Strategic visual memory allocation during complex mental rotation: Evidence for adaptive cognitive resources

Alexandria Grey Sage[1,*]

[1]Explore Science

[*]Corresponding author: ags@explorescience.ai

March 31, 2025

Abstract

The interplay between maintenance and transformation processes in visuospatial cognition remains poorly understood despite its fundamental role in human cognition. We examined how visual working memory for orientation information relates to mental rotation efficiency, and whether imagery vividness moderates this relationship. In a sample of 155 participants, we measured orientation-selective visual working memory across varying set sizes and delays, mental rotation performance across angular disparities (0°-150°), and self-reported imagery vividness. Results revealed modest correlations between visual working memory sensitivity and mental rotation performance, suggesting these processes maintain partial independence despite conceptual overlap. Task complexity did not significantly modulate the strength of these relationships as initially predicted. Imagery vividness correlated with overall visual working memory performance but did not moderate the effects of either set size or angular disparity, indicating that individuals with different imagery abilities may employ distinct strategic approaches to visuospatial processing. These findings illuminate the cognitive architecture underlying complex spatial cognition, demonstrating that resource allocation during visuospatial processing involves both shared mechanisms and process-specific components. This work provides insights for developing targeted interventions to enhance visuospatial skills in educational and clinical contexts, particularly for populations that exhibit specific deficits in spatial cognition.

Keywords: visual working memory, mental rotation, cognitive resources, imagery vividness, individual differences

1 Introduction

The ability to maintain and manipulate visual information underlies a remarkable range of human cognitive achievements—from navigating complex environments to solving abstract spatial problems in mathematics, engineering, and design (Uttal et al., 2013). Two fundamental cognitive processes form the cornerstone of these visuospatial capabilities: visual working memory and mental rotation. Visual working memory (VWM) enables the temporary maintenance of visual information in an accessible, "active" state (Baddeley and Hitch, 1984), with well-documented capacity limitations that vary substantially across individuals (Vogel et al., 2001). Mental rotation, first systematically investigated by Shepard and Metzler (1971), involves the cognitive manipulation of mental representations to predict how objects would appear from different perspectives, with reaction time increasing linearly with angular disparity between stimuli. Together, these processes support our ability to reason about and interact with the visual world, yet the strategic interplay between them remains a significant unresolved question in cognitive psychology.

Despite extensive research examining visual working memory and mental rotation as separate processes, we have limited understanding of how these cognitive functions interact—particularly regarding how limited cognitive resources are allocated between them under varying task demands. A critical linkage was established by Hyun and Luck (2005), who demonstrated that object working memory—not spatial working memory—provides the buffer in which object representations are stored during mental rotation. However, evidence regarding the strength of this relationship remains mixed, with some studies finding limited correlation between VWM and mental rotation abilities (Christie et al., 2013). This empirical ambiguity underscores the need for systematic investigation into how these processes interact, especially since both draw upon shared attentional resources that are subject to strategic allocation based on task demands, individual differences, and cognitive load. The present study addresses this gap by examining how selective visual working memory performance relates to mental rotation efficiency across systematically varied task demands.

Two competing theoretical frameworks have emerged to explain the capacity limitations of visual working memory, with important implications for understanding its relationship to mental rotation. The slot model proposes that VWM consists of a fixed number of discrete slots (typically 3-4 items), with each slot capable of storing a single object with relatively fixed precision (Cowan, 2001; Vogel et al., 2001). In contrast, the resource model conceptualizes VWM as a continuous, limited resource that can be flexibly distributed among visual representations (Bays and Husain, 2008). While some studies have observed discrete slot-like mechanisms (Zhang and Luck, 2008), recent evidence increasingly favors hybrid perspectives that incorporate elements of both models (Tomić and Bays, 2024). The resource framework is particularly relevant to our investigation as it predicts that performance correlations between tasks should strengthen under higher cognitive load when shared attentional resources become constrained. Moreover, object complexity can critically modulate storage capacity beyond simple item count (Alvarez and Cavanagh, 2004), suggesting that the attentional demands of mental rotation may specifically interact with the resource limitations of visual working memory in a load-dependent manner.

Substantial individual differences exist in both VWM capacity (Vogel et al., 2005; Unsworth et al., 2014) and mental rotation ability (Peters et al., 2006), including strategic differences in resource allocation. Visual imagery vividness represents another important dimension of individual difference that may influence both cognitive processes. Originally measured with the Vividness of Visual

Imagery Questionnaire (VVIQ; Marks 1973; McKelvie 1986), imagery vividness has shown mixed relationships with spatial abilities. While some studies suggest a positive relationship between imagery vividness and spatial performance (Hou et al., 2021), others find minimal correlation between subjective ratings and objective measures (Davis et al., 2004). These inconsistent findings suggest that imagery vividness may not directly determine performance levels but rather influence strategic approaches to visuospatial tasks (Keogh and Pearson, 2018). Individual differences in imagery vividness may therefore predict not just overall performance but differential patterns of resource allocation across varying memory loads and transformation demands.

Cognitive load theory provides an important framework for understanding how attentional resources are strategically allocated during complex cognitive tasks. Originally developed by Chandler and Sweller (1991) and refined over time, this theory suggests that cognitive performance deteriorates when task demands exceed available mental resources. Similarly, perceptual load theory established that attentional selection mechanisms become increasingly critical under higher cognitive load (Lavie, 2005). Applied to the interaction between visual working memory and mental rotation, these frameworks suggest that attentional selection mechanisms become more important as task difficulty increases, with performance correlations likely strengthening as shared resources become constrained. This theoretical perspective is particularly relevant when considering complex orientation judgments that require precise spatial resolution during mental rotation tasks, as these judgments may interact with VWM resource limitations in ways that simple presence/absence judgments do not, especially under increased memory load or longer retention intervals.

The current study addresses existing limitations in our understanding by systematically examining the relationship between selective visual working memory performance and mental rotation efficiency across varying task demands. We manipulated both VWM demands (set sizes 3 and 5, delays of 1s and 3s) and mental rotation demands (angular disparities of $0°$, $50°$, $100°$, $150°$) within a single experimental design to test three specific hypotheses. First, we predicted that better selective visual working memory performance would correlate with more efficient mental rotation ($r \geq 0.30$) because both tasks require precise object-based attentional selection mechanisms (Vogel et al., 2005; Cooper and Shepard, 1973) to filter irrelevant information and maintain detailed representations. Second, we hypothesized that this relationship would strengthen under higher cognitive load conditions as shared attentional resources become increasingly taxed, forcing strategic trade-offs that reveal underlying resource dependencies. Third, we predicted that imagery vividness would predict differential strategic approaches, with more vivid imagers showing both enhanced detail encoding (leading to stronger set-size effects due to greater resource allocation per item) and more systematic transformation strategies (steeper RT-by-angle slopes with higher accuracy). This comprehensive approach allows us to examine not only if these cognitive processes share resources, but how these resources are strategically allocated based on task demands and individual differences.

2 Method

2.1 Participants and Procedure

Participants were recruited through Prolific (Peer et al., 2015) using pre-screening criteria established to maximize task validity in an online environment (Kuuramo et al., 2022): ages 18-40 years, balanced gender distribution, normal or corrected-to-normal vision, English proficiency, access to a desktop/laptop computer, and no self-reported neurological or psychiatric conditions. All participants provided informed consent, and ethics approval was obtained from Bellberry Limited.

A priori power analysis was conducted based on a repeated measures ANOVA with within-factors interaction, assuming a medium effect size ($f = 0.25$) (Hyun and Luck, 2005; Shepard and Metzler, 1971). With parameters set to $\alpha = 0.05$, power = 0.95, and correlation among repeated measures = 0.5, the required sample size was 70 participants, increased to 91 to account for anticipated exclusions. The final analyzed sample comprised 198 participants for the Visual Arrays Task, 232 for the Mental Rotation Task, and 374 for the VVIQ2 questionnaire after applying pre-defined exclusion criteria. Cross-task analyses were conducted on the 155 participants with valid data across both cognitive tasks. The final sample had a mean age of 27.23 years (SD = 4.83), with 52.93% female, 46.56% male, and 0.51% preferring not to say.

All experimental tasks were implemented online using PsychoJS (v2021.1.4) and hosted on Pavlovia.org. The order of the two cognitive tasks (Visual Arrays Task and Mental Rotation Task) was counterbalanced across participants, while the VVIQ2 questionnaire was always administered last. Average task completion time was 41.97 minutes (SD = 16.10, range = 20.42-105.38 minutes). The online testing environment did not include automated data quality checks such as browser compatibility verification or attention check trials.

