Massive Effects of Saliency on Information Processing in Visual Working Memory

Martin Constant[1,2] and Heinrich René Liesefeld[1,3]
[1]Department Psychologie, Ludwig-Maximilians-Universität, München, Germany
[2]Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität, München, Germany
[3]Munich Center for Neurosciences – Brain & Mind, Ludwig-Maximilians-Universität, München, Germany

Abstract

Limitations in the ability to temporarily represent information in visual working memory (VWM) are crucial for (visual) cognition. Whether VWM processing is dependent on an object's saliency (i.e., how much it stands out) has been neglected in VWM research. Therefore, we developed a novel VWM task design that allows direct control over saliency. In three experiments with this task (on 10, 31, and 60 healthy adults, respectively), we consistently find that VWM performance is strongly and parametrically influenced by saliency and that both an object's relative (to concurrently presented objects) and absolute saliency influence VWM processing. We also demonstrate that this effect is indeed due to bottom-up saliency, rather than differential fit between each object and the top-down attentional set. A simple computational model assuming that VWM performance is determined by the weighted sum of absolute and relative saliency accounts well for the observed data patterns.

Statement of Relevance

The amount of visual information arriving each moment via the eyes is impossible to process to any reasonable extend by any limited system and human visual processing abilities are severely limited indeed; the major bottleneck for visual processing is called visual working memory (VWM). Using a novel task design, we demonstrate that the selection problem is solved in part by processing preferably the most prominent (salient) objects within a scene. How well an object is processed in VWM is determined both by how much it stands out and by how strong the other competitors in the scene are. This study brings VWM research one step closer to the (highly complex) real world and reveals that saliency has a major impact on VWM processing that is easily overlooked in the traditionally very abstract VWM paradigm.

Keywords: visual short-term memory, priority map, attention, visual search, visual perception

Visual working memory (VWM) is a crucial hub in the processing of visual information and its limitations are strongly related to general cognitive ability (Fukuda et al., 2010). Variation in VWM performance is typically interpreted in terms of some limited commodity (slots or resources, Cowan, 2001; Liesefeld & Müller, 2019a; Luck & Vogel, 2013; Ma et al., 2014), but alternative interpretations exist (Emrich et al., 2017; Oberauer & Lin, 2017). Due to its central role in (theories of) visual cognition, identifying influences on VWM performance is of high applied and theoretical relevance.

It has been extensively demonstrated that how well an object is memorized hinges on its behavioral relevance, that is, on the explicit intention to favor one or several objects (top-down influences; Emrich et al., 2017; Souza & Oberauer, 2016). It has been largely neglected, though, how VWM processing might differ for equally relevant objects due to (contextual) features of these objects themselves (bottom-up influences). In fact, all current models assume that, apart from random variation, all equally relevant objects within a display are processed equally well or have the same chance of being processed. This assumption seems reasonable for highly-controlled, abstract stimuli, but might not hold for (somewhat) more naturalistic stimuli and for the everyday use of VWM in complex real scenes.

From the visual-attention literature it is well known that beyond top-down goals other factors influence the allocation of processing resources (Awh et al., 2012; Liesefeld et al., 2018; Wolfe & Horowitz, 2017). A particularly strong influence on object processing that is largely neglected in the VWM literature is bottom-up saliency. An object is salient if (at least) one of its features stands out, like the blackness of a black sheep in a flock of white sheep. More technically, saliency is largely determined by local feature contrast (Nothdurft, 1993): via lateral inhibition (i.e., at the same hierarchical level of visual processing), neurons with overlapping tuning curves (i.e., coding similar features) mutually suppress each other (*lateral iso-feature suppression*; Z. Li, 2002); the resulting net activity is highest for features that differ maximally from their immediate surround, because the respective neuronal activity receives little suppression. As saliency has a strong and parametric influence on object processing in visual search (Liesefeld et al., 2016), it seems likely that salient objects are also prioritized for VWM processing.

In the rare cases in which the influence of object saliency on VWM processing has been studied, the design did not allow manipulating each object's saliency independently (Rajsic et al., 2016) or confounded saliency with the discriminability of the to-be-remembered feature. Klink et al. (2017), for example, had participants remember the orientation of Gabor gratings and manipulated saliency by varying the Gabor contrast (Figure 1a; see also, Knops et al., 2014). In line with an effect of saliency, the lower the contrast, the worse VWM performance was. However, varying the contrast also influences the discriminability of the to-be-remembered orientation, because the Gabor increasingly merges with the background for lower contrasts. In fact, in psychophysical studies, Gabor contrast is often used to scale discrimination difficulty (e.g., Alvarez & Cavanagh, 2008). These and other confounds also affect studies using (quasi) natural stimuli (which are by definition not well-controlled; for a review, see Santangelo, 2015). Nevertheless, these studies indicate that saliency has some influence on VWM processing.

To study the influence of saliency on VWM encoding under controlled conditions, we developed a task that deconfounds saliency of target objects and discriminability of to-be-remembered features and allows manipulating each object's saliency continuously and independently (Figure 1b). With this novel task, we conducted three experiments in which participants had to remember the color of three target objects. These three targets were always equally likely to be probed but differed in saliency either within or across displays. Our results show a strong impact of bottom-up saliency on how well *equally relevant* objects are stored in VWM.

Experiment 1

Methods

In many VWM studies, participants hold the colors of a bunch of isolated objects in mind for a short retention period and then have to decide whether one of the objects changed color in a second display (*change detection*) or reproduce the color of a probed object (*continuous report*). A wide variety of versions of this basic design exist, but the focus on isolated (i.e., highly salient) objects is common to virtually all of them (see Figure 1d). To open up the VWM paradigm to the well-controlled examination of saliency effects, we developed a novel VWM task in which we can directly, gradually, and independently manipulate each object's saliency, while keeping the discriminability of the to-be-remembered features and the objects' behavioral relevance untouched. This design also enables the use of modern computational models and neuroimaging methods.

Task development built upon our previous experience from visual-attention research. In particular, Liesefeld et al. (2016) devised a visual-search task that allowed a gradual manipulation of the search target's saliency (see also, Nothdurft, 1993) and showed that search becomes faster as a continuous function of target saliency. By placing a tilted target bar into a dense array of vertical non-target bars and adapting the tilt of the target bar (and therefore the contrast between target and non-targets), we could control target saliency to any desired precision. Liesefeld et al. (2017) showed that in this design processing priority (measured by the order of attention allocations) is (almost) perfectly determined by object saliency.

Translating this design to the study of VWM, the memory displays employed here featured a dense array of vertical non-target bars into which three differently tilted and randomly colored target bars were placed (Figure 1b). Participants had to remember the target bars' colors in order to later reproduce one of them. In order not to make color dominate the contrast (and therefore determine saliency), the non-target bars were also drawn in random (completely irrelevant) colors.

