

Properties of a flexible visual short-term memory resource

Holly A. Lockhart, MA

Psychology (Behavioural and Cognitive Neuroscience)

A thesis submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

Faculty of Social Sciences, Brock University
St Catharines, Ontario

© Lockhart 2024

General Abstract

For the last few decades, there has been considerable debate as to whether visual short-term memory (VSTM), the capacity limited memory system for the short-term storage of visual information, is a continuous or discrete resource. One property that has been identified that is consistent with a continuous resource is flexible allocation; in a delayed-recall task, when cues manipulate the priority (task relevance) of the sample items to decouple resource allocation from set-size, performance has been found to vary with the proportion of allocated resources. However, the extent and limits of this property have yet to be identified. Moreover, discrete resources could account for some previous findings through probabilistic encoding.

The current thesis aimed to examine the properties and limits of flexible resource allocation in VSTM. In Chapter 2, I examined whether VSTM resources could be allocated to three levels of attentional priority. Although possible, examining individual differences in the strategies participants used revealed that the majority of participants do not use all three priority levels.

Chapter 3 investigated an alternative resource allocation strategy, whereby the strategic use of a discrete memory resource to store the most relevant memory items would be encoded probabilistically. Response precision better matched predictions of flexible allocation. Further, I directly tested probabilistic encoding by estimating the proportion of “in-memory responses” and comparing it with individuals’ estimated capacity. Results again did not support a probabilistic encoding strategy.

A criticism of flexible allocation is that very low precision memories are indistinguishable from out-of-memory responses. In Chapter 4, I examined flexible

Flexible Properties of VSTM

allocation using a two alternative forced choice (2AFC) task intermixed with continuous response trials, as 2AFC can show evidence of weak memory through better-than-chance recognition performance. Results demonstrated that participants performed better than chance for very low-priority items. This suggests that these items are stored as low-resolution memory representations, rather than being out of memory altogether.

Collectively, these studies reveal properties of a flexible VSTM resource. Taken together, these data further suggest that any model that cannot accommodate a dynamic, flexible resource should be abandoned.

Key words: Visual short-term memory, attention, flexible allocation, capacity, individual differences, visual working memory

Acknowledgments

I would like to thank my advisor, Dr. Stephen Emrich, for his guidance, patience, motivation, and giving me the opportunity to study the brain – Brains are weird. I often tell potential students how important the match with a graduate supervisor is; I would not be at this stage without you. Thank you. I am thankful to my committee members, Dr. Karen Campbell, and Dr. Karen Arnell, for their guidance, questions, and suggestions. Miraculously, in the final edit, none of their suggestions conflicted. I would also like to thank to Dr. Naseem Al-Aidroos and Dr. Blaire Dube, who I was fortunate enough to work with on several projects throughout the years, including Chapter 2.

This endeavor would not have been possible without funding or participants. I would like to acknowledge my gratitude for the external grants, scholarships, fellowships, and direct funding I received from my supervisor's grants in addition to covering research costs and conference travel. I would also like to recognize all of the participants, especially those who were willing and excited to do fMRI and EEG experiments.

I would like to express many thanks to my fellow graduate students for their support, advice, and time both in-person and online: Christine Salahub, Joel Robitaille, Kari Lustig, Lindsay Platter, Kristen Baker, Brent Pitchford, Andrew Chung, Rosa Torres Hernandez, Alyssa Thibeault, and others too many to name over the years. Thank you to the research assistants and lab managers who helped with data collection, especially Joseph Capozza, Mallory Duprey, and Keynen Lynett. Thank you to Thomas Nelson for helping program the first versions of the colour wheel experiments I worked on in 2014 and Tyler Kennedy Collins for helping to code and debug the most recent versions.

My sincerest gratitude extends to my partner, Kevin MacDonald: for every coding error you helped debug, for every grammatical conundrum you helped resolve, for every dinner you made alone, for listening and encouraging me... Thank you.

Lastly, I would not have completed this without the support I received outside of academia. I would also like to express my gratitude to Alexandra Medland, my sisters Kyrinne and Emily, my parents Michael and Esther, the MacDonald Clan, my cats Calcifer and Ventress, my therapist, my dance partners, and my public library.

Table of Contents

General Abstract	
Acknowledgments	
List of Tables	
List of Figures	
List of Abbreviations and Symbols	
1. Chapter 1: General Introduction on the flexible allocation of memory resources	1
The Debated Nature of VSTM Resources	1
Flexible Allocation of Memory Resources	6
Summary of the present thesis.	7
References	11
2. Chapter 2: Limitations on flexible allocation of visual short-term memory resources with multiple levels of goal-directed attentional prioritization	15
Abstract	16
Introduction	17
Methods	20
Participants	20
Participants Experiment 1a	20
Participants Experiment 1b	21
Participants Experiment 2	21
General Procedure	21
Procedure and Stimuli Experiments 1a and 1b	23
Procedure and Stimuli Experiment 2	23
General Analysis	24
Results	24
Experiments 1a and 1b	24
Experiment 2	27
Combined Experiment Data Analyses	28
Temporal Analysis	29
Method	29
Results	30
Bayesian Individual Differences Analysis	31
Method	32
Results	32
Discussion	33
Temporal Analysis	36
Individual Differences Analysis	37
Limitations	39
Conclusion	40
References	41
Appendix 2A: Supplementary analysis: Absolute error.	47
3. Chapter 3: In-memory items are not equal: multiple item report reveals differential memory precision and not proportional memory access for prioritized items.	50
Abstract	51
Introduction	52
Disentangling the mechanism of response precision.	55
Predictions of flexible allocation on multiple responses' precision	56
Predictions of probabilistic encoding on multiple responses' precision	57
In-memory low-priority responses: an alternative test of the predictions of probabilistic encoding.	58
The current study.	58
Methods	59
Participants	59
Procedure and Stimuli:	60
Apparatus.	63

