Flexible Resource Allocation in Visual Working Memory: Insights from Pupillary Responses and Behavioral Precision

Thomas Wahlund

Department of Psychology
Degree Project 30 Higher Education credits
Psychology
International Master’s Program in Psychology (120 credits)
Spring term 2023
Supervisor: Ronald Van den Berg (SU) and Mikael Lundqvist (KI)
Abstract

An ongoing debate exists on how visual working memory (VWM) utilizes its limited resources. This study examined how VWM resource allocation affects memory precision and pupil size. The first aim was to explore if VWM encodes items with a precision relative to their importance according to a flexible resource allocation model or encodes all items equally according to a discrete capacity model. A second aim was to investigate if phasic pupil dilation (PPD) reflected manipulations of item importance. In two experiments, 37 participants briefly memorized four colored items – circles and shapes – and reported on one randomly probed item in each trial. Before presenting items, a cue indicated the probing probability of a circle (50%, 75%, or 100%). The primary outcomes were pupil size during maintenance and response accuracy. As hypothesized, colors of more important items were reported with higher precision, which favors the flexible resource allocation model. Mixture model analysis indicated that differences in guessing rates contributed to the variations, and more important items benefited from more encoding resources. PPD during memory maintenance revealed nuanced patterns in line with the hypotheses, albeit with non-significant group-level differences across conditions. Interestingly, individual-level analyses exposed mixed results, possibly reflecting differences in PPD sensitivity or task strategies across participants. In sum, this study supports a flexible resource allocation model of VWM and underscores the complexity of measuring resource allocation using PPD.
Visual working memory (VWM; all abbreviations available in Appendix B) is a system that actively and temporarily stores visual information for access and processing by other cognitive systems, thus guiding behavior. In addition to the storage of visual items and their features, VWM also includes attentional control of what items to store and strategies to maintain the items in storage (Baddeley & Hitch, 1994; Unsworth & Robinson, 2018).

VWM is limited in capacity, and there is an ongoing debate amongst researchers on how such limitations manifest. The discrete capacity model (Figure 1d) proposing VWM storage of items in slots as a whole and not individual features has dominated the scene (Luck and Vogel, 1997). In this model, one expects an all-or-nothing remembrance, assuming items to be perfectly encoded when within the individual’s estimated VWM capacity – denoted as K. These models were constructed mainly on data from experiments using change-detection tasks – where subjects compare a sample array with a test array and report if they differ (Figure 1a). The hypothesis has been that if individuals encode an item, one expects them to have near-perfect knowledge of any changes between the sample and test array; Detected differences beyond an individual’s VWM capacity are explained as successful guesses (Zhang & Luck, 2008).

The discrete capacity model describes VWM capacity limitations as a function of set size (memory load) – the number of items to memorize. Cowan (2001) suggested a capacity limitation to working memory of around four (ranging from two to six in individuals) chunks – groups of strongly associated concepts. Although not all agree on a “hard” four-chunk limit, there has been a consensus that working memory includes some memory limits, e.g., time and scheduling conflicts, or that memory limits are not general but task and material-specific.

In contrast to the change-detection tasks, Palmer (1990) introduced a partial-discrimination task (Figure 1b) where subjects had to report if a line in a specific position of a memory array was shorter or longer than the line previously displayed in the same position. By altering the length differences, Palmer could measure the mean differential length that subjects could detect as a function of memory load. He concluded that the mean length difference doubled when the memory load increased from one to four items. Replicating the results using changes in orientation and shape, Palmer suggested that the effects could generalize to any changes in continuous visual characteristics. With his study, Palmer challenged the discrete capacity model by showing that a model sharing attention across items could best explain his results.

Flexible resource allocation models (Figure 1d) have developed since Palmer presented his results. Wilken and Ma (2004) suggested that a signal detection model – where noise affects feature encoding – explains data from change-detection tasks with set size and number of stimuli as independent variables better than discrete capacity models. They also introduced a continuous-recall task (Figure 1c), like Palmer’s partial-discrimination task, where participants report the memorized object features on a continuous scale – choosing a color on a color wheel or the angle of a rotated bar. Their results from the continuous-recall task confirmed that noise in encoded items increased with memory load, as the signal detection theory predicted. These results have been replicated in partial-discrimination tasks (Bays & Husain, 2008), showing that the response precision, plotted against memory load, follows a power law function. Signal detection models assuming a continuous memory allocation are problematic since they cannot account for guessing (Zhang & Luck, 2008). To address this, Van den Berg et al. (2012) suggested a modified model in which there is some random variability in the assignment of resources across items. Due to this variability, items sometimes end up being encoded with very low precision, which may look like pure guessing at a behavior level. This flexible-resource model tackled the guessing problem and outperformed previous models on data from partial-discrimination and continuous-recall tasks without needing to postulate an explicit guessing component. However, in a follow-up study (Van den Berg et al., 2014), the same authors found
that a hybrid model – that combines variable, flexible resources with a guessing component – accounted better for the data than a ’pure’ slot model or a ’pure’ resource model.

Traditionally, studies using the continuous-recall paradigm (Figure 1) were designed such that each item was equally likely to be probed in the response stage. However, a key prediction of the flexible-resource model is that if one item is more likely to be probed than the others, then this item should be encoded with higher precision – by allocating more resources to it. Moreover, the magnitude of this difference should increase as the difference in probing probability is made more prominent. This prediction of gradual changes in precision is unique to flexible-resource models because discrete-capacity models predict encoding in an all-or-none fashion.

Lately, studies have used continuous-recall tasks to experimentally verify this prediction (Dube et al., 2017; Emrich et al., 2017). Experiments using continuous-recall tasks allow for estimating the precision with which participants remember a feature and the nature of errors, e.g., guess rate or misattribution within the memory array. When varying and informing participants of the probing likelihood of an item – cue validity (CV) –, experiments show that selective attention based on probing probability better predicts the reporting precision than the memory load, and higher probing probability was associated with fewer guesses (Dube et al., 2017; Emrich et al., 2017). These results suggest that the distribution of limited VWM capacity better fits a flexible resource allocation model than a discrete capacity model. In a more recent study, Salahub et al. (2019) showed that neural activity corresponding to the flexible allocation of VWM resources could be measured using the contralateral delay activity (CDA) – a delayed event-related potential in the contralateral hemisphere in the parietal and occipital regions. They found that the best model fit of CDA amplitude included the weighted sum of proportional resources and memory load. They also measured another event-related potential – N2pc – and found an increased amplitude for high-probability compared to low-probability items. This difference suggests that participants enhance attention towards items more likely to be probed.

In sum, new models based on signal detection theories and experiments using continuous-recall tasks with pre-cued probing probabilities suggest VWM as a flexible resource of temporary storage governed by attention.

**Figure 1**  
**VWM Experimental tasks and capacity models**

a) Change-detection  
Sample array  
Delay  
Did the array change?  

b) Partial-discrimination  
Sample array  
Delay  
What was the color of the probed item?  

(c) Continuous-recall  
Sample array  
Delay  
Is the line longer or shorter?  

**d) Allocation models**

Relative item importance (%)  
10%  
40%  
Discrete Capacity  
Flexible Allocation  

Note. Experimental tasks to measure VWM capacity: a) change detection task where change detection ratio is measured and the number of displayed items manipulated, b) partial-discrimination task where change-detection ratio is measured and a continuous independent variable, e.g., length difference, manipulated, and c) continuous-recall task where precision can be measured as a continuous variable. Fig d) illustrates how items with different importance allocate VWM resources according to the two theoretical models: discrete capacity and flexible allocation.
Attention in VWM

Cowan (2001) suggested that limitations in attention is a prime candidate for explaining working memory capacity limitations and that the possibility to use attentional strategies partly explains differences in working memory capacities between experimental settings. Attention affects VWM both in early – sensory selection for encoding – and late – maintenance rehearsal – stages of memory processing (Awh et al., 2006). In the later stage, attention focuses on internal representations of the memorized items. Awh and Jonides (2001) suggest that covert shifts of spatial attention – spatial rehearsal – may assist memory maintenance. Retro-cues – cues during memory maintenance – have been used experimentally to study the effect of late-stage attention on VWM (Souza & Oberauer, 2016). Such cues draw attention to already encoded items, improving VWM performance for those items by strengthening and protecting them from subsequent interference.

Palmer (1990) investigated the attentional effects on VWM by comparing memory performance for different set sizes with or without distractors. That is if the set-size effect on memory performance is contingent upon the number of attended items or displayed items. The results indicate that the attended items determine the effects of memory load on memory performance and that the number of displayed but not attended items has little effect. Individual differences in working memory capacity may reflect the ability to attend to information and maintain it in working memory when facing distraction and interference (Engle, 2002). Individuals with low working memory capacity show greater difficulties handling proactive interference – similar conflicting patterns in long-term memory – than high-capacity individuals. Using change-detection tasks, Fukuda et al. (2015) found that performance differences between high and low VWM capacity individuals diverged more at higher memory load than what could be explained by pure storing capacity differences. They found these differences to reflect abilities to regulate attentional control of storage capacities.

When investigating the relationships between arousal regulation, task-unrelated thoughts, working memory capacity, and attentional control, Robison and Brewer (2022) found that individuals with a tendency for task-unrelated thoughts had poorer attentional control and working memory capacity. Arousal dysregulation – operationalized as trial-to-trial pupil size variability – was also associated with more task-unrelated thoughts and lower attentional control.

In sum, attention affects VWM in the early and late stages. Individual differences in VWM capacity may reflect differences in storage capacities and attentional control in the face of distractions and task-unrelated thoughts.

Pupil dilation as a measure of attention in VWM

In 1966, Kahneman and Beatty found that pupil size increased with the number of digits subjects were to remember and decreased when reporting the digits. Their results indicated a relationship between pupil size and working memory load. Kahneman (1973) later suggested that task-evoked phasic pupil dilation (PPD) – changes in pupil dilation – is a psychophysiological marker of the mental effort – e.g., attention – required to solve a task.