2.2 Visual Arrays Task

2.2.1 Task Overview and Stimuli

The Visual Arrays Task employed a selective attention orientation change detection paradigm to assess participants' ability to maintain orientation information in visual working memory under varying set sizes and delay durations. The visual arrays consisted of blue and red rectangles presented against a gray background. Each rectangle had dimensions proportional to 2.4% of the horizontal and 7.5% of the vertical display space and could appear in one of four possible orientations ($0°$, $45°$, $-45°$, or $90°$) with equal probability. Rectangles were positioned at predefined coordinates to create distinct, easily discriminable stimuli while maintaining consistency with established visual working memory paradigms (Chen et al., 2020; Vogel et al., 2001).

2.2.2 Trial Structure and Procedure

Each trial in the Visual Arrays Task followed a structured sequence (Figure 1). After a 1-second fixation, a color cue ("BLUE" or "RED") was presented centrally for 300 ms, indicating which colored rectangles required attention. The initial memory array appeared for 250 ms, containing either three or five rectangles of each color (set size manipulation). Following a variable retention interval of either 1 or 3 seconds (delay manipulation), a test array was presented with a white dot overlaid on one of the target-colored rectangles. Participants indicated whether this cued rectangle's orientation had changed from the memory array by pressing designated keys. The 2×2 factorial design (Set Size: 3 vs. $5 \times$ Delay: 1s vs. 3s) enabled systematic investigation of how visual working memory performance varies with both memory load and maintenance duration. The experiment began with practice blocks providing feedback, followed by 7 experimental blocks with 16 trials per block, totaling 112 trials with balanced condition representation.

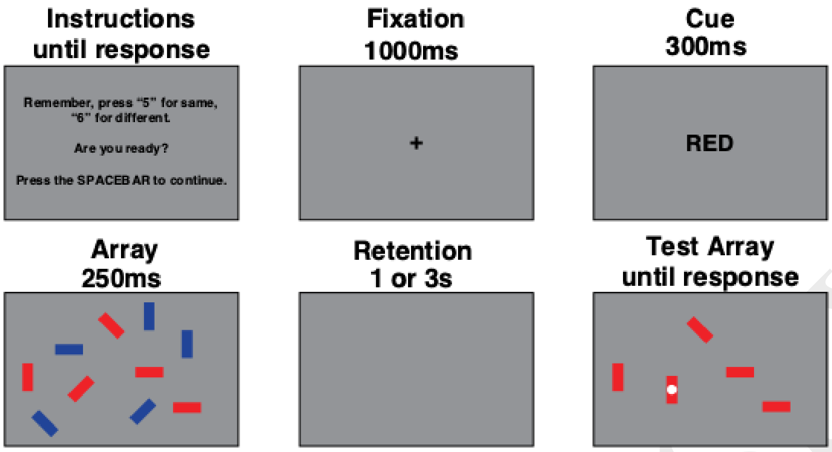

Figure 1: Schematic representation of the visual arrays task trial sequence. The trial sequence begins with instructions displayed until participant response, followed by a fixation cross presented for 1000ms. A color cue ("RED" shown) appears for 300ms, indicating which colored rectangles to attend to. Next, an array of red and blue rectangles at various orientations is displayed for 250ms, followed by a blank retention interval of either 1s or 3s. Finally, a test array appears with a white dot marking one rectangle of the cued color; participants must indicate whether this rectangle's orientation is the same as or different from its orientation in the initial array. Participants respond by pressing "5" for same or "6" for different orientation, and the test array remains visible until response. Set sizes of 3 or 5 rectangles per color were used, with all stimuli presented against a gray background.

2.2.3 Performance Metrics

Performance was quantified using signal detection theory metrics (Green and Swets, 1966; Macmillan and Creelman, 2004), specifically d-prime sensitivity scores calculated for each condition combination (Set Size $\times$ Delay) as $z(\text{hit rate}) - z(\text{false alarm rate})$. For perfect performance, values were adjusted using established methods to prevent infinite d-prime values (Oberauer, 2021). To examine individual differences in resource allocation strategies, we calculated two derived indices: (1) set size effect (difference in d-prime between set size 3 and set size 5), and (2) delay effect (difference in d-prime between 1s and 3s delays). These difference scores were selected to reflect sensitivity to capacity limitations (Vogel et al., 2005) and temporal decay of representations, respectively, though we acknowledge their potentially lower reliability compared to absolute performance measures.

2.3 Mental Rotation Task

2.3.1 Task Overview and Stimuli

The Mental Rotation Task was adapted from Shepard and Metzler's (1971) classic paradigm to measure participants' ability to mentally manipulate 3D objects. Stimuli consisted of 3D block figures obtained from a standardized database of mental rotation stimuli. Each trial presented two 3D block figures that differed in their angular relationship (0°, 50°, 100°, 150°), with an equal number of "same" trials (identical but rotated objects) and "different" trials (mirror images). These angular disparities were selected to systematically evaluate the mental rotation function, allowing assessment of how reaction time varies with increasing angular disparity (Shepard and Metzler, 1971; Grikšienė et al., 2019).

2.3.2 Trial Structure and Procedure

Each trial began with a 250 ms intertrial interval, followed by presentation of two 3D block figures in the center of the display for up to 7.5 seconds. Participants determined whether the figures were the same (identical but rotated, requiring a "b" key press) or different (mirror images, requiring an "n" key press). The experiment began with 20 practice trials including feedback, followed by two experimental blocks of 48 trials each (12 trials per rotation angle), totaling 96 test trials. A 30-second break with a 5-second countdown preceded the second block. This design provided sufficient statistical power for within-subject analyses while keeping the task duration reasonable to avoid fatigue effects (Stark et al., 2024).

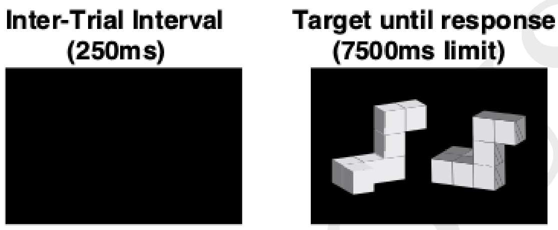

Figure 2: Mental rotation task experimental design. The mental rotation task comprised a sequence of trials designed to assess participants' ability to mentally manipulate 3D objects. Each trial began with an inter-trial interval (250ms) followed by presentation of two 3D block figures displayed simultaneously until response (7500ms maximum). Participants judged whether the objects were identical but rotated (same) or different objects by pressing designated keys. The task included angular disparities of 0°, 50°, 100°, and 150° between object pairs, with equal numbers of same and different trials. Stimuli consisted of 3D block figures selected from a standardized database, presented in the center of the display on a black background. The experiment included 20 practice trials followed by 96 experimental trials divided into two blocks of 48 trials each (12 trials per rotation angle per block).

2.3.3 Performance Metrics

Mental rotation performance was assessed using two primary measures: accuracy (proportion of correct responses) for each angular disparity and reaction time (RT) for correct trials only. Responses faster than 200ms were excluded Ratcliff (1993) to remove anticipatory responses. RT-by-angle slopes were calculated for each participant using linear regression with angular disparity as the predictor (Hale and Jansen, 1994), reflecting the rate of mental rotation with steeper slopes indicating slower rotation (Shepard and Metzler, 1971). Participants with negative RT-by-angle slopes combined with $< 60\%$ accuracy at larger angles were excluded, as this pattern indicates non-compliance with task instructions or use of guessing strategies rather than genuine mental rotation processes.

2.4 VVIQ2 Questionnaire

The Vividness of Visual Imagery Questionnaire-2 (VVIQ2) is an extended version of the original VVIQ developed by Marks (1973), consisting of 32 items measuring the vividness of visual imagery across eight scenarios: familiar person, sunrise, shop front, countryside scene, driving scenario, beach scene, railway station, and garden scene. Each scenario contains four items assessing different aspects of imagery. Participants rated each item on a 5-point scale from 5 (perfectly clear and vivid) to 1 (no image at all). Total scores (range: 32-160) and eight subscale scores (range: 4-20) were calculated, with higher scores indicating greater vividness. For participants with minimal missing data ($< 10\%$), values were imputed using the mean of completed items within the same scenario. Internal consistency was excellent for the total scale ($\alpha = 0.951$) and good to excellent for all subscales (α ranging from 0.776 to 0.879).