Flexible Properties of VSTM

Analysis of Response Precision.	63
Analysis of discrete resource in-memory responses: observed versus predicted.	64
Results	69
Response Precision	70
First versus Second Responses: condition by order interactions	70
Mixed priority versus same-priority responses	72
Probabilistic Encoding: Predicting High-resolution In-memory Responses	73
Discussion	76
Response Precision	77
Variable Precision.	78
Predicted High resolution In-memory Response Proportions	81
The meaning of “Not in memory”	82
References	85
Appendix 3A: Equivalent precision for high and medium priority items in Experiment 2.	89
Response Precision	89
Response Order	92
Appendix 3B: Alternate assumptions for predicting the proportion of in-memory responses.	94
Appendix 3C: Alternate definition of “in-memory”	96
4. Chapter 4: Recall procedure reveals low-priority objects are stored in memory.	97
Abstract	98
Introduction	99
Flexible allocation of VSTM Resources	102
Memory retrieval	105
The current study	106
Method	106
Participants	106
Procedure	107
Analysis	111
Continuous Response	111
Two Alternative Forced Choice	112
Comparing performance by response type	114
Results	114
Continuous Report Precision	114
2AFC Precision	116
Comparing Precision	117
Psychometrics of AFC Performance	118
Response Time	121
Discussion	121
Limitations	126
Conclusion	128
References	129
Appendix 4A: Supplementary Figure 2AFC Response Density	134
Appendix 4B: 2AFC Miss RT Results	135
5. Chapter 5: General Discussion	136
Investigating Response Behaviours	138
Predicting High-Threshold Responses	139
What Recognition Memory can say about Recall Memory	141
Signal Detection	142
Understanding Error Distributions; a Limitation to Interpretation	144
Conclusion	148
References	150

List of Tables

Table 2-1 Summary of Response Strategy Model Parameters	32
Table 3-1 Probability that the probed item was in-memory for ‘low-low’ test trials by capacity estimates	67
Table 4-1 Summary of mixed-effect models of 2AFC accuracy	119

List of Figures

Figure 2-1 Example trials from each experiment	22
Figure 2-2 Results of Experiment 1a	26
Figure 2-3 Results of Experiment 1b	26
Figure 2-4 Results of Experiment 2	27
Figure 2-5 Circular standard deviation of error from combined data by priority condition	28
Figure 2-6 Absolute error for each priority condition by trial number (time)	30
Figure 2-A1 Mean absolute error for all experiments.	48
Figure 3-1 Flexible Allocation versus probabilistic encoding	54
Figure 3-2 Continued from Figure 3-1, now showing two items tested	56
Figure 3-3 Example trial for Experiments 1 and 2	61
Figure 3-4 Raincloud plots of response precision for first and second responses by condition	71
Figure 3-5 Predicted versus observed proportions of low-low in-memory responses	75
Figure 3-A1 Bar graphs for standard deviation of error for Experiment 2 by priority conditions and response order	90
Figure 3-A2 Response density distributions of Experiment 2.	91
Figure 3-A3 Proportion of first responses made to either priority level by condition	93
Figure 3-B1 Predicted In-memory responses by whole number capacity estimates	94
Figure 4-1 Sample trials of continuous response and 2AFC test	111
Figure 4-2 Continuous report performance by probe probability conditions	115
Figure 4-3 2AFC performance by probe probability conditions	116
Figure 4-4 Comparison of circular SD performance between response types	118
Figure 4-5 2AFC performance measures by task difficulty	120
Figure 4-A1 2AFC response density showing non-continuity of responses	134
Figure 4-B1 Miss response times by discrimination bin for probe probability conditions	135

List of Abbreviations and Symbols

Abbreviation	Definition
2AFC	Two alternative forced choice
ANOVA	Analysis of Variance
<i>BF₀₁</i>	Bayes Factor in favour of the null hypothesis, presented in italics to more easily distinguish from BF ₁₀
BF ₁₀	Bayes Factor in favour of the alternate hypothesis, also given as log(BF ₁₀) for strong or very strong evidence.
BOLD	Blood Oxygen Level Dependent signal
CDA	Contralateral Delay Activity
CI	Confidence Interval (95% unless otherwise stated)
Δ	Delta, difference
°	degree
EEG	Electroencephalography
ERP	Event Related Potential (in EEG)
fMRI	Functional Magnetic Resonance Image
<i>K</i> , <i>K</i> -estimate	Memory capacity
s, ms	Seconds, milliseconds
SD	Standard Deviation
RT	Retrieval Time
vs.	versus
VSTM	Visual short-term memory
VWM	Visual working memory, also referred to as VSTM
X^2	Chi-square

Chapter 1: General Introduction on the flexible allocation of memory resources

How do we store information for short-term use? Visual short-term memory (VSTM) is a system for storing visual information for short periods, however, this system has a very limited capacity. Some theories propose a discrete item-based limit of 3-4 items (Luck & Vogel, 1997), but other theories propose a limited pool of a continuous resource (Bays et al., 2008). These theories make distinctly separate claims on the flexibility of VSTM resource allocation: where a discrete resource leads to two possible states (i.e., in memory, or not in memory), a continuous resource allows for more possibilities where greater numbers of items can be in memory at a cost to memory quality. The question of how we store the information in VSTM is an ongoing debate of VSTM models. The focus of this thesis is the properties of flexible allocation of the VSTM resource. This flexibility can be demonstrated through the maintenance of low-resolution memories at high memory loads, and also by the fine control of memory resource allocation in accordance with task relevance. The prevailing debate on the nature of VSTM memory resources has focused on whether resources could be flexibly distributed, but not yet on the limits of flexibility. Through three studies presented in this thesis, I will establish that the flexibility of resource allocation cannot be overlooked in the debate of VSTM models.