During memory maintenance, PPD can track memory load and predict individual VWM capacity in change-detection tasks (Unsworth & Robison, 2015, 2018). Pre-trial pupil size has predicted memory in partial discrimination tasks but not PPD during memory maintenance (Robison & Unsworth, 2018). During encoding, larger PPD has differentiated high-value words from low-value words, impacting memory performance on these words (Ariel & Castel, 2014).

PPD may be sensitive to cognitive but not perceptual load (Chen et al., 2022). Unsworth and Robison (2018) found that PPD only tracked actively attended items but were unaffected by displayed distractors for set sizes above four items. They also found that when asking participants to drop already encoded items from memory, PPD decreased. However, for set
sizes below four items, distractors did induce PPD responses, although smaller than for attended items. Their findings suggest that when the number of displayed items are below an individual’s VWM capacity, distractors may be partly encoded using fewer resources.

Unsworth and Robison (2015, 2018) also found differences in PPD patterns during memory maintenance between subjects with low vs. high VWM capacity. PPD for low-capacity individuals peaks at a lower set size and then decreases compared to high-capacity individuals; for high-capacity individuals, PPD remains flat until approximately a set size of four. These differences indicate that PPD is higher when close to one’s VWM capacity, suggesting sensitivity in an optimal performance range.

A neural candidate mediating PPD and effortful attention is the locus coeruleus (LC) – a nucleus in the dorsal PONS responsible for all norepinephrine released in the cortex (Alnaes et al., 2014). According to the Adaptive Gain Theory (AGT), LC neurons exhibit phasic activation related to exploitation – engagement in the current task – and tonic activation related to disengagement in the current task in favor of exploration – seeking reward in new tasks (Aston-Jones & Cohen, 2005). Studies have shown that LC phasic activation is related to task-evoked PPD, while tonic activation is related to baseline pupil size (Glizenrat et al., 2010; Murphy et al., 2011). AGT predicts that phasic and tonic activation relate to each other as the u-shaped curve of the Yerke-Dodson law (Yerke & Dodson, 1908); Task performance and phasic activation are high when tonic activation is intermediate (Aston-Jones & Cohen, 2005). Murphy et al. (2011) verified the u-shaped relationship in an oddball task, showing optimized performance for intermediate levels of tonic activity measured as pretrial pupil size. In contrast to the ATG predictions, PPD was low when task performance was high. Further analyzing consecutive trials, Murphy and colleagues noted that performance improvements followed high PPD trials in coming trials. This indicates that PPD may signal an attentional challenge leading to re-engagement in the task.

In addition to the LC-pupil size association, Alnaes et al. (2014) found that changes in pupillary size also correlated with superior colliculus activity during a multiple object tracking task. The authors suggested that the superior colliculus association with pupil size may indicate covert shifts of spatial attention – supporting the attention-based rehearsal hypothesis by Awh and Jonides (2001).

In sum, evidence from behavioral and neurological studies suggests that PPD can be used as a physiological marker to track the attended items in VWM during maintenance. As such, it may act as a proxy for VWM load like CDA.

Aim and research questions

The general purpose of this study was to unravel further how VWM allocates resources – attention, encoding, and storage – to visual items that may differ in importance for later recall and processing and how this process can be tracked by pupil size. To that end, this study had four main aims.

Firstly, I aimed to replicate the earlier reported effects of item importance on encoding precision (Dube et al., 2017; Emrich et al., 2017, Salahub et al., 2019). This will be done by manipulating cue validity (CV), which indicate to participants the probability that certain items will be probed in the response stage. By such manipulation, I could compare behavioral outcomes of memory precision in a continuous recall task based on probing likelihood relative to predictions of flexible resource allocation and discrete-capacity models.

Secondly, I aimed to investigate if PPD during maintenance of VWM, acting as a proxy for attention, would measure differences in CV according to a flexible resource allocation model. Specifically, I looked at how CV (50%, 75%, and 100%) predict PPD during memory maintenance. Although PPD, in contrast to CDA, cannot measure attention to specific items, it tracks overall memory load (Unsworth & Robison, 2015, 2018). Comparing PPD between the
three CV conditions would indicate if PPD-tracked attentional resources are distributed proportional to the probing probability and determined by set size. If PPD for the 50% and 75% CV conditions are similar but differ from the 100% CV condition, it would indicate that PPD reflects the set size because the 50% and 75% CV conditions require attention to all items, and the 100% condition attention to the cued items. If PPD for all the conditions is the same, it would indicate that either PPD does not track VWM resources or that PPD reflects a proportional distribution of resources to items in relation to the probing probability. The latter would be because the sum of probing probability for all items is always 100%. If PPD for all items differs, it would indicate that PPD reflects a distribution of resources non-proportional to the probing probability, e.g., logarithmically distributed.

Thirdly, I aimed to investigate if the spatial distribution of important items would affect PPD during memory maintenance. A premise of the attention-based rehearsal hypothesis (Awh & Jonides, 2001) is that rehearsal through spatial attention during memory maintenance improves memory recall. PPD during memory maintenance could indicate such spatial rehearsal. Thus, I included two conditions where cued items are grouped or separated.

Fourthly, I aimed to investigate if trial-to-trial PPD during memory maintenance predicts memory performance in a continuous recall task. In other words, greater attention during memory maintenance reflects precision, for example, reflecting less encoding distortion.

For replication purposes, I based the study on a previous work (experiment 2 in Salahub et al., 2019), which showed that item importance affects memory precision using a continuous recall task. However, that experiment had two shortcomings that I wanted to address. Firstly, the cue in Salahub et al. (2019) always referred to the same two of the four displayed items, which may have induced an attentional bias, e.g., through priming effects. Secondly, they used only 50% and 100% CV, which caused a confound between set sizes: on trials with the 50% cue, all four items had to be remembered, while on trials with a 100% cue, it was safe for the participant to drop the two non-cued items, because there was a 0% chance that they would be probed.

To address the first shortcoming, I conducted two experiments for comparison. In the first experiment, I followed the design used by Salahub et al. (2019), while in the second experiment, the two items that the cue referred to (circles or squares) changed between blocks. To address the second shortcoming, I introduced a 75% CV condition with a four-item set size but a different cue validity than the 50% CV condition.

Hypotheses

1. Regarding memory accuracy, I hypothesized that memory precision would increase as a function of probing probability as in earlier studies (Dube et al., 2017; Emrich et al., 2017). Possibly following a power law function, as Salahub et al. (2019) showed and predicted by Bays and Husain (2008).
   a. Based on previous work (Dube et al., 2017; Emrich et al., 2017), I also expected increased probing probability to reduce the guessing rates, i.e., items with higher probing probability will receive more encoding resources and thus be less likely to result in guesses.
   b. For items successfully encoded – non-guesses – I expected the precision to be more accurate for higher probing probabilities in line with Dube and colleagues’ (2017) findings.

2. Based on Unsworth and Robison's (2018) findings that attention operationalized as PPD during memory maintenance increases with higher memory load, I hypothesized that:
a. PPD would be highest for trials with 50% CV – corresponding to equal probing probability for all attended items.

b. PPD would be lowest for trials with 100% CV – assuming two attended items and two distractor items inducing lower PPD.

c. PPD for trials with 75% CV would be lower than PPD for trials with 50% CV and higher than PPD for trials with 100% CV. Here, I predicted that the difference in PPD would reflect the continuous resource model of VWM. Considering the earlier findings that guesses and precision follow a power-law function (Dube et al., 2017; Emrich et al., 2017; Salahub et al., 2019), I hypothesized that if PPD-measured resources follow the same function, the sum of resources for the 75% CV condition would be in-between the sum of resources for the 50% CV and 100% CV conditions.

3. I hypothesized that trials with greater spatial separation between cued items would induce a more significant PPD response. I base this hypothesis on Awh and Jonides’ (2001) suggestion that VWM maintenance includes attention-based spatial rehearsal, which would induce a larger PPD if encoded high-value items are far apart. Another more speculative reason would be that grouped items use fewer VWM resources because they are easier to chunk (Cowan, 2001).

4. Regarding the association between PPD and memory precision, earlier findings are inconsistent. Murphy and colleagues (2011) found a negative association between PPD and task performance, while Unsworth and Robisons (2018) found no associations between PPD and memory precision. Thus, I did not have any prior hypothesis of such associations and the investigation is exploratory.

Method

Participants

In total, 37 participants (Age = 18 - 45, M_age = 29.3, SD_age = 6.8, Male = 22) performed a visual memory task in two experiments. The first experiment included 10 participants (M_age = 35.2, SD_age = 6.3, Male = 7) and the second experiment included 27 participants (M_age = 27.0, SD_age = 5.5, Male = 15).

I recruited participants through Accindi – a Swedish website for participant recruitment to research projects (https://www.accindi.se/) – and word of mouth. All participants had normal or corrected-to-normal eyesight and were free from strabismus vision disorder. As a pre-screening, participants conducted an online Ishihara self-test to confirm they were not color blind (https://www.colorblindnesstest.org/ishihara-test/). Informed written consent was obtained from all participants before participation. Participants conducted the experiments in two sessions on one occasion, which took approximately 3 hours including a break. They received a 400 SEK gift certificate (https://gogift.io/) as compensation for their time.

Materials and Procedure

Experimental sessions in both experiment 1 and experiment 2 were conducted in a lit room (approximately 50 lux with the screen turned off) at Campus Albano, Stockholm University. All tasks were presented on a 60,5 cm (diagonally) IPS LED backlight screen (EIZO FlexScan EV2451) at 1920x1080 pixels (60Hz) using a Windows PC. The experiment was controlled and rendered using a modified Python script inherited from Salahub et al. (2019; experiment 2). Pupillometry data was recorded using a Tobii Pro Spectrum eye tracker at 1200 Hz. Participants were seated with their heads semi-fixated using a chin rest 59 cm from the screen.
Stimuli and Procedure

I performed two experiments. In both, participants viewed an array of four colored shapes – “circles” and “squares” – and after a brief memory period, they were asked to report the color of one of the earlier presented shapes.

