Effects of Retro-cue Reliability on Visual Working Memory and Attentional Template Efficiency in Visual Search

Fortuna Pacheco, Alexandre; Constant, Martin; Kerzel, Dirk

Effects of Retro-cue Reliability on Visual Working Memory and Attentional Template Efficiency in Visual Search

Alexandre Fortuna*, Martin Constant & Dirk Kerzel

Faculty of Psychology and Educational Sciences, University of Geneva

Abstract

In the visual search literature, an attentional template refers to a mental representation of target features stored in visual working memory. The dual-state model (Olivers et al., 2011) argues that visual working memory can only be focused on a single representation at a time, and therefore, we can only search for one target at a time. Consequently, to detect two potential targets, we must rapidly alternate our attention between the two corresponding representations. By contrast, the resource hypothesis (Huynh Cong & Kerzel, 2021) posits that we can simultaneously activate more than one representation for visual search, although this requires sharing the available resources. In three pre-registered experiments, we compared the predictions from the state-based model and the resource hypothesis in a combined search and memory task. The results suggest that participants allocate their resources based on the reliability of the retro-cue and therefore favor the resource hypothesis. In contrast, the results are less consistent with the state-based model, which predicts that the cost in terms of response time and accuracy should remain identical regardless of the retro-cue's reliability.

Public Significance Statement

In everyday life, we often search for things like our phone or keys using a mental image of what we are looking for. This study shows that people focus better on what they try to memorize when they believe a hint is reliable. When the hint correctly pointed to what mattered, people were faster and more accurate in finding or remembering it. This means our attention and memory do not work in fixed ways, but that they adapt based on how helpful we expect the information to be. This helps us better understand the interaction between memory and attention, and how expectations influence performance in simple tasks.

Keywords: visual search, memory task, attentional template, retro-cues, visual working memory

Supplemental materials: https://osf.io/q8d47/

This work was presented at the Vision Sciences Society Conferences (2025). This research was supported by the grant FNS 219517 from the Swiss National Science Foundation (SNSF) to Dirk Kerzel. We have no known conflict of interest to disclose.

Alexandre Fortuna https://orcid.org/0009-0008-5090-6923
Martin Constant https://orcid.org/0000-0001-9574-0674
Dirk Kerzel https://orcid.org/0000-0002-2466-5221

Correspondence should be addressed to Alexandre Fortuna, Faculty of Psychology and Educational Sciences, University of Geneva, 40 Boulevard du Pont d'Arve, 1205 Genève, Switzerland. E-mail: Alexandre.Fortuna@unige.ch

People devote a significant part of our daily lives to search for familiar objects in complex, densely packed visual scenes. In some situations, one needs to search for several objects simultaneously. A common example is cooking dinner, where we often search for two utensils at the same time. Working memory plays a key role in both perceptual attention and action control: it maintains target templates to guide perceptual selection, and stores task sets that support goal-directed behavior (Myers et al., 2017). In these cases, stored feature representations serve to guide attentional selection during visual search (e.g., the shape of a spoon and the color of a bowl), known as attentional templates (Duncan & Humphreys, 1989) or attentional control sets (Folk et al., 1992).

An attentional template is a mental representation of one or more target features used to guide visual search. Several models of visual search (Schneider, 2013; Wolfe, 2021) argue that attentional templates are stored in visual working memory (VWM) and are activated shortly before visual search to prioritize objects with matching attributes. This is supported by electrophysiological studies (Carlisle et al., 2011; Grubert & Eimer, 2018, 2020; Woodman et al., 2013). Consequently, attention is directed toward items that match the stored target features through a top-down process (Chelazzi et al., 1998). Once candidate target objects have been selected, they are compared to the attentional templates to determine whether they are a target (Eimer, 2014; Wolfe, 2021).

Nevertheless, how multiple attentional templates are maintained and used in VWM remains unclear. Olivers et al. (2011) proposed the single-item-template hypothesis, which suggests that VWM operates with two representational states: an active state and an accessory state. The "active" state, where the item representation serves as an attentional template. In this framework, only a single representation at a time can be kept in the active state. All other items are maintained in an "accessory" state, where they cannot interact with visual search. When searching for multiple items, representations sequentially switch from the accessory state to the active state. Olivers et al. (2011) based their proposal on attentional capture effects (Theeuwes, 1991, 1992) observed in dual-task settings. In these tasks, participants maintained a single feature (e.g., the color red) in VWM while searching for an unrelated target (e.g., a diamond among circles). A distractor matching the memorized feature would often capture attention and slow response times (RTs).

However, recent studies have demonstrated that VWM does not always influence attentional guidance, and that this influence may be present or absent depending on the structure of the visual search task. When the search target remains constant over multiple trials, it is thought that the corresponding attentional template can be transferred to "activated long-term memory" (Gunseli et al., 2014; Reinhart & Woodman, 2014, 2015; Wolfe, 2021). This offloading reduces demands on VWM, allowing accessory memory representations to become active and interfere with search performance. Conversely, when the search target changes from trial to trial, the corresponding attentional template continuously needs to be updated in VWM, forcing it to remain in the active state. As a result, the other representations remain accessory, and no memory-based interference occurs. (Downing & Dodds, 2004; Hollingworth & Hwang, 2013; Olivers, 2009). However, Foerster and Schneider (2018) showed that even search-irrelevant features of the active VWM template can still capture attention and cause interference, even when it would be beneficial to ignore them.

Moreover, storing several items can introduce competition effects, further constraining attentional guidance (Carlisle & Woodman, 2019). Although VWM can store approximately 3–4 items (Cowan, 2001; Luck & Vogel, 2013), competition may arise particularly when multiple items share the same feature values relevant for the search task, such as color or shape. This similarity increases the likelihood that none of the items reaches the active status required to guide attention effectively. For instance, holding two identically shaped items in memory creates strong competition so that neither can become fully active to bias attention. In contrast, when the remembered items differ in shape, competition is reduced and one item is more likely to achieve active status and guide attention (Woodman & Luck, 2007).

Although the single-item-template hypothesis has received considerable support (Grubert et al., 2016; Olivers et al., 2011; van Moorselaar et al., 2014), several elements call it into question. Many studies have demonstrated that several attentional templates can be activated simultaneously to guide visual search (Carlisle & Woodman, 2019; Hollingworth & Beck, 2016; Zhou et al., 2020). That is, participants can simultaneously search for two targets at the same time. This capacity was evidenced both by behavioral (Irons et al., 2012; Kerzel & Witzel, 2019) and electrophysiological (Berggren et al., 2020; Grubert & Eimer, 2015, 2016) studies. The single item template hypothesis is however unfit to explain participants' ability to activate several attentional templates at the same time. Therefore, to address these issues, the multiple item template hypothesis (Beck et al., 2012) was proposed. It assumes that several items stored in VWM can concurrently serve as attentional templates, allowing attention to be directed by multiple targets simultaneously.

Similarly, Bahle et al. (2020) proposed an activation-based model, in which all items stored in VWM exist on a continuum of activation, and the item with the highest activation has the greatest influence on attentional guidance. According to this model, rather than being restricted to binary states (active vs. accessory) or to models assuming fixed, all-or-nothing resource allocations, VWM representations dynamically vary in their ability to bias attention depending on their activation strength. That is, stronger activations enhance the probability that items guide search behavior, while lower activations weaken their influence without entirely removing them from attentional competition.

Within this notion of a flexibly distributed resource, Huynh Cong and Kerzel (2021) approached attentional templates not in terms of activation strength, as suggested by Bahle et al. (2020) nor as distinct representational states that require deactivation, as proposed by Olivers et al. (2011), but rather as depending on how much working memory resource is available. They also assume that representations that achieve the status of attentional template, either during encoding or maintenance, receive an amount of resources proportional to their relevance in visual search. Similar to the activation model, the more resources are allocated to a template, the more visual search should be enhanced. However, the amount of resources alone does not determine whether a representation interacts with visual search. That is, the representation receiving the largest amount of resources does not automatically become the target of visual search (Huynh Cong & Kerzel, 2021). In other words, an object can be recalled very precisely without influencing visual search.

This *resource hypothesis* builds on resource-based models of VWM (Bays et al., 2009; Bays & Husain, 2008; Ma et al., 2014) and is supported by several studies examining how memory resources relate to visual search performance (Hollingworth & Hwang, 2013; Kerzel & Witzel, 2019; Rajsic et al., 2017). One core prediction is that memory precision and search efficiency must covary: if allocating more resources to a template does not improve search performance, the hypothesis would be invalidated. Increasing the priority of a template for search should also increase the resources allocated to its memory representation. Based on the resource models of VWM (Bays et al., 2009; Bays & Husain, 2008; Ma et al., 2014), this increase is expected to enhance the precision of recall. In general, memory resources can be flexibly distributed among stored representations depending on their relevance (Franconeri et al., 2013; Ma et al., 2014). Because internal representations are subject to random fluctuations, and the magnitude of this noise rises with the number of stored representations, memory performance declines as the number of items in working memory increases (Ma et al., 2014).

The resource hypothesis is also related to the approach suggested by Kristjánsson (2023) who argues that attentional and VWM templates are not fixed, perfectly tuned entities, but rather dynamic representations modulated by spatial and temporal context as well as the characteristics of stimuli. This approach aligns with modern theories such as predictive coding and Bayesian inference (Chetverikov & Kristjansson, 2022), which postulate that the visual system continuously adjusts its predictions based on discrepancies between expectations and sensory signals.