The critical deviation from previous research on VWM is that our displays are cluttered with irrelevant vertical non-target bars. As explained above, this is necessary to control the saliency of the relevant bars, because saliency of an object depends on its relationship to the immediate surround. This is not an artificial change to the task, though, but mim-

---

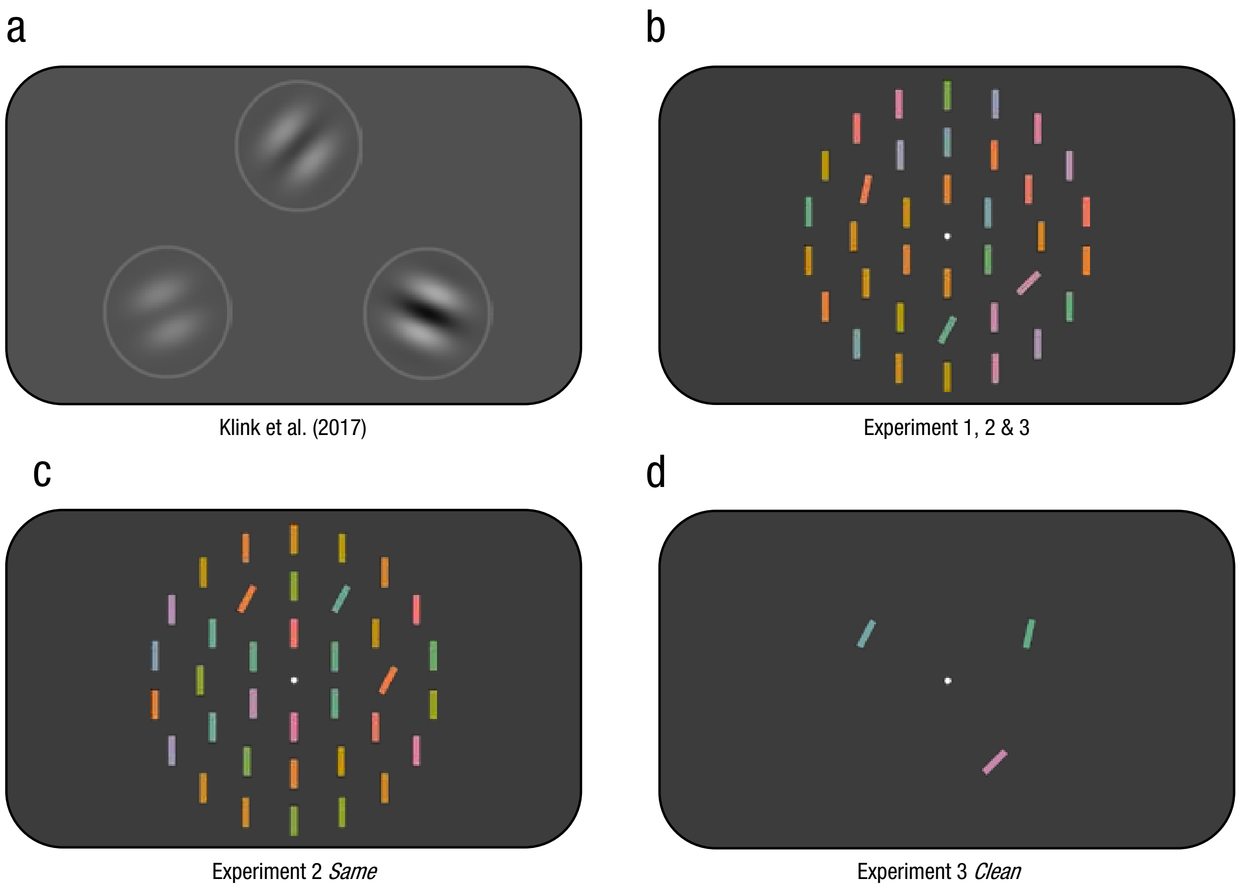
September 7, 2020. This version has not been peer-reviewed.
Martin Constant https://orcid.org/0000-0001-9574-0674
Heinrich R. Liesefeld https://orcid.org/0000-0003-4551-8607
All experiments reported here were preregistered on OSF. The preregistrations, experimental programs, analysis scripts, and data files can be found at: https://osf.io/pfb67/

This work was supported by the German Research Foundation (DFG) under Grant LI 2868/3-1 awarded to HRL. We thank Dimana Balcheva and Lena Stemmer for their assistance in data collection.

Correspondence concerning this article should be addressed to Heinrich René Liesefeld, Department Psychologie, Ludwig-Maximilians-Universität, Leopoldstr. 13, D-80802 Munich, Germany, Email: Heinrich.Liesefeld@psy.lmu.de

Figure 1

A typical example of previous manipulations of saliency and design of the present memory displays.



Note. Klink et al. (2017) had participants remember the orientations of Gabor gratings and manipulated saliency via the gratings' contrasts (panel a); note how the contrast also influences the discriminability of the to-be-remembered orientations. In our novel task design (panels b and c; Experiments 1-3), participants have to remember the colors of three tilted target bars to later reproduce one of these colors and saliency is manipulated via target tilt. Using the same tilt (panel c) equates the bars' relative saliency within each display. Removing the vertical non-target bars in Experiment 3 (panel d) renders all target bars highly salient (leaving only the isolated colored objects that are often used in VWM studies).

ics a feature of the real world: hardly any natural environment consists of well-isolated relevant objects, but the real world is utterly cluttered with many objects that are irrelevant for the task at hand (e.g., Hollingworth, 2008). Also note that in Liesefeld et al. (2016), even the smallest tilt employed in the present study (12°) produced clear pop out, that is, participants were able to almost exclusively process the target bar and completely ignore the vertical non-target bars. Thus, the vertical bars are sufficiently less salient than even the 12°-tilted bars, so that they likely do not significantly compete for (VWM) processing as distractors in other designs would (Liesefeld et al., 2014; Vogel et al., 2005; for a review see, Liesefeld et al., 2020).

Participants

For each experiment, sample size was determined via sequential testing with Bayes factors, following the recommendations by Schönbrodt and Wagenmakers (2018, for details see Supplement). The preregistered critical directed (one-tailed) tests determining the stopping rule for Experiment

1 examined whether VWM performance (the mean absolute angular distance between correct and selected response, henceforth: *recall error*) would decrease with object saliency (tilt). This resulted in a sample of 10 healthy human adults (Mean age: 26.3 ± 3.37, 4 females, all right-handed).

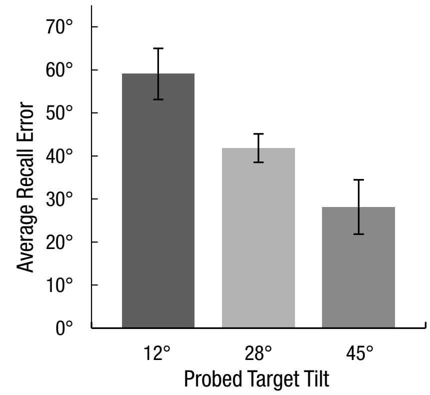
Stimuli, procedure & design

After a 1s fixation dot, the memory display (Figure 1b) was presented for 350 ms, followed by another 1s fixation dot (delay period). A response display was then presented containing a color wheel and outlined placeholder bars at the locations of each bar from the memory display. One of the placeholders was filled in black to indicate which bar to report (hereafter: *probe*), and participants were instructed to report the color they remembered for that bar by using the computer mouse to select a point on the color wheel. After each response, a feedback line appeared at the correct location on the color wheel to show to the participant the correct response (and, by implication, how far off the actual response was).