2.5 Data Quality and Exclusions

Stringent data quality criteria were established a priori to ensure reliable measurements in the online testing environment (Giraudier et al., 2022). Participant-level exclusion criteria included: performance thresholds ($< 60\%$ overall accuracy on either task), response patterns indicating disengagement (identical responses for > 10 consecutive trials, response times < 200ms on $> 10\%$ of trials, response time standard deviation < 100ms), missing data ($> 25\%$ of trials in any condition), and completion time (< 15 minutes for the entire study). Trial-level exclusions were also applied: for the Visual Arrays Task, trials with RTs < 200ms or $> 10,000$ms were excluded; for the Mental Rotation Task, trials with RTs < 200ms or $> 7,500$ms were excluded; and trials with RTs > 2.5SD from the participant's mean within each condition were removed.

The exclusion rates exceeded our anticipated 30% rate. For the Visual Arrays Task, 188 participants were excluded (48.1%), primarily due to insufficient accuracy (128 participants) or suspicious response patterns (99 participants). For the Mental Rotation Task, 155 participants were excluded (41.0%) based on similar criteria plus negative RT-by-angle slope combined with poor accuracy at larger angles. For the VVIQ2, 11 participants were excluded due to problematic response patterns or missing data. Comparison of demographic characteristics between included and excluded participants revealed no significant differences in age, gender distribution, or education level, suggesting that exclusions did not systematically bias the sample composition.

2.6 Statistical Approach

2.6.1 Within-Task and Cross-Task Analyses

For the Visual Arrays Task, a 2×2 repeated-measures ANOVA with set size (3, 5) and delay (1s, 3s) as within-subject factors was conducted with d-prime as the dependent variable. Log transformation with empirically determined offset was applied to address non-normality. For the Mental Rotation Task, separate repeated-measures ANOVAs with angular disparity ($0°$, $50°$, $100°$, $150°$) as the within-subject factor were conducted for accuracy and reaction time. Relationships between tasks were examined using Pearson correlations among the 155 participants with valid data on both tasks. Specifically, d-prime scores from each Visual Arrays condition were correlated with Mental Rotation performance metrics at each angular disparity, yielding 16 correlations. False Discovery Rate correction ($q = 0.05$) was applied to control for multiple comparisons.

2.6.2 Moderation and Reliability Analyses

The moderating role of imagery vividness was assessed through Pearson correlations between VVIQ2 scores and key performance metrics: set size effects in the Visual Arrays Task, RT-by-angle slopes in the Mental Rotation Task, and overall accuracy in the Mental Rotation Task. Additionally, repeated-measures ANCOVAs were conducted with VVIQ2 z-score as a continuous covariate, examining potential interactions with task conditions. Split-half reliability with Spearman-Brown correction was calculated for both tasks. The Visual Arrays Task showed acceptable overall reliability ($r = 0.770$), though condition-specific reliability varied substantially. The Mental Rotation Task demonstrated stronger reliability, particularly for reaction time measures ($r = 0.976$ for overall mean RT), while accuracy measures showed more variable reliability. Attenuation-corrected correlations were calculated for key cross-task relationships to account for measurement error effects on observed correlations. All analyses were conducted in Python, with data and analysis scripts available on GitHub.

3 Results

3.1 Visual Working Memory Performance Shows Strong Set Size Effects and Modest Delay Effects

Visual working memory for orientation information was primarily constrained by memory load (set size), with retention interval (delay) exerting a considerably smaller influence (Figure 3A). A repeated-measures ANOVA on d-prime sensitivity scores revealed a robust main effect of set size ($F(1, 197) = 1304.83$, $p < 0.001$, partial $\eta^2 = 0.269$), demonstrating that participants remembered orientation information significantly more accurately when maintaining three items compared to five items (mean difference = 0.278, Cohen's $d = 1.428$, 95% CI [1.116, 1.741]). This finding aligns with established capacity limits in visual working memory, which typically accommodates only 3-4 simple items. The effect of delay was also significant but substantially smaller ($F(1, 197) = 9.19$, $p < 0.01$, partial $\eta^2 = 0.011$), indicating a modest decline in memory precision over longer retention intervals. Importantly, the interaction between set size and delay was not significant ($F(1, 197) = 0.63$, $p = 0.428$, partial $\eta^2 < 0.001$), suggesting these factors influence performance independently. As shown in Figure 3B, while nearly all participants exhibited decreased performance with increased set size, the magnitude of this effect varied considerably across individuals.

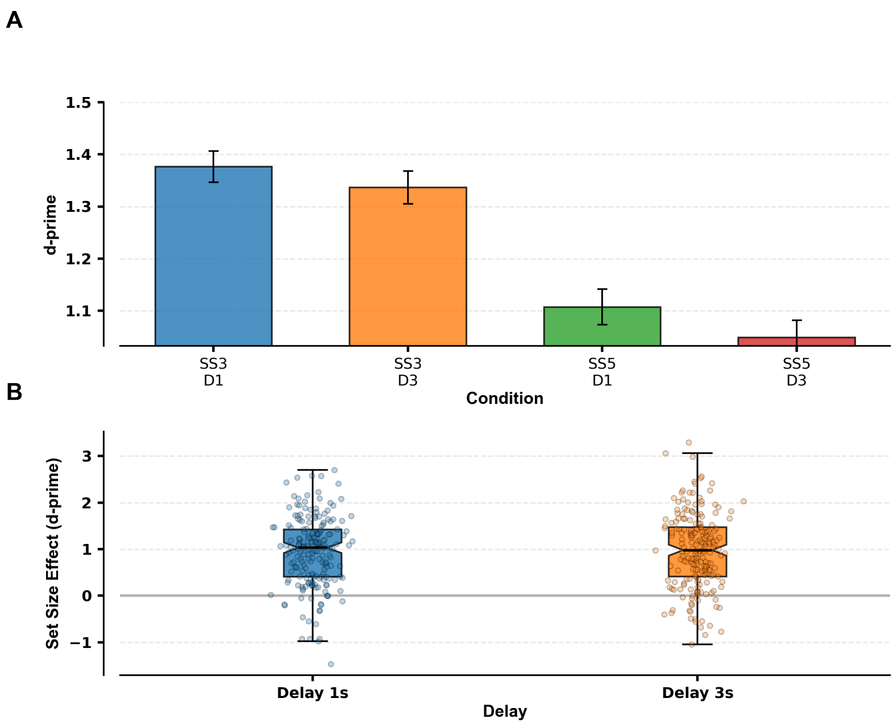

Figure 3: Visual working memory for orientation information is significantly impaired by increased memory load and longer retention intervals. Performance in the Visual Arrays Task demonstrates robust set size effects and more subtle delay effects, with minimal interaction between these factors. The consistent positive set size effect across participants suggests a fundamental capacity limitation in orientation working memory that persists regardless of retention interval. Note that while nearly all participants show performance decrements with increased set size, the magnitude of this effect varies considerably across individuals. (A) Mean d-prime sensitivity scores across set size (3 vs. 5 items) and delay (1s vs. 3s) conditions. Blue and orange bars represent set size 3 conditions; green and red bars represent set size 5 conditions. Horizontal brackets with asterisks indicate significant differences between set sizes at each delay condition. (B) Set size effect (d-prime difference between set size 3 and 5) for 1s (blue) and 3s (orange) delay conditions. Horizontal gray reference line at zero indicates no effect. Notched boxes show median (center line) and interquartile range; colored dots represent individual participants. D-prime was calculated using $z(\text{hit rate}) - z(\text{false alarm rate})$ for each condition. Repeated measures ANOVA revealed significant main effects of set size ($F(1, 198) = 482.90$, $p < 0.001$, $\eta_p^2 = 0.269$) and delay ($F(1, 198) = 14.48$, $p < 0.001$, $\eta_p^2 = 0.011$). Error bars in panel A represent 95% confidence intervals; $n = 198$ participants. $***p < 0.001$.

3.2 Mental Rotation Performance Demonstrates Classic Angular Disparity Effects

Mental rotation performance exhibited the characteristic pattern (Shepard and Metzler, 1971) of decreasing accuracy and increasing reaction time with greater angular disparity (Figure 4), confirming the engagement of analog mental transformation processes. Accuracy decreased systematically across angular disparities (0°: 92.97%, 50°: 88.64%, 100°: 84.80%, 150°: 81.82%), yielding a significant main effect ($F(3, 693) = 154.61$, $p < 0.001$, partial $\eta^2 = 0.161$, 95% CI [0.123, 0.198]). Concurrently, reaction times increased linearly with angular disparity (0°: 1911ms, 50°: 2480ms, 100°: 2938ms, 150°: 3071ms), producing a large main effect ($F(3, 693) = 659.03$, $p < 0.001$, partial $\eta^2 = 0.273$, 95% CI [0.223, 0.323]). The mean RT-by-angle slope was 7.87ms/degree (SD = 3.81), with angular disparity explaining approximately 78% of the variance in reaction times at the group level ($R^2 = 0.78$). Notably, only 40.95% of participants exhibited individually significant slopes, suggesting considerable variability in mental rotation strategies or task engagement. This proportion is lower than typically reported in laboratory-based studies and may reflect the challenges of online testing environments or individual differences in spatial processing approaches.