The Debated Nature of VSTM Resources

Visual short-term memory is the system for holding visual information “in mind” for a brief period when it is no longer in front of you (Postle, 2016). VSTM has a very limited memory capacity, it is often equated to only being able to store 3-4 items (Cowan, 2001; Luck & Vogel, 1997). The limited capacity of VSTM can be demonstrated through the set-size effect on response accuracy, as in Luck and Vogel (1997) where performance

accuracy declines as memory load increases. Specifically, in a change detection task with set-sizes between one and twelve, response accuracy declined systematically after set-size 3 in a manner that suggested roughly four items were held in memory storage (Luck & Vogel, 1997). This model has been described as a “slot” of memory or a quantized unit, similar to a juice box that can be distributed but only in whole units (Zhang & Luck, 2008). This discrete memory resource is conceptualized as an all-or-none process, whereby the features of items stored in memory will be remembered accurately but if the number of items exceeds capacity some items will not be in memory at all.

Discrete resource theories found support in an interpretation of event-related potentials (ERPs) from the lateral posterior occipital channels called the contralateral delay activity (CDA). The CDA is a difference wave between the ipsilateral and contralateral sides of a balanced perceptual display measured during the maintenance period of a recall task. Critically, although the displays are perceptually balanced, one side is indicated to be attended and memorized without moving the eyes from central fixation. Thus, the only intended difference is the cognitive processing given to the memorized side which is measured as a difference in the ipsilateral and contralateral waveforms. The CDA was found to get progressively more negative in a stepwise fashion as item load was increased up to typical memory capacity and then plateaued (Vogel & Machizawa, 2004). Similarly, univariate BOLD signal from the intraparietal sulcus (IPS) was found to increase in a stepwise manner up to capacity as load increased (Todd & Marois, 2004). The asymptote points of these neural correlates of VWM load were found to strongly correlate with individual participants’ VWM item limit (i.e., memory capacity, K) providing evidence that these signals were markers of VWM storage itself (Fukuda &

Vogel, 2009; Todd & Marois, 2005; Vogel et al., 2005). However, recently this interpretation has been challenged, suggesting that the CDA is actually indicative of attentional and executive processes (Emrich et al., 2022).

At the same time that discrete models were finding support in neural signals, Wilken and Ma (2004) published a new way to measure VSTM performance that would lead to revolutionary changes in the interpretation of memory recall errors. They argued that the predominant testing paradigm at the time, the change detection task, was not theory-neutral in that it assumes that memory “is encoded as a discrete unit within the brain” (p. 1121, Wilken & Ma, 2004). From a signal detection perspective, the change display is compared to the memorized display where it must overcome internal noise to achieve a discrete *hit* response. Wilken and Ma reasoned that the discrete nature of high-threshold responses in change detection masks the continuous nature of perceptual encoding of memorized information. In fact, Pashler (1988) stated that the calculation of K-estimates with hit rates and false alarms that he proposed “admit[s] no partial information and attribute[s] all errors to maintenance rather than the comparison process” (p. 370). Instead, Wilken and Ma (2004) proposed that memory could be better modelled from a signal detection framework using a continuous resource with gaussian noise.

To overcome the limitation of high-threshold responses in change detection, Wilken and Ma (2004) had participants report the colour feature of memorized items on a continuous colour wheel known as a continuous recall task or delayed-estimation. In continuous recall, the response error is measured as the difference between the reported value and the target value, usually in degrees. Each response error value can be plotted on a frequency distribution that takes the shape of an assumed circular normal distribution

which can then be summarized by the circular standard deviation to measure memory precision. These response distributions demonstrated the innate noise in memory expected from a continuous resource. Further, Wilken and Ma suggested that the internal noise of the memory representation increased monotonically with higher set sizes which explained the lower change detection performance at higher set sizes since the noise in the memory representation would be more difficult to overcome. Thus, the analysis of response error distributions demonstrated that memory precision is not set uniformly in an all-or-none fashion as predicted by discrete-limited theories of VSTM.

Response error distributions have since become the gold standard method of measuring memory performance; beyond memory precision these responses can be mathematically decomposed into component parts that are purported to represent different kinds of memory responses. In a landmark study, Zhang & Luck (2008) demonstrated that the response error from continuous report tasks could be broken down into two underlying distributions which represented the capacity and resolution of memory independently. Termed the two-component mixture model, response distributions were separated into a uniform guessing distribution and a normally distributed memory distribution. Instead of memory precision being continuously variable, the underlying uniform distribution took away responses that were not from memory, leaving a normal distribution that represented the in-memory responses.