Experiment 1. The visual memory task consisted of text-based instructions and feature-based visual stimuli repeatedly presented in a trial cycle. All stimuli were generated on a screen with a black background (RGB: 0, 0, 0). The trial cycle included four phases (Figure 2) and an intertrial fixation period where a central fixation dot (radius 0.1°) was present. During fixation, eye positions were checked to ensure participants were looking at the fixation dot. Successful fixation required two-thirds of position checks within a predefined fixation window (0.5° x 0.5°) during a 300 ms window. After initial fixation validation, fixation was continuously validated until the start of the response phase. If fixation failed after the fixation period, the trial was terminated and scheduled to be replayed at the end of a 48-trial block. The maximum number of replayed trials was set to 20% of the total block to avoid the experiment getting too long. Once this 'budget' was fully used, breaking fixation would mark the trial as 'invalid' in the data file rather than be terminated. After successful fixation, a cueing phase started. During this phase, a text-based cue paired the word "square" or "circle" with a validity cue (50%, 75%, or 100%), indicating the probability of probing the cued shape in the response phase.

After the cueing phase, trials entered an encoding phase presenting stimuli for memorization – two color-filled visual stimuli on the horizontal axis and two on the vertical axis. Stimuli always consisted of two circles (1.15° radius, area = 1.33) and two squares (0.65° side, area = 1.33) balanced between trials for presentation on vertical and horizontal positions. Circles and squares alternated between being grouped or split on the vertical and horizontal axis in a balanced and randomized order (see Figure 2 for examples of a “grouped” and a “split” stimulus). Stimuli colors were picked pseudo-randomly from a 360-degree CIE color wheel (radius = 49, x = -6, y = 14) with at least 30 30-degree difference between them. After the encoding phase, the trial entered a delayed memory maintenance phase (delay period) with a blank screen and a fixation dot. Finally, in the response phase, "squares" and "circles" were presented in the same positions as in the encoding phase but without colors. One stimulus was probed and identified by a thicker line. The selection of probed stimuli was according to the cues and the respective validity, e.g., the cue "75% square" gave a 75% probing probability for one of the squares and a 25% probing probability for one of the circles. During the response phase, participants tried to match the color of the probed stimuli with the memorized color of the corresponding stimuli in the encoding phase using a color wheel positioned outside the stimuli (12° radius). The color wheel in the response screen was rotated at a random angle on each trial.

The timing of the intertrial fixation period varied due to the waiting time for fixation; the minimum intertrial interval was 550 ms. Unsuccessful validation within 2 seconds led to a reminder of participants to look at the center. The cueing phase was 750 ms, while the encoding phase and the delay period were 1500 ms long. The time of the response phase depended on the participants' response time.

The stimuli shapes presented during the encoding phase could be presented as upper and lower at four different positions – horizontally at 3 and 9 o'clock and vertically at 6 and 12 o'clock. Moreover, the stimuli were presented pairwise allowing for the same shapes to form a group, i.e., in the same vertical or horizontal positions, or split, i.e., in different vertical or horizontal positions. The shapes were equally presented in all eight pairwise combinations and matched those with eight split positions. Such presentation balanced the possible effect of differential spatial attention (group vs. split), giving 16 positional combinations in a trial mini-block. Furthermore, the three validity cues (50%, 75%, and 100%) represented three conditions forming a 48-trial block to represent all positions for all cues.
For the 50% and 75% validity cues, either the target shape – the cued shape – or the non-target shape could be probed in the response phase. That gave a 50% and 25% probability for probing the non-target shapes. Target and non-target probing occurrences were pseudorandomized to ensure balance over the positions and avoid forming probing priors over time that may affect the encoding strategy. Since cued shapes were always in pairs, the selection of the actual item finally probed was randomized.

To compare memory precision for the different shape probing frequencies (25%, 50%, 75%, and 100%) and get sufficient power to detect differences individually, each participant did 20 trial blocks. Hence, the number of trials per participant was 960, distributed on cue validity (CV) and shape probing frequency (Table 1). Participants conducted the experiment in two sessions of 480 trials, separated by a more extended break. Shorter breaks were also introduced between blocks to reduce fatigue exposure.

The cued shape was held constant within participant sessions and balanced between sessions. Thus, the only change between trials was CV, which was randomized within blocks following the original design of Salahub et al. (2019). Besides replicating earlier experiments, holding the cue shape constant made it possible to analyze the priming effects of memory strategies.

**Experiment 2.** The methods of Experiment 2 were identical to those of Experiment 1, except for the following differences. The cued shape and CV were held constant within and balanced between mini-blocks of 16 trials. The experimental design ensured that shape and CV alternated between mini-blocks by pseudorandomizing the mini-block order. The 50% CV was replaced with an “all equal” cue to reduce priming effects further.

**Table 1**

*Number of trials per cue and probing condition in the two experiments*

<table>
<thead>
<tr>
<th>Cue validity</th>
<th>Probe</th>
<th>Shape probing frequency</th>
<th>Number of trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>Target</td>
<td>50%</td>
<td>160</td>
</tr>
<tr>
<td>50%</td>
<td>Non-target</td>
<td>50%</td>
<td>160</td>
</tr>
<tr>
<td>75%</td>
<td>Target</td>
<td>75%</td>
<td>240</td>
</tr>
<tr>
<td>75%</td>
<td>Non-target</td>
<td>25%</td>
<td>80</td>
</tr>
<tr>
<td>100%</td>
<td>Target</td>
<td>100%</td>
<td>320</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td><strong>960</strong></td>
</tr>
</tbody>
</table>

**Figure 2**

*Trial cycle sequence*
Measurement and Analysis

I used Rstudio version 2021.9.0.351 (Rstudio Team, 2021) for statistical precision and pupil data analysis. Preprocessing – filtering, interpolation, and smoothing – of eye data was done in Python using PyCharm 2022.3 CE. Statistical analysis is primarily based on frequency statistics adding Bayes factors to evaluate the credence of the null hypothesis when p-values are high.

Independent variables

The experimental design included two similar independent variables – CV and shape probing frequency (SPF) – adapted for the two measured outcomes – memory precision and pupil dilation. SPF is the probing probability for the probed shape given the cue (Table 1). For example, given a 75% CV, there was a 75% SPF of the target shape – cued shape – and a 25% SPF of the non-target shape. Given a 50% CV, both the target and non-target shapes had a 50% SPF, and for the 100% CV, target shapes had a 100% SPF. In the reported experiments, the cue indicated the probing probability of a shape, thereby signaling attention allocation towards two items to the participant. Thus, the theoretical attentional resource one would expect participants to allocate to a particular item is SPF/2.

Since the probe was first known to the participants in the response phase, the independent variable used to predict change in pupil size during the trial cycle was the CV.

A third independent variable – Split – was used to investigate the effects of spatial attention on pupil size and memory precision. The Split variable distinguishes between trials with the same or different shapes in the vertical and horizontal positions.

Dependent variables

Estimation Error

The graded report on the color wheel was an indirect measurement of memory precision. I calculated a raw error score ($\epsilon_{\text{raw}}$) as the difference in degrees on the color wheel (Figure 2) between the response color ($C_{\text{response}}$) in the response phase with the original color ($C_{\text{origin}}$) in the encoding phase. Since the response colors followed a circular distribution, the maximum difference was 180º, and the raw error score was adjusted to be less than 180º. Thus, $\epsilon_{\text{raw}}$ corresponded to the error distribution in degrees ranging from -180º to +180º centered around $C_{\text{origin}}$ (Formula 1).

Further calculating the standard deviation of the $\epsilon_{\text{raw}}$ over the trials within a condition gave a precision index for the participant on that condition. Like Salahub et al. (2019), I used this precision index ($\epsilon_{\text{raw}}$ SD) as the primary dependent variable to compare memory precision data (note that higher memory accuracy result in a lower precision index). For statistical reasons, I also used the absolute value of $\epsilon_{\text{raw}}$ as an approximation for memory precision when comparing the differences in SPF within individuals.

$$\epsilon_{\text{raw}} = (C_{\text{response}} - C_{\text{origin}} + 180) \mod 360 - 180$$

(1)

Since the number of included trials varied across participants, I calculated a weighted mean of the standard deviation of the $\epsilon_{\text{raw}}$ ($\epsilon_{\text{raw}}$ SD) rather than an ordinary mean. The weights were proportional to the number of trials that a participant performed. The group level weighted mean of $\epsilon_{\text{raw}}$ SD was fitted with SPF as a predictor using both linear (Model 1) and power law (Model 2) models. In addition, a mixture model (Model 3; Zhang & Luck, 2008) combining a uniform distribution and a von Mises distribution – equivalent to the normal distribution for circular data – was used to fit a probability density function with the $\epsilon_{\text{raw}}$ distribution for each SPF. The assumption behind Model 3 was that the von Mises distribution would model the
memory precision distribution of memorized items, and the uniform distribution would model guesses for non-memorized items.

Model 1: \[ y = \beta_0 + \beta_1 x_i + \epsilon \]
Model 2: \[ y = \beta x - k y = \beta x - k \]
Model 3: \[ y = p_{\text{guess}} U(-180^\circ, 180^\circ) + (1 - p_{\text{guess}}) VM(\mu, \kappa), \] where \( \mu \) and \( \kappa \) are the mean and concentration parameter, respectively, of the von Mises distribution. \( p_{\text{guess}} \) is the probability that the participant is guessing due to the probed item not being in memory. The model assumes that when a participant guesses, each response value is equally likely (uniform distribution on \(-180\) to \(+180\) degrees).

The weighted mean of \( E_{\text{raw}} \) SD was separately calculated with 95% compatibility intervals for experiment 1 and experiment 2 and plotted together to compare the effect on estimated error for target and non-target shapes to explore cue priming effects.

**Response Delay**
Response delay was the time from the presentation of the probe during the response phase until participants selected a color on the color wheel with a mouse click. SPF as a predictor of response delay on the group level was analyzed by regressing a linear model on the data.

**Phasic pupil dilation**
The Tobii Pro Spectrum eye tracker measures pupil dilation during the trial. Before statistical analysis in R, I preprocessed the eye tracker data in a Python script through five steps. First, I averaged pupil position and size from the left and right pupils for each time point. Second, invalid data points – blinks and other non-observed data — were filtered out. Third, a linear interpolation filled in within-trial missing data points based on the trial time series. Fourth, a Savitzky-Golay filter smoothened the within-trial data. Fifth, data was resampled from 1200 Hz to 100 Hz using a mean over the sample period method.