Each participant completed a total of 600 trials divided into blocks of 30 trials. Each condition (i.e., tilt of the probe) was randomly presented 200 times (10 times per block).

Figure 2

Recall error (i.e., mean absolute difference between correct and given response) in Experiment 1.



Note. Using memory displays as illustrated in Figure 1b, which required participants to remember the colors of three bars tilted 12°, 28°, and 45°, respectively, and later reproduce one of these colors. Error bars represent 95% within-participant *CI*s

Results

As expected, our manipulation of saliency had a huge and reliable impact on VWM performance (see Figure 2): Despite all three objects being equally relevant, recall error was higher for 12° (M ± between-participants 95% CI; 59.07° ± 11.97) than for 28° probes (41.84° ± 10.06), $t(9) = 6.56$, $p < .001$, $d_z = 2.07$ [0.93, 3.19], $BF_{+0} = 551.51$, and higher for 28° than for 45° probes (28.14° ± 5.24), $t(9) = 4.66$, $p < .001$, $d_z = 1.47$ [0.54, 2.37], $BF_{+0} = 70.6$. Effect sizes were so huge that, despite the relatively small sample size (which we had defined as the minimum in our preregistration), the Bayes factors indicated overwhelming evidence for both differences. This finding demonstrates that VWM performance is strongly and parametrically dependent on saliency.

Fitting to the Zhang and Luck (2008) model revealed that p_{mem} differed significantly between 12° (44.08% ± 16.71) and 28° probes (68.89% ± 12.59), $t(9) = -6.37$, $p < .001$, $d_z = -2.01$ [−3.10, −0.89], $BF_{10} = 227.57$; and between 28° and 45° probes (86.41% ± 6.08), $t(9) = -4.10$, $p = .003$, $d_z = -1.30$ [−2.14, −0.42], $BF_{10} = 18.18$. However, sd did not significantly differ between 12° (26.93° ± 7.88) and 28° probes (25.99° ± 3.33), $t(9) = 0.315$, $p = .760$, $d_z = 0.10$ [−0.52, 0.72], $BF_{01} = 3.10$; or between 28° and 45° probes (23.91° ± 2.06), $t(9) = 1.29$, $p = .230$, $d_z = 0.41$ [−0.25, 1.04], $BF_{01} = 1.68$. Even though this evidence for the absence of an effect on (im)precision (sd) is only moderate or indecisive, respectively, it is clear that potential effects on precision cannot explain the overwhelming evidence for an effect of saliency on recall error ($BF_{+0} = 551.51$ and $BF_{+0} = 70.6$).

Discussion

Using a novel design, providing high experimental control over object saliency while mimicking the visual complexity of real-world scenes, Experiment 1 confirmed that the allocation of a limited VWM resource is strongly and parametrically dependent on saliency. Moreover, fitting a standard model of VWM performance to the data revealed that this effect is mainly due to salient objects being remembered more likely rather than more precisely. Even though we observed only moderate ($BF_{01} = 3.10$) and indecisive ($BF_{01} = 1.68$) evidence for the absence of an effect on precision, it is clear that potential effects on precision cannot explain the overwhelming evidence for an effect of saliency on recall error ($BF_{+0} = 551.51$ and $BF_{+0} = 70.6$).

Experiment 2

Saliency might influence VWM processing in two, nonexclusive ways: First, objects compete for (VWM) processing, so that the most salient object within a display is (eventually) remembered best. This effect depends on the object's relation to other objects in the display and we therefore refer to it as an effect of *relative saliency*. Second, processing of more salient objects might be enhanced regardless of what else is in the display – an effect of absolute saliency. In visual search, the *absolute saliency* of a single target affects processing difficulty (Liesefeld et al., 2016; Nothdurft, 1993), but only little is known regarding effects of relative saliency with multiple target objects.

Methods

To disentangle the two potential effects of saliency, we ran an experiment that compared the *mixed* displays of Experiment 1 with displays containing three bars of the same tilt (*same* displays). An effect of absolute saliency would predict that even in displays with only 12°-tilted bars (12°-*same* displays), each 12°-tilted bar is remembered worse than each 45°-tilted bar in 45°-*same* displays. If relative saliency contributes to the effect of saliency observed in Experiment 1, the 45° tilted object was processed particularly well (beyond the effect of absolute saliency) by virtue of the other two tilted bars being less salient. Correspondingly, the 12° tilted object then was processed particularly badly due to the other two tilted bars being more salient. By contrast, when all targets within a display are equally salient, the degree of VWM processing should be equal for all of them. This means that each 45°-tilted object in a display with only 45°-tilted objects among vertical bars would be remembered worse than the 45°-tilted object competing with the 28°- and 12°-tilted object in *mixed* displays. Conversely, each 12°-tilted object in a display with only 12°-tilted objects would be remembered better than the 12°-tilted object competing with the 28°- and 45°-tilted object in Experiment 1. Thus, demonstrating that performance decreases from *mixed* to *same* displays for 45°-tilted objects and increases for 12°-tilted objects would constitute proof of an influence of relative saliency on VWM performance.

In Experiment 2, the preregistered critical directed tests determining the stopping rule for the sequential testing procedure examined whether recall error would decrease with object saliency (as in Experiment 1) for both *same* and *mixed* displays and (additionally) whether recall error would differ, for the same probe tilt, between *mixed* and *same* displays, whereby we predicted an increase for 45° and a decrease for 12° probes. This resulted in a sample of 31 healthy human adults (Mean age: 26.4 ± 5.44, 25 female, 4 left-handed). Experiment 2 was modeled after Experiment 1 with the crucial difference being that one of two types of memory displays could be presented on each trial:

1. *Mixed* displays were identical to the displays of Experiment 1 (Figure 1b) in all relevant aspects.

2. *Same* displays were similar to *mixed* displays except that the tilted bars all shared the same tilt (either 12°, 28° or 45°).

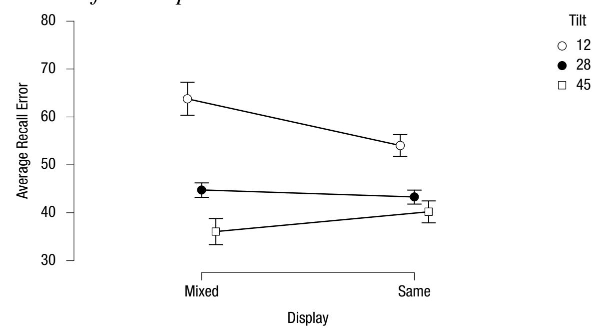
Each participant completed a total of 600 trials divided into blocks of 30 trials each. Each condition (i.e., type of display × tilt of the probe) was randomly presented 100 times.