However, recent findings suggest that the link between precision and attentional guidance may not be straightforward as the resource hypothesis suggests. According to resource-based accounts, increasing the amount of resources allocated to an item in VWM should improve not only its precision but also its ability to serve as an attentional template. In this view, a more precise representation should be more effective in biasing attention during tasks like visual search. However, recent findings suggest that this assumption may not always hold. For example, Dube and Al-Aidroos (2019) showed that prioritizing an item for enhanced memory precision via retro-cueing did not increase its ability to bias attention during search. Similarly, Dube et al. (2019) demonstrated that while deterministic retro-cues (100% valid) led to attentional capture by the cued item, probabilistic retro-cues (70% valid) did not—even though they improved memory performance. These findings support the idea that memory precision and attentional template status can be dissociated, and that increased resource allocation is not always sufficient for attentional guidance.

In the present study, we aimed to resolve a central theoretical disagreement concerning the mechanisms through which VWM representations guide attention.

Although prior research has established that retro-cues can improve recall accuracy (Gunseli et al., 2015; Murray et al., 2013), it remains an open question whether the reliability of these cues can continuously modulate both memory fidelity and attentional efficiency (see Dube & Al-Aidroos, 2019). Addressing this question is critical for adjudicating between binary state-switching accounts and resource-based frameworks. We sought to determine whether attentional templates in VWM are governed by all-or-none dynamics or by flexible, graded distributions of resources, thereby clarifying the architecture of the current models of attention–memory interactions.

This investigation seeks to elucidate the mechanisms governing resource allocation during memory maintenance and visual search. Resource-based models posit that retro-cue reliability modulates the prioritization of information within VWM: highly reliable cues should facilitate both enhanced memory accuracy and more effective guidance of attention during search. This also means that reduced reliability should lead to a more distributed allocation of resources, leading to diminished recall performance and less efficient search behavior. Some studies (Gunseli et al., 2015; Shimi et al., 2014) reported that retro-cues with higher reliability (80% vs. 50%) yielded greater improvements in recall accuracy at the cued location. Applying this rationale to the present task, it follows that greater retro-cue reliability should be associated not only with improved recall accuracy, but also with faster and more accurate visual search performance for the cued attentional template.

In three preregistered experiments, participants remembered colors and then performed a visual search task or a continuous-recall task. Both tasks were combined with retro-cueing to manipulate the relevance of attentional templates. The continuous-recall paradigm examines VWM by capturing both item recall and the accuracy with which features are remembered, based on response variability around the correct value (Zhang & Luck, 2008) though it has recently been argued that AFC paradigms might also be sufficient (see, Schurgin et al., 2020). In Experiment 1 participants had to remember two colors, and the cue was less reliable in some blocks than in others. A similar manipulation was performed in Experiment 2 but with three colors in three different shapes. Finally, in Experiment 3, two colors were presented in two different shapes, and each shape was associated with either the continuous-recall task or the visual search task. We expected that the increased allocation of resources created by the retro-cue would enhance visual search efficiency and improve memory performance.

Across three experiments, cue reliability was used to manipulate resource allocation. In Experiments 1 and 3, there were two memory items, and the cue was less reliable in some blocks than in others. A similar manipulation was performed in Experiment 2, but with three items to memorize. Experiment 3 extended this by using a single-target search task where participants maintained an attentional template for the search task along with a representation for a memory task.

Experiment 1

The aim of Experiment 1 was to demonstrate that participants allocate resources to the cued target depending on the retro-cue's reliability. When the retro-cue is highly reliable, we expect that participants prioritize the cued representation more strongly than when it is less reliable, showing that search efficiency is parametrically influenced by the retro-cue reliability.

We assume that response times (RTs) in the search task and recall accuracy in the memory task improve depending on the amount of resources allocated to the probed stimulus based on its priority (i.e., higher-priority items receive more resources). In line with the resource hypothesis, we expect shorter responses times and smaller recall errors after a valid than after an invalid retro-cue. More importantly, we expect the difference between valid and invalid conditions to increase with the reliability of the retro-cue. That is, with a retro-cue reliability of 50%, the difference between valid and invalid retro-cues is expected to be smaller than with a retro-cue reliability of 70%. The reason is that the allocation of resources should follow the relevance of the attentional template. If the cue reliability is higher, the retro-cued template becomes more relevant, and more resources are allocated to it. As a result, performance should differ more strongly between valid and invalid cues in both the search and the memory task.

Methods

Transparency and openness

Experiments' preregistrations are available at:

- https://osf.io/tgbdc (E1)
- https://osf.io/jnwvt (E2)
- https://osf.io/js9yp (E3)

The cleaned data, experimental scripts, analysis code and colors used are available on OSF at:

https://osf.io/q8d47/. Data were analyzed using CPython 3.12.4 with the following packages: *numpy* 2.0.0 (Harris et al., 2020), seaborn 0.13.2 (Waskom, 2021), *pingouin* 0.5.4 (for the repeated-measures ANOVAs; Vallat, 2018), *matplotlib* 3.9.1 (Hunter, 2007) and *pandas* 2.2.3 (The pandas development team, 2024). To apply the swap model (Bays et al., 2009) to the data, we used the MemToolbox 1.1.0 (Suchow et al., 2013) on MATLAB R2017a. The experiment was programmed with PsychoPy version 2024.2.1 (Peirce et al., 2019). Post-hoc analyses were conducted using R (version 4.5.0; R Core Team, 2025) with the effectsize package version 1.00 (Ben-Shachar et al., 2020). Chronologically, we conducted Experiment 1 first, then Experiment 3 and finally Experiment 2. Experiments were conducted in 2024 and 2025.

Participants

32 participants took part in the experiment to acquire class credits. Six participants were excluded because they did not meet the inclusion criteria specified in our preregistration (for more information on inclusion criteria, see procedure section). We decided to exclude one participant from all analyses (and not only from the model-fit analysis) due to a guess rate of .78. The remaining 25 participants (age: M = 22.0 years, SD = 5.8, 4 males) all reported normal or corrected-to-normal vision. A power sensitivity analysis with G*Power 3.1 (Faul et al., 2007) revealed that 24 participants would allow us to detect effect sizes as small as Cohen's $d_z = 0.60$ ($\alpha = .050$, power = .80). Although the initial plan, as stated in the pre-registration, specified a sample of 24 participants, the final dataset comprised 25 individuals, as the application of exclusion criteria left one more eligible participant than expected. Ethical approval for the study was granted by the Ethics Committee of the Faculty of Psychology and Educational Sciences (CUREG n° 20231004-331-2). All participants provided informed consent prior to the start of the experiment.

Apparatus

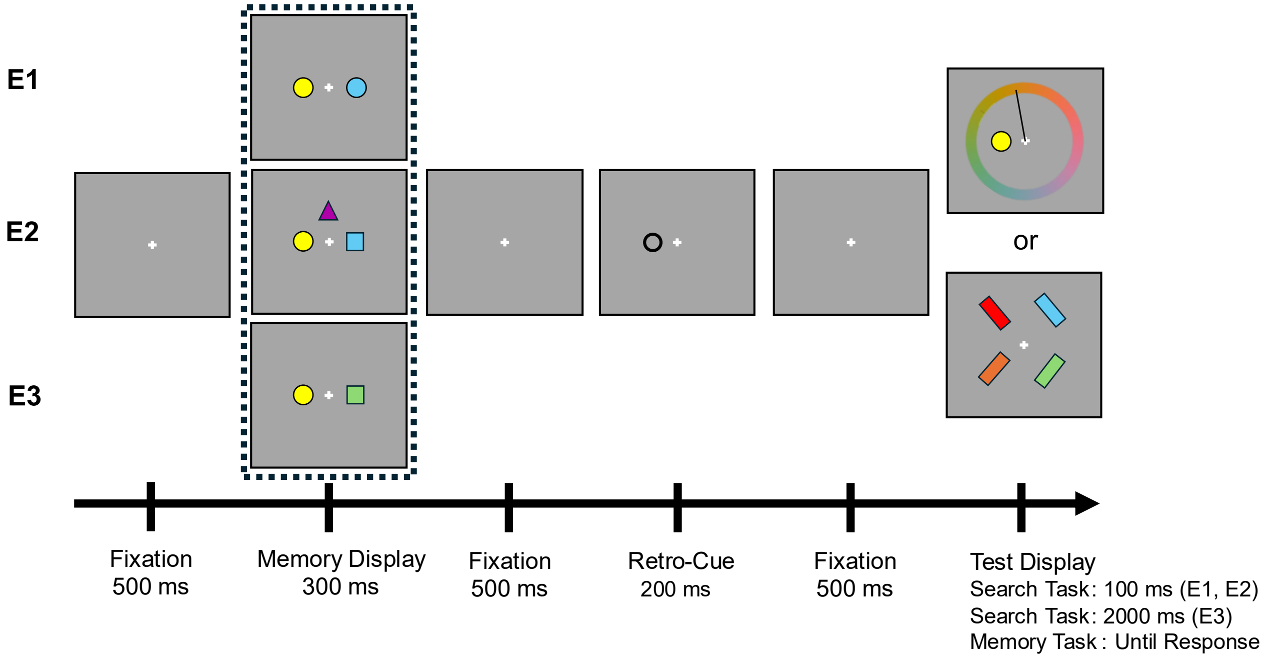
The stimuli were presented on a 22.5-inch IPS-LCD monitor (VIEWPixx Lite monitor 100 Hz, 1,920 × 1,200 pixels, standard backlight; VPixx Technologies Inc., Saint-Bruno, Canada). Colors were measured with an i1Display Pro (VPixx Edition) colorimeter by X-Rite (Grand Rapids, Michigan, United States) and color coordinates are available in the OSF repository. Head position was stabilized with a chin/forehead rest at a viewing distance of 66 cm. Responses were collected using a standard wired mouse with two buttons.