This mathematic model of response error prompted two possible modifications to the discrete-limit resource model to explain the variable resolution of memory. Zhang and Luck first proposed a variant, the *slot + resource* model which theorized two types of memory resources: one discrete resource that determines memory capacity; and one

continuous resource pool that is distributed among in-memory items to determine the resolution of memory. Alternatively in the second variant, the *slot + averaging* model, achieves this with a single quantized discrete resource. In this theory, the number of memory slots determines both the capacity and resolution: when information stores are below capacity, items may be stored in multiple slots independently and the average of the repeated representations increases the resolution of memory (Zhang & Luck, 2008).

Alternatively, another parsimonious model conceptualizes memory as a limited pool of continuous memory resources. These models interpret the observed decline of precision to suggest that there is a shared pool of continuous resources (Bays et al., 2009; Fougne et al., 2012; van den Berg et al., 2012). Importantly, discrete and continuous resource models differ in the predicted amount of flexibility in their allocation. The Zhang and Luck *slot + averaging* model strictly limits the flexibility of resource allocation, whereas a continuous resource model allows for extremely flexible allocation of memory resources, as this pool of resources is not limited by discrete “slots”.

Returning to the same neural data that seemed to plateau at individuals’ memory capacity (Vogel et al., 2005; Todd & Marois, 2005), these data can be reinterpreted not as a plateau but as fitting a saturation function which infinitely approaches a limit (Ma et al., 2014). Interestingly, although the plateau was suggested to happen at the individual’s capacity, this question wasn’t actually tested until recently. Bays (2018) formally tested a capacity plateau model against a saturation function using the fMRI data from Todd & Marois (2005) and Xu & Chun (2006) to find that the saturation model was a better fit to the data. The saturation model is consistent with neural resource models (Bays, 2014)

which use a signal detection framework to explain how neural population activity can predict the observed patterns of memory errors.

As well, interpreting the CDA as an index for the number of items in memory due to the stepwise amplitude pattern does not fit within either the continuous resource models or the updated discrete resource models such as slot + averaging (Bays, 2018; Emrich et al., 2022). Alternatively, the CDA amplitude is indicative of attentional and executive processing (Emrich et al., 2022). In traditional memory studies, the number of items and the allocation of memory resources has been yoked one-to-one, but by directing attention with a cue, Salahub and colleagues (2019) demonstrated that the CDA amplitude was not dependent on the number of items in memory, instead it tracked the lateral allocation of memory resources. These new interpretations of the neural data suggest that instead of a discrete resource with a fixed item limit there is a quantized continuous resource pool that is allocated to memoranda.

Flexible Allocation of Memory Resources

Contrary to discrete resource models, a continuous resource allows for allocation to a large number of to-be remembered items at the cost of memory resolution. This property has been criticized as allocating a single drop of memory resources becomes indistinguishable from having no memory representation for the item (Adam et al., 2017; Luck & Vogel, 2013). However, this hypothetical flexibility is a critical difference that can be experimentally tested.

In theory, the allocation of continuous memory resources can be finely controlled due to the flexible nature of the VSTM resource. Evidence for the flexibility of memory resources has been shown by manipulating the task relevance of items. The precision of

memory recall predictably aligns with the likelihood of an item to be probed in a manner that would be similarly predicted by an item-based assignment of memory resources (Emrich et al., 2017). Specifically, in two experiments, items were cued to indicate the probe probability at levels to match the resource allocation at equivalent memory loads: For instance, 50% probe probability of a single item is equivalent to a two-item memory load where each item would receive 50% of memory resources. These data demonstrated that probe probability was a better predictor of performance than memory load (i.e., the number of items), suggesting that memory resources were being allocated according to the task-relevance of the items. Further, comparing model fits of four leading error distribution decomposition models (such as the *slot + averaging* model, further models are described below in Chapter 5: *Understanding error distributions*) found that the best fitting model was a variable precision continuous resource model of VSTM (Emrich et al., 2017). This model is congruent with a process of attention allocation of a flexible and continuous VSTM resource; thus, it was concluded that VSTM resources are flexible in nature. While Emrich and colleagues (2017) used simultaneously presented spatial cues to direct attention allocation to more likely probe targets, this effect has also been found with feature-based cues (Dube et al., 2017; Salahub et al., 2019), as well as when using reward to manipulate goal relevance (Klyszejko et al., 2014).

Summary of the present thesis.

The present thesis focuses on attributes of the flexible allocation of VSTM resources. Each of the presented experiments manipulates the task relevance of to-be remembered objects through cues presented simultaneously with the memory set to initiate the flexible allocation of memory. The motivation for this work was to investigate

the potential properties and limitations of flexible allocation of VSTM resources by attentional prioritization.

The first of these investigations asks: ‘Can flexible allocation go beyond two levels of priority?’ In the foundational work on flexible allocation described above there were two priority levels of items in memory: high (cued), and low (uncued) (Dube et al., 2017; Emrich et al., 2017). It was demonstrated that the degree that the cued items could be prioritized above the uncued items was extremely flexible. The assigned probe probability was found to predict response precision at equivalent memory loads (i.e. 50% is comparable to load 2), but further assigned probability was a better predictor than the actual memory load (Emrich et al., 2017). In Chapter 2, I examined whether the allocation of VSTM resources could be extended to three levels of priority. Although other work has demonstrated some ability to allocate memory resources to three or more levels with a cue given before the memory sample (Allen & Ueno, 2018; Yoo et al., 2018), this had not been demonstrated under simultaneous cuing conditions as had been used in Emrich et al., (2017). Further, it was examined whether the ability is universal or an individual difference by using Bayesian model comparison to categorize response patterns into different resource allocation strategies. Finally, the temporal dynamics of response behaviour was investigated moving beyond means-based analysis to investigate trial-by-trial response precision. Together these approaches robustly explored the limits of flexible allocation.