Results

The *mixed* condition of Experiment 2 replicated the results of Experiment 1 (see Figure 3): Recall error was higher for 12° (63.77° ± 5.21) than for 28° probes (44.74° ± 4.95), $t(30) = 10.57$, $p < .001$, $d_z = 1.90$ [1.30, 2.49], $BF_{+0} = 1.44e+9$, and higher for 28° than for 45° probes (36.06° ± 4.08), $t(30) = 5.83$, $p < .001$, $d_z = 1.05$ [0.60, 1.48], $BF_{+0} = 1.68e+4$.

Figure 3

Results from Experiment 2



Note. The change in recall error from *mixed* to *same* displays (decrease for 12° and increase for 45°) demonstrates an effect of relative saliency (the various targets compete for processing resources) in *mixed* displays; the effect of tilt in *same* displays demonstrates an effect of absolute saliency (less salient items are processed less well; see text for details). Error bars represent 95% within-participant *CI*s.

Crucially, and as expected, performance was better for 12° probes, $t(30) = 6.02$, $p < .001$, $d_z = 1.08$ [0.63, 1.52], $BF_{+0} = 2.69e+4$ and worse for 45° probes, $t(30) = -2.88$, $p = .004$, $d_z = -0.52$ [−0.89, −0.13], $BF_{-0} = 11.56$, in *same* compared to *mixed* displays. This difference was only weak and unreliable for 28° probes (for which we had no specific hypotheses as mentioned in our pre-registration), $t(30) = 1.57$, $p = .128$, $d_z = 0.28$ [−0.08, 0.64], $BF_{01} = 1.75$. Indeed, VWM recall performance for a particular object depends on the object's relative saliency with respect to the other objects in the scene.

Even though the effect of probed-target tilt was weaker for *same* than for *mixed* displays, it was still present, indicating an effect of absolute saliency on top of the effect of relative saliency. Recall error was higher for 12°- (54.02° ± 4.74) than for 28-° same displays (43.29° ± 4.88), $t(30) = 7.79$,

$p < .001$, $d_z = 1.40$ [0.90, 1.89], $BF_{+0} = 2.39\text{e}+6$, and higher for 28°- than for 45°- same displays (40.19° ± 5.40), $t(30) = 3.10$, $p = .002$, $d_z = 0.56$ [0.17, 0.93], $BF_{+0} = 18.85$ (see Figure 3). Replicating Experiment 1, results from the Zhang and Luck (2008) mixture model again showed that salience influenced mainly p_{mem} in both *mixed* and *same* displays (see Supplement).

Computational modeling

One might argue that the observed effects of target tilt are not due to differential bottom-up saliency, but rather to differential fit between each object and the top-down attentional set (or template) used to select target objects (e.g., Duncan & Humphreys, 1989; Geng & Witkowski, 2019). In particular, looking for tilted objects, participants' attentional set in our study might be matched best by the 45°-tilted object, followed by the 28°-tilted object, despite all objects being equally relevant. Such an attentional template might be optimal, because it minimizes the match between the search template and the vertical (0°) non-target objects, thus potentially minimizing interference (Geng & Witkowski, 2019).

Methods

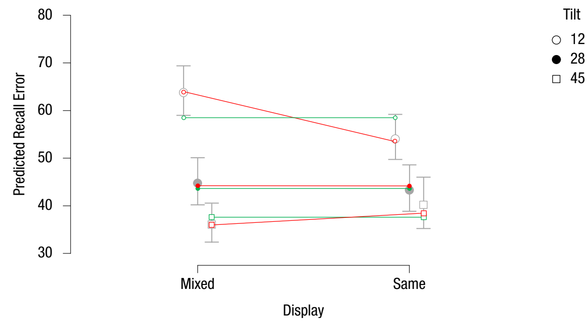
To test how well the two conflicting explanations can account for the data from Experiment 2, we used computational modeling. In particular, we devised two novel models that implement the two potential accounts for the observed data pattern: (i) The saliency model attempts to account for the data by a mixture of absolute and relative saliency, whereby the degree to which relative saliency has an influence is a free parameter estimated from the data. (ii) The alternative optimal-template model posits that the different target bars differentially match with the top-down template. Importantly, rather than deciding a priori on the value of the template, we included template tilt as a free parameter, so that the optimization algorithm could estimate the (unobservable) template tilt from the observed behavioral data (see Supplement for a detailed description of both models).

Results

Comparing the fit of both models to the data of Experiment 2 (see Figure 4), the saliency model well outperformed the optimal-template model. In particular, the optimal-template model fails to account for the difference between *mixed* and *same* displays. Thus, performance in Experiment 2 is best explained by variation in saliency. Notably, to account for the data, the saliency model has to assume a positive influence of relative saliency ($w_{rel} > 0$), thus providing further support for this novel assumption.

Figure 4

Predictions of our preferred saliency model (red) and the alternative optimal-template model (green) for Experiment 2.



Note. For comparison, empirical data and ± 1 standard error of the mean (*SE*) are plotted in addition (grey). While the saliency model is in line with the data (all predictions are within 1 *SE* of the empirical data), the optimal-template model cannot account for the difference between *mixed* and *same* displays.

Experiment 3

The model was devised post-hoc to rule out a potential alternative explanation in terms of an attentional set. To additionally provide an empirical test with a-priori hypotheses, we preregistered and ran Experiment 3. We reasoned that if differential fit between the objects and the attentional set explains our results, the effect of tilt should remain when the vertical bars are removed (*clean* displays; Figure 1d)[1], because the tilted bars still differentially fit to this assumed attentional set. By contrast, our explanation in terms of saliency predicts that removing the task-irrelevant vertical bars renders all tilted bars highly and almost equally salient, because local feature contrast is high for all three bars when presented in isolation (see Methods of Exp. 1). In contrast to the *cluttered* displays of Experiment 1, the effect of tilt should therefore be strongly decreased or even completely absent in *clean* displays.

Methods

Experiment 3 was conducted online (see Supplement for details). The preregistered critical directed t tests determining the stopping rule for the sequential testing procedure examined whether recall error would decrease with object saliency in displays with vertical non-target bars (*cluttered*

[1] We thank Nelson Cowan for suggesting these displays at the Virtual Working Memory Symposium 2020.

displays; as in Experiment 1 and 2 *mixed*), and whether the effect of tilt was lower in *clean* compared to *cluttered* displays. A third non-critical hypothesis was that the effect of tilt might be fully absent in *clean* displays. This sequential-testing procedure (see Supplement and preregistration for details) resulted in a sample of 60 healthy human adults (Mean age: 25.6 ± 6.20, 23 female, 4 left-handed).

Experiment 3 was modeled after Experiment 1 with the critical difference being that one of two types of memory displays could be presented on each trial:

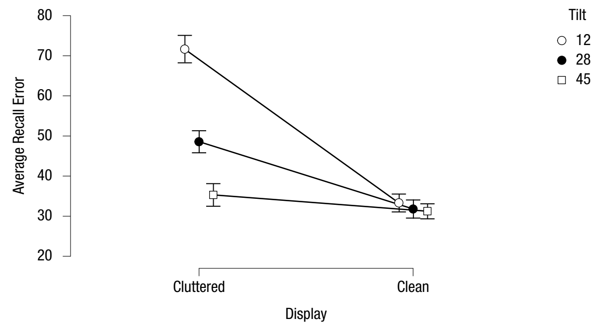
1. *Cluttered* displays were identical to the displays of Experiment 1 (Figure 1b) in all relevant aspects.