I operationalized phasic pupil dilation (PPD) as the relative change in dilation from the trial baseline – the mean pupil size during the last 50 ms of the intertrial interval (fixation period). Thus, I calculated PPD as the pupil dilation at the measured time point (denoted \( P_t \)) divided by the mean pupil dilation during the baseline (denoted \( P_{\text{baseline}} \)) minus one. Multiplying by 100 gave the relative percental increase or decrease in pupil size (Formula 2).

\[
\Delta P = \left( \frac{P_t}{P_{\text{baseline}}} - 1 \right) \cdot 100\%
\]  

(2)

The PPD was analyzed using time series plots over the trial cycle, calculating the mean and 95% compatibility intervals per CV over trials per individual. Individual means per time point were then pooled into a group mean to compare the effect of CV on PPD. Furthermore, the PPD was compared between CVs using Welch’s t-tests for each time point to identify during what periods of the trial cycle the differences in PPD were determined to be significantly different (\( \alpha = .05 \)).

Finally, to investigate the effect of PPD during the delay period on memory precision, \( E_{\text{raw}} \) was converted to standardized z-scores and correlated against PPD per SPF on both individual and group levels.
Eye position

The right and left eye position was measured during data collection but not further analyzed in this thesis.

Ethics

This study was part of a larger project approved by the Ethical Review Authority (Dnr 2022-05444-01). I collected informed written consent from all participants before participation and experimented without registration of any personal data. The experiment was conducted in a way expected to be psychologically and physiologically safe.

Results

I conducted two visual working memory experiments to investigate how memory precision is affected by CV. Three participants – one in the first and two in the second – could not complete the experiment in time. Their results were excluded from further analysis, as were the results of two participants with less than 10% valid trials (Table A1). Due to a programming error in the second experiment, the memory delay phase for non-fixation trials – non-terminated trials without proper eye fixation – was shortened. Such trials were eliminated and not used in the analysis (Table A1). Thus, the final analysis was performed on 9 participants (M age = 35.1, SD age = 6.6, Male = 6) for experiment 1 and 23 participants (M age = 26.6, SD age = 5.5, Male = 12) for experiment 2.

Memory accuracy

To test my hypothesis that memory precision increases as a function of probing probability, I first used a two-way ANOVA on $E_{raw}$ with SPF and Split as predictor variables to investigate the potential main and interaction effects of SPF and position on memory accuracy. The ANOVA showed a significant main effect of SPF on $E_{raw}$ ($F(3, 504) = 98.7, p < .001$). In contrast, the main effect of Split was not significant ($F(1, 504) = 0.61, p = .43$). Neither was the interaction effect of SPF and Split significant ($F(3, 504) = 0.35, p = .79$). Furthermore, I calculated the Bayes factor for the effect of Split on memory accuracy for each SPF to test the relative strength of the null hypothesis – a null effect of Split on memory accuracy. Bayes factors of 0.23 for 25% SPF and 0.19 for 50%, 75%, and 100% SPF indicate moderate to strong (Wagenmakers et al., 2018) evidence favoring the null hypothesis. In further analysis of memory accuracy, I do not include Split as a predictor since the ANOVA and Bayes factors did not indicate any effect of spatial attention on $E_{raw}$.

Memory accuracy and probed items

Next, to test for a possible attentional bias in a previous study (Salahub et al., 2019), I compared the relationship of memory precision and probed items between experiments 1 and 2. The cued shape was held constant in the first experiment, and only the CV changed between trials. This design allowed for comparing memory precision on target items – the cued shape – and non-target items at equal probing probability. When a CV is 50%, the probability of probing target and non-target items is the same. Thus, they should be given the same attentional and memory resources for optimal task performance. When plotting the weighted mean of individual $E_{raw}$ SD for different SPF and probed item with 95% compatibility intervals (CI; figure 3), a difference between experiment 1 and 2 emerge. For the 50% SPF in experiment 1, there is a significant difference in precision between probed non-target and target items $t(8) = 3.76 (p < .01)$. In comparison, for the 50% SPF in experiment 2, there is no significant difference in precision between probed non-target and target items $t(22) = -.71 (p = .48)$. These results
suggest that in experiment 1 participants are biased by consistently cueing the same shape in all trials.

**Figure 3**
*Precision by probed item for experiment 1 and 2*

![Graph showing precision by probed item for experiment 1 and 2](image)

*Note.* Weighted mean and 95% CI for mean precision by probed item. Precision for probing non-target items is indicated in blue, and precision for target items in yellow.

**Memory accuracy and shape probing frequency**

Considering the possible priming bias of experiment 1, memory accuracy data was further analyzed using experiment 2 data only. To test my hypothesis that memory precision increases as a function of probing probability, I calculated a one-way ANOVA with SPF as a predictor of $E_{\text{raw}}$ SD on the experiment 2 data. The ANOVA showed a significant main effect of SPF on $E_{\text{raw}}$ SD ($F(3, 88) = 22.42, p < .001$) on a group level. I then plotted memory accuracy – precision – as a function of SPF with 95 % CI. Figure 4 shows a non-overlapping difference between each SPF group (note that higher memory precision results in a lower $E_{\text{raw}}$ SD). When fitting a linear and a power-law model, neither model perfectly fits the data, somewhat contrasting with earlier findings (Salahub et al., 2019). Significant main effects of SPF on memory accuracy were also detected for all individuals (Table A2), although the differential pattern varies (Figure A1).

**Figure 4**
*Precision per SPF*

![Graph showing memory accuracy per SPF](image)

*Note.* Group level weighted mean of $E_{\text{raw}}$ SD with 95 % CI and fit of a linear and power-law model for comparison. Lower $E_{\text{raw}}$ SD indicates more accurate responses.
Memory accuracy and guessing rates

To further analyze the differences in memory accuracy for different SPFs, the density functions of $E_{\text{raw}}$ were plotted (Figure 5). I then used mixture models to test my hypotheses that increasing probing probabilities will reduce guessing rates (H1a) and that precision for successfully encoded items will be more accurate for higher probing probabilities (H1b). Mixture models combining a uniform distribution with a Von Mises distribution show the relative influence of each distribution. As indicated in Figure 5, the uniform distribution's relative influence ($P_{\text{uniform}}$) is higher for the lower SPF, assuming to reflect a higher degree of guesses for lower SPF. The guessing rate ($P_{\text{uniform}}$) for each SPF is plotted in Figure 6a and, as hypothesized, shows a decrease in guesses for higher SPFs. The kappa ($\kappa$) value in a Von Mises distribution measures the concentration – where $1/\kappa$ is analogous to the variance in the normal distribution. This variance – assumed to be a measure of the precision of the encoded items – is plotted for each SPF in Figure 6b. The decrease in variance for encoded items with higher SPF indicates, as hypothesized, that improvement in memory accuracy for high-value items is not only related to fewer guesses.

**Figure 5**

Density function of $E_{\text{raw}}$ per SPF

**Note.** Density distributions (radians from $-\pi$ to $\pi$) of $E_{\text{raw}}$ for different shape probing frequencies with fitted mixture distribution models (red). Mixture models include a uniform distribution and a Von Mises distribution according to model 3 described above.
Figure 6
**Guessing rate and precision for encoded items per SPF**

*Note.* Results from the mixture distribution model per SPF with a) the proportion of the fitted mixture distribution model related to the uniform distribution for different SPF and b) precision calculated as $1/\kappa$ (variance in the modelled von mises distribution) for different SPF.

**Memory accuracy and response time**

To control for the effects on response time, I plotted the weighted mean response time for each SPF (Figure 7a). Comparing the differences in response times between the SPF groups in a one-way ANOVA did not render significant results ($F(3, 88) = 0.19, p = .90$). While these results indicate that SPF may not affect the response time, there is a weak but significant negative correlation between response time and memory accuracy (Figure 7b; $r = -.21; p < .05$). These results indicate that the memory accuracy differences reported are not a speed-accuracy tradeoff.

Figure 7
**Response time as a function of SPF and memory accuracy**

*Note.* Figure 7 a) show the weighted mean response time is plotted for each SPF. The bars represent 95 % CI and the dotted green line a fitted linear model. Figure 7 b) show mean response time as a function of mean precision ($E_{raw}$ SD) per SPF and participant. The response time outliers – over 4 seconds – belong to the same participant, suggesting a longer average response time in all SPF conditions for that participant. Thus, these outliers were not removed.
Pupil Dilation

My second research question concerned how CV, reflecting the relative importance of different items, affects attention during memory maintenance, as measured by PPD. Before analyzing this relationship, I investigated my third hypothesis, whether spatial attention influenced PPD. According to the attention-based rehearsal hypothesis, covert spatial attention shifts occur during memory maintenance (Awh & Jonides, 2001). When high-value items are separated in space, that might increase pupil size.

To analyze spatial attention’s influence on pupil dilation, I plotted the mean PPD over the trial cycle for different trial types – Split – and conditions – CV (Figure 8). In contrast to my hypothesis (H3), these plots indicate that spatial attention had no or little effect on trials when CV was 50% or 75% and a small but non-significant effect on trials when CV was 100%. Furthermore, a two-way ANOVA comparing the difference between the mean PPD during the delay period for trial type X condition resulted in non-significant main effects for condition (F(2, 186) = 1.62, p = .20) and trial type (F(1, 186) = 0.043, p = .84). Neither did the ANOVA result in condition X trial type interaction effects (F(2, 186) = 0.174, p = .84). Calculating the effect of the trial type on average PPD during the delay period as Bayes factors indicate moderate evidence for the null hypothesis – no effect of split vs. grouped – for 50% (Bayes factor = .262), 75% (Bayes factor = .256), and 100% (Bayes factor = .303) CV. Since the ANOVA and Bayes factors did not indicate any effect of spatial attention on average PPD in the delay period, I pooled together data from split and grouped trials in further analysis of pupil dilation.

Figure 8
Comparison of mean PPD over the trial cycle for split vs grouped trials

a) Fix Cue Stimuli Delay Response b) Fix Cue Stimuli Delay Response

Note. The change in mean pupil size compared with the baseline over the trial cycle plotted for grouped (blue) and split (red) trials when the CV is a) 50%, b) 75%, and c) 100%. The shaded areas indicate the 95% CI.