3.3 Relationship Between Visual Working Memory and Mental Rotation is Task-Specific

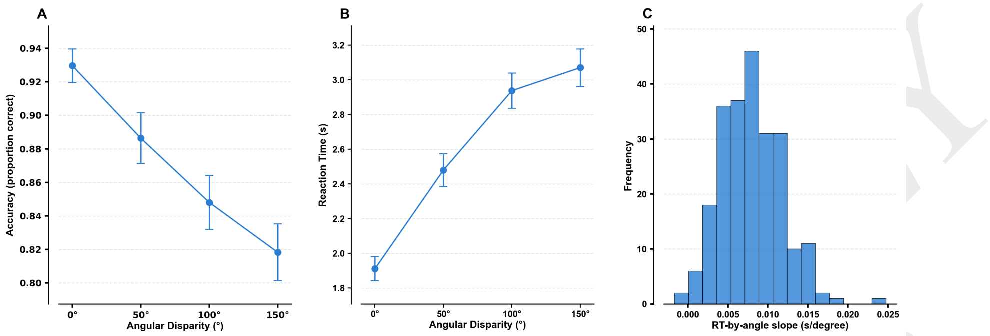

Figure 4: Mental rotation performance demonstrates systematic effects of angular disparity on both accuracy and processing time. Mental rotation task performance reveals the classic pattern of decreasing accuracy and increasing reaction time as angular disparity increases, confirming the engagement of analog mental transformation processes. The distribution of RT-by-angle slopes indicates substantial individual variability in rotation efficiency, with less than half of participants showing statistically significant linear relationships between angle and response time, suggesting potential strategy differences in spatial processing. (A) Mean accuracy (proportion correct) decreases systematically as angular disparity increases from $0°$ to $150°$, with performance dropping from 93% at $0°$ to 82% at $150°$. (B) Mean reaction time (s) shows a characteristic linear increase with angular disparity, rising from 1.9s at $0°$ to 3.1s at $150°$. (C) Histogram showing the distribution of individual RT-by-angle slopes (s/degree), with most participants exhibiting slopes between 0.005-0.015 s/degree. Error bars in panels A and B represent 95% confidence intervals. Blue lines in panels A and B connect mean values across angular disparities. Repeated measures ANOVA revealed significant main effects of angular disparity on both accuracy ($F(3, 693) = 93.56$, $p < 0.001$, $\eta_p^2 = 0.132$) and reaction time ($F(3, 693) = 659.03$, $p < 0.001$, $\eta_p^2 = 0.273$). Only 40.9% of participants exhibited statistically significant RT-by-angle slopes ($p < 0.05$). Data from $n = 232$ participants after exclusions.

The association between visual working memory and mental rotation performance was more modest and task-specific than hypothesized (Figure 5A). The correlation between average visual arrays d-prime and overall mental rotation accuracy ($r = 0.25$, $p < 0.05$) fell below our predicted threshold of $r \geq 0.30$, failing to support Hypothesis 1. Similarly, the correlation with RT-by-angle slope was negligible and directionally opposite to our prediction ($r = 0.049$, $p > 0.05$). These modest associations align with theoretical frameworks suggesting partial independence between spatial storage and spatial transformation processes (Shah and Miyake, 1996). A more detailed analysis of condition-specific relationships revealed sixteen correlations between visual arrays d-prime and mental rotation accuracy across conditions. After applying False Discovery Rate correction, only two correlations remained significant: d-prime in the set size 5, 1s delay condition with mental rotation accuracy at $0°$ ($r = 0.261$, 95% CI [0.109, 0.400], $p = 0.001$) and $100°$ ($r = 0.253$, 95% CI [0.100, 0.393], $p = 0.002$), as shown in Figure 5B. Since the FDR correction with $q = 0.05$ would lead us to expect approximately 0.8 (5% of 16) significant correlations by chance, finding only two significant correlations suggests highly specific relationships rather than a domain-general mechanism connecting these cognitive abilities.

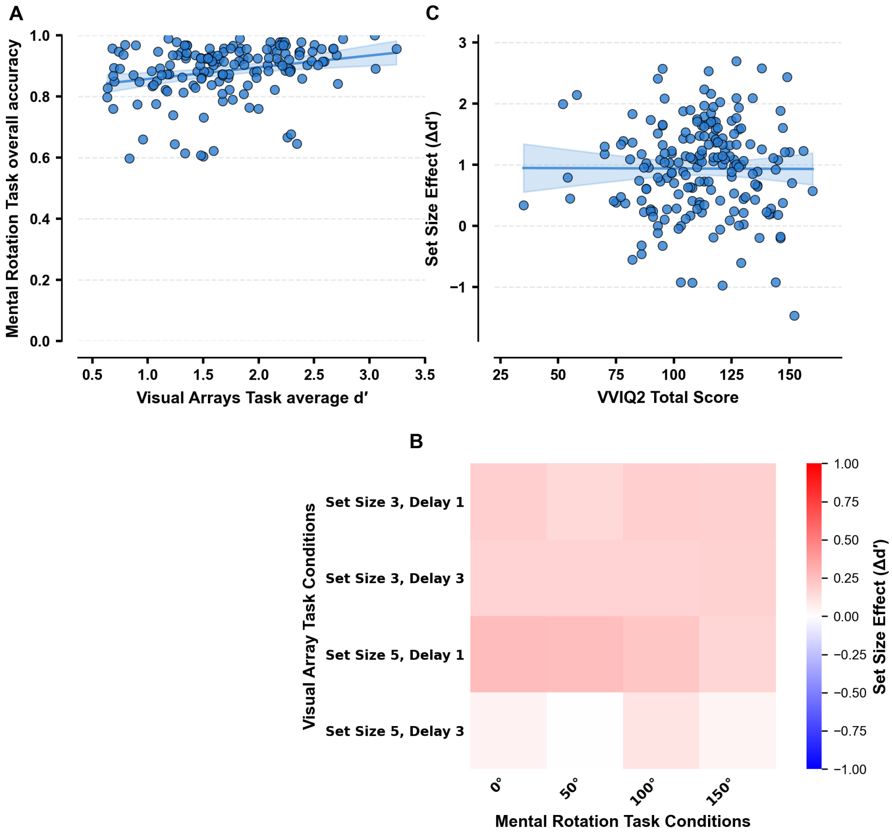

Figure 5: Relationship between visual working memory for orientation information and mental rotation ability shows modest task-specific correlations. The association between visual working memory and mental rotation performance varies across task conditions, with selective correlations between specific task parameters rather than a strong domain-general relationship. While visual arrays task performance positively correlates with mental rotation accuracy (panel A), this relationship is primarily driven by specific visual array conditions correlating with performance at particular angular disparities (panel B). Individual differences in imagery vividness (VVIQ2) show minimal association with the set size effect in visual working memory (panel C). (A) Scatter plot showing the relationship between Visual Arrays Task average d-prime and Mental Rotation Task overall accuracy. Blue dots represent individual participants, with the blue line indicating the best-fit regression line and light blue shading showing the 95% confidence band ($r = 0.193$, $p = 0.002$, $n = 155$). (B) Correlation matrix between Visual Arrays Task conditions (rows) and Mental Rotation Task conditions (columns). Color intensity represents correlation strength from negative (blue) to positive (red), with only Set Size 5, Delay 1 condition showing significant positive correlations with 0° and 100° angular disparities after FDR correction. (C) Scatter plot showing the relationship between VVIQ2 Total Score and Set Size Effect (difference in d-prime between set sizes 3 and 5). Blue dots represent individual participants, with the blue regression line and light blue confidence band indicating no significant correlation ($r = -0.004$, $p = 0.960$, $n = 193$). Performance was measured using d-prime sensitivity index for the Visual Arrays Task and proportion correct for the Mental Rotation Task. Set size effects were calculated as the difference in d-prime between set size 3 and set size 5 conditions. All correlations were assessed using Pearson's r with 95% confidence intervals, and multiple comparisons were corrected using the False Discovery Rate method.

