even for the simple condition of two items rotating only 10°, color swap trials still generated an error rate of approximately 27%, suggesting that participants focused almost entirely on the easier tasks of detecting the introduction of new colors or incorrect rotation magnitudes, and gave up on tracking which items went where.

If people are unable to rotate >1–2 unique items to new locations, then why do participants rate the task of rotating 4 unique circles across 110° at *only* a subjective 3.4 out of 5 average on a difficulty scale? One possibility is that participants employed strategies that made the task subjectively feel easier, even if the objective performance did not receive the same benefits. When asked about potential strategies during Experiment 2's debriefing, 81.3% of the 16 participants reported using a discernible strategy (see Fig. 5 for breakdown of strategies). Of these participants, 53.8% reported using a strategy where they focused their attention on only *one* colored circle and often memorized the colors of the remaining circles without too much attention to their relative locations (see *Bind One Item* in Fig. 5). 38.5% of participants, on the other hand, reported using a strategy where they only focused on *two* of the colored circles at a time and then memorized the remaining colors, not retaining much location information about these other colors (see *Bind Two Items* in Fig. 5). Finally, 7.7% of participants reported using a strategy where they imposed a non-stimulus cue to help them imagine the rotation (e.g., "Imagined they were chasing or that they were splotches of paint leaving a trail"; see *Imposed Non-Stimulus Cue* in Fig. 5 for a different strategy example previously mentioned in Experiment 1)).

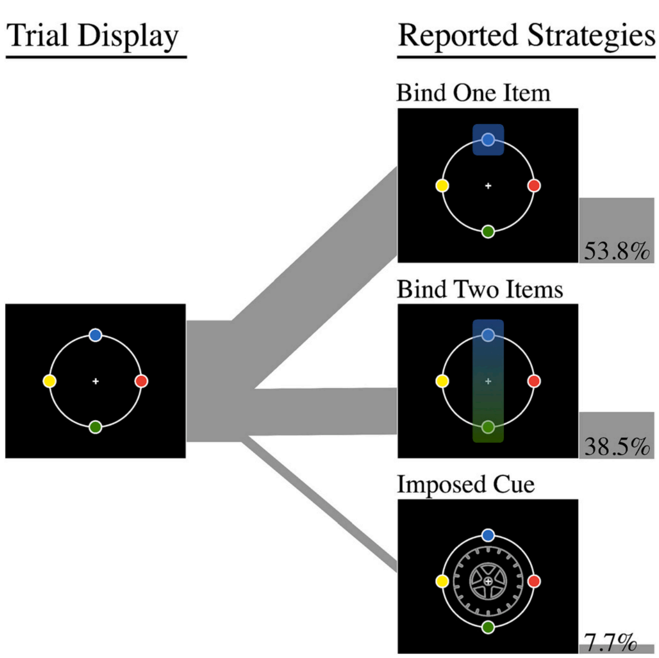

Fig. 5. Self-reported strategy breakdown for Experiment 2. The first column (*Trial Display*) is an example of the rotating colored circles that a participant may have seen during a trial (display not to scale). The second column (*Reported Strategies*) shows the three main strategies that participants noted when asked which strategies they used during the experiment, including when they focused on only one item (*Bind One Item*), focused on only two items (*Bind Two Items*), and when they used a non-stimulus cue to aid in rotation (*Imposed Cue*). Line thickness and percentages indicate what total percentage of participants that did report a strategy responded with one of the following strategies.

These self-report strategies may indicate that participants may have felt the mental imagery task was much easier than it was objectively because they relied on perceived shortcuts, increasing subjective performance without necessarily increasing objective performance. Across the two most popular strategies reported ("Bind One Item" and "Bind Two Items"), overall difficulty ratings in the subjective task were lower for participants that used the "Bind One Item" strategy ($M = 2.44$, $SE = 0.05$) versus those that reported using the "Bind Two Items" strategy ($M = 2.82$, $SE = 0.07$). However, performance in the objective task was higher for participants that used the "Bind Two Items" strategy ($M = 88.06\%$, $SE = 1.71\%$) than for those that used the "Bind One Item" strategy ($M = 80.75\%$, $SE = 1.76\%$) (see Supplementary Materials for a complete breakdown of subjective and objective performance based on the "Bind One Item" and "Bind Two Items" strategies). Given the low sample size for each strategy, we hesitate, however, to compute statistics based on these strategies across their subjective ratings and objective performance, as this post-hoc analysis would not be sufficiently powered. However, it does seem plausible that employing a strategy that requires attending primarily to one item could have made the mental rotation tasks seem subjectively easy to perform (with an average subjective rating of 2.44 out of 5 on a difficulty scale), but did not aid in also increasing objective performance. Future work should explore whether these types of strategy differences could truly affect mental imagery performance (e.g., studying differences in performance when forcing strategies for certain tasks, observing which strategies emerge when focusing on certain transformations or manipulations, etc.).