Experiment 1 vs Experiment 2
As reported above, the differences in memory accuracy for probed targets and nontargets in the 50% CV condition suggest that the consistent reference to the same shape in all cues may bias participants in Experiment 1. To check if this bias is reflected in the pupil data, mean PPD over the trial cycle in the 50% CV condition was plotted for Experiments 1 and 2 (Figure 9). As shown in Figure 9, the PPD curves and their 95% CI for both experiments overlap in the delay period.

I conducted a Welch’s t-test to compare the Experiment 1 mean PPD during the maintenance phase (m = 1.47%, sd = 3.07%) with the Experiment 2 mean (m = 1.48%, sd = 3.02%). The t-test provides no significant support for a statistical difference between the groups (t(14.451) = -0.011, p = .99). A Bayes factor of 0.363 when using the experiment as a predictor of the mean PPD during the maintenance phase indicates moderate support for the null hypothesis – no group difference. Given no significant difference in pupil dilation between the Experiment 1 and 2 groups, I pooled them together for further analysis.

Figure 9
Comparison of mean PPD over the trial cycle for Exp 1 vs Exp 2 when CV is 50%

Note. The change in mean pupil size compared with baseline over the trial cycle plotted for experiment 1 (blue) and experiment 2 (red) trials when CV is 50%. The shaded areas indicate the 95% CI.

Cue validity and pupil dilation

To investigate the hypotheses of how CV may affect pupil dilation (H2a, H2b and H2c), I plotted the mean PPD over the trial cycle for different cue validities with 95% CI on group (Figure 10) and individual levels (Figure 11). In addition, I calculated Welch's t-tests to compare the PPD mean for each combination of conditions – 50%, 75%, and 100% CV – at each time sample. Significant t-test results (α = .05) for a given time sample have been plotted in Figures 10 and 11 to visualize mean PPD differences between conditions. Although a visual difference in mean PPD emerged in the delay period on a group level (Figure 10), such differences were insignificant at all time samples. However, p-values with a strict cut-off (α = .05) may be misleading. To verify the evidence's strength, mean PDD was analyzed over the entire and latter 750 ms part of the delay period (Table 2). Mean PPD during the entire delay period was 1.5% (sd = 2.99%) for 50% CV, 1.0% (sd = 2.90%) for 75% CV, and 0.6% (sd = 2.48%) for 100% CV while during the last 750 ms of the maintenance period 1.9% (sd = 2.92%) for 50% CV, 1.3% (sd = 2.78%) for 50% CV, and 0.7% (sd = 2.31%) for 50% CV. I also calculated Bayes factors comparing the difference in mean PPD over the delay period between the different
combinations of CVs. These analyses show that while the evidence for the null hypothesis – no difference – is moderate when comparing 50% with 75% CV (Bayes factor = .30) and 75% with 100% CV (Bayes factor = .30), the evidence for the null hypothesis is weak when comparing 50% with 100% CV (Bayes Factor = .53). Restricting the analysis to the last 750 ms of the delay period the evidence for the null hypothesis when comparing 50% with 100% CV become even weaker (Bayes Factor = .93) while comparing 50% with 75% CV (Bayes factor = .33) and 75% with 100 % CV (Bayes factor .37) remains moderate.

**Figure 10**

*Group level mean PPD over the trial cycle for different cue validities*

![Graph showing PPD over trial per CV for different cue validities](image)

*Note.* The change in mean pupil size compared with baseline over the trial cycle plotted for 50% CV (blue), 75% CV (red), and 100% CV (green). The color shaded areas indicate the 95% CI. Significant differences between condition PPD means using Welch’s t-test with $\alpha = .05$ are plotted for each time sample in the lower part of the graph. Although, no significant differences were detected for any time sample on the group-level (compare with individual results in Figure 11).

**Table 2**

*Differences in mean PPD during delay period for different cue validities*

<table>
<thead>
<tr>
<th>Cue Validity</th>
<th>BF full delay period 1500 ms</th>
<th>BF last 750 ms of delay period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>50%</td>
</tr>
<tr>
<td>50%</td>
<td>1.5%</td>
<td>-</td>
</tr>
<tr>
<td>75%</td>
<td>1.0%</td>
<td>0.30</td>
</tr>
<tr>
<td>100%</td>
<td>0.6%</td>
<td>0.53</td>
</tr>
</tbody>
</table>

*Note.* Bayes factors when comparing the PPD mean for different combinations of cue validities. Bayes factor less than 1 is in favor of the null hypothesis – no difference.

*Individual differences*
The above analyses suggest no effect of CV on pupil dilation at the group level. Effects may still be visible in a subset of participants. Therefore, as a more exploratory analysis, I also looked at pupil data at the level of individual participants. Interestingly, there seem to be two kinds of participants: most participants show no effect at all of CV on PPD, e.g., participants 7, 16, 19, 23, 26, 29, 36, and 37, while the remaining subjects show a pattern broadly consistent with the non-significant, but emerging, pattern on the group level, e.g., participants 8, 10, 14, 17, 24, 31, 33, and 35 (Figure 11). When analyzing the differences in mean PPD for different cue validities over the last 750 ms of the delay period using one-way ANOVA and Bayes factors, the individual differences also become statistically evident where 13 of 32 individuals show a significant (p < .05) effect of CV on PPD in a one-way ANOVA (Table 3).

**Figure 11**  
*Individual mean PPD over the trial cycle for different cue validities*

A) Experiment 1 participant 4 – 13

B) Experiment 2 participant 14 – 40
Note. The change in mean pupil size compared with baseline over the trial cycle plotted for 50% CV (blue), 75% CV (red), and 100% CV (green). The color shaded areas indicate the
95% CI. Significant differences between condition PPD means using Welch’s t-test with α = .05 are plotted for each time sample in the lower part of the graph.

Table 3
Differences in mean PPD during last 750 ms of delay period for different cue validities

<table>
<thead>
<tr>
<th>Id</th>
<th>50%</th>
<th>75%</th>
<th>100%</th>
<th>50 vs 100</th>
<th>75 vs 75</th>
<th>100</th>
<th>f-value</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>-2.3%</td>
<td>-1.8%</td>
<td>-2.6%</td>
<td>0.11</td>
<td>0.12</td>
<td>0.15</td>
<td>0.41</td>
<td>2, 695</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4.9%</td>
<td>3.6%</td>
<td>3.5%</td>
<td>0.78</td>
<td>0.68</td>
<td>0.10</td>
<td>2.79</td>
<td>2, 782</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-1.3%</td>
<td>-1.4%</td>
<td>-0.7%</td>
<td>0.28</td>
<td>0.11</td>
<td>0.52</td>
<td>1.82</td>
<td>2, 637</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1.0%</td>
<td>-1.6%</td>
<td>0.1%</td>
<td>0.22</td>
<td>10.39</td>
<td>0.75</td>
<td>4.49</td>
<td>2, 413</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3.4%</td>
<td>3.3%</td>
<td>3.7%</td>
<td>0.14</td>
<td>0.11</td>
<td>0.17</td>
<td>0.51</td>
<td>2, 647</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>5.2%</td>
<td>4.3%</td>
<td>-0.2%</td>
<td>5.33</td>
<td>0.21</td>
<td>7.25</td>
<td>33.68</td>
<td>2, 940</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.5%</td>
<td>-0.5%</td>
<td>-1.1%</td>
<td>0.25</td>
<td>0.42</td>
<td>0.18</td>
<td>4.15</td>
<td>2, 796</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>3.3%</td>
<td>3.3%</td>
<td>4.3%</td>
<td>0.32</td>
<td>0.30</td>
<td>0.11</td>
<td>1.41</td>
<td>2, 596</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>-0.8%</td>
<td>-0.4%</td>
<td>-0.6%</td>
<td>0.10</td>
<td>0.14</td>
<td>0.11</td>
<td>0.53</td>
<td>2, 919</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>6.3%</td>
<td>5.4%</td>
<td>1.6%</td>
<td>6.16</td>
<td>0.91</td>
<td>1.89</td>
<td>74.69</td>
<td>2, 927</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>-2.7%</td>
<td>-3.8%</td>
<td>-4.4%</td>
<td>1.62</td>
<td>0.36</td>
<td>0.14</td>
<td>3.33</td>
<td>2, 956</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>-4.4%</td>
<td>-4.4%</td>
<td>-4.6%</td>
<td>0.10</td>
<td>0.09</td>
<td>0.11</td>
<td>0.17</td>
<td>2, 858</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>5.4%</td>
<td>1.8%</td>
<td>0.2%</td>
<td>5.45</td>
<td>18.08</td>
<td>0.29</td>
<td>12.52</td>
<td>2, 848</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>-0.008</td>
<td>-0.011</td>
<td>-1.3%</td>
<td>0.17</td>
<td>0.10</td>
<td>0.11</td>
<td>0.65</td>
<td>2, 946</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.022</td>
<td>0.018</td>
<td>1.9%</td>
<td>0.12</td>
<td>0.14</td>
<td>0.10</td>
<td>0.46</td>
<td>2, 835</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>1.8%</td>
<td>-0.6%</td>
<td>0.04%</td>
<td>167.25</td>
<td>1.17</td>
<td>0.22</td>
<td>15.63</td>
<td>2, 952</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>3.4%</td>
<td>3.3%</td>
<td>3.0%</td>
<td>0.15</td>
<td>0.12</td>
<td>0.14</td>
<td>0.27</td>
<td>2, 507</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>6.0%</td>
<td>3.9%</td>
<td>2.3%</td>
<td>1.06</td>
<td>2.28</td>
<td>119.88</td>
<td>43.51</td>
<td>2, 907</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>-0.8%</td>
<td>-0.7%</td>
<td>0.2%</td>
<td>0.61</td>
<td>0.10</td>
<td>0.42</td>
<td>2.21</td>
<td>2, 705</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>2.5%</td>
<td>3.1%</td>
<td>2.5%</td>
<td>0.10</td>
<td>0.17</td>
<td>0.14</td>
<td>0.72</td>
<td>2, 836</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>-0.9%</td>
<td>-1.3%</td>
<td>-0.4%</td>
<td>0.17</td>
<td>0.14</td>
<td>0.39</td>
<td>1.35</td>
<td>2, 600</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>1.1%</td>
<td>0.5%</td>
<td>0.7%</td>
<td>0.22</td>
<td>0.59</td>
<td>0.11</td>
<td>2.04</td>
<td>2, 937</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>0.2%</td>
<td>-0.2%</td>
<td>-0.5%</td>
<td>0.36</td>
<td>0.16</td>
<td>0.13</td>
<td>1.18</td>
<td>2, 649</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>3.6%</td>
<td>4.1%</td>
<td>4.1%</td>
<td>0.11</td>
<td>0.12</td>
<td>0.09</td>
<td>0.34</td>
<td>2, 952</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>5.4%</td>
<td>4.5%</td>
<td>3.2%</td>
<td>19.99</td>
<td>0.19</td>
<td>0.53</td>
<td>5.32</td>
<td>2, 905</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>3.3%</td>
<td>2.9%</td>
<td>2.0%</td>
<td>0.36</td>
<td>0.11</td>
<td>0.19</td>
<td>1.50</td>
<td>2, 775</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>6.3%</td>
<td>6.2%</td>
<td>3.0%</td>
<td>9.38</td>
<td>0.09</td>
<td>3.00</td>
<td>63.07</td>
<td>2, 957</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>2.0%</td>
<td>1.1%</td>
<td>1.4%</td>
<td>0.46</td>
<td>2.79</td>
<td>0.15</td>
<td>3.59</td>
<td>2, 724</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>3.4%</td>
<td>3.0%</td>
<td>0.8%</td>
<td>1.03</td>
<td>0.20</td>
<td>1.30</td>
<td>33.32</td>
<td>2, 612</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>1.1%</td>
<td>1.2%</td>
<td>0.2%</td>
<td>0.53</td>
<td>0.09</td>
<td>0.78</td>
<td>2.64</td>
<td>2, 872</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>-2.6%</td>
<td>-2.6%</td>
<td>-2.7%</td>
<td>0.09</td>
<td>0.09</td>
<td>0.10</td>
<td>0.08</td>
<td>2, 926</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>3.8%</td>
<td>3.8%</td>
<td>2.8%</td>
<td>0.97</td>
<td>0.10</td>
<td>1.09</td>
<td>3.17</td>
<td>2, 725</td>
<td></td>
</tr>
</tbody>
</table>

