

Distractor Filtering and Working Memory: Effects of Saliency at Encoding and Maintenance



EIRLYS MORGAN
SUPERVISED BY DR FIONA MCNAB

ACKNOWLEDGMENTS

The author would like to thank Lord Laidlaw and the Laidlaw Scholarship for providing the funding for this research project

Appreciation is further extended to Dr Fiona McNab for the ongoing support and input throughout the project

Abstract

A person's ability for distractor-filtering has the potential to limit working memory capacity (WMC). Considering attention constraints on memory, the current study aims to bridge the gap between attention and working memory research. With a working memory task, we investigate the effects of saliency on distractor-filtering at encoding and maintenance. We demonstrate unique contributions to WMC of salient encoding and salient delay distractors however fail to replicate this effect for non-salient stimuli, finding a unique contribution of only non-salient encoding distractors. We also find a unique contribution of non-salient encoding distractor-filtering when controlling for salient encoding distractors. We thus find mixed support for previous research. The limitations of the study are discussed in relation to these mixed findings, and it is concluded that whilst there is some evidence of separate mechanisms for salient and non-salient distractor-filtering at encoding vs maintenance, more research is necessary to untangle these.

Introduction

Working Memory is the ability to temporarily hold small amounts of information in mind for use in cognitive tasks (Baddeley & Hitch, 1974). This limited capacity storage system acts as a buffer between past knowledge and future action making it critical for language comprehension, problem-solving, planning, and many other aspects of cognition (Cowan, 2014). Decline in working memory is associated with psychological conditions such as ADHD and Schizophrenia (Luck & Vogel, 2013) as well as natural ageing and understanding the neural mechanisms that form the basis of working memory could aide future understanding of such deficits, leading to the development of better strategies for managing the impact of such a decline.

Attention constraints on WMC are a key area for exploration within cognitive psychology (Ricker et al., 2018) and have produced a wealth of research leading to these two cognitive domains increasingly being viewed as overlapping constructs. Given the importance of working memory in cognition, and its limited capacity, selectively attending to and encoding task-relevant information is critical. Neuroimaging research has established that a person's ability to ignore distractors can limit WMC, a finding that benefits our understanding of the interindividual variation in WMC. Vogel et al. (2005) suggest that people with a high WMC

demonstrate greater distractor-filtering efficiency. Participants completed a visual memory task requiring them to selectively remember items from an array. Electroencephalography (EEG) recordings from participants with a high WMC revealed smaller event related potential (ERP) waveforms in brain areas associated with the encoding and maintenance of visual memory (Vogel & Machizawa, 2004) for arrays containing both relevant and irrelevant stimuli than participants with smaller WMC. Since these results indicate the distractors were not consuming memory space, researchers concluded that participants with a high WMC exhibit not only an increased capacity for remembering, but a greater ability for distractor-filtering. This was supported by research measuring ERP's during distractor-filtering which found that high WMC is associated with the suppression of distracting information (Gaspar et al., 2016; Zanto & Gazzaley, 2009). These findings appear to confirm that WMC is limited not only by enhancing attention to task-relevant information, but more so the active suppression of task-irrelevant information.

McNab and Klingberg (2008) extended this finding, aiming to isolate the control mechanisms associated with the selection of information for storage in memory using fMRI. Participants completed a visuospatial memory with distractors presented during the encoding of information. Results showed a negative correlation between activity in the left basal ganglia, and brain activity reflecting the unnecessary storage of distracting information (as studied by Vogel et al., (2005)). A positive correlation was found between WMC and basal ganglia activity. The time course of this brain activation suggests that the basal ganglia are involved in the preparation of selecting information to be encoded. Consistent with previous findings that WMC is limited by distractor-filtering ability, McNab and Klingberg (2008) present evidence for a neural mechanism by which control is exerted over the selective encoding of information. It seems therefore, that performance in working memory tasks depends on the ability to filter information in stages of processing as early as encoding to prevent relevant information being overlooked by a limited capacity system.