The second question asked is: ‘could an alternative mechanism explain past results that appear to demonstrate flexible allocation?’ In Chapter 3, an alternative mechanism for the parametric response precision was investigated. Although flexible

memory resources are a prediction of continuous models of memory that are not easily compatible with quantized discrete models of memory resources, these results could be a by-product of a discrete allocation mechanism termed ‘probabilistic encoding’. This theory posits that the discrete memory resource is assigned to to-be remembered objects in a strategic manner. Specifically, the more likely to be probed items are encoded into memory at a frequency that depends on the probe probability. Over the course of the experiment, this creates bi-modal source of response error with precise in-memory responses and randomly guessed out-of-memory responses which appears to match the predictions of flexible allocation of memory resources when averaged. This chapter challenges this interpretation by stressing the limits of strategic discrete allocation against a flexible continuous allocation with two responses recorded per trial. The predictions of each mechanism are tested by comparing the response precision of multiple high- and low-priority memory items that were simultaneously held in memory. Further, probabilistic encoding is tested by comparing a value of predicted proportion of trials with in-memory responses to the observed proportion of trials for each participant’s memory capacity.

Finally, is flexible allocation dependant on the test paradigm? In Chapter 4 the analysis of flexible allocation is extended to another type of VSTM response paradigm: two alternative forced choice. In a typical delayed estimation paradigm, the continuous nature of the response is sensitive to the predicted continuous nature of memory error. Whereas in a typical discrete response paradigm the responses are dichotomous which can lead to the interpretation that an item can only either be in memory or not. This chapter presents an experiment where both two alternative force choice responses and

delayed estimation trials are intermixed. The alternative choice presented was varied in the magnitude difference from the target colour which impacts the discriminability of the choice. Combined with various flexible allocation conditions manipulating the probe probability the response choice is dependent on both the memory fidelity and the difficulty of the decision. Importantly, the alternative forced choice paradigm is a recognition task rather than a recall task which are much more sensitive to identifying memory signal from noise – more typically used in long-term memory tasks. This property allows for a clearer investigation of low-priority memories which can be mistaken for random guessing in the typical response distribution analyses. The results of this experiment are in line with predictions of continuous resource models of VSTM representations that have (predictably) variable signal strength (alternatively variable noise), including very low-precision memory representations.

Together, this thesis presents evidence for flexible allocation in a variety of VSTM paradigm variations, but flexibility is not an unlimited property of VSTM resources. This property was found to be limited to only two levels of priority under simultaneous cuing conditions for the majority of participants. Evidence of flexible allocation is extended by testing the predictions with multiple responses, and in another response paradigm. The evidence from the experiments presented here add to the debated nature of VSTM resources by further establishing the importance of the property of flexible allocation.

References

- Adam, K. C. S., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology*, *97*, 79–97.
doi.org/10.1016/j.cogpsych.2017.07.001
- Allen, R. J., & Ueno, T. (2018). Multiple high-reward items can be prioritized in working memory but with greater vulnerability to interference. *Attention, Perception, & Psychophysics*, *80*(7), 1731–1743. doi.org/10.3758/s13414-018-1543-6
- Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience*, *34*(10), 3632–3645.
doi.org/10.1523/jneurosci.3204-13.2014
- Bays, P. M. (2018). Reassessing the evidence for capacity limits in neural signals related to working memory. *Cerebral Cortex*, *28*(4), 1432–1438.
doi.org/10.1093/cercor/bhx351
- Bays, P. M., Catalao, Raquel. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 7.1–7.11. doi.org/10.1167/9.10.7
- Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*, 851–854. doi.org/10.1126/science.1158023
- 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.
doi.org/10.1017/S0140525X01003922

- Dube, B., Emrich, S. M., & Al-Aidroos, N. (2017). More than a filter: Feature-based attention regulates the distribution of visual working memory resources. *Journal of Experimental Psychology: Human Perception and Performance*, 43(10), 1843–1854. doi.org/10.1037/xhp0000428
- 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. doi.org/10.1037/xhp0000398
- Emrich, S. M., Salahub, C., & Katus, T. (2022). Sensory Delay Activity: More than an Electrophysiological Index of Working Memory Load. *Journal of Cognitive Neuroscience*, 35(1), 135–148. doi.org/10.1162/jocn_a_01922
- Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, 3, 1229. doi.org/10.1038/ncomms2237
- Fukuda, K., & Vogel, E. K. (2009). Human variation in overriding attentional capture. *Journal of Neuroscience*, 29(27), 8726–8733. doi.org/10.1523/JNEUROSCI.2145-09.2009
- Klyszejko, Z., Rahmati, M., & Curtis, C. E. (2014). Attentional priority determines working memory precision. *Vision Research*, 105, 70–76. doi.org/10.1016/j.visres.2014.09.002
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279–281. doi.org/10.1038/36846