Note. The table presents the mean PPD for each CV during the last 750 ms in the trial cycle – timepoint 3200 ms to 3950 ms. Bayes factors were calculated to compare the relative statistical support for CV constituting a predictor for the difference in mean PPD during
pairwise comparison. Bayes factor less than one favors the null hypothesis, i.e., no difference. F-value, degrees of freedom, and significance represent the result of a one-way ANOVA with CV as a predictor of mean PPD. Significance levels presented as * p < .05, ** p < .01, and *** p < .001.

**Memory accuracy and phasic pupil dilation**

To analyze the trial-by-trial relationship between mean PPD in the last 750 ms of the delay period and memory accuracy, $E_{raw}$ was transformed to z-scores on an individual basis and plotted against mean memory maintenance PPD for each SPF on a group level (Figure 1). For all SPF, the Pearson correlation was insignificant (p < .05) with a correlation coefficient close to zero (25% SPF: r = -.006, p = .71; 50% SPF: r = -.009, p = .40; 75% SPF: r = .024, p = .06; 100% SPF: r = .007, p = .51). These results suggest that PPD could not predict $E_{raw}$ during the delay period on a trial-by-trial basis.

**Figure 12**

*Mean PPD in the delay period vs. z-transformed $E_{raw}$*

*Note.* Mean phasic pupil dilation compared with z-transformed $E_{raw}$ in scatterplots for 25%, 50%, 75%, and 100% shape probing frequency.
Discussion

How the VWM’s limited capacity is manifested is an open debate between researchers. The discrete capacity model suggesting that VWM has a limited number of slots for temporary storage of encode items has for long been the most prevalent hypothesis. This model predicts that items and their features are encoded with high accuracy and any variance in memory precision is due to guesses for non-encoded items (Luck & Vogel, 1997; Zhang & Luck, 2008). In the last decades, several researchers have suggested alternative models based on flexible resource allocation schemes predicting VWM as a flexible storage where items are encoded relative to their importance for future processing and recall (Palmer, 1990; Wilken & Ma, 2004; Bays & Husain, 2008; Van den Berg et al., 2012).

My aim with this study was to test the flexible resource allocation model and pupil dilation as an indirect measure of flexible resources allocation. In two experiments I investigated if relative importance between visual stimuli affected memory precision in a continuous-recall task and if changes in pupil dilation during memory maintenance would track such differences. Across the two experiments I found evidence supporting the flexible resource allocation model in the memory accuracy data, while the pupil data was not fully consistent with my hypotheses.

Memory accuracy

As hypothesized, item importance significantly affected memory precision, such that colors of the more important items were, on average, more accurately remembered and reported with higher precision. The significant differences in memory precision between all SPFs (25%, 50%, 75%, and 100%; Figure 4) are difficult to reconcile with a discrete capacity model, suggesting that VWM encodes all items into separate slots with equal precision. Such a model would predict memory precision differences between the 50% CV and 100% CV conditions resulting in different memory loads – four vs. two items – but the model would also predict non-significant memory precision differences between the 50% CV and 75% CV condition as they theoretically would induce the same memory load – four items. Including the 75% CV condition in this study corrected for a possible confound in a previous study (experiment 2 from Salahub et al., 2019), making the results from the current study more robust in favor of a flexible resource allocation model compared with the results from the previous study.

Zhang and Luck (2008) argued that differences in variance could be explained by participants occasionally making successful guesses. Applying a mixture model – combining a uniform and Von Mises distribution – to the current data suggests that guessing can account for a significant proportion of the variation (Figure 6a). As expected, the estimated proportion of guesses decreases with higher cued probing probability, i.e., with increased importance, items are more likely to be encoded in memory, resulting in fewer guesses. Interestingly, when analyzing the items encoded in memory, cued probing probability still influences memory precision (Figure 6b), such that participants remember the color of more important items with higher accuracy. These findings suggest that high-value items, as indicated by cues, benefit from more VWM encoding resources, a result that strongly favors a flexible resource allocation model.

In contrast to previous studies (Dube et al., 2017; Emrich et al., 2017, Salahub et al., 2019), my data for the relationship between memory accuracy and SPF did not fully fit a power-law model (Figure 4). A theoretical account for a power-law relationship between precision and neural resources following a maximum likelihood decoding has previously been suggested (Bays & Husain, 2008). Although the data in the current study does not entirely exclude a power-law relationship between memory precision and probing probability, differences between the studies may explain the slightly different results in at least two cases. Like in this
study, Salahub et al. (2019) and Dube et al. (2017) used feature-based cues – circles or squares. However, both those studies held feature cues constant within subjects. As shown in experiment 1, holding the feature cue constant induces an attentional bias, which affects the memory precision negatively for non-cued items and slightly positively for cued items (Figure 3). These findings suggest that these studies’ memory precision results may be induced by errors that could make fitted models inaccurate.

Another finding in this study was that the speed-accuracy tradeoff could not explain the memory accuracy differences. There were no indications that cued probing probability would affect response time. Thus, participants’ higher response precision for more important items seems to be because of more precise memories and not due to spending more time responding.

In sum, the memory accuracy data from this study replicate and add robustness to earlier findings by addressing existing confounds. These findings add to the evidence of a flexible resource allocation model of VWM and strongly challenge discrete capacity models.

The predictiveness of PPD

Although the PPD differences for different CVs during memory maintenance were in my hypothesized directions – highest for 50% CV trials, followed by 75% CV trials, and lowest for 100% CV trials – these differences were small and non-significant on a group level with overlapping compatibility intervals over the whole maintenance cycle (Figure 1). Further analyzing the individual results indicated mixed results, with 40% of participants showing significant PPD differences for different CVs during the later parts of the memory maintenance period. In contrast, a majority showed no significant differences. These results are intriguing. What could explain such individual differences? At least two categories of explanations could be identified: 1) individuals are differently sensitive to PPD in ways not captured by the current study, and 2) individuals interpret the task differently and use different strategies to tackle it.

An explanation in the first category could be that the task is differently challenging for different individuals. Pupil dilation is a psychophysiological marker of mental effort (Kahneman, 1973) and is sensitive to arousal. A classic theory in psychology is the Yerkes-Dodson law, which postulates that performance follows a U-shaped curve relative to arousal such that intermediate levels of arousal are optimal for performance. Similarly, the Adaptive Gain Theory (AGT; Aston-Jones & Cohen, 2005) proposes that tonic and phasic activation of neurons in the locus coeruleus (LC) follow the same u-shaped pattern. Engagement in the current task is high when phasic LC is high and tonic LC is intermediate. When tonic LC is either high or low, phasic LC attenuates. Tonic LC has been associated with pretrial pupil dilation and phasic LC with phasic dilation during a task (Murphy et al., 2011; Gilzenrat et al., 2010). Murphy and colleagues (2011) suggested that task-evoked PPD signals attentional challenge. Based on these theories, one may speculate that different individuals find the task in this study differently challenging. PPD would only be high enough to induce differences between CV conditions when the challenge is within an optimal window – neither easy nor difficult.

While an optimal performance window may be more theoretically speculative, some experimental evidence suggests that the task might not have been challenging enough for all individuals. Unsworth and Robison (2015, 2018) found PPD changes at set sizes below four to be high for individuals with low VWM capacity and low for individuals with high VWM capacity. In the current experiment, the maximum set size was four, which may indicate that it was not challenging enough to induce PPD in high-capacity individuals. However, if only high-capacity individuals would have non-significant PPD responses relative to CV conditions, one would expect individuals with non-significant PPD results to have better memory accuracy performance. Such a pattern is not evident comparing individuals in Table 3 and Table A2, suggesting that this could not be the only explanation.
I base the second category of explanations on the assumption that different individuals take on the task using different strategies. Baddeley and Hitch (1994) separate working memory into the visual sketchpad and phonological loop in their memory model, suggesting that the latter stores verbalized information. Verbalizing visual information can be transformed from the visual sketchpad to the phonological loop and repeated for better recall. Such transfer can be prevented by experimentally suppressing articulation, which was not explicitly done in this experiment.