McNab and Dolan (2014) studied the attentional mechanisms behind distractor-filtering at encoding and maintenance. Results from both a laboratory-based experiment, and a smartphone game revealed that distractor-filtering at both encoding and maintenance uniquely predicted WMC. This indicates separate mechanisms for the contribution to WMC of distractor-filtering at encoding vs maintenance, and the researchers propose that they

constitute two separate bases for working memory.

The attention literature recognises that the allocation of attention is based on more than just top-down control mechanisms (Wolfe & Horowitz, 2017). One such influence is saliency which, whilst largely unstudied in working memory research, can have a dramatic effect on attention. Early research proposed the stimulus-driven account of attentional capture. A visual search task by Theeuwes, (1992) found that attention was drawn automatically to unique colour singletons within an array. Researchers argue this as evidence that attention is controlled primarily by bottom-up features such as saliency and is automatically allocated to items with unique ‘pop-out’ features. This supports work by Pashler (1988) who found that items with single unique features can capture attention.

However, this idea that attentional capture can be driven purely by stimulus salience, irrelevant of task goals, is a topic of debate. Mounting evidence suggests that the salience based attentional capture evidenced above is contingent on top-down goal-driven processes. The biased competition model of attention proposes that relevant and non-relevant stimuli must compete for processing space, and that top-down control mechanisms automatically bias attention to these bottom-up features based on stimulus attributes and task goals to process the most relevant information (Desimone & Duncan, 1995). Accordingly, physically salient stimuli will capture the attention of an observer only if they match the attentional bias stemming from the current goal. Consequently, non-targets that have similar features may get caught in the attentional bias and will be attended to despite their irrelevance to the current goal. A pupillometry study by Olmos-Solis et al. (2018) appears to support this theory. Participants completed a visual search task with distractors shown prior to beginning the search. Longer pupil restriction was observed during the delay when the distractor matched the search target, indicating that target matching but not non-matching stimuli capture attention.

Sawaki and Luck (2010) dispute the validity of the biased competition theory’s claim that items matching the attentional bias template are automatically encoded. Results from their study indicated that whilst salient stimuli were detected by the visual system, they were not always attended to. The researchers therefore present the signal suppression hypothesis. This claims that salient items have an intrinsic ability to attract attention however, active top-down

control can suppress the signal and prevent task-irrelevant items being processed. Whilst this hypothesis shares similarities with the biased competition theory, it eliminates the automatic element of attentional capture instead suggesting an effortful top-down suppression mechanism. Sawaki and Luck (2011) extended this hypothesis into the mechanism of working memory. Participants were shown a memory array of two stimuli followed by two distractors. EEG results showed activation of brain areas associated with attention suppression when participants viewed memory-matching distractors. This again indicates that attention is not automatically captured by salient information, even when it matches an attentional template. Despite criticism by Wang and Theeuwes (2020) that the small array size resulted in the distractors not being salient enough to capture attention, Stilwell et al. (2022) were able to replicate the results of Sawaki and Luck (2011) using an increased set size and confirmation via a computational model that the stimuli were highly salient.

Considering the large effects of saliency on attention described above and recent research establishing saliency effects on WMC (Constant & Liesefeld, 2021), as well as the variation in WMC following distraction at encoding vs maintenance, the current study aims to explore whether the impacts of saliency on WMC differ between distractor-filtering at maintenance and encoding.

It is hypothesised that:

Distractor-filtering at encoding and maintenance will uniquely predict WMC with both salient and non-salient distractors

Distractor-filtering in salient and non-salient conditions will uniquely predict WMC

Method

Participants

41 participants (29 male, 12 female) aged 18-40 ($M=27.667$, $SD=7.256$) gave informed consent to take part in this study. Participants were recruited via online platform Prolific and received a monetary reward. All participants had normal or corrected to normal vision, normal colour vision, and no history of neurological impairment or psychiatric illness.

Data points from 2 participants were removed as outliers (values ≤ 0), leaving a final sample size of 39.