Stimuli

Figure 1 illustrates the stimuli and shows the time course of a trial. Stimuli were presented on a light gray background (15.88 cd/m²) and a white fixation cross measuring 0.18 degrees of visual angle (dva) was shown in the center of the display. Five distinct colors were needed in each trial. We describe colors by their angular distance on the color wheel. The color wheel consisted of 360 colors drawn from the CIELAB colorspace, (L* = 63, center a* = 9, center b* = 27, radius = 40) and were converted to RGB using the cielab2rgb function in PsychoPy using the sRGB transfer function. Five colors were selected on each trial: one initial color was randomly selected and based on this one, four others each spaced 72° apart were selected. Each color was then jittered by up to ±6°, introducing slight variations while preserving similar color differences. In each trial, randomly, two of the five colors were presented in the memory display, but only one of these was later probed, either in the search or the memory task. The remaining colors were used for the three non-targets in the search task.

The remaining colors were used for the three non-targets in the search task. Each trial started with the fixation cross presented alone for 500 ms. Then the memory display consisting of two-colored disks (0.37 dva radius; 0.43 dva² area), each outlined by a 0.5 pixels black outline and positioned 1.5 dva to the left and right of the fixation cross (center-to-center), appeared for 300 ms. The retro-cue appeared 500 ms after the offset of the memory display and stayed for 200 ms. It was a black unfilled circle (0.6 dva radius) shown at the location of one of the two disks in the memory display. Another 500-ms interval separated the offset of the retro-cue and the onset of one of two response displays. In the search task, the response display appeared for 100 ms and consisted of four rectangular bars (measuring 1.30 x 0.33 dva) that were displayed in the corners of a virtual square at 2 dva from fixation. Two bars were tilted 45° to the left and two 45° to the right. The target bar shared its colors with a stimulus from the memory display and the other bars were in colors unseen that trial. Participants had to indicate the orientation of the target bar by clicking the left (for left-tilted) or right (for right-tilted) mouse button.

Figure 1. *Experimental procedure.*



Note. E1, E2 and E3 refer to Experiments 1, 2, and 3. Stimulus sizes and colors were adjusted for clarity. In all experiments, the retro-cue matched the position and shape of a memory item. On each trial, participants performed either a memory task (selecting the remembered color on a color wheel) or a search task (reporting the tilt of the bar matching the color of one item from the memory display). In Experiment 3, each target shape was associated with a task: the square was associated with the search task and the disk with the memory task.

In the memory task, an empty circle was presented at one of the two positions from the memory display. Participants had to select the color previously shown at that location. The circle was surrounded by a color wheel with an inner radius of 4.5 dva and an outer radius of 5.5 dva. The color selected by mouse movement was highlighted by a line between the central fixation mark and the color wheel. In addition, the outline circle was dynamically filled with the currently selected color. Participants confirmed their selection by pressing the left mouse button. On each trial, the orientation of the color wheel was randomly selected among multiples of 30° to avoid motor biases. This response display stayed until a response was made.

Procedure

Participants were invited to the laboratory twice and received one course credit for each session. During the first session, the task was explained to them, and they carried out a preliminary session consisting of 4 blocks of 40 trials each. Cue validity in the first two blocks was either 50% or 70%, counterbalanced across participants, and changed for the last two blocks. Participants were informed of the retro-cues' reliability at the start of each block and were asked to remember two colors. At the end of this session, participants' performance was reviewed. To continue with the main experiment, participants had to reach an accuracy above 80% in each condition of the search task and, in each condition of the memory task, an average of less than 60° of absolute angular deviation (henceforth: recall error) between the selected color and the target color. If they reached the criterion, participants completed 320 experimental trials (4 blocks of 80 trials each) and were scheduled for a follow-up session of 640 trials (4 blocks of 160 trials each). Overall, participants completed 960 trials. Participants who did not meet the criterion were dismissed but still received the credit for the full session. Another seven participants completed the first part of the experiment but did not come back for the follow-up session for unknown reasons. Their data were not included in the analyses.

Design

We employed a 2 × 2 repeated-measures design. The two within-participant factors were retro-cue reliability (50%, 70%) and retro-cue validity (valid, invalid). The retro-cue either correctly indicated the target (valid) or indicated the non-target (invalid). We use the term reliability to refer to the probability that the retro-cue was valid. In other words, a 70%-reliable cue is a cue that is valid in 70% of the trials. Retro-cue reliability changed between blocks of 160 trials in the first session and between blocks of 320 in the second session. The retro-cue reliability of the first block was counterbalanced across participants. In a block of trials, the search and the memory task as well as the possible target locations were equally likely to occur. Overall, before any trial exclusion there were 480 trials for each task (search or memory). Of these 480 trials, there were 120 valid and 120 invalid in the 50%-reliability condition, and 168 valid and 72 invalid in the 70%-reliability condition.

Analysis

The repeated-measures ANOVAs reported for Experiments 1, 2 and 3 used the same design. That is, they were 2 x 2 ANOVAs with validity and reliability as factors and varied only by their dependent variables.

For post-hoc analyses, all p values reported in the text were corrected using the Holm–Bonferroni method (Holm, 1979) to control for multiple comparisons. We also report Cohen's d_z (Cohen, 1988), with 95% confidence intervals calculated using the standard parametric method based on the noncentral t distribution (Cumming, 2013).

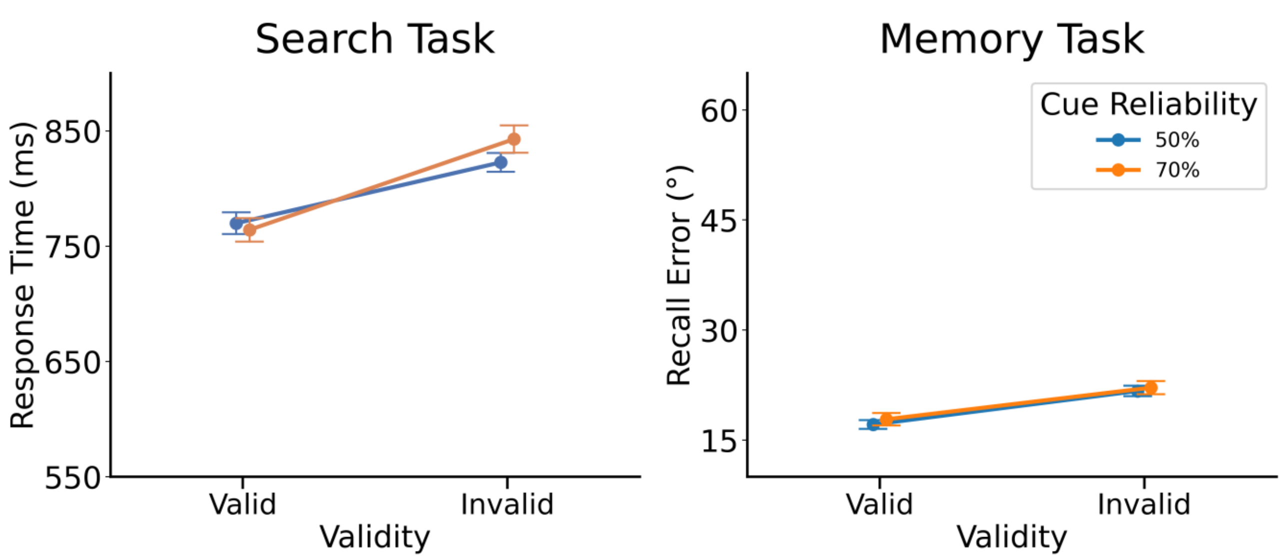
Results

RTs in the search task

Deviating from the preregistrations, we excluded trials with RTs shorter than 200 ms or longer than 2000 ms (3% of total trials) and trials with choice errors (8% of total trials). Next, as specified in the preregistration, we removed trials with RTs deviating more than 2.5 *SD* from the respective condition mean (3% of total trial). Although the preregistration mentioned the analysis of both median and average RTs, we ultimately decided to analyze only average RTs (Figure 2, left). For completeness, we nonetheless report the results of median-based analyses in the OSF repository. All median-derived results are similar to the mean-derived results.

The 2 x 2 ANOVA revealed that the main effect of retro-cue validity, $F(1, 24) = 25.61$, $p < .001$, $\eta_p^2 = .516$, was modulated by an interaction between retro-cue validity and reliability, $F(1, 24) = 7.30$, $p = .012$, $\eta_p^2 = .233$, showing that the difference between valid and invalid cues was smaller with 50% than 70% cue

Figure 2. *Results from Experiment 1.*



Note. Left panel shows average RTs (expressed in ms) as a function of cue reliability and validity. Right panel shows average recall error (expressed in degrees) as a function of cue reliability and validity. Error bars represent 95% within-participant *CIs*, calculated using Cousineau's (2005) method with Morey's (2008) correction.

validity (53 vs. 79 ms difference). To further investigate the interaction, we conducted post hoc paired-samples *t* tests. These *t* tests and *p* values from post-hoc comparisons were corrected for multiple testing using the Holm–Bonferroni method (Holm, 1979). The correction was applied separately for each experiment, across all post hoc comparisons relevant to the same dependent variable (e.g., RTs in the search task, recall error, etc.). Therefore, all these *p* values are corrected for a family of 2 comparisons.

The difference between valid and invalid cues was significantly different from zero for both levels of cue reliability (50% cue reliability: 764 ms vs. 843 ms, $t(24) = 4.70$, $p_{\text{holm}} < .001$, $d_z = 0.94$, 95% *CI* [0.47, 1.41]; 70% cue reliability: 770 ms vs. 823 ms, $t(24) = 4.91$, $p_{\text{holm}} < .001$, $d_z = 0.98$ [0.50, 1.46]). There was no main effect of retro-cue reliability, $F(1, 24) = 0.50$, $p = .488$, $\eta_p^2 = .020$.