Furthermore, long-term memory affects working memory (Cowen, 2001). In essence, working memory chunks point to long-term associations, which may be unlimited in size. Verbalization strategies, e.g., repeated rehearsals of a group of items, can be viewed as a long-term association between chunks, thus forming a new working memory chunk. Other examples of strategies include recoding, e.g., when naming the colors that are visually displayed, or hierarchical shifting, e.g., grouping the colors in association with a country flag. Such strategies may influence the number of chunks in working memory and affect what one can recall from an array. Long-term memory associations between chunks should be limited to estimate working memory capacity. One way to limit such associations in the current study was to randomize the colors shown in each trial and pseudo-randomize the shape positions. Using verbalization and other strategies was not controlled for and could explain some PPD differences between individuals. As an anecdotal example, one participant explained to the experimenter how she had utilized the fact that she was multilingual, which gave access to more words for color nuances. That participant’s PPD data was almost flat, with no differences between CV conditions and high mean memory precision for all probing probability conditions. This anecdotal example suggests that controlling for verbalization and other strategies may be essential in future studies of VWM and PPD.

In sum, on a group level, the PPD differences for different CVs during memory maintenance were non-significant, although the results were in the hypothesized direction. On an individual level, the results contradicted, with about a third of participants showing significant differences in PPD between conditions. I propose two general explanations for such individual differences – different PPD sensitivity and the use of different strategies – both should be controlled for in future studies.

PPD and spatial attention

In contrast to my hypothesis, spatial separation did not influence PPD nor response accuracy. Although the attention-based rehearsal hypothesis (Awh & Jonides, 2001) suggests covert spatial shifts during memory maintenance, the results do not support that such shifts would affect changes in pupil size. It is important to note that this finding does not challenge the attention-based rehearsal hypothesis per se. A more plausible interpretation is that PPD measurements in this experiment did not detect such effects. There could be several reasons for this. One reason is that PPD may not be sensitive enough to detect the relatively small spatial differences. Figure 9c indicates a slight difference between grouped and separated trials for the 100% CV, although with largely overlapping compatibility intervals. This indicates that a difference might be possible to detect using larger sample sizes, making the spatial difference bigger or including a more extensive set size – increasing memory load. It is also possible that the individual differences detected for PPD relative to CV apply to spatial attention. Such differences have not been analyzed.

The attention-based rehearsal hypothesis focuses on the covert spatial shifts of attention. Van Ede et al. (2019) found evidence for such shifts as a gaze bias towards covertly attended stimuli using a retro-cue paradigm. The superior colliculus is involved in orienting towards environmental stimuli. Although Alnaes et al. (2014) found a correlation between pupil size and superior colliculus activity, Van Ede and colleagues’ results suggest that eye position, e.g.,
gaze bias or microsaccades, may be better than pupil dilation to measure the effects of spatial attention. Analyzing eye position was not part of the aim of this thesis. However, the data was collected using a 1200 Hz sample rate, allowing for further analysis of microsaccades and gaze bias to investigate the effect of spatial separation.

In sum, I did not find evidence that PPD during the memory maintenance phase would track differences in spatial separation. However, such findings do not invalidate the theories underlying my hypothesis.

**PPD and memory accuracy**

Finally, pupil dilation did not predict accuracy on a trial-per-trial basis for any SPF condition. It is somewhat surprising to find that correlations are insignificant and close to zero, considering that the AGT postulates phasic LC associated with PPD (Anaes et al., 2014) to track task engagement and performance (Aston-Jones & Cohen, 2005). On the other hand, Robison and Unsworth (2018) found pre-trial pupil size, but not PPD, to predict memory accuracy. Murphy and colleagues showed that changes in PPD tracked improved performance in succeeding, but not current, trials. That would indicate that PPD changes re-engage participants in tasks and that the memory improvement effects may be delayed. Time series analysis to investigate whether such a phenomenon exists in the current data was not performed, but it could be an endeavor worth pursuing.

**Limitations and Future Directions**

The current study had several limitations that affect the possibilities to analyze and interpret the results. These are connected to the two primary purposes of the study: to investigate the effect of different cue validities on memory precision to bring light into the debate of VWM models and to investigate to what extent PPD could track resource allocation of items with different importance. The PPD results were most challenging to interpret and call for follow-up experiments to bring clarity.

A critical finding in this study was the individual differences in PPD relative to CV condition. These results clearly show that one must be careful when interpreting group-level effects from individual psychophysiological measurements, like pupil size unless controlling for such individual differences. The study had several limitations that could have improved the ability to interpret the individual results and make it possible to control for individual differences. Firstly, individuals were not limited in using strategies like verbalization, nor was any data collected about what strategies individuals used. Unfortunately, this makes any discussion about individual differences speculative and difficult to analyze. Future studies measuring PPD relative to VWM should consider embedding post-trial inquiries tracking the individual strategies and perceived challenge level. That would provide a better means to analyze individual data based on two hypotheses: 1) that the use of strategy affects PPD, and 2) that PPD may track the level of task challenge on a trial-per-trial basis (Murphy et al., 2011). Considering the individual difference future studies may consider using single-N designs to examine how within-individual differences over time, e.g., daily psychometrical fluctuations, affect PPD.

A second limitation was that the maximum set size was four items. Considering Unsworth and Robison’s (2015; 2018) findings that high VWM capacity individuals have marginal PPD response for lower set sizes, this limitation may have affected the ability to detect PPD effects of CV conditions.

A third limitation was the relatively short inter-trial intervals (ITI) used for fixation and calculating a trial baseline. Looking closely at Figure 1, one can see that the mean PPD curve continues to decline through the fixation interval and into the cueing phase. This decline could suggest that pupil size only partially reaches an accurate baseline between trials, which would
further affect the PPD, calculated as the relative increase relative to the baseline. Analyzing PPD as a relative change in percentage to the pre-trial pupil size can also be questioned since it somewhat combines indirect measures of two theoretically dependent variables – tonic and phasic LC activity (Aston-Jones & Cohen, 2005). On the other hand, analyzing PPD as absolute pupil dilation changes would be problematic when comparing individuals.

The study also has some limitations that affect the memory precision data. As was evident from the response precision data when comparing Experiment 1 and Experiment 2, non-cued items were not given the same attention in the 50/50 case. Unequal attention could also affect response precision for the 25% SPF, which in this study was assumed to derive from the rest of the 75% CV logically. Future studies should also include trials with 25% cue validities to understand the effect of low-probability cues.

A general challenge with the current study is the long experimental time – three hours – and the repetitiveness of the task, which may be tiring. It could have led some participants not fully to use the cues and their validities as intended. Future studies could include a reward system, e.g., introducing a feedback system on memory accuracy, to achieve more ecologically valid cues and ensure participants' engagement.

A final general limitation is that the current experiment cannot sufficiently separate if the effects of CV on memory precision and PPD are mediated by differences in encoding efficiency or attention to encoded items during the maintenance phase. Expanding the experiment to include retro-cues at the beginning of the maintenance phase would allow for measurements of attention after encoding and compare those with the effects of pre-cues in this experiment (Souza & Orberauer, 2016).

In sum, the study had several limitations, many limiting the possibilities to analyze and interpret the PPD data, particularly the individual differences. For future studies aiming to use pupillometry to measure the processes of VWM, understanding what constitutes the differences on individual levels would be essential. Such studies could investigate how individual differences in perceived challenge, use of different memory strategies, and estimated VWM capacity levels affect measured PPD during memory maintenance. While this and earlier studies have provided evidence for a flexible resource allocation model of VWM, separating the different phases of the VWM process will be essential to further our understanding of what resources – encoding, storage, and attention – are flexibly allocated. In such studies, taking advantage of continuous-recall tasks, possibly combining pre- and retro-cues with different validities, could be a way forward.

**Conclusion**

In this study, I have shown that the color of items assigned with a higher probing probability is reported with higher precision in a continuous-recall task. These findings favor a model where VWM utilizes its scarce resources through flexible allocation to optimize recall over a model equally assigning discrete resources to each memorized item. The study contributes to the knowledge of VWM behavior by replicating earlier studies while eliminating some previous confounds.

Moreover, the study has shown that for a few individuals, but not all, change in pupil size during memory maintenance is affected by cue validity and possibly tracking the unequal allocation of attentional resources in a lab setting. The present results suggest that the data are affected by individual differences, for which I have provided two speculative explanations. Many questions remain, however, and answering those would require follow-up experiments with different designs. Based on the present findings, I have provided several directions for such follow-ups.
The study failed to show that spatial separation of items affects the change in pupil size during memory maintenance. Likewise, there was no evidence for a trial-by-trial relationship between change in pupil size and reported memory precision. These failures to find significant results may have no direct implications for current VWM and attentional theories. However, they can be used as indicative boundaries for their explanatory power on phasic pupil dilation. Overall, the study has expanded the evidence in favor of a flexible resource allocation model of VWM and provided some intriguing indications of how individual differences color phasic pupil dilation as a measurement during memory maintenance.
References


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


Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453(7192), 233-235. [https://doi.org/10.1038/nature06860](https://doi.org/10.1038/nature06860)
# Appendix A