2. *Clean* displays contained only the three tilted bars (i.e., the task-irrelevant vertical non-target bars were removed) but were otherwise identical to *cluttered* displays.

Each participant completed a total of 150 trials divided into blocks of 50 trials. Each condition (i.e., type of display × tilt of the probe) was randomly presented 25 times.

Figure 5

Results from Experiment 3



Note. The drastic decrease in the effect of tilt from *cluttered* to *clean* displays demonstrates that the effect is due to saliency rather than an attentional set/template (see text for details). That performance is generally better in *clean* displays demonstrates that objects in typical VWM displays are of high saliency. Error bars represent within-participant 95% *CI*s.

Results

For *cluttered* displays, we replicated the results of Experiment 1 and 2 (*mixed* displays; see Figure 5): Recall error was higher for 12° (71.64° ± 4.39) than for 28° probes (48.56° ± 4.38), $t(59) = 11.74$, $p < .001$, $d_z = 1.52$ [1.14, 1.88], $BF_{+0} = 2.63e + 14$, and higher for 28° than for 45° probes (35.30° ± 3.47), $t(59) = 6.11$, $p < .001$, $d_z = 0.79$ [0.50, 1.08], $BF_{+0} = 3.18e{+}5$. Crucially, and as expected, the effect of tilt decreased in *clean* compared to *cluttered* displays, for 12° compared to 28° probes, $t(59) = -10.01$, $p < .001$, $d_z = -1.29$ [–1.63, -0.95], $BF_{-0} = 2.69e{+}4$, and for 28° compared to 45° probes, $t(59) = -5.06$, $p < .001$, $d_z = -0.65$ [0.93, –0.37], $BF_{-0} = 8024.60$. Finally, there was no significant effect of tilt and there was even some evidence for the absence of this effect in *clean* displays for 12° (33.31° ± 2.94) compared to 28° probes (31.79° ± 3.17), $t(59) = 1.17$, $p = .247$, $d_z = 0.15$ [–0.10, 0.40], $BF_{01} = 3.71$, and moderate evidence for absence for 28° compared to 45° probes (31.23° ± 2.82), $t(59) = 0.46$, $p = .650$, $d_z = 0.06$ [–0.19, 0.31], $BF_{01} = 6.41$. This pattern indicates that the effect of target tilt is not due to differential match between the objects and an attentional set/template, but rather due to variation in saliency.

General Discussion

We set out to demonstrate an influence of saliency on performance in a VWM task, an influence that has not yet been acknowledged by any current model of VWM processing. Experiment 1 indeed provided overwhelming evidence for the existence of this effect by showing that how well an object's color is remembered is largely determined by how much it differs in tilt from its immediate surround (local feature contrast). Experiment 2 demonstrated that relative and absolute saliency both contribute to the effect of saliency. Finally, a newly devised computational model and Experiment 3 demonstrated that the effect of target tilt is indeed explained by saliency rather than differential fit between each object and some attentional set. How saliencies of multiple relevant objects interact has – to the best of our knowledge – not yet been systematically examined and an observation of an effect of relative saliency is therefore not only new for the VWM community, but for the visual-cognition community in general.

Many theories of visual search (e.g., Duncan & Humphreys, 1989; Fecteau & Munoz, 2006; Liesefeld & Müller, 2019b, in press; Wolfe, 2007) assume a pre-attentive spatial representation of the visual scene coding for relevance at each location and informing a second, attentive processing stage. This assumption is needed to explain how second-stage focal attention can be allocated to the most promising objects in view without analyzing each object in detail first. This pre-attentive priority map is thought to be influenced by task goals and experiences (top-down) as well as saliency (bottom-up). We propose that the very same *priority map* supporting visual search might also determine VWM processing (Bundesen et al., 2011; Liesefeld et al., 2020). Findings from the present study and those manipulating each object's relevance (e.g., Emrich et al., 2017) can be integrated using the priority-map concept: while previous studies manipulated top-down influences, we are the first to systematically manipulate bottom-up contributions (i.e., saliency) to pre-attentive priority-map activations in a VWM task.

There are many potential mechanisms by which first-stage priority (and, thus, saliency) could theoretically impact second-stage VWM processing: (*a*) it might influence VWM encoding directly (in particular without the allocation of focal attention) or via the allocation of an attentional resource (Emrich et al., 2017). (*b*) Encoding and attention allocation could be conceived of as serial (one object is processed/attended after the other) or parallel (all objects are processed/attended at once; Bundesen et al., 2011; Sewell et al., 2014). (*c*) Priority might affect how much (if any) information about each object is processed and/or how much of a limited (quantized or continuous) VWM resource it receives (Ma et al., 2014; Vogel et al., 2006); (*d*) Priority might additionally influence (third-stage) post-selective/post-encoding processes, such as how fast attention is disengaged from a processed object to continue with the next object in line (see Sauter et al., 2020, preprint) or how well a processed object is stored (e.g., by attaining a special state, Oberauer & Lin, 2017; Olivers et al., 2011). All kinds of combinations between these mechanisms seem theoretically possible and we will speculate on some in turn.

Our exploratory Zhang and Luck (2008) mixture-model analysis indicated that saliency mainly affects whether an item is encoded (p_{mem}) rather than the precision of encoding (*sd*). If mixture modeling is valid (for a critical view, Ma, 2018, preprint), this finding somewhat constrains the range of potential mechanisms by which saliency (as represented on a first-stage, pre-attentive priority map) is translated into VWM performance: If, at the second stage, all objects are processed in parallel one would assume that information on each object accrues continually with a slower rate for less salient objects (e.g., Moran et al., 2016). The mixture-modeling finding would then indicate that an object is stored in full once a certain amount of information is accumulated (Bundesen et al., 2011), because otherwise we should have observed an effect on *sd*. Alternatively, second-stage encoding might proceed serially starting at the most salient and sometimes not reaching the least salient target object (Wolfe, 2007, e.g., because focal attention needs to be allocated sequentially to encode each object).