- 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. doi.org/10.1016/j.tics.2013.06.006
- Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, 17(3), 347–356. doi.org/10.1038/nn.3655
- Pashler, H. (1988). Familiarity and visual change detection. *Attention, Perception, & Psychophysics*, 44(4), 369–378. doi.org/10.3758/BF03210419
- Postle, B. R. (2016). How Does the Brain Keep Information “in Mind”? *Current Directions in Psychological Science*, 25(3), 151–156. doi.org/10.1177/0963721416643063
- Salahub, C., Lockhart, H. A., Dube, B., Al-Aidroos, N., & Emrich, S. M. (2019). Electrophysiological correlates of the flexible allocation of visual working memory resources. *Scientific Reports*, 9, 19428. doi.org/10.1038/s41598-019-55948-4
- Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, 428, 751–753. doi.org/10.1038/nature02466
- Todd, J. J., & Marois, R. (2005). Posterior parietal cortex activity predicts individual differences in visual short-term memory capacity. *Cognitive, Affective, & Behavioral Neuroscience*, 5(2), 144–155. doi.org/10.3758/cabn.5.2.144
- 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. doi.org/10.1073/pnas.1117465109

- Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 748–751.
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*, *438*, 500–503. doi.org/10.1038/nature04171
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 1120-1135. doi.org/10.1167/4.12.11
- Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*(7080), 91–95. doi.org/10.1038/nature04262
- Yoo, A. H., Klyszejko, Z., Curtis, C. E., & Ma, W. J. (2018). Strategic allocation of working memory resource. *Scientific Reports*, *8*, 16162.
doi.org/10.1038/s41598-018-34282-1
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. doi.org/10.1038/nature06860

Chapter 2: Limitations on flexible allocation of visual short-term memory resources with multiple levels of goal-directed attentional prioritization

First published by Springer Nature as:

Lockhart, H. A., Dube, B., MacDonald, K. J., Al-Aidroos, N., & Emrich, S. M. (2024).

Limitations on flexible allocation of visual short-term memory resources with multiple levels of goal-directed attentional prioritization. *Attention, Perception, & Psychophysics*, 86(1), 159-170.

Acknowledgments:

Reproduced with permission by Springer Nature.

This project was funded by a Natural Sciences and Engineering Research Council (NSERC) Discovery Grant awarded to SME [435945], NSERC Discover Grant awarded to NA [RGPIN-2018-04689], and supported by NSERC Postgraduate Scholarship Doctoral Program awarded to HAL, and a NSERC Canada Graduate Scholarship awarded to BD. We would like to thank Joseph Capozza for their assistance programming Experiment 1a, and Joel Robitaille for providing some of the code used to run the Bayesian model comparison analysis.

Abstract

Studies suggest that visual short-term memory (VSTM) is a continuous resource that can be flexibly allocated using probabilistic cues that indicate test likelihood (i.e., goal-directed attentional priority to those items). Previous studies using simultaneous cues have not examined this flexible allocation beyond two distinct levels of priority. Moreover, previous studies have not examined whether there are individual differences in the ability to flexibly allocate VSTM resources, as well as whether this ability benefits from practice. The current study used a continuous report procedure to examine whether participants can use up to three levels of attentional priority to allocate VSTM resources via simultaneous probabilistic spatial cues. Three experiments were performed with differing priority levels, cues, and cue presentation times. Group level analysis demonstrated flexible allocation of VSTM resources, however, there was limited evidence that participants could use three goal-directed priority levels. A temporal analysis suggested that task fatigue, rather than practice effects, may interact with item priority. A Bayesian individual-differences analysis revealed that a minority of participants were using three levels of attentional priority, demonstrating that, while possible, it is not the predominant pattern of behavior. Thus, we provided evidence that flexible allocation to three attention levels is possible under simultaneous cuing conditions for a minority of participants. Flexible allocation to three categories may be interpreted as a skill of high performing participants akin to high memory capacity.

Introduction

Visual short-term memory (VSTM), the system used to hold visual information on-line for short periods of time, has received considerable attention (Luck & Vogel, 2013); In particular, the nature of the representational limitations has been the subject of much debate. Many previous studies examining the nature of VSTM representations have focused on how different models explain changes in performance across memory loads (e.g., van den Berg et al., 2014). Although the debate surrounding the precise nature of the limits of VSTM is on-going (Fougnie et al., 2016; Schurgin et al., 2020; Zhang & Luck, 2008), there are other properties of VSTM beyond working memory capacity limits that must be accounted for by these models. For example, a key prediction of continuous-resource models is that individuals should be able to flexibly allocate resources to objects according to their goals. This prediction has been supported using simultaneously presented spatial-cues (i.e., presented at the same time as the memory array; Emrich et al., 2017), feature-based cues (Dube et al., 2017) as well as feature-based monetary reward (Klyszejko et al., 2014), and frequency-based priority for memory of attentional templates in single and dual-target visual search (Huynh Cong & Kerzel, 2022). In each of these studies, memory performance for individual items was proportional to the level of task-relevant priority or reward, representing a goal directed aspect of attentional priority (Awh et al., 2012). Thus, these studies suggest that the reported precision of an item in memory is best characterized by the amount of attentional prioritization, and not by the total number of objects in memory, consistent with a flexible and continuous model of VSTM resources.