## Tables and Figures

### Table A1

<table>
<thead>
<tr>
<th>ID</th>
<th>Experiment</th>
<th>Age</th>
<th>Gender</th>
<th>Terminated trials</th>
<th>Non-fixation trials</th>
<th>Successful trials</th>
<th>Success rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1</td>
<td>33</td>
<td>Male</td>
<td>462</td>
<td>0</td>
<td>698</td>
<td>73%</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>23</td>
<td>Male</td>
<td>374</td>
<td>0</td>
<td>786</td>
<td>82%</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>36</td>
<td>Male</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>38</td>
<td>Female</td>
<td>200</td>
<td>315</td>
<td>645</td>
<td>67%</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>39</td>
<td>Male</td>
<td>140</td>
<td>251</td>
<td>421</td>
<td>44%</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>45</td>
<td>Male</td>
<td>200</td>
<td>306</td>
<td>654</td>
<td>68%</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>38</td>
<td>Female</td>
<td>118</td>
<td>5</td>
<td>955</td>
<td>99%</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>32</td>
<td>Male</td>
<td>200</td>
<td>160</td>
<td>800</td>
<td>83%</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>37</td>
<td>Male</td>
<td>200</td>
<td>359</td>
<td>601</td>
<td>63%</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>25</td>
<td>Female</td>
<td>140</td>
<td>9</td>
<td>951</td>
<td>99%</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>34</td>
<td>Male</td>
<td>38</td>
<td>0</td>
<td>960</td>
<td>100%</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>27</td>
<td>Female</td>
<td>23</td>
<td>0</td>
<td>960</td>
<td>100%</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>19</td>
<td>Female</td>
<td>161</td>
<td>98</td>
<td>862</td>
<td>90%</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>38</td>
<td>Male</td>
<td>205</td>
<td>106</td>
<td>854</td>
<td>89%</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>25</td>
<td>Female</td>
<td>240</td>
<td>906</td>
<td>54</td>
<td>6%</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>28</td>
<td>Female</td>
<td>147</td>
<td>10</td>
<td>950</td>
<td>99%</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>18</td>
<td>Female</td>
<td>215</td>
<td>111</td>
<td>849</td>
<td>89%</td>
</tr>
<tr>
<td>21</td>
<td>2</td>
<td>29</td>
<td>Male</td>
<td>240</td>
<td>922</td>
<td>38</td>
<td>4%</td>
</tr>
<tr>
<td>22</td>
<td>2</td>
<td>22</td>
<td>Male</td>
<td>80</td>
<td>1</td>
<td>959</td>
<td>100%</td>
</tr>
<tr>
<td>23</td>
<td>2</td>
<td>21</td>
<td>Male</td>
<td>240</td>
<td>455</td>
<td>515</td>
<td>54%</td>
</tr>
<tr>
<td>24</td>
<td>2</td>
<td>25</td>
<td>Female</td>
<td>94</td>
<td>50</td>
<td>910</td>
<td>95%</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>31</td>
<td>Male</td>
<td>212</td>
<td>245</td>
<td>715</td>
<td>74%</td>
</tr>
<tr>
<td>26</td>
<td>2</td>
<td>24</td>
<td>Male</td>
<td>181</td>
<td>120</td>
<td>840</td>
<td>88%</td>
</tr>
<tr>
<td>27</td>
<td>2</td>
<td>24</td>
<td>Female</td>
<td>237</td>
<td>354</td>
<td>606</td>
<td>63%</td>
</tr>
<tr>
<td>28</td>
<td>2</td>
<td>24</td>
<td>Male</td>
<td>91</td>
<td>8</td>
<td>952</td>
<td>99%</td>
</tr>
<tr>
<td>29</td>
<td>2</td>
<td>33</td>
<td>Female</td>
<td>240</td>
<td>308</td>
<td>652</td>
<td>68%</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>31</td>
<td>Male</td>
<td>172</td>
<td>2</td>
<td>958</td>
<td>100%</td>
</tr>
<tr>
<td>31</td>
<td>2</td>
<td>24</td>
<td>Female</td>
<td>213</td>
<td>48</td>
<td>912</td>
<td>95%</td>
</tr>
<tr>
<td>32</td>
<td>2</td>
<td>31</td>
<td>Male</td>
<td>235</td>
<td>168</td>
<td>792</td>
<td>83%</td>
</tr>
<tr>
<td>33</td>
<td>2</td>
<td>22</td>
<td>Female</td>
<td>26</td>
<td>0</td>
<td>960</td>
<td>100%</td>
</tr>
<tr>
<td>34</td>
<td>2</td>
<td>31</td>
<td>Female</td>
<td>234</td>
<td>199</td>
<td>761</td>
<td>79%</td>
</tr>
<tr>
<td>35</td>
<td>2</td>
<td>27</td>
<td>Female</td>
<td>240</td>
<td>335</td>
<td>625</td>
<td>65%</td>
</tr>
<tr>
<td>36</td>
<td>2</td>
<td>18</td>
<td>Male</td>
<td>232</td>
<td>79</td>
<td>881</td>
<td>92%</td>
</tr>
<tr>
<td>37</td>
<td>2</td>
<td>18</td>
<td>Male</td>
<td>204</td>
<td>26</td>
<td>934</td>
<td>97%</td>
</tr>
<tr>
<td>38</td>
<td>2</td>
<td>23</td>
<td>Male</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>39</td>
<td>2</td>
<td>37</td>
<td>Male</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>40</td>
<td>2</td>
<td>31</td>
<td>Male</td>
<td>240</td>
<td>225</td>
<td>735</td>
<td>61%</td>
</tr>
</tbody>
</table>

*Note.* Participant 6, 18, 21, 38 and 39 excluded.
### Table A2

**Individual mean error per SPF**

<table>
<thead>
<tr>
<th>ID</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>100%</th>
<th>df</th>
<th>F-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>68.2</td>
<td>47.7</td>
<td>41.6</td>
<td>24.1</td>
<td>3,956</td>
<td>26.70</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>15</td>
<td>22.4</td>
<td>26.0</td>
<td>25.3</td>
<td>19.4</td>
<td>3,956</td>
<td>2.97</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>16</td>
<td>80.5</td>
<td>70.8</td>
<td>74.2</td>
<td>55.7</td>
<td>3,858</td>
<td>8.95</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>17</td>
<td>103.9</td>
<td>58.2</td>
<td>35.9</td>
<td>30.0</td>
<td>3,850</td>
<td>80.14</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>19</td>
<td>63.7</td>
<td>48.6</td>
<td>43.5</td>
<td>26.9</td>
<td>3,946</td>
<td>18.73</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>20</td>
<td>88.8</td>
<td>62.9</td>
<td>46.1</td>
<td>39.0</td>
<td>3,845</td>
<td>37.00</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>22</td>
<td>96.5</td>
<td>40.9</td>
<td>30.0</td>
<td>27.5</td>
<td>3,955</td>
<td>91.47</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>23</td>
<td>89.4</td>
<td>101.8</td>
<td>92</td>
<td>85.9</td>
<td>3,511</td>
<td>4.42</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>24</td>
<td>68.3</td>
<td>34.1</td>
<td>27.8</td>
<td>21.7</td>
<td>3,906</td>
<td>32.25</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>25</td>
<td>87.7</td>
<td>89.4</td>
<td>71.5</td>
<td>50.6</td>
<td>3,711</td>
<td>25.89</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>26</td>
<td>88.6</td>
<td>93.6</td>
<td>88.2</td>
<td>78.0</td>
<td>3,836</td>
<td>6.46</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>27</td>
<td>93.1</td>
<td>61.4</td>
<td>40.4</td>
<td>29.6</td>
<td>3,602</td>
<td>33.63</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>28</td>
<td>107.9</td>
<td>44.8</td>
<td>26.3</td>
<td>20.9</td>
<td>3,948</td>
<td>169.3</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>29</td>
<td>96.9</td>
<td>77.8</td>
<td>60.0</td>
<td>41.2</td>
<td>3,648</td>
<td>30.84</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>30</td>
<td>89.7</td>
<td>87.4</td>
<td>74.8</td>
<td>46.1</td>
<td>3,954</td>
<td>41.35</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>31</td>
<td>84.6</td>
<td>42.9</td>
<td>29.4</td>
<td>20.8</td>
<td>3,908</td>
<td>65.73</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>32</td>
<td>83.2</td>
<td>62.1</td>
<td>40.3</td>
<td>37.8</td>
<td>3,788</td>
<td>34.07</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>33</td>
<td>59.1</td>
<td>39.0</td>
<td>41.1</td>
<td>22.0</td>
<td>3,956</td>
<td>24.08</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>34</td>
<td>101.0</td>
<td>74.0</td>
<td>43.1</td>
<td>36.2</td>
<td>3,757</td>
<td>54.66</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>35</td>
<td>46.4</td>
<td>26.8</td>
<td>29.0</td>
<td>19.1</td>
<td>3,621</td>
<td>8.81</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>36</td>
<td>77.8</td>
<td>71.2</td>
<td>57.9</td>
<td>36.6</td>
<td>3,877</td>
<td>29.46</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>37</td>
<td>108.5</td>
<td>72.5</td>
<td>48.3</td>
<td>41.4</td>
<td>3,930</td>
<td>68.66</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>40</td>
<td>103.4</td>
<td>70.6</td>
<td>52.2</td>
<td>41.2</td>
<td>3,731</td>
<td>43.15</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

*Note.* P-values calculated for differences in the absolute value of $E_{\text{raw}}$. 
Figure A1

*Individual precision per SPF*
Note. Individual level $E_{\text{raw}}$ SD with 95% CI.
Appendix B

Abbreviations

*Adaptive Gain Theory (AGT): A theory suggesting that LC neurons exhibit phasic activation related to exploitation – engagement in the current task – and tonic activation related to disengagement in the current task in favor of exploration – seeking reward in new tasks (Aston-Jones & Cohen, 2005).*

*Contralateral Delay Activity (CDA): A delayed event-related potential in the contralateral hemisphere in the parietal and occipital regions.*

*Cue Validity (CV): The probability that a cued item or shape will be probed.*

*Guessing Probability (p_{guess}, P-uniform): The probability that the participant is guessing due to the probed item not being in memory.*

*Locus Coeruleus (LC): A nucleus in the dorsal PONS responsible for all norepinephrine released in the cortex.*

*Original Color (C_{origin}): The original color of the item in the encoding phase.*

*Phasic Pupil Dilation (PPD): Changes in pupil dilation related to a task, i.e., during a trial.*

*Raw Error Score (E_{raw}): The difference in degrees on the color wheel between the response and the original color.*

*Response Color (C_{response}): The response color given by the participant in the response phase.*

*Shape Probing Frequency (SPF): The probing probability for the probed shape given the cue.*

*Visual Working Memory (VWM): A system that actively and temporarily stores visual information.*