Recall error in the memory task

For the analysis of the color wheel task, our dependent variable was the recall error. Deviating from the preregistrations, we only analyzed the average recall error (Figure 2, right). There was a main effect of validity, $F(1,24) = 17.51$, $p < .001$, $\eta_p^2 = .422$, showing that recall errors were lower with valid than invalid cues (17.48° vs. 21.93°). Contrary to our expectations, there was no interaction between retro-cue validity and reliability, $F(1, 24) = 0.05$, $p = .826$, $\eta_p^2 = .002$. The main effect of retro-cue reliability was also not significant, $F(1,24) = 0.94$, $p = .342$, $\eta_p^2 = .038$.

Exploratory analyses

Accuracy in the search task

We analyzed accuracy in the search task to ensure that potential differences in memory performance could not be attributed to speed-accuracy tradeoffs. Accuracy was lower with invalid than valid retro-cues (0.90 vs. 0.95), $F(1, 24) = 10.18$, $p = .004$, $\eta_p^2 = .298$. However, there was no main effect of retro-cue reliability, $F(1, 24) = 0.34$, $p = .566$, $\eta_p^2 = .014$, and no interaction between retro-cue reliability and validity, $F(1, 24) = 0.60$, $p = .445$, $\eta_p^2 = .025$.

RTs in the memory task

To explore potential factors influencing judgments in the color wheel task, we analyzed the RTs between the onset of the color wheel and the mouse click to confirm the selected color. We retained only values between 500 and 4000 ms (without applying a standard deviation-based exclusion criterion), which eliminated 6% of trials. The interaction between validity and retro-cue reliability, $F(1, 24) = 5.14$, $p = .033$, $\eta_p^2 = .177$ and the main effect of validity, $F(1, 24) = 18.13$, $p < .001$, $\eta_p^2 = .430$ were significant.

We performed post hoc tests on the interaction levels and the difference between invalid and valid cues was smaller with 50% reliability than with 70% reliability (50 vs. 112 ms difference). The difference between invalid and valid at 50% reliability approached significance (2135 vs. 2085 ms), $t(24) = 1.94$, $p_{\text{holm}} = .065$, $d_z = 0.39$ [–0.02, 0.79] but was significant at 70% reliability (2152 vs. 2040 ms), $t(24) = 5.43$, $p_{\text{holm}} < .001$, $d_z = 1.09$ [0.59, 1.58]. The main effect of retro-cue reliability was not significant, $F(1, 24) = 0.11$, $p = .746$, $\eta_p^2 = .004$.

Swap model

We submitted the memory task's results to Bays et al. (2009) mixture-swap model in order to disentangle the sources of error. Note that we used Bays et al. (2009) mixture-swap model, but that other swap models exist (e.g., Bays, 2016; Williams et al., 2022; see Williams et al., 2023 for a comparison of these models). P_{guess} reflects the proportion of random guesses, *SD* reflects the precision of responses directed toward the correct target and P_{swap} reflects the proportion of responses corresponding to the non-probed item. We applied the swap model to move beyond a simple analysis of overall error magnitude and to dissociate the potential underlying sources of error.

This decomposition was critical to determine whether our experimental manipulations specifically affected memory precision, the likelihood of item confusion, or the propensity to guess randomly. In particular, an effect on *SD* indicates a modulation of the quality of memory representations, consistent with flexible resource allocation. An effect on P_{swap}, on the other hand, reflects changes in the probability of selecting the correct item, suggesting differences in selection or binding. These parameters were fit with the MemToolbox (Suchow et al., 2013).

For guess rates (P_{guess}), there was no main effect of the cue reliability, $F(1, 24) = 0.09$, $p = .762$, $\eta_p^2 = .004$, nor was there a main effect of validity, $F(1, 24) = 1.94$, $p = .177$, $\eta_p^2 = .075$. The interaction between the cue reliability and the validity was also not significant, $F(1, 24) = 0.005$, $p = .944$, $\eta_p^2 < .001$. Mean guess rate was .048.

For swap rates (P_{swap}), there was no main effect of the retro-cue reliability, $F(1, 24) = 1.56$, $p = .224$, $\eta_p^2 = .061$, but swap rated were higher with invalid than with valid retro-cues (.045 vs .013), $F(1, 24) = 5.70$, $p = .025$, $\eta_p^2 = .192$. The interaction between the cue reliability and validity was not significant, $F(1, 24) = 0.15$, $p = .706$, $\eta_p^2 = .006$. Mean swap rate was .029.

For the *SD*, there was no main effect of cue reliability, $F(1, 24) = 0.73$, $p = .400$, $\eta_p^2 = .030$, but *SD*s were higher with invalid than valid retro-cues (21 vs. 17), $F(1, 24) = 9.87$, $p = .004$, $\eta_p^2 = .291$. The interaction between cue reliability and validity was not significant, $F(1, 24) = 1.75$, $p = .198$, $\eta_p^2 = .068$. Mean *SD* was 18.2.

Discussion

In Experiment 1, we investigated the effect of retro-cue reliability on visual search and memory performance. In the search task, RTs were shorter when the retro-cue was valid compared to when it was invalid. Moreover, the interaction between retro-cue reliability and validity revealed that the difference in search RTs between valid and invalid cues was larger when cue reliability was high (70%) than when it was low (50%). This finding suggests that participants adapted their resource allocation based on the expected informativeness of the cue, effectively prioritizing relevant memory representations during the upcoming search task. This interaction also underscores that resource allocation in visual search is not fixed but flexibly adapts to the expected reliability of the attentional template.

In contrast, in the memory task, we did not observe an interaction between cue validity and cue reliability despite the clear main effect of cue validity. One plausible explanation is that the memory task was too easy, resulting in a ceiling effect that may have limited the detection of subtle reliability-based modulations. Supporting this assumption, we found a significant interaction in the RTs recorded during the color wheel task, although the effects were smaller than in the search task. Therefore, we conducted Experiment 2, which involved a more challenging search task that should help avoid ceiling effects.

Experiment 2

Experiment 2 was similar to Experiment 1 but more difficult because three colors had to be remembered. Accordingly, the retro-cue reliability changed from 50% to 33% for unpredictive cues but stayed at 70% for predictive cues. Moreover, instead of being presented on two disks, the three colors were presented on three different shapes (disk, square and triangle). We expected performance on both the search and memory tasks to be better following valid cues compared to invalid cues. Critically, we predicted that the difference between valid and invalid trials would be larger in blocks with 70% compared to the 33% retro-cue reliability.

Methods

Participants

In total, 56 participants took part in the experiment to acquire class credits. 31 participants were excluded because they did not meet the inclusion criteria (see Experiment 2 procedure) leaving 25 participants (age: $M = 21.2$ years, $SD = 4.6$, 3 males) for analyses.

Stimuli and procedure

The stimuli and procedure were as in Experiment 1 with the following changes. In the memory display, a disk (radius of 0.37 dva), a square (side length of 0.66 dva), and an equilateral triangle (side length of 1.00 dva) were presented at 2.50 dva of eccentricity, spaced 120° apart around a circular layout. The stimulus positions were fixed on the circular layout, but the shape assigned to each location varied across trials. The search display was presented for 2,000 ms instead of 100 ms in Experiment 1. We also used six instead of five colors, which were therefore 60° apart and jittered by ±6°. The reliability of retro-cues was 33% and 70% in separate blocks of trials. Therefore, out of the 480 trials for each task, there were 80 valid and 160 invalid trials in the 33% cue reliability condition, and 168 valid and 72 invalid trials in the 70% cue reliability condition. Finally, the shape of the retro-cue matched the shape of the cued location. Each shape and each location were cued and probed in one third of the trials.

To be included in Experiment 2, participants were required to keep mean accuracy above 70% in each condition of the search task, which is less strict than in Experiment 1. Additionally, they were required to keep the mean recall error below 60° in each condition of the memory task, which is similar to Experiment 1.

As mentioned in the preregistration, we expected RTs and recall errors to be better in valid compared to invalid conditions, reflecting better performance when attention was effectively oriented. Critically, we hypothesized that these validity effects would interact with cue reliability. Specifically, we anticipated stronger effects of validity with high compared to low retro-cue reliability.

Results

RTs in the search task

As in Experiment 1, we excluded trials with RTs shorter than 200 ms or longer than 2000 ms (0.02% of trials) and trials with choice errors (21% of trials). Next, we removed trials with RTs deviating more than 2.5 *SD* from the respective condition mean (3% of trials). Figure 3 shows average RTs.

There was a main effect of retro-cue validity, $F(1,24) = 85.04$, $p < .001$, $\eta_p^2 = .780$, as well as an interaction effect between retro-cue validity and reliability, $F(1, 24) = 27.25$, $p < .001$, $\eta_p^2 = .532$. The interaction shows that the difference between invalid and valid cues was larger with 70% retro-cue validity (790 vs. 640 ms), $t(24) = 9.08$, $p_{holm} < .001$, $d_z = 1.82$ [1.18, 2.46], than with 33% (748 vs. 671 ms), $t(24) = 6.73$, $p_{holm} < .001$, $d_z = 1.35$ [0.81, 1.89]. There was no main effect of cue reliability, $F(1, 24) = 0.59$, $p = .450$, $\eta_p^2 = .024$.

Recall error in the memory task.