The question remains, however, whether there are limits to the number of distinct

attention levels that can be implemented to effectively allocate VSTM resources. In other words, at a given time, how many levels of goal-directed prioritization can be maintained between concurrently memorized representations. Previous studies using simultaneously presented cues have only tested two levels of attention prioritization (Emrich et al., 2017; Dube et al., 2017; Klyszejko et al., 2014; Salahub et al., 2019). By contrast, some recent studies using pre-cues (cues presented before the memory array) have shown memory accuracy and precision could be influenced by up to four levels of priority (e.g., Allen & Ueno, 2018; Yoo et al., 2018). However, cues appearing at different times relative to the memory array have different effects on attention and memory. For instance, retro-cues have been distinguished from other types of cues because they direct attention to internal representations rather than affecting encoding at perception, like a pre-cue (Griffin & Nobre, 2003; Souza & Oberauer, 2016). Similarly, simultaneous cues limit the contribution of preparatory attentional control and are therefore significantly different from pre-cues, which allow participants to plan locations to be attended before the memory array appears. Thus, it is unclear if the evidence from pre-cue studies regarding the number of priority levels would extend to flexible allocation with simultaneous cues.

In addition, while numerous studies have identified that it is possible to flexibly allocate VSTM resources, these effects have primarily been examined at the group level, potentially ignoring important individual differences. The attentional control literature has established that there are large individual differences in the ability to inhibit distractors and focus on goal-relevant information (e.g., Unsworth et al., 2021). As well, working memory capacity is cemented as an individual characteristic central to theories of cognition that has been correlated with other cognitive abilities such as fluid

intelligence and learning (Luck & Vogel, 2013). Even further, it is likely a direct relationship between attentional control and working memory capacity that determines memory performance (Emrich & Busseri, 2015), furthering the importance of an investigation into individual differences in flexible allocation ability. In the present study we report three similar experiments that allow for combined analyses on a full sample of 99 unique participants to investigate individual differences in flexible allocation ability, with experiment as a between-subject factor.

Finally, to fully investigate the limits of goal-directed flexible allocation, meta-task related effects on performance should be considered (e.g., practice effects, fatigue). Practice with feedback (Adam & Vogel, 2018) may allow participants to better differentiate priority categories within the attentional set, leading to improved task-performance over time. Alternatively, performance may worsen across the task as even short periods of time can lead to fatigue, impairing response precision, especially if participants did not nap before testing (MacDonald et al., 2018). While traditional studies of learning, practice effects, and fatigue have used multi-session designs, the current study uses a novel approach comparing performance over the full length of the task. By looking at trial-by-trial changes in response precision utilizing a mixed-effect model, it is possible to isolate the effect of time-on-task between the priority conditions.

Accordingly, the aim of the current study was to examine whether it is possible for three levels of priority to guide attention during a simultaneous cueing memory task. Participants were shown two distinct cues to indicate the high- and medium-priority items, with no cue indicating low-priority items – but notably these low-priority items are not distractors and are still test candidates. In Experiment 1a and 1b the high-priority

object was tested 50% of the time, a medium-priority object tested 25% (1a) or 30% (1b) of the time, and two low-priority objects with no spatial cues each tested 12.5% (1a) or 10% (1b) of the time. In Experiment 2, the difference between high- and lower priority conditions were exaggerated (70%, 20%, 5%). If it is possible for participants to allocate memory resources using three attention levels, I predicted response error—the inverse of precision and a proxy for resource allocation—to scale with each item’s respective priority, as has been previously shown with two levels of priority (Emrich et al., 2017; Dube et al., 2017). We also examined individual differences through a Bayesian model analysis of task strategy and meta-task effects by analysing whether performance across the different cue conditions changes over the course the experiment (temporal analysis).

Methods

Participants

Participants were recruited from Brock University using an online research pool or through posters around campus (Experiment 1a), and from the University of Guelph using an online research pool (Experiment 1b and 2). Each experiment had independent samples. All participants reported normal or corrected to normal vision and were tested for normal colour vision. All procedures were approved by either Brock University’s Research Ethics Board or the University of Guelph’s Research Ethics Board. In accordance with REB procedures, informed consent was obtained from all participants.

Participants Experiment 1a

Participants were 20 students from Brock University (7 males, 13 females) with a mean age of 20.38 (range 17-29). Participants were given the option to listen to non-lyrical music during the task on their own devices. Four participants chose to listen to

music. Participants were offered a choice of research credit for courses or paid at a rate of \$10/hour of participation.

Participants Experiment 1b

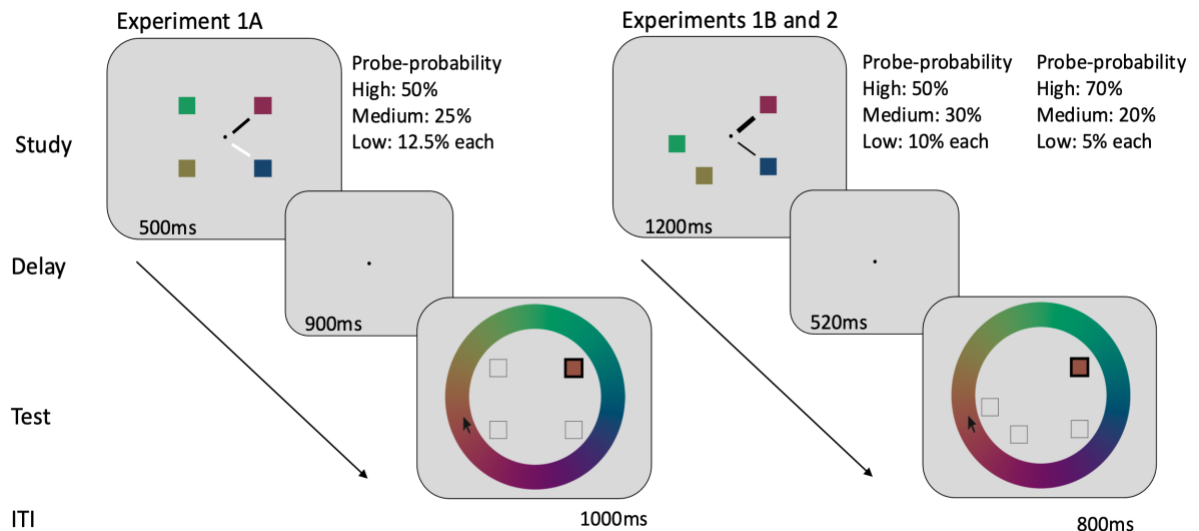
Participants were forty-one students from the University of Guelph (16 males, 25 females) with mean age 20.22 (range 18-26) who participated for research credit. One participant was excluded before statistical analysis for having average absolute recall error $> 90^\circ$ in two conditions, suggesting they were guessing randomly most of the time.