3.4 Imagery Vividness Shows Limited Relationship with Spatial Task Performance

Individual differences in imagery vividness (Marks, 1995) did not moderate task performance as predicted in our hypotheses. The correlation between VVIQ2 scores and set size effect in visual working memory was essentially zero ($r = -0.004$, $p = 0.960$), contradicting our prediction of $r \geq 0.25$ (Figure 5C). Similarly, VVIQ2 scores did not significantly correlate with mental rotation RT-by-angle slopes or overall accuracy. More detailed analyses using repeated-measures ANCOVAs with VVIQ2 as a continuous covariate revealed a significant interaction between imagery vividness and set size in the visual working memory task ($F = 19.03$, $p < 0.001$, partial $\eta^2 = 0.037$, 95% CI [0.010, 0.084]), but no significant main effect of VVIQ2 or interactions with delay. For mental rotation, we observed only a marginally significant main effect of VVIQ2 on accuracy ($p < 0.10$, partial $\eta^2 = 0.020$, 95% CI [0.000, 0.048]) and no significant interactions with angular disparity for either accuracy or reaction time. These results indicate that while imagery vividness may influence how individuals allocate memory resources across different set sizes, it does not specifically modulate performance variations with increasing task demands in the manner predicted by Hypothesis 3.

3.5 Measurement Reliability Contextualizes Correlation Findings

Measurement reliability analyses provide critical context for interpreting the modest relationships observed between tasks, with comprehensive effect sizes and statistical power presented in Table 1. Split-half reliability coefficients (Cronbach, 2025) varied substantially across measures, with excellent reliability for mental rotation reaction times ($r = 0.976$) but lower reliability for visual arrays d-prime measures (ranging from $r = 0.534$ to $r = 0.713$ across conditions). The condition-specific visual arrays measures showed particularly concerning reliability, with the set size 3, delay 3s condition falling well below conventional psychometric standards. Similarly, the reliability of derived effect measures varied considerably, with acceptable reliability for the set size effect ($r = 0.731$) but poor reliability for the delay effect ($r = 0.544$). When applying attenuation correction (Spearman, 2010) to account for measurement error, the estimated "true" correlation between visual working memory and mental rotation performance increased from $r = 0.300$ to $r = 0.367$. These reliability limitations not only constrain the observable correlations between tasks but also highlight important methodological challenges in studying relationships between these cognitive abilities. Table 1 further details the effect sizes for both tasks, providing a comprehensive assessment of the statistical robustness of our primary experimental manipulations across all task conditions.

Table 1: Effect sizes and statistical power for visual working memory and mental rotation tasks. The table presents effect size metrics with 95% confidence intervals for experimental manipulations in the Visual Arrays Task (VA; n=198) and Mental Rotation Task (MRT; n=232), with partial eta-squared (^{2}p) reported for repeated-measures ANOVA effects and Cohen's d for pairwise comparisons. For VA, effect sizes quantify performance differences in d-prime sensitivity scores across set size (3 vs. 5 items) and delay duration (1s vs. 3s) manipulations. For MRT, effect sizes reflect differences in both accuracy (proportion correct) and response time (ms) across angular disparity conditions (0°, 50°, 100°, 150°), with negative values in RT comparisons indicating faster responses at smaller angular disparities. All values derive from within-subjects designs with all participants completing all conditions. Effect size interpretations follow conventional thresholds: small (^{2}p: 0.01-0.05, d: 0.2-0.5), medium (^{2}p: 0.06-0.13, d: 0.5-0.8), or large (^{2}p: 0.14, d: 0.8), with observed power calculated post-hoc based on the obtained effect sizes and sample size.

Effect	Test Type	Effect Size (95% CI)	Interpretation	Observed Power
Visual Arrays Task				
Set size	**ANOVA**	**0.27 [0.18–0.37]**	**Large**	**0.95**
Delay	**ANOVA**	**0.01 [0.00–0.06]**	**Small**	**0.32**
Set size × Delay	**ANOVA**	**0.00 [0.00–0.03]**	**Very small**	**0.06**
Set size 3 vs set size 5	**Pairwise comparison**	**1.43 [1.12–1.74]**	**Large**	**1.00**
Delay 1 vs delay 3	Pairwise comparison	0.25 [-0.03–0.53]	Small	0.94
Mental Rotation Task - Accuracy				
Angular disparity	**ANOVA**	**0.13 [0.08–0.18]**	**Medium**	**0.80**
0° vs 50°	**Pairwise comparison**	**0.42 [0.16–0.68]**	**Small**	**1.00**
0° vs 100°	**Pairwise comparison**	**0.76 [0.49–1.02]**	**Medium**	**1.00**
0° vs 150°	**Pairwise comparison**	**1.02 [0.74–1.29]**	**Large**	**1.00**
50° vs 100°	**Pairwise comparison**	**0.35 [0.09–0.61]**	**Small**	**1.00**
50° vs 150°	**Pairwise comparison**	**0.62 [0.35–0.88]**	**Medium**	**1.00**
100° vs 150°	**Pairwise comparison**	**0.27 [0.01–0.53]**	**Small**	**0.98**
Mental Rotation Task - RT				
Angular disparity	**ANOVA**	**0.27 [0.22–0.32]**	**Large**	**0.95**
0° vs 50°	Pairwise comparison	-1.37 [-1.66—1.09]	Large	1.00
0° vs 100°	Pairwise comparison	-2.05 [-2.37—1.73]	Large	1.00
0° vs 150°	Pairwise comparison	-2.30 [-2.63—1.97]	Large	1.00
50° vs 100°	Pairwise comparison	-1.12 [-1.39—0.84]	Large	1.00
50° vs 150°	Pairwise comparison	-1.30 [-1.58—1.01]	Large	1.00
100° vs 150°	Pairwise comparison	-0.28 [-0.54—0.02]	Small	0.99

4 Discussion

4.1 Summary of Key Findings

Our investigation into the relationship between visual working memory precision and mental rotation efficiency reveals a more nuanced picture than simple resource-sharing theories would predict. Three core empirical phenomena emerged from our experimental results. First, visual working memory demonstrated strong set size effects ($p^2 = 0.269$) but relatively modest delay effects ($p^2 = 0.011$), suggesting that capacity constraints manifest primarily as a function of item quantity rather than maintenance duration—a finding consistent with established capacity limits in working memory (Cowan, 2001). Second, mental rotation performance exhibited robust angular disparity effects on both accuracy ($p^2 = 0.273$) and reaction time ($p^2 = 0.273$), confirming the canonical mental rotation effect first documented by Shepard and Metzler (1971) and further developed by Cooper and Shepard (1973). Third, and perhaps most intriguingly, we observed modest correlations between visual working memory and mental rotation performance ($r = 0.25$ between average VA d-prime and MRT accuracy) that remained relatively stable across varying task demands, contrary to our expectations that these relationships would strengthen under increased cognitive load.

4.2 Evaluation of Pre-registered Hypotheses

When evaluated against our pre-registered hypotheses, our findings present a complex picture of visuospatial cognition. Hypothesis 1, which predicted strong correlations ($r \geq 0.30$) between visual working memory and mental rotation efficiency, was not supported by the observed correlation ($r = 0.25$), though this estimate increased to $r = 0.367$ after correction for measurement attenuation. More notably, Hypothesis 2—predicting stronger correlations under more demanding conditions (larger set sizes, longer delays, and greater angular disparities)—was not supported, as correlations remained stable across varying difficulty levels. This stability challenges the intuition that higher cognitive load would reveal stronger dependencies between these processes. Finally, Hypothesis 3, which predicted that imagery vividness would moderate performance patterns, was not supported. The correlation between VVIQ2 scores and set size effect in visual working memory was negligible ($r = -0.004$, $p = 0.960$), and while a significant interaction between set size and VVIQ2 emerged in our ANCOVA ($F = 19.03$, $p < 0.001$), the overall pattern did not align with our hypothesized moderating effects. These findings suggest that visual working memory and mental rotation rely on distinct yet partially overlapping mechanisms that maintain relative independence even under high cognitive load—a pattern more aligned with multi-component models of visuospatial processing (Baddeley and Hitch, 1984; Cornoldi and Mammarella, 2021) than with unitary resource accounts.