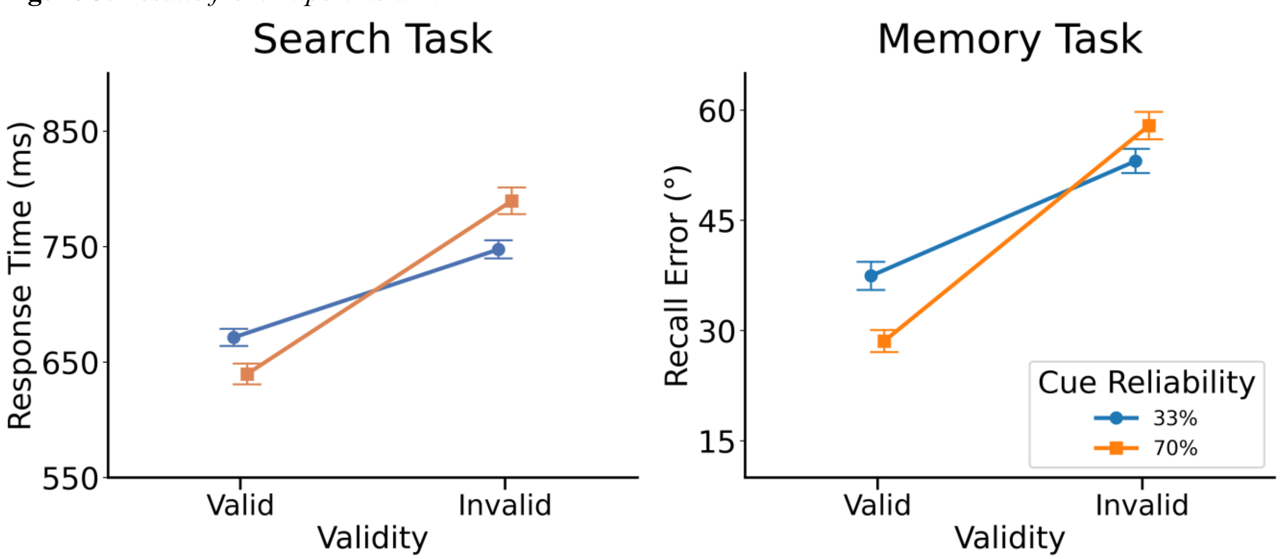
The average recall errors are also shown in Figure 3. There was a main effect of retro-cue validity, $F(1, 24) = 77.36$, $p < .001$, $\eta_p^2 = .763$, as well as an interaction between retro-cue validity and reliability, $F(1, 24) = 38.93$, $p < .001$, $\eta_p^2 = .619$. The difference between invalid and valid trials was significant with 33% reliability (53° vs. 37°), $t(24) = 5.44$, $p_{holm} < .001$, $d_z = 1.09$ [0.59, 1.58] but was larger with 70% reliability (58° vs. 29°), $t(24) = 10.92$, $p_{holm} < .001$, $d_z = 2.18$ [1.46, 2.91]. The main effect of retro-cue reliability was not significant, $F(1, 24) = 3.17$, $p = .088$, $\eta_p^2 = .117$.

Exploratory analyses

Accuracy in the search task

Similar to Experiment 1, we explored accuracy in the search task to control for potential differences in overall task performance across conditions. There was a main effect of retro-cue validity, $F(1, 24) = 89.04$, $p < .001$, $\eta_p^2 = .788$, which was modulated by an interaction between retro-cue validity and reliability, $F(1, 24) = 19.93$, $p < .001$, $\eta_p^2 = .454$.

Figure 3. *Results from Experiment 2.*



Note. Left panel shows average RTs (expressed in ms) as a function of cue reliability and validity. Right panel shows average recall error (expressed in degrees) as a function of cue reliability and validity. Error bars represent 95% within-participant *CI*s, calculated using Cousineau's (2005) method with Morey's (2008) correction.

The interaction shows that the difference between invalid and valid cues was larger with 70% cue validity (.68 vs. .87), $t(24) = -9.56$, $p_{\text{holm}} < .001$, $d_z = -1.91$ [–2.57, –1.25], than with 33% (0.74 vs. 0.82), $t(24) = -4.92$, $p_{\text{holm}} < .001$, $d_z = -0.98$ [–1.46, –0.51]. There was no main effect of retro-cue reliability, $F(1, 24) = 1.34$, $p = .259$, $\eta_p^2 = .053$.

RTs in the memory task

As in Experiment 1, we analyzed RTs to explore whether retro-cue reliability and validity influenced response speed during the color recall task. We again retained only values between 500 and 4000 ms, which eliminated 2% of trials. There was an interaction between retro-cue validity and reliability, $F(1, 24) = 32.66$, $p < .001$, $\eta_p^2 = .576$. That is, the difference between invalid and valid was larger with 70% reliability (1747 vs. 1420 ms), $t(24) = 11.98$, $p_{\text{holm}} < .001$, $d_z = 2.40$ [1.63, 3.17], than with 33% reliability (1613 vs. 1454 ms), $t(24) = 6.80$, $p_{\text{holm}} < .001$, $d_z = 1.36$ [0.82, 1.90]. There was also a main effect of validity, $F(1, 24) = 141.78$, $p < .001$, $\eta_p^2 = .855$, but the main effect of retro-cue reliability was not significant, $F(1, 24) = 2.99$, $p = .097$, $\eta_p^2 = .111$.

Swap Model

For guess rates (P_{guess}), and similar to the results of Experiment 1, there were no main effects of retro-cue reliability, $F(1, 24) = 0.02$, $p = .878$, $\eta_p^2 = .001$, validity, $F(1, 24) < 0.01$, $p = .983$, $\eta_p^2 < .001$, and no interaction between the retro-cue reliability and validity, $F(1, 24) = 0.17$, $p = .687$, $\eta_p^2 = .007$. Mean guess rate was .290.

For swap rates (P_{swap}), there was no main effect of retro-cue reliability, $F(1, 24) = 3.58$, $p = .071$, $\eta_p^2 = .130$ and, different from Experiment 1, no main effect of validity either, $F(1, 24) = 4.01$, $p = .057$, $\eta_p^2 = .143$. The interaction between the retro-cue reliability and validity was also not significant, $F(1, 24) = 2.10$, $p = .161$, $\eta_p^2 = .080$. Mean swap rate was .310. Taken together with the guess rate (.290), this suggests that approximately 60% of responses did not correspond to the target item, either due to random guessing or confusions with non-target items.

Concerning the *SD*, there were no effects of retro-cue reliability, $F(1, 24) = 0.37$, $p = .547$, $\eta_p^2 = .015$, or validity, $F(1, 24) = 0.06$, $p = .804$, $\eta_p^2 = .003$. However, the interaction between cue reliability and validity was significant, $F(1, 24) = 6.03$, $p = .022$, $\eta_p^2 = .201$. However, the difference between valid and invalid trials was not significant at 33% retro-cue reliability (24.9 vs. 25.7), $t(24) = -0.91$, $p_{\text{holm}} = .371$, $d_z = -0.18$ [–0.56, 0.21], nor at 70% cue reliability (26.1 vs. 25.1), $t(24) = 1.90$, $p_{\text{holm}} = .130$, $d_z = 0.37$ [–0.02, 0.76].

Discussion

Experiment 2 was designed to address a limitation of Experiment 1, where the lack of evidence for an interaction between cue validity and reliability in the memory task raised the possibility of a ceiling effect. We reasoned that the relatively low task demands in Experiment 1 may have limited our ability to detect subtle effects of cue reliability. By increasing the difficulty of the memory task, we aimed to reveal modulations in memory performance based on cue reliability.

In the search task, trials with valid retro-cues resulted in significantly faster RTs compared to invalid cues. Notably, there was no evidence for an effect of retro-cue reliability alone, suggesting that the reliability factor exerts its influence primarily through its interaction with cue validity. Importantly, an interaction between retro-cue reliability and validity was observed. This interaction indicates that the RT difference between valid and invalid cues was larger when retro-cue reliability was high: at 70% reliability, valid cues improved RTs by 150 ms on average while they only improved by 77 ms at 33% reliability. Interestingly, these results imply that even non-predictive retro-cues can facilitate faster responses when valid, but that this benefit is substantially amplified as the cues become more reliable.

For the memory task, the results similarly revealed a clear impact of retro-cue validity on recall error. Specifically, valid retro-cues resulted in significantly lower recall errors compared to invalid cues. At 33% reliability, the average recall error was 37° for valid trials versus 53° for invalid ones, while at 70% reliability, the error further decreased to 29° for valid trials and increased to 58° for invalid ones. Again, although the main effect of cue reliability alone was not significant, the interaction between cue validity and reliability was significant, indicating that the benefit of valid retro-cues on recall accuracy was enhanced when retro-cues were more reliable. This pattern of results reinforces the conclusion that retro-cues not only direct attention effectively but also significantly enhances the accuracy of VWM when deemed reliable.

These results indicate that resources are flexibly allocated, showing that the attentional and memory benefits from valid retro-cues strongly depend on the reliability of the retro-cue. Hence, our results favor a model with a more flexible allocation of working memory resources rather than one with an all-or-none activation mechanism.

Experiment 3

In Experiments 1 and 2, one of the two colors in the memory display was randomly selected for either the memory or the search task, and participants did not know in advance which would be probed. This uncertainty may have required them to maintain both items in a state capable of guiding attention, potentially increasing competition and leading to more diffuse allocation of memory resources. In Experiment 3, we reduced this ambiguity by introducing fixed associations between item shape and task identity, such that one item served the search task and the other the memory task. This allowed us to distinguish between an attentional template and a non-template representation.

According to the resource hypothesis, both types of representations can receive flexible amounts of memory resources based on their expected relevance. By manipulating retro-cue reliability for both items, we tested whether this flexibility extends equally to attentional templates and simple representations. If so, both should show similar patterns of cue validity effects as a function of cue reliability, supporting the view that graded resource allocation operates across different types of memory representations.

Methods

Participants

In total, 30 participants took part in the experiment to acquire class credits. Six participants were excluded because they did not reach the inclusion criterion for the experiment, which was the same as in Experiment 1. The remaining 24 participants (age: *M* = 22.6 years, *SD* = 9.2, 3 males) all reported normal or corrected-to-normal vision.

Stimuli and Procedure

The stimuli and procedure were similar to Experiment 2, with a few modifications. There were only two shapes, a disk and a square that were presented 1.5 dva to the left and right of the fixation cross. Crucially, the square was associated with the search task and the disk with the memory task. In other words, when the search task occurred, the target shared its color with the square and, when the memory task occurred, the to-be-remembered color was always that of the disk. Moreover, since there were only two targets, the cue reliabilities were similar to Experiment 1 (50% and 70%).