Design

Repeated measures design. Independent variables were saliency, operationalised as salient (pink) and non-salient (yellow) distractors, and distractor presentation period (encoding or maintenance). The dependent variable was estimated WMC, represented by Cowan’s K value.

Trial and condition presentation was counterbalanced across participants to control for order and practice effects.

Materials and Procedure

The experiment was built using Gorilla experiment builder and run online via a URL. Participants completed two working memory tasks, the first to gain a WMC estimate and the other to assess the impact of saliency on distraction at encoding and maintenance. See figure 1 for a visual depiction of the tasks

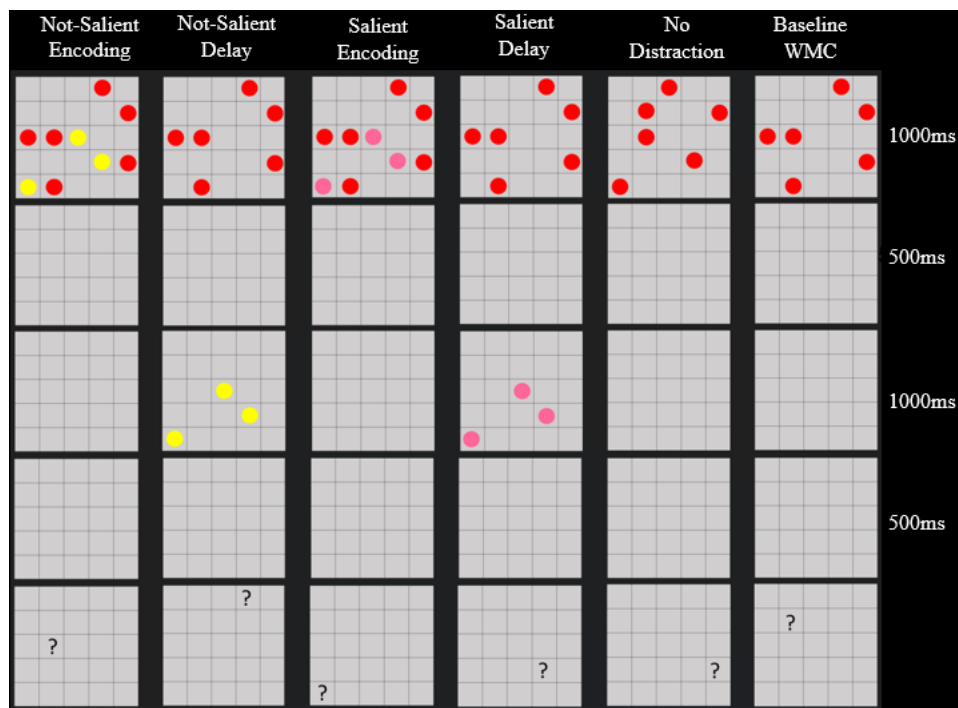


Figure 1. example of the task used to obtain working memory estimates

WMC Estimate

To measure WMC, participants were asked to remember the location of 6 red circles displayed on a 5x5 grid presented for 1000ms. After a 2000ms delay, in which a blank grid was displayed, participants responded to a probe stimulus (a question mark) by pressing a key to indicate whether the probe was in the same position as one of the targets. There were 40 trials in this condition, probes were presented in, and adjacent to a target position for 20 trials each.

Distraction at Encoding and Maintenance

This task was comprised of 5 conditions with 40 trials each: No distraction (ND), Non-salient delay distraction (NSDD), Non-salient encoding distraction (NSED), Salient delay distraction (SDD), Salient encoding distraction (SED). 6 red target circles were presented simultaneously on a 5x5 grid for 1000ms followed by a 2000ms delay in which the grid remained on the screen. A probe (a question mark) was shown on the grid after the delay period either in or adjacent to a target position.

In the ND condition, no distractor circles were displayed at any point during the trial.

