Supplemental Online Material (SOM) for *Massive Effects of Saliency on Information Processing in Visual Working Memory*

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. For the laboratory Experiments 1 and 2, we set a minimum of 10 and a maximum of 60 participants. 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; see preregistration for details). 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.[1] 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 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 three randomly chosen positions 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

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

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 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. Bayes factors (*BF*s) quantify 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-tailed) tests, the subscript "1" refers to the alternative hypothesis (H_1). When conducting directed (one-tailed) 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% CIs in brackets. Within-participants CIs displayed in the graphs were calculated according to the method of Morey (2008). Finally, as an exploratory analysis, we fitted the data from Experiment 1 and 2 – separately per participant and condition – to the mixture model of Zhang and Luck (2008)[2]. This model (which is not without critiques, see Ma, 2018) 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

				95% *CI*	
Parameter	Condition	*M*	*SD*	Lower	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

[2] 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

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 ($s_{\text{abs(i)}}$) and relative ($s_{\text{rel(i)}}$) saliency:

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

To keep the model as simple as possible, we assume that the degree of tilt (t_{i}) (with respect to the non-targets) directly translates into an object's individual saliency ($s_{\text{ind(i)}}$). 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):

$$s_{\text{rel}} = \frac{s_{\text{ind(i)}}}{\sum_{\text{j=1}}^{\text{k}} s_{\text{ind(j)}}} = \frac{t_{\text{i}}}{\sum_{\text{j=1}}^{\text{k}} t_{\text{j}}};\ i, j = 1, \dots, k$$

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):

$$s_{\text{abs(i)}} = \frac{s_{\text{ind(i)}}}{s_{\text{max}}} = \frac{t_{\text{i}}}{90}$$

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}):

$$d_{\text{i}} = |t_{\text{t}} - t_{\text{i}}|;\text{ with } 0 \leq t_{\text{t}} \leq 180 \text{ and } d_{\text{i}} \leq 90$$

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):

$$re_{\mathrm{i}} = \alpha \cdot s_{\mathrm{total(i)}}{}^{\beta}$$

If we had used the same transfer function for the template model, a $d_{\mathrm{i}} = 0$ (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:

$$re_{\mathrm{i}} = \alpha \cdot d_{\mathrm{i}}{}^{\beta} + \gamma$$

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 40.35°, 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	–	33.32	-0.42	–	3.61
Template	–	40.35	0.39	1.22	35.56	57.11

Additional References

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

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

JASP Team. (2020). *JASP* (0.13.1) [Computer software]. https://jasp-stats.org/

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), 1–11. https://doi.org/10/ggpbsf

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

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/ggbnjg

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

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

The analyses and figures in the main article provide only information on the average recall error. As providing only the mean can sometimes hide irregularities in the data, we provide here the error distribution for all experiments and conditions (Figure S1).

Figure S1. *Error distribution for all experiments*