Results

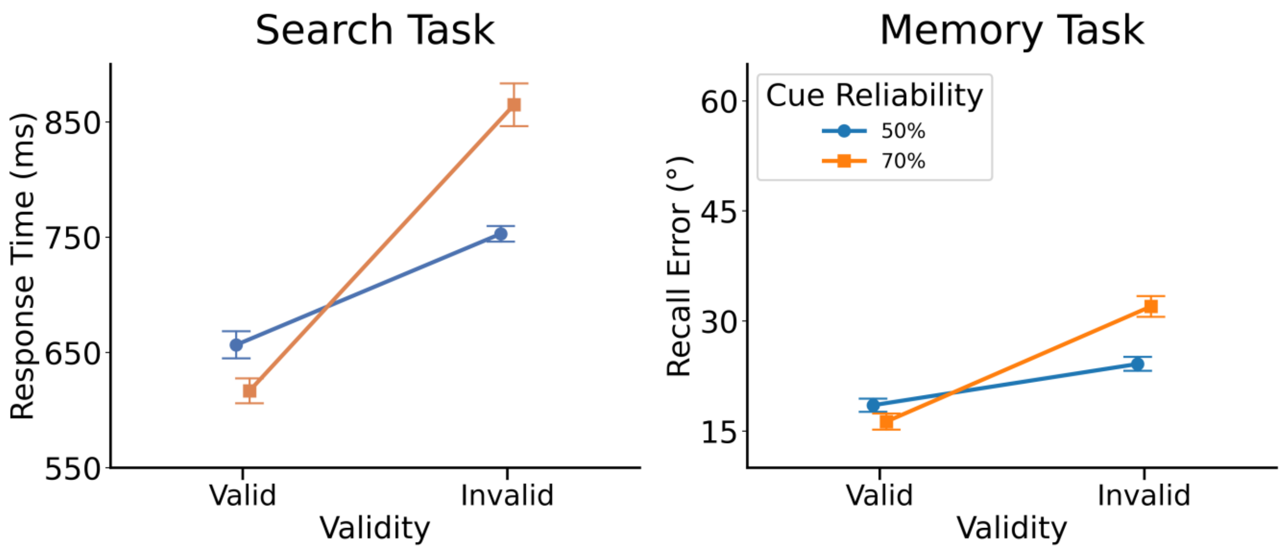
RTs in search task

As in Experiment 1 and 2, trials with anticipations or late responses (2% of the trials) were excluded. Incorrect responses accounted for 9% of the trials. We also removed trials with RTs deviating more than 2.5 *SD* from the respective condition mean (3% of total trial). Figure 4 shows average RTs. There was a main effect of validity, $F(1, 23) = 110.41$, $p < .001$, $\eta_p^2 = .828$, and an unexpected main effect of cue reliability, $F(1, 23) = 10.24$, $p = .004$, $\eta_p^2 = .308$. There was also an interaction effect between cue reliability and validity, $F(1, 23) = 65.37$, $p < .001$, $\eta_p^2 = .740$, showing that the difference between valid and invalid cues was smaller with 50% retro-cue validity than with 70% (96 vs. 248 ms). The difference between invalid and valid trials was significant with 50% reliability (753 vs. 657 ms), $t(23) = 7.85$, $p_{holm} < .001$, $d_z = 1.60$ [1.00, 2.21], and with 70% reliability (865 vs. 617 ms), $t(23) = 10.46$, $p_{holm} < .001$, $d_z = 2.13$ [1.41, 2.86].

Recall error in the memory task

Figure 4 shows average recall error. There was a main effect of validity, $F(1, 23) = 42.45$, $p < .001$, $\eta_p^2 = .649$, an unexpected main effect of cue reliability, $F(1, 23) = 14.12$, $p = .001$, $\eta_p^2 = .380$, and an interaction between retro-cue validity and reliability, $F(1, 23) = 56.18$, $p < .001$, $\eta_p^2 = .710$, demonstrating that the difference between valid and invalid cues was smaller with 50% retro-cue validity than with 70% (5° vs. 16°). The difference between invalid and valid trials was significant with 50% reliability (24° vs. 19°), $t(23) = 3.99$, $p_{holm} < .001$, $d_z = 0.81$ [0.35, 1.28], and with 70% reliability (32° vs. 16°), $t(23) = 7.60$, $p_{holm} < .001$, $d_z = 1.55$ [0.96, 2.14].

Figure 4. *Results from Experiment 3.*



Note. Left panel shows average RTs (expressed in ms) as a function of cue reliability and validity. Right panel shows average recall error (expressed in degrees) as a function of cue reliability and validity. Error bars represent 95% within-participant *CIs*, calculated using Cousineau's (2005) method with Morey's (2008) correction.

Exploratory analyses

Accuracy in the search task

As in previous experiments, we analyzed accuracy in the search task to ensure that RTs were not traded off against response accuracy. There was a main effect of validity, $F(1, 23) = 47.27$, $p < .001$, $\eta_p^2 = .673$, but no main effect of cue reliability, $F(1, 23) = 7.74$, $p = .011$, $\eta_p^2 = .252$. The interaction between retro-cue validity and reliability was significant, $F(1, 23) = 27.10$, $p < .001$, $\eta_p^2 = .541$, revealing that the difference between valid and invalid cues was smaller with 50% retro-cue validity than with 70% (0.07 vs. 0.13). The difference between valid and invalid trials was significant with 50% reliability (0.95 vs. 0.88), $t(23) = –5.38$, $p_{holm} < .001$, $d_z = –1.10$ [–1.61, –0.59] and with 70% reliability (0.96 vs. 0.83), $t(23) = –7.35$, $p_{holm} < .001$, $d_z = –1.50$ [–2.08, –0.92].

RTs in the memory task

We analyzed RTs in the color wheel task to explore whether retro-cue reliability and validity influenced the speed of memory-based responses. As before, we excluded trials with RTs below 500 and above 4000 ms (5% of trials). RTs were shorter with valid than invalid retro-cues (1844 ms vs. 2008 ms), $F(1, 23) = 58.76$, $p < .001$, $\eta_p^2 = .719$. The main effect of cue reliability, $F(1, 23) = 0.57$, $p = .459$, $\eta_p^2 = .024$, was not significant, but the interaction was, $F(1, 23) = 44.02$, $p < .001$, $\eta_p^2 = .657$. The difference between valid and invalid cues was smaller with 50% (1921 vs. 1992 ms), $t(23) = 4.09$, $p_{holm} < .001$, $d_z = 0.84$ [0.37, 1.30], than with 70% retro-cue validity (1792 vs. 2038 ms), $t(23) = 8.43$, $p_{holm} < .001$, $d_z = 1.72$ [1.09, 2.35].

Swap Model

For guess rates (P_{guess}), there was no main effect of the retro-cue reliability, $F(1, 23) = 0.75$, $p = .394$, $\eta_p^2 = .032$, and no interaction between the retro-cue reliability and validity, $F(1, 23) = 0.41$, $p = .527$, $\eta_p^2 = .018$. However, contrary to the previous two experiments, there was a main effect of validity, $F(1, 23) = 21.53$, $p < .001$, $\eta_p^2 = .484$, indicating larger guess rates for invalid than valid retro-cues (.120 vs .040). The mean guess rate was .08.

For swap rates (P_{swap}), there was no main effect of retro-cue reliability, $F(1, 23) = 0.90$, $p = .353$, $\eta_p^2 = .038$, and the interaction between cue reliability and cue validity, $F(1, 23) = 3.76$, $p = .065$, $\eta_p^2 = .141$ was not significant either. Similar to Experiment 1, the main effect of validity was significant, $F(1, 23) = 6.84$, $p = .016$, $\eta_p^2 = .229$, suggesting larger P_{swap} for invalid than valid retro-cues (.04 vs .02). The mean swap rate was .03. Taken together with the mean guess rate (.08), this indicates that approximately 11% of responses were not attributed to the target item, either due to random guessing or confusion with a non-target.

For *SD*s, results again converge with those of Experiment 1 in that there was no effect of retro-cue reliability, $F(1, 23) = 3.52$, $p = .073$, $\eta_p^2 = .133$, but a main effect of cue validity, $F(1, 23) = 28.55$, $p < .001$, $\eta_p^2 = .554$, and an interaction between the cue reliability and cue validity, $F(1, 23) = 10.03$, $p = .004$, $\eta_p^2 = .304$. For the 50% cue reliability, invalid trials had a higher *SD* than valid trials (19 vs. 17), $t(23) = 2.55$, $p_{\mathrm{holm}} = .018$, $d_z = 0.52$ [0.09, 0.95], and this difference increased with 70% cue reliability (22 vs. 16), $t(23) = 5.29$, $p_{\mathrm{holm}} < .001$, $d_z = 1.08$ [0.58, 1.58].

Discussion

In Experiment 3, we examined how retro-cue reliability modulates performance when participants retained explicit color–task associations. In this paradigm, the retro-cue signaled not only the color that would be probed but also the task that was to be executed (i.e., a square indicated the search task, and a disk indicated the color wheel task). This dual information allowed us to investigate whether increasing cue reliability would systematically bias VWM resources toward the cued item and task.

For the search task, our results revealed a main effect of retro-cue validity indicating a benefit of valid retro-cues on search RTs. There was also a main effect of retro-cue reliability, and, critically, an interaction between cue reliability and validity. Specifically, the performance benefit for valid cues was enhanced when reliability increased: the RT difference between valid and invalid trials was larger at 70% cue reliability compared to 50%. This difference seems mainly driven by a larger cost of invalid cues in the 70% condition compared to the 50% condition, but there was also a modest increase in performance for valid trials. The interaction highlights that cue reliability plays a powerful role in dynamically modulating the prioritization of items in working memory, rather than merely boosting performance across the board.