In the NSED condition, 3 yellow distractor circles were presented alongside the targets

In the NSDD condition, 3 yellow distractor circles were presented for 1000ms during the delay period 500ms after the targets had disappeared.

In the SED condition, 3 pink distractor circles were presented alongside the targets

In the SDD condition, 3 pink distractor circles were presented for 1000ms during the delay period 500ms after the targets had disappeared.

Within conditions, the location of the target and distractor circles differed on every trial. At least one distractor was placed adjacent to a target on every trial. Trials were matched between conditions to ensure an equal level of difficulty for each condition.

Participants completed 5 blocks, each comprising of 40 trials. Within each block, trials were divided equally between each condition (8 trials per condition per block). The location of the probe was in and adjacent to a target position for 20 trials each. In adjacent trials, the probe was placed in the position of a distractor 50% of the time.

Prior to the task, participants were shown a pink, yellow, and red circle to determine whether they identified them as different colours. This was to ensure a level of consistency across participants as the task was completed using their own desktop so lighting, brightness, and resolution levels could not be monitored.

Participant were required to remember the location of only the red circles, ignoring the yellow and pink circles and to indicate with a yes/no response whether the location of the probe had previously been occupied by a red circle.

Results

WMC was estimated using Cowan’s K value. This is a standard formula for estimating the number of items that can be held in working memory (Cowan, 2001; McNab & Dolan, 2014); $K=S(H-F)$. K is WMC, S is array size, H is the hit rate, and F is the false alarm rate. 6 K values were calculated per participant (1 per condition). Means for each condition are shown in table 1. Zero-order correlations for conditions are shown in table 2, all K values from all conditions correlate significantly.

Table 1.

Mean WMC (K values) for each condition

No Distraction	WMC	Not-salient Encoding	Not-salient Delay	Salient Encoding	Salient Delay
5.112 (0.969)	4.755 (1.186)	4.883 (1.042)	5.071 (1.068)	5.118 (0.978)	4.989 (1.104)

Table 2.

Results of the zero-order correlations (R) between each variable

	ND	WMC	NSED	NSDD	SED	SDD
ND	-	.672**	.740**	.736**	.556**	.787**
WMC	-	-	.729**	.594**	.610**	.615**
NSED	-	-	-	.796**	.826**	.804**
NSDD	-	-	-	-	.605**	.86**
SED	-	-	-	-	-	.548**

***correlation is significant at the .01 level (two-tailed)*

ANOVA

Data met the assumption of normality but not the assumption of sphericity using Mauchley's test $W(9) = 4.05, p = <.001$.

The results of a repeated measures ANOVA applying the Greenhouse-Geisser correction revealed no significant differences between conditions $F(2.76, 104.79) = 0.523, p = .652$.

Regression

Separate regression analyses revealed that K values for NSED and NSDD both significantly predict WMC (NSED: standardised $\beta = .729, p = <.001$, NSDD: standardised $\beta = .594, p = <.001$).

Hierarchical regression controlling for performance in the NSDD condition revealed K values for the NSED condition uniquely predicted WMC (standardised $\beta = .699, p = <.001$). When controlling for the variance associated with the NSED condition, K values for the NSDD condition did not uniquely predict WMC (standardised $\beta = .037, p = .845$).

Two separate regression models revealed that K values for SED and SDD both significantly predict WMC (SED: standardised $\beta = .610, p = <.001$, SDD: standardised $\beta = .615, p = <.001$).

A further hierarchical regression model showed that K values for SED and SDD uniquely and significantly predict WMC when controlling for the other condition (SED: standardised $\beta = .391, p = .010$, SDD: standardised $\beta = .401, p = .008$).

As there was a unique contribution to WMC of NSED but not NSDD, a hierarchical regression model was computed for SED and NSED. This revealed that K values for NSED uniquely predict WMC when controlling for SED (standardised $\beta = .706, p = .001$), but no unique contribution of SED when controlling for NSED (standardised $\beta = .027, p = .896$).