4.3 Resource Allocation in Visual Working Memory

The pronounced set size effect in our visual working memory task provides important insights into resource allocation mechanisms. This robust effect ($p^2 = 0.269$) aligns with resource-based models of visual working memory, where a limited pool of resources must be distributed across all maintained items, resulting in decreased precision as the number of items increases (Bays and Husain, 2008). By contrast, the substantially smaller delay effect ($p^2 = 0.011$) suggests that once information is successfully encoded in visual working memory, it can be maintained relatively well over several seconds. Notably, we observed no significant interaction between set size and delay, indicating these factors affect performance independently rather than compounding—a finding inconsistent with models predicting greater maintenance-related decay with increasing set size due to resource competition. Our d-prime sensitivity scores, calculated using classical signal detection theory (Green and Swets, 1966), revealed these patterns despite variable reliability in our visual working memory measures (ranging from $r = 0.534$ to 0.770). This measurement variability represents an important consideration when interpreting the relationship between visual maintenance and spatial transformation processes.

4.4 Individual Differences in Mental Rotation Strategies

The mental rotation task results confirmed the canonical angular disparity effect, with response times increasing linearly with angular disparity (mean slope = 7.87ms/degree) and accuracy decreasing as angular disparity increased. However, an unexpected finding emerged when examining individual performance patterns: only 40.95% of participants exhibited statistically significant positive RT-by-angle slopes. This surprisingly low proportion reveals considerable heterogeneity in how individuals approach mental rotation tasks, with the majority not demonstrating the linear relationship between angular disparity and response time characteristic of analog mental transformation. Given the excellent reliability of our mental rotation RT measures ($r = 0.976$), this finding likely reflects genuine strategic diversity rather than measurement error. Some participants may employ feature-based analytical strategies less sensitive to angular disparity than the holistic rotation strategy originally proposed by Shepard and Metzler (1971). This strategic diversity has significant implications for theories of spatial cognition that assume uniformity in mental transformation processes and highlights the importance of examining individual differences rather than relying solely on aggregate performance patterns.

4.5 Cross-Task Relationships and Resource Allocation

The pattern of correlations between visual working memory and mental rotation tasks provides critical insights into the nature of resource allocation across visuospatial processes. The modest correlation between average visual arrays d-prime and overall mental rotation accuracy ($r = 0.25$) suggests partial overlap in the cognitive resources supporting these tasks, but falls short of indicating a strong, unified resource pool. After applying attenuation correction to account for measurement error (Spearman, 2010), the estimated "true" correlation increased to $r = 0.367$—a moderate relationship that still indicates substantial independence between processes. This measurement correction is particularly important given the asymmetric reliability between our tasks, with the visual arrays task showing much lower reliability ($r = 0.534$ to 0.770) than the mental rotation task ($r = 0.976$ for RT measures). Even after correction, the modest correlation suggests that visual working memory and mental rotation rely on substantially dissociable mechanisms, a finding that aligns with specialized component models of visuospatial working memory (Cornoldi and Mammarella, 2021) rather than unitary resource accounts. Moreover, the stability of these correlations across different levels of task demand (set size, delay duration, and angular disparity) further supports the view that these processes maintain their relative independence even under increased cognitive load.

4.6 The Role of Imagery Vividness in Visuospatial Processing

Contrary to our predictions, individual differences in imagery vividness, as measured by the VVIQ2 questionnaire (building on the subjective imagery assessment tradition established by Marks (1973)), demonstrated minimal relationship with performance patterns across tasks. The correlation between VVIQ2 scores and set size effect in visual working memory was effectively zero ($r = -0.004$), contradicting our prediction of a moderate relationship ($r \geq 0.25$). Similarly, VVIQ2 scores did not significantly correlate with mental rotation RT-by-angle slopes or overall accuracy. While we observed a significant interaction between set size and VVIQ2 in our ANCOVA analysis, the overall pattern indicates that subjective imagery vividness does not moderate performance in the straightforward manner we had hypothesized. This disconnect between subjective imagery experiences and objective task performance aligns with previous research showing limited relationships between self-reported imagery capacity and behavioral measures (Davis et al., 2004). Rather than directly determining performance, imagery vividness may influence cognitive strategy selection in ways not captured by our primary performance metrics, or alternatively, may reflect phenomenological aspects of imagery experience that operate independently from the mechanisms supporting visual working memory maintenance and mental transformation processes.

4.7 Methodological Considerations and Limitations

Several methodological factors warrant careful consideration when interpreting our findings. Most notably, our online testing paradigm resulted in unexpectedly high exclusion rates—48.1% for the Visual Arrays Task and 40.97% for the Mental Rotation Task—substantially exceeding the 30% anticipated in our pre-registration. While consistent with known challenges in online cognitive research (Crump et al., 2013), these exclusion rates raise important questions about sample representativeness. Participants meeting our inclusion criteria may systematically differ from the broader population in ways that affect generalizability, potentially selecting for individuals with better sustained attention, stronger motivation, or more suitable testing environments. Additionally, the differential reliability between our key measures represents a fundamental constraint on detecting true relationships between these processes. The excellent reliability of mental rotation measures ($r = 0.976$ for RT) contrasts sharply with the more modest reliability of visual working memory measures (as low as $r = 0.534$), potentially attenuating observed cross-task correlations. Finally, online testing introduced uncontrolled variability in display characteristics, timing precision, and environmental factors that may have differentially affected our tasks, particularly the more perceptually demanding visual arrays paradigm, which relies on precise stimulus presentation and timing.

4.8 Theoretical Implications and Future Directions

Our findings contribute to a more nuanced framework of visuospatial cognition that accommodates both shared and distinct components in visual working memory and mental rotation. Rather than supporting either strictly unitary or fully modular accounts, our results suggest partially overlapping yet functionally distinct mechanisms that maintain relative independence even under increased cognitive load. This perspective has several important implications for future research. First, neuroimaging investigations could identify the neural substrates of these partially overlapping mechanisms, potentially revealing both shared activation patterns and process-specific neural signatures. Second, eye-tracking methodologies could illuminate strategy differences that might explain the surprising individual variability in mental rotation performance patterns we observed. Third, the development of adaptive procedures that optimize task difficulty at the individual level could enhance measurement precision and reveal more nuanced relationships between maintenance and transformation processes. From an applied perspective, our findings suggest that training programs targeting either visual working memory or mental rotation might show limited transfer effects, necessitating more comprehensive approaches to enhance visuospatial cognition. The strategic diversity revealed in our mental rotation task also highlights the importance of considering individual differences in cognitive approach when designing interventions to support spatial thinking in educational contexts, particularly in STEM fields where spatial visualization abilities predict achievement.

4.9 Conclusion and Practical Implications

The partial independence between visual working memory and mental rotation processes demonstrated in our study has significant implications for educational and applied contexts. In educational settings where spatial visualization abilities strongly predict achievement in STEM fields, our findings suggest that training interventions targeting either visual maintenance or spatial transformation abilities in isolation may yield limited transfer effects. Instead, comprehensive training approaches addressing multiple components of visuospatial cognition may prove more effective. The strategic diversity we observed in mental rotation performance further underscores the need for instructional approaches that accommodate different cognitive styles and strategies in spatial reasoning. From a methodological perspective, our results highlight the critical importance of addressing measurement reliability in cognitive research, particularly when examining individual differences and cross-task relationships. The asymmetric reliability between visual working memory and mental rotation measures in our study exemplifies a common challenge in cognitive research that requires careful consideration in both research design and data interpretation. By advancing methodological rigor, acknowledging strategic diversity, and developing more sophisticated models of visuospatial resource allocation, future research can deepen our understanding of the foundations of spatial cognition and inform more effective interventions for enhancing these essential cognitive abilities across educational, professional, and clinical domains.

Acknowledgments

We thank all participants who contributed their time to this study.

Funding

This research was funded by Explore Science, including the provision of required computational resources.

References

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, doi:10.1111/j.0963-7214.2004.01502006.x.

Baddeley, A. D. & Hitch, G. (1984). Working memory. *Psychological Medicine*, *14*(2), 265–271, doi:10.1017/S0033291700003548.

Bays, P. M. & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854, doi:10.1126/science.1158023.

Chandler, P. & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*(4), 293–332, doi:10.1207/S1532690XCI0804_2.

Chen, S., Kocsis, A., Liesefeld, H., Müller, H., & Conci, M. (2020). Object-based grouping benefits without integrated feature representations in visual working memory. *Attention, Perception Psychophysics*, *83*, 1357 – 1374, doi:10.3758/s13414-020-02153-5.

Christie, G. J., Cook, C. M., Ward, B. J., Tata, M. S., Sutherland, J., Sutherland, R., & Saucier, D. (2013). Mental rotational ability is correlated with spatial but not verbal working memory performance and p300 amplitude in males. *PLoS ONE*, doi:10.1371/journal.pone.0057390.