A similar pattern emerged in the memory task. Valid retro-cues led to lower recall error than invalid retro-cues. Higher retro-cue reliability also had a significant main effect on memory precision, but both main effects were modulated by an interaction. At 70% reliability, the difference between valid and invalid trials in recall error was much more pronounced than at 50% reliability, showing that when the retro-cue was highly reliable, participants not only remembered the cued color more precisely in valid trials but also suffered a larger decrement in accuracy in invalid trials. These results confirm that increasing cue reliability systematically intensifies the bias in resource allocation in VWM.

Overall, results from both search and memory tasks are consistent with the predictions of the resource hypothesis of working memory. They suggest that retro-cue benefits are not merely about enhancing the representation of a particular color but also about dynamically reallocating limited VWM resources in accordance with task-specific demands. When cue reliability is high, the cost-benefit tradeoff becomes more extreme, improving performance in valid trials while impairing it in invalid trials. This graded modulation of performance by cue reliability contrasts with models that do not predict an influence of retro-cue reliability on VWM allocation. Thus, Experiment 3 extends our previous findings by demonstrating that the dual role of the retro-cue (indicating both stimulus identity and task relevance) leads to a more pronounced bias in VWM resources, in line with the predictions of the resource hypothesis.

General Discussion

Across three experiments, we tested whether the resource hypothesis proposed by Huynh Cong and Kerzel (2021) offers a more powerful account of how attentional templates operate in VWM compared to the state-based model put forward by Olivers et al. (2011). Overall, our findings converge on the conclusion that attentional resources can be flexibly shared among multiple items, with the amount of resources allocated depending on the probabilistic validity (i.e., relevance) of those items during search or memory tasks. This flexible, gradable allocation is not easily captured by a strict dual-state framework, which posits that only one representation can be in an active, template-like state at a time, whereas all other representations remain in an accessory state until switched into the active one.

The first experiment examined the effect of retro-cue validity and cue reliability on both a visual search task and a continuous recall task. While the results showed that participants performed better under valid cue conditions than invalid ones, the effect of cue reliability manifested itself only in the search task through its interaction with validity. This pattern aligns with the resource hypothesis (Huynh Cong & Kerzel, 2021) indicating that VWM resources are preferentially allocated to the more likely relevant representation.

We believe that the continuous-recall task in Experiment 1 was too easy, therefore leading to a potential ceiling effect that may have erased the need to rely on the cue. Therefore, to address this limitation we increased task difficulty in Experiment 2 in order to allow us to detect potential effects. As in the first experiment, participants performed significantly better under valid retro-cue conditions, and higher cue reliability produced a more pronounced advantage for valid cues. Conversely, invalid cues hindered performance more when they were informative, highlighting a selective, probability-based resource allocation. The interaction between cue reliability and validity further corroborated the resource hypothesis. This challenges the assumption of the state-based framework that only one template at a time can guide attention effectively. Instead, they suggest that multiple representations can simultaneously guide search, with performance scaled by the distribution of VWM resources (Franconeri et al., 2013; Ma et al., 2014; Huynh Cong & Kerzel, 2021).

Finally, in Experiment 3, participants had to maintain two task-relevant items in working memory: one associated with the search task and the other with the memory task, while using the retro-cue to estimate which item would most likely be tested. The known mapping between shape and task allowed participants to anticipate the role of each item and distribute their memory resources accordingly. Once again, both the search and memory tasks revealed a highly significant interaction between cue reliability and cue validity. Higher cue reliability (70%) yielded better performance under valid cues and poorer performance under invalid cues, which was also true for lower cue reliability (50%), but to a lesser degree. This pattern was observed not only in the search task but also in the memory task, supporting the view that retro-cue reliability drives a flexible and graded allocation of memory resources to representations interacting with search and representations without such interactions.

Taken together, the results from all three experiments provide converging evidence for a resource-based account of VWM, in which attentional templates are not constrained by a strict dichotomy of “active” versus “accessory” states. Instead, resources are flexibly allocated based on the probabilistic relevance of items, yielding fine-grained effects on both attention and memory performance.

Nevertheless, it is worth examining whether some aspects of our findings could also be interpreted within the dual-state framework proposed by Olivers et al. (2011). One could argue that, depending on cue reliability, participants may alternate more or less rapidly between different representations in an effort to prioritize potential targets. In this view, a cued representation may have a higher probability of being switched into the active state when cue reliability is high.

While this account may accommodate performance differences in the search task, it struggles to explain the differences observed in the memory task, particularly those related to the precision of recall. The state-based model does not include a mechanism for allocating memory resources in a probabilistic or continuous manner; rather, it assumes a discrete switch between active and accessory states. Although our design tested only two levels of cue reliability (50% vs. 70%, or 33% vs. 70%), the observed performance patterns are more consistent with resource hypothesis, which predicts flexible and continuous allocation of VWM resources based on the inferred relevance of each item.

Another limitation is that task difficulty varied simultaneously across both tasks, making it difficult to disentangle the effects of difficulty on each task independently. Moreover, we only employed two reliability levels per experiment (i.e., non-predictive vs. moderately predictive) and we did not have a neutral- or no-retro-cue condition.

In addition, while our results strongly indicate that multiple attentional templates can be maintained in parallel, further work is needed to explore the boundary conditions of this flexibility. For instance, exploring the role of perceptual similarity between targets, individual differences in VWM capacity, or the timing of retro-cue presentation could shed more light on how resources are dynamically reallocated over time.

More broadly, our findings reinforce the theoretical strength of the resource hypothesis by aligning it with contemporary frameworks of cognition. The idea that VWM templates are dynamically prioritized based on contextual probabilities is consistent with recent views proposing that attentional and working memory templates are flexible, context-sensitive representations rather than fixed entities (Kristjánsson, 2023). Our findings align with theories such as predictive coding and Bayesian inference (Chetverikov & Kristjánsson, 2022; Ma et al., 2014), which suggest that internal representations are continuously updated based on the integration of expectations and sensory evidence. From this perspective, retro-cues do not merely switch items into a fixed "active" state but serve as probabilistic signals that bias the distribution of limited cognitive resources. These insights not only advance our understanding of attentional control in VWM but also align with broader theoretical trends in cognitive science that emphasize probabilistic, flexible, and context-sensitive mechanisms of internal representation.

Constraints on Generality

The participants in our study were undergraduate psychology students from the University of Geneva, all of whom were adults, with a majority being young women. As such, the generalizability of our findings to broader age groups, educational backgrounds, or cultural contexts remains uncertain. Moreover, our experimental design employed a specific form of retro-cueing based on color-location associations and manipulated cue reliability within a limited range (33% to 70%). Different cueing modalities, feature dimensions, or levels of cue reliability might yield different results. Thus, caution is warranted when extending these conclusions to other populations, task formats, or attentional cueing paradigms.

Acknowledgements

Thanks to Anaë Motz for helping with data collection. The Swiss National Science Foundation (FNS 10001E_219517) supported the present research.

Contribution

Conceptualization: Alexandre Fortuna, Dirk Kerzel
Data curation: Alexandre Fortuna
Formal analysis: Alexandre Fortuna
Funding acquisition: Dirk Kerzel
Investigation: Alexandre Fortuna, Martin Constant, Dirk Kerzel
Methodology: Alexandre Fortuna, Martin Constant, Dirk Kerzel
Software: Alexandre Fortuna
Supervision: Martin Constant, Dirk Kerzel
Validation: Alexandre Fortuna, Martin Constant, Dirk Kerzel
Visualization: Alexandre Fortuna
Writing – original draft: Alexandre Fortuna
Writing – review & editing: Alexandre Fortuna, Martin Constant, Dirk Kerzel

References

Bahle, B., Thayer, D. D., Mordkoff, J. T., & Hollingworth, A. (2020). The architecture of working memory: Features from multiple remembered objects produce parallel, coactive guidance of attention in visual search. *Journal of Experimental Psychology: General*, *149*(5), 967–983. https://doi.org/10.1037/xge0000694

Bays, P. M. (2016). Evaluating and excluding swap errors in analogue tests of working memory. *Scientific Reports, 6*, 19203. https://doi.org/10.1038/srep19203

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 7.1-11. https://doi.org/10.1167/9.10.7

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

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, *5*(56), 2815. https://doi.org/10.21105/joss.02815

Beck, V. M., Hollingworth, A., & Luck, S. J. (2012). Simultaneous Control of Attention by Multiple Working Memory Representations. *Psychological Science*, *23*(8), 887–898. https://doi.org/10.1177/0956797612439068

Berggren, N., Nako, R., & Eimer, M. (2020). Out with the old: New target templates impair the guidance of visual search by preexisting task goals. *Journal of Experimental Psychology: General*, *149*(6), 1156–1168. https://doi.org/10.1037/xge0000697

Carlisle, N. B., Arita, J. T., Pardo, D., & Woodman, G. F. (2011). Attentional templates in visual working memory. *Journal of Neuroscience*, *31*(25), 9315–9322. https://doi.org/10.1523/JNEUROSCI.1097-11.2011

Carlisle, N. B., & Woodman, G. F. (2019). Quantifying the attentional impact of working memory matching targets and distractors. *Visual Cognition*, *27*(5–8), 452–466. https://doi.org/10.1080/13506285.2019.1634172

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, *80*(6), 2918–2940. https://doi.org/10.1152/jn.1998.80.6.2918

Chetverikov, A., & Kristjánsson, Á. (2022, November). Probabilistic representations as building blocks for higher-level vision. *Neurons, Behavior, Data Analysis, and Theory, 1*, Article 1–32. https://doi.org/10.51628/001c.24910

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203771587

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

Cumming, G. (2013). *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis* (1st ed.). Routledge. https://doi.org/10.4324/9780203807002