Discussion

The positive correlation between WMC, encoding distraction, and delay distraction is consistent with previous findings (McNab & Klingberg, 2008; Vogel et al., 2005) suggesting high WMC is associated with increased distractor-filtering ability. The lower mean WMC estimate compared to the ND condition is likely attributable to practice effects as it was the

first task all participants completed.

The current study supports the hypothesis that SED and SDD both uniquely predict WMC, suggesting separate mechanisms for encoding and delay distractor-filtering for salient distractors. However, the results fail to replicate the findings of McNab and Dolan (2014). Despite a unique contribution of NSED to WMC, NSDD did not uniquely predict WMC. We posit this replication failure may be due to the inclusion of salient and non-salient stimuli within the same task. Previous research has shown differences in brain activity when anticipating distractors (Bonnefond & Jensen, 2012), it is therefore conceivable that the uncertainty surrounding what colour distractor to expect may interfere with the distractor-filtering mechanisms proposed by McNab and Dolan (2014). Another explanation could be that this uncertainty affects the processing of the bottom-up ‘pop-out’ features often associated with these yellow distractors (see Pashler (1988); Theeuwes (1992); Wang and Theeuwes (2020)). Considering this suggestion, future research could separate tasks for salient and non-salient distractors. Under this circumstance, participants will know what colour distractor may appear, and researchers might find results more comparable to those of McNab and Dolan (2014).

In accordance with the smartphone, but not laboratory, study by McNab and Dolan (2014), we found a positive correlation between encoding and distraction conditions for both salient and non-salient distractors suggesting shared variability between the two mechanisms. In their paper, McNab and Dolan (2014) theorise this variability may represent a common mechanism that acts differently in conditions of encoding and distraction. Considering the grid stimuli used in the smartphone, and this, study which may have made labelling positions easier than the circular grid used for the laboratory study, we tentatively suggest this shared mechanism may also contribute differently dependent on task difficulty.

The hypothesis that salient and non-salient distractor-filtering will each uniquely predict WMC is not supported. Due to the failure to replicate previous results, only the encoding conditions were input into a regression analysis. The unique contribution of NSED but not SED to WMC indicates similar mechanisms for distractor-filtering. It could be that the mechanism for both salient and non-salient distractor-filtering requires competition resolution

for the allocation of attention between distractors and targets, but that NSED requires an additional mechanism for preventing the capture of attention by ‘pop-out’ features. However, further study is required to unravel the complexities of this mechanism.

One thing to consider on reflection of this experiment is the definition of saliency. As the targets in our experiment were red, salient distractors were pink as it is thought their similarity to red may result in them being captured by an attentional bias (Desimone & Duncan, 1995; Olmos-Solis et al., 2018; Sawaki & Luck, 2010). However, it may be the case that the non-salient (yellow) stimuli could also be considered salient due to their contrast with the surrounding scene. This pop-out feature may attract attention due to its bottom-up saliency (Theeuwes, 1992; Wang & Theeuwes, 2020), and potentially due to its semantic meaning. Yellow is often used to signal important information e.g., caution signs. Given this, it would perhaps be prudent to replicate the experiment using different colours: for example, dark blue targets, light blue salient distractors, and green non-salient distractors. This may help to separate similarity vs saliency effects.

Conclusion

This study presents evidence for differences in the contribution to WMC of salient and non-salient distractor-filtering. While it provides mixed support for previous findings it is, to the best of our knowledge, the first to analyse the effect of saliency at encoding and maintenance and highlights the need for further research.