Another implication from our study is that previous studies might have unintentionally induced and misinterpreted disguised effects of saliency. Data from the *same* condition of Experiment 2 indicate that less information was remembered in low-saliency compared to high-saliency displays (the effect of absolute saliency). One could easily misinterpret this effect as a decrease in VWM capacity from high- to low-saliency displays. However, this would be theoretically awkward, because a fixed limit is the core assumption behind both slot- and flexible resource theories of VWM alike (for an alternative, see Oberauer & Lin, 2017). Actually, this effect reminds of other findings that processing difficulty of an object class correlates with how many objects of that class can be hold in VWM: manipulating object complexity, Alvarez and Cavanagh (2004) showed that visual-search rate (as a measure of processing difficulty) predicts VWM capacity for the respective object class. They argued that search rate and VWM capacity were related by the objects' informational content, which would affect how long it takes to process each item in visual search and how much of the limited VWM capacity it consumes. In light of the present results, it seems equally likely, though, that the two measures are more directly related by the saliency-dependent ease of processing each object. For example, processing of the first low saliency/high complexity object(s) might take so long (see mechanism *d* above) that on some trials no time is left to process the remaining object(s) in the display (e.g., in our *same* displays, only two out of the three 12° objects might have been processed on some trials). Crucially, in our study, this cannot be explained by the to-be-encoded informational content (which was the same for each object) but must be due to the saliency of the object carrying that information. Thus, effects of object complexity on VWM performances observed in earlier studies might alternatively be explained as effects of saliency. More complex objects might be less salient in their respective displays and therefore take longer to process (irrespective of their informational content). Along similar lines, our findings might trigger re-evaluations of further influential findings from VWM studies using (relatively) complex stimuli.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111. https://doi.org/10/b88gs7

Alvarez, G. A., & Cavanagh, P. (2008). Visual short-term memory operates more efficiently on boundary features than on surface features. *Perception & Psychophysics*, *70*(2), 346–364. https://doi.org/10/dbs6k7

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443. https://doi.org/10/f34nps

Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience*, *34*(10), 3632–3645. https://doi.org/10/gfwfbw

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2011). A neural theory of visual attention and short-term memory (NTVA). *Neuropsychologia*, *49*(6), 1446–1457. https://doi.org/10/b8863z

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/ddq83h

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458. https://doi.org/10/ck4p2t

Emrich, S. M., Lockhart, H. A., & Al-Aidroos, N. (2017). Attention mediates the flexible allocation of visual working memory resources. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(7), 1454–1465. https://doi.org/10/gbn3xj

Fairchild, M. D. (2013). *Color appearance models* (3rd ed.). John Wiley & Sons, Inc. https://doi.org/10/dc5n

Fecteau, J., & Munoz, D. (2006). Salience, relevance, and firing: A priority map for target selection. *Trends in Cognitive Sciences*, *10*(8), 382–390. https://doi.org/10/fb66m9

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review*, *17*(5), 673–679. https://doi.org/10/dwkkc6

Geng, J. J., & Witkowski, P. (2019). Template-to-distractor distinctiveness regulates visual search efficiency. *Current Opinion in Psychology*, *29*, 119–125. https://doi.org/10/gg5vtp

Hollingworth, A. (2008, August 12). Visual Memory for Natural Scenes. In S. J. Luck & A. Hollingworth (Eds.), *Visual Memory* (pp. 123–162). Oxford University Press. https://doi.org/10/ck2gnr

JASP Team. (2020). *JASP* (Version 0.13.1). https://jasp-stats.org/

Klink, P. C., Jeurissen, D., Theeuwes, J., Denys, D., & Roelfsema, P. R. (2017). Working memory accuracy for multiple targets is driven by reward expectation and stimulus contrast with different time-courses. *Scientific Reports*, *7*(1), 9082. https://doi.org/10/gbtwp7

Knops, A., Piazza, M., Sengupta, R., Eger, E., & Melcher, D. (2014). A shared, flexible neural map architecture reflects capacity limits in both visual short-term memory and enumeration. *Journal of Neuroscience*, *34*(30), 9857–9866. https://doi.org/10/f6fskg

Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for Participants' Viewing Distance in Large-Scale, Psychophysical Online Experiments Using a Virtual Chinrest. *Scientific Reports*, *10*(1), 904. https://doi.org/10/ggpbsf

Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, *6*(1), 9–16. https://doi.org/10/bjs32m

Liesefeld, H. R., Liesefeld, A. M., Pollmann, S., & Müller, H. J. (2018). Biasing allocations of attention via selective weighting of saliency signals: Behavioral and neuroimaging evidence for the dimension-weighting account. In T. Hodgson (Ed.), *Processes of Visuospatial Attention and Working Memory* (pp. 87–113). Springer International Publishing. https://doi.org/10/dc5q

Liesefeld, H. R., Liesefeld, A. M., Sauseng, P., Jacob, S. N., & Müller, H. J. (2020). How visual working memory handles distraction: Cognitive mechanisms and electrophysiological correlates. *Visual Cognition*, 1–16. https://doi.org/10/gg5vsv

Liesefeld, H. R., Liesefeld, A. M., Töllner, T., & Müller, H. J. (2017). Attentional capture in visual search: Capture and post-capture dynamics revealed by EEG. *NeuroImage*, *156*, 166–173. https://doi.org/10/gbsjqj

Liesefeld, H. R., Moran, R., Usher, M., Müller, H. J., & Zehetleitner, M. (2016). Search efficiency as a function of target saliency: The transition from inefficient to efficient search and beyond. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 821–836. https://doi.org/10/ggbnjc

Liesefeld, H. R., & Müller, H. J. (2019a). Current directions in visual working memory research: An introduction and emerging insights. *British Journal of Psychology*, *110*(2), 193–206. https://doi.org/10/gfvm2p

Liesefeld, H. R., & Müller, H. J. (2019b). Distractor handling via dimension weighting. *Current Opinion in Psychology*, *29*, 160–167. https://doi.org/10/ggbnjd

Liesefeld, H. R., & Müller, H. J. (in press). Modulations of Saliency Signals at Two Hierarchical Levels of Priority Computation Revealed by Spatial Statistical Distractor Learning. *Journal of Experimental Psychology: General*.

Liesefeld, A. M., Liesefeld, H. R., & Zimmer, H. D. (2014). Intercommunication between prefrontal and posterior brain regions for protecting visual working memory from distractor interference. *Psychological Science*, *25*(2), 325–333. https://doi.org/10/f5344z

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/10/39v

Ma, W. J. (2018, preprint). Problematic usage of the Zhang and Luck mixture model. *bioRxiv*. https://doi.org/10/ggbpq3

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356. https://doi.org/10/gd8cbs

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10/ft2dgc

Moran, R., Zehetleitner, M., Liesefeld, H. R., Müller, H. J., & Usher, M. (2016). Serial vs. parallel models of attention in visual search: Accounting for benchmark RT-distributions. *Psychonomic Bulletin & Review*, *23*(5), 1300–1315. https://doi.org/10/gbgdk6

Nelder, J. A., & Mead, R. (1965). A Simplex method for function minimization. *The Computer Journal*, *7*(4), 308–313. https://doi.org/10/bkqf

Nothdurft, H.-C. (1993). The role of features in preattentive vision: Comparison of orientation, motion and color cues. *Vision Research*, *33*(14), 1937–1958. https://doi.org/10/d6fw5n

Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review*, *124*(1), 21–59. https://doi.org/10/f9mhb2

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/fcfxg6

Rajsic, J., Sun, S. Z., Huxtable, L., Pratt, J., & Ferber, S. (2016). Pop-out and pop-in: Visual working memory advantages for unique items. *Psychonomic Bulletin & Review*, *23*(6), 1787–1793. https://doi.org/10/f9hf36

Santangelo, V. (2015). Forced to remember: When memory is biased by salient information. *Behavioural Brain Research*, *283*, 1–10. https://doi.org/10/f67dgp