Cooper, L. A. & Shepard, R. N. (1973). *CHRONOMETRIC STUDIES OF THE ROTATION OF MENTAL IMAGES*, (pp. 75–176). Elsevier.

Cornoldi, C. & Mammarella, I. C. (2021). *The organization of visuospatial working memory*, (pp. 7–26). Routledge, 1 edition.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114, doi:10.1017/S0140525X01003922.

Cronbach, L. J. (2025). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334, doi:10.1007/BF02310555.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating amazon's mechanical turk as a tool for experimental behavioral research. *PLoS ONE*, *8*(3), e57410, doi:10.1371/journal.pone.0057410.

Davis, M., McMahon, M., & Greenwood, K. (2004). The role of visual imagery in the enhanced cognitive interview: guided questioning techniques and individual differences†‡. *Journal of Investigative Psychology and Offender Profiling*, *1*, 33–51, doi:10.1002/JIP.6.

Giraudier, M., Ventura-Bort, C., Wendt, J., Lischke, A., & Weymar, M. (2022). Memory advantage for untrustworthy faces: Replication across lab- and web-based studies. *PLoS ONE*, *17*, e0264034, doi:10.1371/journal.pone.0264034.

Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*.

Grikšienė, R., Arnatkevičiūtė, A., Monciunskaite, R., Koenig, T., & Ruksenas, O. (2019). Mental rotation of sequentially presented 3d figures: sex and sex hormones related differences in behavioural and erp measures. *Scientific Reports*, *9*, doi:10.1038/s41598-019-55433-y.

Hale, S. & Jansen, J. (1994). Global processing-time coefficients characterize individual and group differences in cognitive speed. *Psychological Science*, *5*, 384 – 389, doi:10.1111/j.1467-9280.1994.tb00290.x.

Hou, J., Chen, C., Dong, Q., Prabhakaran, V., & Nair, V. A. (2021). Superior pitch identification ability is associated with better mental rotation performance. *Musicae Scientiae*, doi:10.1177/10298649211013409.

Hyun, J. & Luck, S. J. (2005). Visual working memory as the substrate for mental rotation. *Journal of Vision*, *5*(8), 425–425, doi:10.1167/5.8.425.

Keogh, R. & Pearson, J. (2018). The blind mind: No sensory visual imagery in aphantasia. *Cortex*, *105*, 53–60, doi:10.1016/j.cortex.2017.10.012.

Kuuramo, C., Saarinen, J., & Kurki, I. (2022). Forgetting in visual working memory: Internal noise explains decay of feature representations. *Journal of Vision*, *22*(8), 8, doi:10.1167/jov.22.8.8.

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, *9*(2), 75–82, doi:10.1016/j.tics.2004.12.004.

Macmillan, N. A. & Creelman, C. D. (2004). *Detection theory: A user's guide, 2nd ed.*

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, *64*(1), 17–24, doi:10.1111/J.2044-8295.1973.TB01322.X.

Marks, D. F. (1995). *New directions for mental imagery research.*

McKelvie, S. (1986). Effects of format of the vividness of visual imagery questionnaire on content validity, split-half reliability, and the role of memory in test-retest reliability. *British Journal of Psychology*, doi:10.1111/J.2044-8295.1986.TB01997.X.

Oberauer, K. (2021). Measurement models for visual working memory-a factorial model comparison. *Psychological review*, *128*(8), 757–782, doi:10.17605/OSF.IO/ZWPRV.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2015). Beyond the turk: An empirical comparison of alternative platforms for online behavioral research. *SSRN Electronic Journal*, doi:10.2139/ssrn.2594183.

Peters, M., Lehmann, W., Takahira, S., Takeuchi, Y., & Jordan, K. (2006). Mental rotation test performance in four cross-cultural samples (n = 3367): Overall sex differences and the role of academic program in performance. *Cortex*, *42*(7), 1005–1014, doi:10.1016/S0010-9452(08)70206-5.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532, doi:10.1037/0033-2909.114.3.510.

Shah, P. & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125*(1), 4–27, doi:10.1037/0096-3445.125.1.4.

Shepard, R. N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703, doi:10.1126/science.171.3972.701.

Spearman, C. (2010). The proof and measurement of association between two things. *International Journal of Epidemiology*, *39*(5), 1137–1150, doi:10.1093/ije/dyq191.

Stark, P., Bozkir, E., Sójka, W., Huff, M., Kasneci, E., & Göllner, R. (2024). The impact of presentation modes on mental rotation processing: a comparative analysis of eye movements and performance. *Scientific Reports*, *14*, doi:10.1038/s41598-024-60370-6.

Tomić, I. & Bays, P. (2024). A dynamic neural resource model bridges sensory and working memory. *eLife*, *12*, doi:10.1101/2023.03.27.534406.

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26, doi:10.1016/j.cogpsych.2014.01.003.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, *139*(2), 352–402, doi:10.1037/a0028446.

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), 500–503, doi:10.1038/nature04171.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 92–114, doi:10.1037/0096-1523.27.1.92.

Zhang, W. & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235, doi:10.1038/nature06860.

5 Supplementary Material

5.1 Measurement Reliability Analysis: Task Performance Consistency

We conducted comprehensive reliability analyses across both cognitive tasks to evaluate measurement robustness and contextualize our main findings. This methodological assessment is critical for interpreting the limited cross-task correlations observed in our primary analyses. Using split-half reliability (correlating performance on odd versus even trials with Spearman-Brown correction), we identified a substantial disparity in measurement consistency between tasks. As illustrated in Figure 6A, mental rotation response time measures demonstrated excellent reliability ($r > 0.9$) across all angular disparity conditions, while visual arrays d-prime sensitivity metrics showed predominantly questionable to unacceptable reliability ($r < 0.7$).