REFERENCES

- Baddeley, A. D., & Hitch, G. (1974). Working Memory. *Psychology of Learning and Motivation - Advances in Research and Theory*, 8(C), 47–89.
[https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Bonnefond, M., & Jensen, O. (2012). Alpha Oscillations Serve to Protect Working Memory Maintenance against Anticipated Distracters. *Current Biology*, 22(20), 1969–1974.
<https://doi.org/10.1016/J.CUB.2012.08.029>
- Constant, M., & Liesefeld, H. R. (2021). Massive Effects of Saliency on Information Processing in Visual Working Memory. *Psychological Science*, 32(5), 682–691.
https://doi.org/10.1177/0956797620975785/ASSET/IMAGES/LARGE/10.1177_0956797620975785-FIG2.JPEG
- 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>
- Cowan, N. (2014). Working Memory Underpins Cognitive Development, Learning, and Education. *Educational Psychology Review*, 26(2), 197. <https://doi.org/10.1007/S10648-013-9246-Y>
- Desimone, R., & Duncan, J. (1995). NEURAL MECHANISMS OF SELECTIVE VISUAL ATTENTION. *Annu. Rev. Neurosci*, 18, 193–222. www.annualreviews.org
- Gaspar, J. M., Christie, G. J., Prime, D. J., Jolicoeur, P., & McDonald, J. J. (2016). Inability to suppress salient distractors predicts low visual working memory capacity. *Proceedings of the National Academy of Sciences of the United States of America*, 113(13), 3693–3698.
https://doi.org/10.1073/PNAS.1523471113/SUPPL_FILE/PNAS.201523471SI.PDF
- 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>
- McNab, F., & Dolan, R. J. (2014). Dissociating distractor-filtering at encoding and during maintenance. *Journal of Experimental Psychology: Human Perception and Performance*, 40(3), 960. <https://doi.org/10.1037/A0036013>
- McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *NATURE NEUROSCIENCE VOLUME*, 11(1).
<https://doi.org/10.1038/nn2024>
- Olmos-Solis, K., van Loon, A. M., & Olivers, C. N. L. (2018). Pupil Dilation Reflects Task Relevance Prior to Search. *Journal of Cognition*, 1(1). <https://doi.org/10.5334/JOC.12>
- Pashler, H. (1988). Cross-dimensional interaction and texture segregation. *Perception Cl Psychophysics*, 43(4), 307–318.

- Ricker, T. J., Nieuwenstein, M. R., Bayliss, D. M., & Barrouillet, P. (2018). Working memory consolidation: Insights from studies on attention and working memory. *Annals of the New York Academy of Sciences*, 1424(1), 8–18. <https://doi.org/10.1111/NYAS.13633>
- Sawaki, R., & Luck, S. J. (2010). Capture versus suppression of attention by salient singletons: Electrophysiological evidence for an automatic attend-to-me signal. *Attention, Perception, & Psychophysics* 2010 72:6, 72(6), 1455–1470. <https://doi.org/10.3758/APP.72.6.1455>
- Sawaki, R., & Luck, S. J. (2011). Active suppression of distractors that match the contents of visual working memory. *Http://Dx.Doi.Org/10.1080/13506285.2011.603709*, 19(7), 956–972. <https://doi.org/10.1080/13506285.2011.603709>
- Stilwell, B. T., Egeth, H., & Gaspelin, N. (2022). Electrophysiological Evidence for the Suppression of Highly Salient Distractors. *Journal of Cognitive Neuroscience*, 34(5), 787–805. https://doi.org/10.1162/JOCN_A_01827
- Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics* 1992 51:6, 51(6), 599–606. <https://doi.org/10.3758/BF03211656>
- Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature* 2004 428:6984, 428(6984), 748–751. <https://doi.org/10.1038/nature02447>
- Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature* 2005 438:7067, 438(7067), 500–503. <https://doi.org/10.1038/nature04171>
- Wang, B., & Theeuwes, J. (2020). Saliency Determines Attentional Orienting in Visual Selection. *Journal of Experimental Psychology: Human Perception and Performance*. <https://doi.org/10.1037/XHP0000796>
- Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour* 2017 1:3, 1(3), 1–8. <https://doi.org/10.1038/s41562-017-0058>
- Zanto, T. P., & Gazzaley, A. (2009). Neural Suppression of Irrelevant Information Underlies Optimal Working Memory Performance. *Journal of Neuroscience*, 29(10), 3059–3066. <https://doi.org/10.1523/JNEUROSCI.4621-08.2009>