Sauter, M., Hanning, N. M., Liesefeld, H. R., & Müller, H. J. (2020, preprint). Statistical learning of frequent distractor locations in visual search: A role for post-selective distractor rejection processes? *PsyArXiv*. https://doi.org/10/gg5vtw

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, *25*(1), 128–142. https://doi.org/10/gcsk2r

Sewell, D. K., Lilburn, S. D., & Smith, P. L. (2014). An information capacity limitation of visual short-term memory. *Journal of Experimental Psychology. Human Perception and Performance*, *40*(6), 2214–2242. https://doi.org/10/gg5fn7

Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory: 13 years of the retrocue effect. *Attention, Perception, & Psychophysics*, *78*(7), 1839–1860. https://doi.org/10/f83nvs

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/ggbnjj

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*(22), 8780–8785. https://doi.org/10/f3trn3

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), 500–503. https://doi.org/10/drbpg5

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1436–1451. https://doi.org/10/c2vrp7

Wolfe, J. M. (2007). Guided Search 4.0: Current Progress with a Model of Visual Search. *Integrated Models of Cognitive Systems* (pp. 99–119). Oxford University Press. https://doi.org/10/gftcmp

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 0058. https://doi.org/10/gftcnc

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10/c693sk

Supplement

Supplementary Methods

Participants

For each experiment, sample size was determined via sequential testing with Bayes factors, following the recommendations by Schönbrodt and Wagenmakers (2018). This recently developed sequential testing procedure with preregistered hypotheses continues data collection until a pre-defined level of evidence in terms of Bayes factors in favor of or against each preregistered hypothesis is reached and thereby ensures that strong evidence for either the presence or the absence of each relevant effect is gained. In our preregistration, we set a minimum of 10 and a maximum of 60 participants in the laboratory Experiments 1 and 2. For Experiment 3, which was conducted online and was shorter, we set a minimum of 20 and a maximum of 100 participants (*BF*s were evaluated after each batch of 20 participants). We stopped testing when sufficient evidence for either the null or the alternative ($BF \geq 6$) was reached, which was achieved for each critical test.

All participants provided informed consent prior to the respective experiment, reported normal or corrected-to-normal visual acuity and normal color vision and were naïve as to the purpose of the study. They received either course credits or monetary remuneration (9 €/h) in Experiments 1 and 2. Experiment 3 was run online and recruitment was done via Prolific (https://prolific.co/). Participants were paid 1.50£ for around 15 minutes of their time. All experimental procedures were approved by the ethics committee of the Department Psychology and Pedagogics at LMU. In Experiments 1 and 2, no participant was excluded from the analyses and two trials of one participant (0.33%) were dropped in each experiment because of a delay in memory-display offset. Three participants of Experiment 2 had already participated in Experiment 1 and three others had participated in another related experiment.[2] In Experiment 3, eight participants were excluded. As specified during recruitment, these eight participants were not compensated and were replaced.

Stimuli

For Experiment 1 and 2, stimuli were displayed on a 24" TFT-LCD monitor (ASUS VG248QE, 1920x1080 pixels, 60 Hz) at a viewing distance of 70 cm. The testing room was pitch dark and there were between one and four participants in each testing session. For Experiment 1, OpenSesame 3.2.7 (Mathôt et al., 2012) with the PsychoPy backend was used for stimulus presentation. For CIE L*a*b* conversion to sRGB, the colormath Python package was used. Experiment 2 and 3 were written in JavaScript and HTML5, using the d3.js library for color conversion. Experiment 2 was run in Mozilla Firefox (68.0) and the online Experiment 3 was run on participants' computers using various browsers. For Experiment 3, participants' display size and distance from the screen were estimated via the methods of Q. Li et al. (2020). We used a central fixation dot (white; 0.18° in Experiments 1 and 0.16° in Experiments 2 and 3) against a gray background (RGB: [60, 60, 60], L* = 25.3, 14.2 cd/m² for Experiment 1 and 2). The sample display consisted of 33 vertical and 3 differently tilted (12°, 28° and 45°) colored bars subtending a visual angle of 1.30 × 0.33° each. The bars were arranged in three concentric rings (2°, 4° and 6° radius) with respectively 6, 12 and 18 bars on each. The relevant (tilted) bars were presented at a randomly chosen position on the middle ring. Colors were randomly drawn from a circle in a luminance plane of the CIE 1976 L*a*b* color space (L* = 63, center: a* = 9, b* = 27, illuminant: D65, 2° standard observer) with a radius of 40 (Mean ΔE2000 between two adjacent colors: 0.43). These parameters were chosen to ensure that all colors could be mapped onto the 24-bits sRGB color space. CIE L*a*b* is a device-independent color space based on the opponent color theory that aspires to be perceptually uniform, taking into account the specificities of the human color vision system (for a more detailed overview, see Fairchild, 2013). The color wheel (360 colors; randomly rotated in 30° steps) used to give the response had a width of 0.66° and a radius of 8°, 7.8°, or 7.1° in Experiments 1, 2 and 3, respectively. While the mouse hovered on the color wheel, the probe dynamically changed color according to the mouse position.

Analysis

Our analyses focus on the mean average absolute distance between the correct and the selected color (*recall error*). For statistical analyses, JASP 0.13.1 (JASP Team, 2020) was used with default settings for the Bayesian priors. Directed Bayesian *t* tests were conducted to analyze the differences between the different tilts. The *BF* quantifies the support for a hypothesis (first subscript) over another (second subscript), regardless of whether these models are correct. The subscript "0" always refers to the null hypothesis (H_0). When conducting undirected (two-sided) tests, the subscript "1" refers to the alternative hypothesis (H_1). When conducting directed (one-sided) tests, instead of "1", the subscripts "+" or "−" were used depending on the direction of the hypothesis (H_+ or H_-, respectively). Throughout the results, we will report the BF for the most favored hypothesis (e.g., if the null is more probable, BF_{01} will be reported), as we find it most intuitive to interpret.

We also conducted the traditional (frequentist) significance tests for reference and report effect sizes (Cohen's d_z) followed by their respective 95% *CI*s in brackets. Finally, as an exploratory analysis, we fitted the data from Experiment 1 and 2 – separately per participants and condition – to

[2] Withholding these participants from the analyses did not influence the pattern of results.

the mixture model of Zhang and Luck (2008).[3] This model (which is not without critiques, see Ma, 2018, preprint) assumes that the recall error arises from two sources represented by two parameters. The first parameter is the probability that the probed object is present in memory (p_{mem}). If the probed object is not in memory, the response will be made randomly. If the probed object is in memory, the second parameter reflects the precision of its representation (*sd*; higher *sd*s indicate lower precision). We extracted these parameters (Table S1; using MemToolbox; Suchow et al., 2013, https://memtoolbox.org/) and ran statistical analyses on them (Table S2). The below tables show the results for Experiment 2; the respective analyses for Experiment 1 are described in the main article. Due to the low number of trials per condition (25), we did not apply mixture-modeling to the data of Experiment 3.