Finally, we also suspect that, without external feedback to inform them of mistakes, some people *feel* that they can rotate 4 unique circles. Mental imagery could rely on the same summary statistics that are argued to support an illusion of detail in online perception (Brady & Alvarez, 2011; Cohen, Dennett, & Kanwisher, 2016). Rapidly accessing details on-demand can support the feeling that they were always there (Rensink, O'Regan, & Clark, 1997), leaving us surprised when we miss large changes, even in natural scenes (Simons & Rensink, 2005). Those statistics fail us within the cold objectivity of a visual memory test with computer-validated correct and incorrect answers. But in our mental imagery we can remain blissfully unaware that this detail is an illusion.

Funding

This work was supported in part by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1842165, and in part by Grant No. DRL-1661264 from the National Science Foundation.

Open practices statement

The data for all experiments are available at https://osf.io/kavtj, and none of the experiments were pre-registered.

CRediT authorship contribution statement

Cristina R. Ceja: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration, Funding acquisition. **Steven L. Franconeri:** Conceptualization, Methodology, Writing – review & editing, Visualization, Supervision, Funding acquisition.

Declaration of Competing Interest

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Data availability

Data for all experiments are available at https://osf.io/kavtj.

Acknowledgments

We thank Maksim Giljen, Victoria Kam, and Kensei Maeda for assistance with data acquisition, and Caitlyn McColeman and Evan Anderson for helpful comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2023.105436.

References

Ahmad, J., Swan, G., Bowman, H., Wyble, B., Nobre, A. C., Shapiro, K. L., & McNab, F. (2017). Competitive interactions affect working memory performance for both simultaneous and sequential stimulus presentation. *Scientific Reports, 7*(1), 1–16.

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.

Alvarez, G. A., & Thompson, T. W. (2009). Overwriting and rebinding: Why feature-switch detection tasks underestimate the binding capacity of visual working memory. *Visual Cognition, 17*(1–2), 141–159.

Andrade, J., Kemps, E., Werniers, Y., May, J., & Szmalec, A. (2002). Insensitivity of visual short-term memory to irrelevant visual information. *The Quarterly Journal of Experimental Psychology Section A, 55*(3), 753–774.

Bates, K. E., & Farran, E. K. (2021). Mental imagery and visual working memory abilities appear to be unrelated in childhood: Evidence for individual differences in strategy use. *Cognitive Development, 60*, Article 101120.

Berger, G. H., & Gaunitz, S. C. (1979). Self-rated imagery and encoding strategies in visual memory. *British Journal of Psychology, 70*(1), 21–24.

Borst, G., Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2012). Representations in mental imagery and working memory: Evidence from different types of visual masks. *Memory & Cognition, 40*(2), 204–217.

Borst, G., & Kosslyn, S. M. (2010). Individual differences in spatial mental imagery. *The Quarterly Journal of Experimental Psychology, 63*(10), 2031–2050.

Borst, G., Niven, E., & Logie, R. H. (2012). Visual mental image generation does not overlap with visual short-term memory: A dual-task interference study. *Memory & Cognition, 40*(3), 360–372.

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

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

Bundesen, C., & Larsen, A. (1975). Visual transformation of size. *Journal of Experimental Psychology: Human Perception and Performance, 1*(3), 214.

Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in Cognitive Sciences, 20*(5), 324–335.

Finke, R. A., & Kosslyn, S. M. (1980). Mental imagery acuity in the peripheral visual field. *Journal of Experimental Psychology: Human Perception and Performance, 6*(1), 126.

Finke, R. A., & Pinker, S. (1982). Spontaneous imagery scanning in mental extrapolation. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 8*, 142–147.

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.

Golomb, J. D., Chun, M. M., & Mazer, J. A. (2008). The native coordinate system of spatial attention is retinotopic. *Journal of Neuroscience, 28*(42), 10654–10662.

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature, 458*(7238), 632–635.

Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance, 33*(1), 31.

Horowitz, T. S., Klieger, S. B., Fencsik, D. E., Yang, K. K., Alvarez, G. A., & Wolfe, J. M. (2007). Tracking unique objects. *Perception & Psychophysics, 69*(2), 172–184.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(3), 683–702.

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review, 92*, 137–172.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8*, 679–685.

Keogh, R., & Pearson, J. (2011). Mental imagery and visual working memory. *PLoS One, 6*(12). e29221.

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.

Keogh, R., & Pearson, J. (2017). The perceptual and phenomenal capacity of mental imagery. *Cognition, 162*, 124–132.

Keogh, R., Wicken, M., & Pearson, J. (2021). Visual working memory in aphantasia: Retained accuracy and capacity with a different strategy. *Cortex, 143*, 237–253.

Kosslyn, S. M. (1975). Information representation in visual images. *Cognitive Psychology, 7*(3), 341–370.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, Mass: MIT Press.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance, 4*(1), 47.