Downing, P., & Dodds, C. (2004). Competition in visual working memory for control of search. *Visual Cognition*, *11*(6), 689–703. https://doi.org/10.1080/13506280344000446

Dube, B., & Al-Aidroos, N. (2019). Distinct prioritization of visual working memory representations for search and for recall. *Attention, Perception, & Psychophysics*, *81*(5), 1253–1261. https://doi.org/10.3758/s13414-018-01664-6

Dube, B., Lumsden, A., & Al-Aidroos, N. (2019). Probabilistic retro-cues do not determine state in visual working memory. *Psychonomic Bulletin & Review*, *26*(2), 641–646. https://doi.org/10.3758/s13423-018-1533-7

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458. https://doi.org/10.1037/0033-295X.96.3.433

Eimer, M. (2014). The neural basis of attentional control in visual search. *Trends in Cognitive Sciences*, *18*(10), 526–535. https://doi.org/10.1016/j.tics.2014.05.005

Fan, L., Ding, C., Guo, R., Xu, M., Diao, L., & Yang, D. (2016). Visual working memory representations guide the detection of emotional faces: An ERP study. *Vision Research*, *119*, 1–8. https://doi.org/10.1016/j.visres.2015.08.021

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Foerster, R. M., & Schneider, W. X. (2018). Involuntary top-down control by search-irrelevant features: Visual working memory biases attention in an object-based manner. *Cognition, 172*, 37–45. https://doi.org/10.1016/j.cognition.2017.12.002

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology. Human Perception and Performance*, *18*(4), 1030–1044. https://doi.org/10.1037/0096-1523.18.4.1030

Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources: competitive content maps for attention and memory. *Trends in Cognitive Sciences*, *17*(3), 134–141. https://doi.org/10.1016/j.tics.2013.01.010

Grubert, A., Carlisle, N. B., & Eimer, M. (2016). The control of single-color and multiple-color visual search by attentional templates in working memory and in long-term memory. *Journal of Cognitive Neuroscience*, *28*(12), 1947–1963. https://doi.org/10.1162/jocn_a_01020

Grubert, A., & Eimer, M. (2015). Rapid parallel attentional target selection in single-color and multiple-color visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(1), 86–101. https://doi.org/10.1037/xhp0000019

Grubert, A., & Eimer, M. (2016). All set, indeed! N2pc components reveal simultaneous attentional control settings for multiple target colors. *Journal of Experimental Psychology. Human Perception and Performance*, *42*(8), 1215–1230. https://doi.org/10.1037/xhp0000221

Grubert, A., & Eimer, M. (2018). The time course of target template activation processes during preparation for visual search. *The Journal of*

Neuroscience, *38*(44), 9527–9538. https://doi.org/10.1523/JNEUROSCI.0409-18.2018

Grubert, A., & Eimer, M. (2020). Preparatory template activation during search for alternating targets. *Journal of Cognitive Neuroscience*, *32*(8), 1525–1535. https://doi.org/10.1162/jocn_a_01565

Gunseli, E., Olivers, C. N. L., & Meeter, M. (2014). Effects of search difficulty on the selection, maintenance, and learning of attentional templates. *Journal of Cognitive Neuroscience*, *26*(9), 2042–2054. https://doi.org/10.1162/jocn_a_00600

Gunseli, E., van Moorselaar, D., Meeter, M., & Olivers, C. N. L. (2015). The reliability of retro-cues determines the fate of noncued visual working memory representations. *Psychonomic Bulletin & Review*, *22*(5), 1334–1341. https://doi.org/10.3758/s13423-014-0796-x

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hollingworth, A., & Beck, V. M. (2016). Memory-based attention capture when multiple items are maintained in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(7), 911–917. https://doi.org/10.1037/xhp0000230

Hollingworth, A., & Hwang, S. (2013). The relationship between visual working memory and attention: Retention of precise colour information in the absence of effects on perceptual selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1628), 20130061. https://doi.org/10.1098/rstb.2013.0061

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70. https://www.jstor.org/stable/4615733

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Huynh Cong, S., & Kerzel, D. (2021). Allocation of resources in working memory: Theoretical and empirical implications for visual search. *Psychonomic Bulletin & Review*, *28*(4), 1093–1111. https://doi.org/10.3758/s13423-021-01881-5

Irons, J. L., Folk, C. L., & Remington, R. W. (2012). All set! Evidence of simultaneous attentional control settings for multiple target colors. *Journal of Experimental Psychology. Human Perception and Performance*, *38*(3), 758–775. https://doi.org/10.1037/a0026578

Kerzel, D., & Witzel, C. (2019). The allocation of resources in visual working memory and multiple attentional templates. *Journal of Experimental Psychology Human Perception & Performance*, *45*(5), 645–658. https://doi.org/10.1037/xhp0000637

Kristjánsson, Á. (2023). Priming of probabilistic attentional templates. *Psychonomic Bulletin & Review*, *30*(1), 22–39. https://doi.org/10.3758/s13423-022-02125-w

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400. https://doi.org/10.1016/j.tics.2013.06.006

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356. https://doi.org/10.1038/nn.3655

Morey, R. D. (2008). Confidence intervals from normalized data: a correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Murray, A. M., Nobre, A. C., Clark, I. A., Cravo, A. M., & Stokes, M. G. (2013). Attention restores discrete items to visual short-term memory. *Psychological Science*, *24*(4), 550–556. https://doi.org/10.1177/0956797612457782

Myers, N. E., Stokes, M. G., & Nobre, A. C. (2017). Prioritizing Information during working memory: beyond sustained internal attention. *Trends in Cognitive Sciences*, *21*(6), 449-461. https://doi.org/10.1016/j.tics.2017.03.010

Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review*, *124*(1), 21–59. https://doi.org/10.1037/rev0000044

Oberauer, K. (2019). Working memory and attention – a conceptual analysis and review. *Journal* *of Cognition*, *2*(1). https://doi.org/10.5334/joc.58

Olivers, C. N. L. (2009). What drives memory-driven attentional capture? The effects of memory type, display type, and search type. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1275–1291. https://doi.org/10.1037/a0013896

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, *15*(7), 327–334. https://doi.org/10.1016/j.tics.2011.05.004

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Rajsic, J., Ouslis, N. E., Wilson, D. E., & Pratt, J. (2017). Looking sharp: Becoming a search template boosts precision and stability in visual working memory. *Attention, Perception, & Psychophysics*, *79*(6), 1643–1651. https://doi.org/10.3758/s13414-017-1342-5

Reinhart, R. M. G., & Woodman, G. F. (2014). High stakes trigger the use of multiple memories to enhance the control of attention. *Cerebral Cortex (New York, N.Y.: 1991)*, *24*(8), 2022–2035. https://doi.org/10.1093/cercor/bht057

Reinhart, R. M. G., & Woodman, G. F. (2014). Enhancing long-term memory with stimulation tunes visual attention in one trial. *Proceedings of the National Academy of Sciences*, *112*(2), 625–630. https://doi.org/10.1073/pnas.1417259112

Shimi, A., Nobre, A. C., Astle, D., & Scerif, G. (2013). Orienting attention within visual short-term memory: Development and mechanisms. *Child Development*, *85*(2), 578–592. https://doi.org/10.1111/cdev.12150

Schneider, W. X. (2013). Selective visual processing across competition episodes: A theory of task-driven visual attention and working memory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1628), 20130060. https://doi.org/10.1098/rstb.2013.0060

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. *Nature Human Behaviour*, *4*(11), 1156–1172. https://doi.org/10.1038/s41562-020-00938-0

Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Attention, Perception, & Psychophysics*, *78*(7), 1839–1860. https://doi.org/10.3758/s13414-016-1108-5

Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision*, *13*(10), 9. https://doi.org/10.1167/13.10.9

The Pandas Development Team. (2024). pandas-dev/pandas: Pandas (V2.2.2) [Computer software]. https://doi.org/10/ggt8bh

Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. *Perception & Psychophysics*, *50*(2), 184–193. https://doi.org/10.3758/BF03212219

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, *51*(6), 599–606. https://doi.org/10.3758/BF03211656

Vallat, R. (2018). Pingouin: statistics in python. *Journal of Open Source Software*, *3*(31), 1026. https://doi.org/10.21105/joss.01026

van Moorselaar, D., Theeuwes, J., & Olivers, C. N. L. (2014). In competition for the attentional template: Can multiple items within visual working memory guide attention? *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1450–1464. https://doi.org/10.1037/a0036229

Waskom, M. L. (2021). seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

Williams, J. R., Brady, T. F., & Störmer, V. S. (2022). Guidance of attention by working memory is a matter of representational fidelity. *Journal of experimental psychology. Human perception and performance*, *48*(3), 202–231. https://doi.org/10.1037/xhp0000985

Williams, J. R., Robinson, M. M., & Brady, T. F. (2023). There is no theory-free measure of "swaps" in visual working memory experiments. *Computational Brain & Behavior*, *6*(1), 159–171. https://doi.org/10.1007/s42113-022-00150-5

Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review*, *28*(4), 1060–1092. https://doi.org/10.3758/s13423-020-01859-9

Woodman, G. F., & Luck, S. J. (2007). Do the contents of visual working memory automatically influence attentional selection during visual search? *Journal of Experimental Psychology: Human Perception and Performance, 33*(2), 363–377. https://doi.org/10.1037/0096-1523.33.2.363

Woodman, G. F., Carlisle, N. B., & Reinhart, R. M. G. (2013). Where do we store the memory representations that guide attention? *Journal of Vision*, *13*(3), 1. https://doi.org/10.1167/13.3.1

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

Zhou, C., Lorist, M. M., & Mathôt, S. (2020). Concurrent guidance of attention by multiple working memory items: Behavioral and computational evidence. *Attention, Perception, & Psychophysics*, *82*(6), 2950–2962. https://doi.org/10.3758/s13414-020-02048-5