Supplementary Results

Table S1

Descriptive Statistics for Mixture-Model Parameters Estimated from Data of Experiment 2

Parameter	Condition	*M*	*SD*	95% *CI* Lower	95% *CI* Upper
	Experiment 2 – *Mixed* Displays				
p_{mem}	12°	40.09%	20.67	32.50	47.67
	28°	66.41%	19.41	59.29	73.53
	45°	75.02%	15.53	69.32	80.72
sd	12°	29.75°	12.95	25.00	34.50
	28°	27.84°	10.43	24.01	31.66
	45°	23.63°	4.65	21.92	25.33
	Experiment 2 – *Same* Displays				
p_{mem}	12°	52.51%	18.71	45.65	59.38
	28°	68.39%	21.67	60.44	76.34
	45°	71.82%	21.94	63.77	79.87
sd	12°	26.53°	8.60	23.37	29.68
	28°	27.74°	8.30	24.70	30.79
	45°	26.76°	6.52	24.37	29.15

Details on Computational Modeling

Saliency model

The core of our saliency model is given by Equation 1, which states that an object *i*'s total saliency (s_{total}) is determined by the weighted (w_{rel}) sum of its absolute ($s_{abs(i)}$) and relative ($s_{rel(i)}$) saliency:

$$s_{total} = s_{abs(i)} + w_{rel} \cdot s_{rel(i)} \tag{1}$$

To keep the model as simple as possible, we assumed that the degree of tilt (t_i) (with respect to the non-targets) directly translates into an object's individual saliency ($s_{ind(i)}$). This sufficiently approximates the true transfer function for the present purposes as demonstrated by the model fit (see Table S3 and Figure 4 in the main document).

We implemented relative saliency as the object's individual saliency divided by the sum of all *k* objects' saliencies (including the object's own saliency; *divisive normalization*, Bays, 2014; Liesefeld & Müller, in press):

$$s_{rel} = \frac{s_{ind(i)}}{\sum_{j=1}^{k} s_{ind(j)}} = \frac{t_i}{\sum_{j=1}^{k} t_j}; i, j = 1, ..., k \tag{2}$$

Absolute saliency was defined as the individual saliency normalized by the maximal saliency (in the present design, saliency would be maximal for 90° tilted bars):

$$s_{abs(i)} = \frac{s_{ind(i)}}{s_{max}} = \frac{t_i}{90} \tag{3}$$

Template Model

Template mismatch (d_i) was defined as the difference between the tilt of the template (as estimated from the data via the free parameter t_t) and the individual tilt of each object (t_i):

$$d_i = |t_t - t_i|;\ \text{with } 0 \leq t_t \leq 180 \text{ and } d_i \leq 90 \tag{4}$$

Model fitting

To relate total saliency to performance in the present task (recall error averaged across participants, *re*) for the purpose of fitting the models to the empirical data, we used (out of convenience and to keep our modeling simple and agnostic with regard to the exact mechanisms linking saliency/template mismatch and VWM recall performance) a power-law function with the free parameters α and β (as we did in other contexts before, Liesefeld et al., 2016):

$$re_i = \alpha \cdot s_{total(i)}{}^{\beta} \tag{5}$$

If we had used the same transfer function for the template model, a di = 0 (i.e., a perfect template match) would predict $re = 0$. Thus, to predict non-perfect performance even for perfect template matches, we had to give this model extra flexibility by including an intercept term as a fourth free parameter:

$$re_i = \alpha \cdot d_i{}^{\beta} + \gamma \tag{6}$$

[3]Due to a technical mistake only the response and the correct answer were stored for Experiment 2, so that we could not apply other, more advanced models (e.g., Bays, 2014; Oberauer & Lin, 2017; van den Berg et al., 2012).

Table S2

Paired Samples t *Tests on Mixture-Model Parameters for Experiment 2.*

Comparison	t	df	d_z	BF	Favors
			Mixed displays		
p_{mem} – 12° vs. 28°	-9.66^{***}	30	−1.73 [−2.29, −1.17]	9.83e+7	H_1
p_{mem} – 28° vs. 45°	-4.71^{***}	30	−0.85 [−1.25, −0.43]	456.21	H_1
sd – 12° vs. 28°	0.87	30	0.16 [−0.20, 0.51]	3.68	H_0
sd – 28° vs. 45°	2.41^{*}	30	0.43 [0.06, 0.80]	2.26	H_1
			Same displays		
p_{mem} – 12° vs. 28°	-6.84^{***}	30	−1.23 [−1.69, −0.75]	1.11e+5	H_1
p_{mem} – 28° vs. 45°	−1.83	30	−0.33 [−0.69, 0.04]	1.19	H_0
sd – 12° vs. 28°	−0.68	30	−0.12 [−0.47, 0.23]	4.23	H_0
sd – 28° vs. 45°	0.73	30	0.13 [−0.22, 0.48]	4.07	H_0
			Mixed vs. *Same* displays		
p_{mem} – 12°	4.38^{***}	30	0.79 [0.38, 1.19]	201.01	H_1
p_{mem} – 28°	−0.73	30	0.13 [−0.48, 0.22]	4.09	H_0
p_{mem} – 45°	1.36	30	0.24 [−0.12, 0.60]	2.27	H_0
sd – 12°	1.04	30	0.19 [−0.17, 0.54]	3.19	H_0
sd – 28°	0.04	30	0.01 [−0.35, 0.36]	5.22	H_0
sd – 45°	-2.26^{*}	30	−0.48 [−0.85, −0.10]	3.70	H_1

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

The values of free parameters (w_{rel}, α, and β, or t_i, α, β and γ, respectively) were determined by a simplex routine (Nelder & Mead, 1965) as implemented as *fminsearch* in MATLAB, minimizing the sum of the squared differences between empirical and predicted recall performance (*SS*) per Tilt × Display Type cell (averaged across participants).

Modeling results and interpretation

As shown in Figure 4 of the main article, our saliency model quite accurately reproduced the observed data pattern. This model also accounts well for the data pattern in Experiment 3 (not shown). Notably, parameters α and β cannot affect the predicted data pattern, because the exact same transformation is applied to each total-saliency estimate from each cell of the respective experimental design. That is, the only free parameter used to account for the observed pattern is w_{rel}. By contrast, the template model failed to account for the difference between *mixed* and *same* displays (i.e., it cannot account for the effect of relative saliency) despite having one more free parameter than the saliency model (i.e., despite being less parsimonious).

Parameter estimates for the two models are given in Table S3. It is interesting to note that the estimated template is 42.40°, thus, quite close to the maximal target tilt (45°). Furthermore, w_{rel} was estimated at 0.57. A w_{rel} considerably above zero confirms an influence of relative saliency beyond the influence of absolute saliency on VWM performance.

Table S3

Estimated Parameters of Two Simple Models Linking Either Saliency (Relative and Absolute) or Match Between Each Object and an (Optimal) Template to Recall Error in Experiment 2.

Model	w_{rel}	t_t	α	β	γ	SS
Saliency	0.57	–	42.12	−0.42	–	3.61
Template	–	42.40	0.08	1.63	37.74	57.11