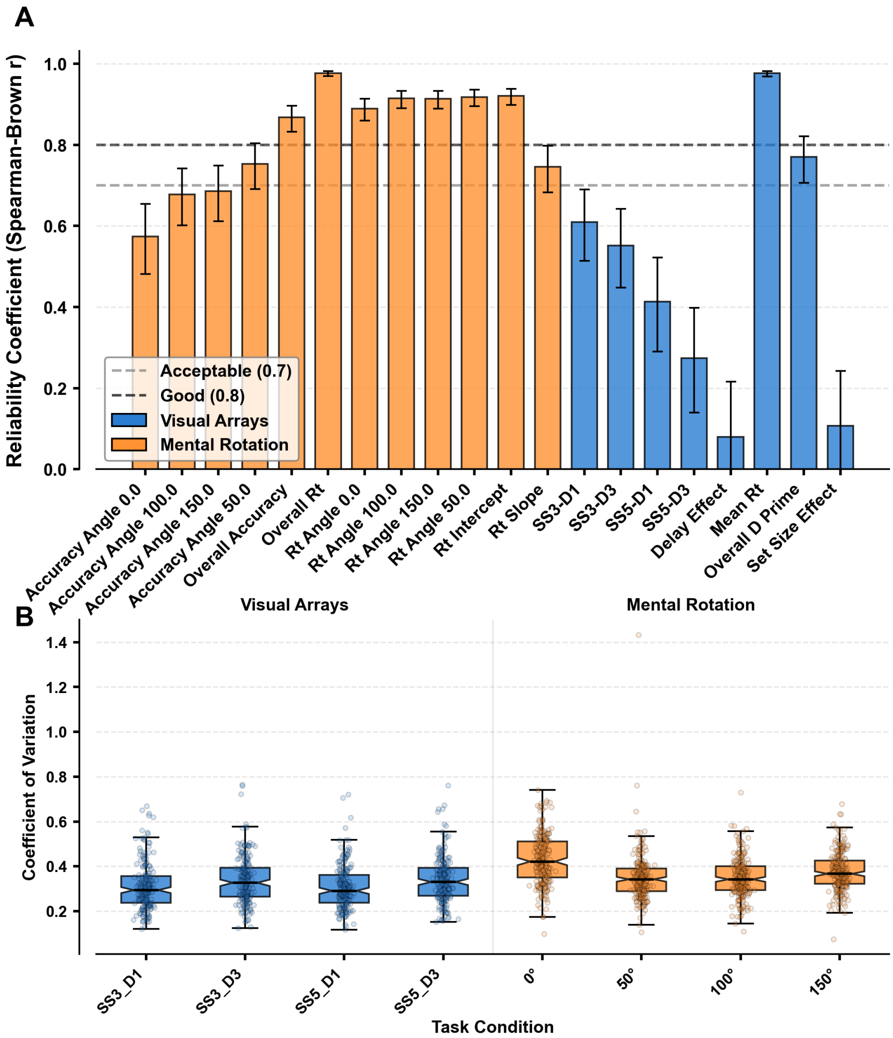

Figure 6: Mental rotation task demonstrates superior measurement reliability compared to visual arrays task across multiple performance metrics. Reliability assessment across cognitive tasks reveals a consistent methodological challenge in measuring visual working memory processes compared to mental rotation abilities. Mental rotation response time measures achieve excellent reliability ($r > 0.9$), while visual arrays d-prime sensitivity metrics show predominantly questionable to unacceptable reliability ($r < 0.7$). Critically, derived measures (set size effect, delay effect) exhibit particularly poor reliability ($r < 0.2$), indicating that difference scores between experimental conditions should be interpreted with substantial caution. These measurement properties may explain the limited cross-task correlations observed in the main analyses. (A) Reliability coefficients (Spearman-Brown corrected) for key performance measures in Visual Arrays (blue) and Mental Rotation (orange) tasks. Horizontal dashed lines indicate thresholds for acceptable (0.7, gray) and good (0.8, black) reliability. Error bars represent 95% confidence intervals. (B) Response time variability (coefficient of variation) across task conditions. Box plots display median (center line), interquartile range (box), and distribution range (whiskers), with small points representing individual participant values. Higher coefficient values indicate greater response time inconsistency within participants. Split-half reliability was calculated by correlating performance on odd versus even trials and applying the Spearman-Brown correction. Visual Arrays conditions: set size (SS) 3 or 5 items with delay (D) of 1s or 3s; Mental Rotation conditions: angular disparities of 0°, 50°, 100°, and 150°. Visual Arrays task $n = 199$ participants; Mental Rotation task $n = 232$ participants.

This striking reliability difference represents a fundamental methodological challenge in measuring visual working memory processes. Of particular concern were the derived measures in the visual arrays task (set size effect, delay effect), which exhibited exceptionally poor reliability ($r < 0.2$). These findings have significant implications for the interpretation of difference scores between experimental conditions, especially when using the visual arrays paradigm. Beyond reliability coefficients, we also assessed response time variability (coefficient of variation) across conditions (Figure 6B), revealing substantial individual differences in performance consistency. This complementary metric provides additional insight into within-participant stability and further explains the disparity in measurement properties between these widely-used cognitive tasks.

5.2 Cross-Task Correlations: Shared Mechanisms Between Visual Working Memory and Mental Rotation

To investigate potential cognitive mechanisms shared between visual working memory and mental rotation processes, we examined cross-task performance correlations across varying levels of task demand. This analysis offers insight into how these distinct visuospatial cognitive abilities might relate and interact. As detailed in Table 2, we observed modest but consistent associations between visual arrays d-prime sensitivity and mental rotation accuracy, with correlation strengths ranging from $r = 0.155$ to $r = 0.261$. Importantly, the pattern of these correlations revealed stronger relationships at smaller angular disparities, suggesting that visual working memory resources may be differentially engaged depending on the complexity of mental transformation required.

Table 2: MRT Accuracy by Angular Disparity. Cross-task correlations between visual working memory and mental rotation performance across varying task demands ($n = 155$). Values represent Pearson's r correlation coefficients between d-prime sensitivity in the Visual Arrays (VA) task (rows) and accuracy/response time measures in the Mental Rotation Task (MRT) at different angular disparities (columns). VA Conditions represent increasing memory demands: I = Set Size 3, Delay 1s; II = Set Size 3, Delay 3s; III = Set Size 5, Delay 1s; IV = Set Size 5, Delay 3s. MRT conditions I-IV represent increasing angular disparities (0°, 50°, 100°, 150°). Bold values with asterisks (*) indicate significance after False Discovery Rate (FDR) correction ($p < .05$), while values with daggers (†) indicate significance before correction only ($p < .05$ uncorrected). Row and column averages quantify overall correlation patterns across conditions, revealing stronger relationships between visual working memory performance and mental rotation accuracy at smaller angular disparities.

	MRT Accuracy by Angular Disparity				Row
VA Condition	I	II	III	IV	Avg.
I	*0.193*†	*0.187*†	0.155	*0.190*†	0.181
II	*0.169*†	*0.177*†	*0.168*†	*0.172*†	0.171
III	**0.261***	*0.159*†	**0.253***	*0.222*†	0.224
IV	0.047	0.045	0.003	0.105	0.050
Column Avg.	0.168	0.142	0.145	0.172	
	MRT Response Time by Angular Disparity				**Row**
VA Condition	I	II	III	IV	Avg.
I	0.081	0.055	0.105	0.065	0.077
II	0.057	−0.039	0.014	0.057	0.022
III	−0.062	−0.118	−0.031	−0.040	−0.063
IV	0.015	0.007	0.045	0.017	0.021
Column Avg.	0.023	−0.024	0.033	0.025	

$^*p < .05$ FDR-corrected. $^\dagger p < .05$ uncorrected. NA = data not available.

These cross-task relationships demonstrated a clear pattern of specificity. Correlations between visual working memory performance and mental rotation accuracy were consistently stronger than those with mental rotation response time, with several reaching significance before correction for multiple comparisons. The most robust associations emerged between visual arrays performance under higher memory load conditions (Set Size 5) and mental rotation accuracy at smaller angular disparities (0° and 100°), with two correlations remaining significant after FDR correction. This differential pattern suggests that visual working memory engagement may be particularly important during less complex mental transformations or when the working memory system is already under higher load. The modest correlation strengths observed (predominantly below $r = 0.3$) align with theoretical accounts suggesting partial but incomplete overlap between these cognitive processes.

5.3 Vividness of Visual Imagery and Task Performance: Examining Cognitive Strategy Differences

To explore how individual differences in cognitive strategy might influence task performance, we examined the relationship between self-reported imagery abilities and objective performance measures. The Vividness of Visual Imagery Questionnaire-2 (VVIQ2) was administered to assess participants' subjective imagery vividness, with the instrument demonstrating excellent internal consistency (Cronbach's $\alpha = 0.951$) and strong subscale reliability (α ranging from 0.776 to 0.879). This psychometric robustness provided a solid foundation for investigating whether imagery vividness might explain individual differences in visual working memory and mental rotation performance.

Our analysis revealed a complex relationship between imagery vividness and cognitive task performance. Contrary to our initial hypotheses, VVIQ2 scores did not show straightforward correlations with either visual arrays set size effects or mental rotation performance metrics at our pre-established threshold ($r \geq 0.25$). However, more nuanced analysis using repeated measures ANCOVA uncovered significant interactions between VVIQ2 scores and experimental manipulations: specifically between imagery vividness and set size in the visual arrays task ($F = 19.03$, $p < 0.001$), and between imagery vividness and angular disparity for mental rotation accuracy ($p < 0.001$). These interaction effects suggest that while imagery vividness may not directly predict overall performance, it appears to modulate how cognitive resources are allocated across varying levels of task difficulty. Further subscale analysis found no specific imagery domains (e.g., familiar person, sunrise, garden) that correlated significantly with task performance after FDR correction, suggesting general imagery vividness, rather than domain-specific abilities, may influence the visuospatial cognitive processes measured in our paradigms.

5.4 Gender Comparisons and Additional Methodological Considerations

We conducted exploratory analyses examining potential gender differences and additional methodological approaches to enrich interpretation of our primary results. When examining gender effects across task performance metrics, we found no significant differences after controlling for overall performance levels. To ensure equitable comparisons, performance metrics were standardized within gender groups, with the resulting dataset of 4,940 observations showing no systematic gender-based differences in efficiency scores (calculated as RT/accuracy) across both visual arrays and mental rotation tasks. This finding contributes to the ongoing discourse regarding potential gender effects in visuospatial cognition, suggesting that when appropriate standardization procedures are applied, performance patterns remain consistent across gender groups in our specific cognitive paradigms.

The comprehensive supplementary analyses presented here provide critical context for interpreting the main findings of our study. The marked difference in reliability between visual arrays and mental rotation tasks highlights the methodological importance of assessing and reporting measurement properties in cognitive research, particularly when examining individual differences or correlations between cognitive measures. Additionally, the complex relationship patterns observed between visual working memory and mental rotation processes inform theoretical frameworks regarding the modularity and interaction of visuospatial cognitive systems. The measurement challenges identified in derived measures (particularly difference scores) underscore the importance of careful methodological consideration when designing and interpreting experiments aimed at isolating specific cognitive processes. Together, these supplementary analyses strengthen the foundation for our primary conclusions while providing valuable methodological insights for future research in this domain.