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. New York: Oxford University Press.

Laeng, B., & Sulutvedt, U. (2014). The eye pupil adjusts to imaginary light. *Psychological Science, 25*(1), 188–197.

Lam, H., Rensink, R. A., & Munzner, T. (2006, July). Effects of 2D geometric transformations on visual memory. In *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization* (pp. 119–126).

Lewis, D. E., O'Reilly, M. J., Khuu, S. K., & Pearson, J. (2013). Conditioning the Mind's eye: Associative learning with voluntary mental imagery. *Clinical Psychological Science, 1*(4), 1–11.

Logie, R. H. (1995). *Visuospatial working memory*. Hillsdale, NJ: Erlbaum.

Logie, R. H. (2003). Spatial and visual working memory: A mental workspace. In D. Irwin, & B. Ross (Eds.)*, Vol. 42. Cognitive vision: The psychology of learning and motivation* (pp. 37–78). New York: Elsevier Science.

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64*, 17–24.

van der Meulen, M., Logie, R. H., & Della Sala, S. (2009). Selective interference with image retention and generation: Evidence for the workspace model. *Quarterly Journal of Experimental Psychology, 62*, 1568–1580.

Meyerhoff, H. S., Jardine, N., Stieff, M., Hegarty, M., & Franconeri, S. (2021). Visual ZIP files: Viewers beat capacity limits by compressing redundant features across objects. *Journal of Experimental Psychology: Human Perception and Performance, 47*(1), 103–115.

Mohr, H. M., Linder, N. S., Dennis, H., & Sireteanu, R. (2011). Orientation-specific aftereffects to mentally generated lines. *Perception, 40*(3), 272–290.

Moriya, J. (2018). Visual mental imagery influences attentional guidance in a visual-search task. *Attention, Perception, & Psychophysics, 80*, 1127–1142.

Olivers, C. N., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology. Human Perception and Performance, 32*(5), 1243–1265.

Pearson, D. G. (2001). Imagery and the visuo-spatial sketchpad. In J. Andrade (Ed.), *Working memory in perspective* (pp. 33–59). Hove, U.K.: Psychology Press.

Pearson, D. G., Logie, R. H., & Gilhooly, K. J. (1999). Verbal representations and spatial manipulation during mental synthesis. *European Journal of Cognitive Psychology, 11*(3), 295–314.

Pearson, J. (2019). The human imagination: The cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience, 20*, 624–634.

Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental imagery: Functional mechanisms and clinical applications. *Trends in Cognitive Sciences, 19*(10), 590–602.

Pearson, J., Rademaker, R. L., & Tong, F. (2011). Evaluating the mind's eye: The metacognition of visual imagery. *Psychological Science, 22*(12), 1535–1542.

Quinn, J. G., & McConnell, J. (1996). Irrelevant pictures in visual working memory. *The Quarterly Journal of Experimental Psychology Section A, 49*(1), 200–215.

Rademaker, R. L., & Pearson, J. (2012). Training visual imagery: Improvements of metacognition, but not imagery strength. *Frontiers in Psychology, 3*, 224.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*(5), 368–373.

Saiki, J. (2003). Feature binding in object-file representations of multiple moving items. *Journal of Vision, 3*(1), 6–21.

Saiki, J., & Miyatsuji, H. (2009). Estimated capacity of object files in visual short-term memory is not improved by retrieval cueing. *Journal of Vision, 9*(3), 16.

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science, 20*(2), 207–214.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*(3972), 701–703.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences, 9*(1), 16–20.

Slotnick, S. D. (2008). Imagery: Mental pictures disrupt perceptual rivalry. *Current Biology, 18*(14), R603–R605.

Slotnick, S. D., Thompson, W. L., & Kosslyn, S. M. (2012). Visual memory and visual mental imagery recruit common control and sensory regions of the brain. *Cognitive Neuroscience, 3*(1), 14–20.

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance, 31*(2), 248–261.

Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and Instruction, 17*(2), 219–234.

Tartaglia, E. M., Bamert, L., Mast, F. W., & Herzog, M. H. (2009). Human perceptual learning by mental imagery. *Current Biology, 19*(24), 2081–2085.

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

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences, 15*(4), 160–168.
Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4*, 1120–1135.
Winawer, J., Huk, A. C., & Boroditsky, L. (2010). A motion aftereffect from visual imagery of motion. *Cognition, 114*(2), 276–284.
Wolfe, J. M., & Cave, K. R. (1999). The psychophysical evidence for a binding problem in human vision. *Neuron, 24*(1), 11–17.
Xu, Y., & Franconeri, S. L. (2015). Capacity for visual features in mental rotation. *Psychological Science, 26*(8), 1241–1251.
Zacks, J. M. (2008). Neuroimaging studies of mental rotation: A meta-analysis and review. *Journal of Cognitive Neuroscience, 20*(1), 1–19.
Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*, 233–235.