Participants Experiment 2

Thirty-nine students from the University of Guelph (11 males, 28 females) with mean age 21.43 (range 18-27) participated for research credit. Two participants did not complete all the trials, missing 10 and 23 trials respectively. Missing 1% and 2.3% of all trials did not affect the priority level of the conditions; thus, they remained in the sample.

General Procedure

All experiments were variations of delayed-estimation memory tasks with four to-be-remembered objects, two distinct spatial cues indicating high and medium probability to be the test item (exact probability differed by experiment, hereafter termed the probe probability), and one tested location, see Figure 2-1. All experiments were presented using PsychoPy (Peirce, 2008). All responses were made using the mouse by clicking on the colour that best matched their memory. Response location was recorded and the circular distance between the reported colour and the target colour was calculated as the error. All participants completed a session of practice trials before the experimental session and were given the opportunity to ask questions.

Figure 2-1*Example trials from each experiment.*

Note. All experiments were delayed estimation tasks consisting of a study phase, delay, and untimed test phase. In experiment 1a spatial cues were differentiated by colour (black and white), in experiments 1b and 2, line thickness differentiated the high- and medium-priority cues. In all experiments the meanings of the cues were counterbalanced.

For Experiment 1a, stimuli were presented on a 20" LCD display approximately 57 cm away. Stimuli were coloured squares of 1° of visual angle with a constant distance from central fixation. Colour stimuli were randomly selected from 360 unique colours from a continuous colour wheel made using CIE $L^*a^*b^*$ colour space with coordinates of $a = -6$ and $b = 14$ with a radius of 49, calibrated to the monitor and sampled with a minimum sampling distance of 30° .

For Experiments 1b and 2, stimuli were viewed on a 1280 x 1024 CRT monitor using a 75 Hz refresh rate, with viewing distance fixed at 57 cm using a head and chin rest. Stimuli were coloured squares of 1.2° of visual angle with a constant distance from central fixation. Stimuli locations were chosen from a determined set of eight possible locations. Colour stimuli were randomly selected from 360 unique colours from a

continuous colour wheel made using CIE L*a*b* colour space with coordinates of $a = -6$ and $b = 14$ with a radius of 49, calibrated to the monitor and sampled with a minimum distance of 50° .

Procedure and Stimuli Experiments 1a and 1b

The experiments were conducted at two institutions simultaneously without planned collaboration, thus some differences between experiments were not theoretically motivated by the current findings. In Experiment 1a, four coloured squares were equally spaced around a central fixation with two spatial cues, one black and one white cue (meaning counterbalanced), on a grey background; while in 1b, four coloured squares were presented around a central fixation at variable locations with a thick (high-priority) and thin (low-priority) spatial cue. See Figure 2-1 for the timing of each presentation. In Experiment 1a the three levels of priority were 50%, 25%, and 12.5%; in Experiment 1b the levels of priority were 50%, 30%, and 10%. Each experiment included 400 total trials.

Participants in Experiment 1a completed the Media Multi-tasking Inventory (Ophir et al., 2009) during the experimental session but the data is not pertinent to the current research question. Participants also performed a standard colour change detection task with set sizes 2, 4, and 6, but the data were not analyzed since only this experiment of the three had this measure.

Procedure and Stimuli Experiment 2

The presentation was the same as Experiment 1b with the following exceptions: the high-priority cued item was tested on 70% of trials, the medium priority cued item was tested on 20% of trials, and the two uncued items were tested on the remaining 10% of trials (5% each). Participants completed a total of 1,000 trials. See Figure 2-1.

General Analysis

Responses were measured as the circular distance between the target and reported colour value in degrees (i.e., circular error). The circular standard deviation of error on a von Mises distribution was used as a measure of (im)precision (calculated by the *circular* package in R, Agostinelli & Lund, 2023). Figures of the error distributions are based on raincloud plots (Allen et al., 2021). In the current analysis, summary statistics such as circular standard deviation of error have been demonstrated to be the most reliable measure for testing differences between conditions (Ma, 2018). Another summary statistic, mean absolute error, has been presented in a supplementary analysis.

Participants were excluded for poor performance measured as having mean absolute error $> 90^\circ$ in at least two conditions, which suggested they were guessing randomly most of the time. Tests between priority conditions were performed using Bayesian repeated-measures ANOVA to quantify evidence for the null hypothesis (Bayes Factor, BF_{01}) or the alternative hypothesis (BF_{10}). A Bayes factor larger than 3 is considered moderate evidence, and a Bayes factor larger than 10 is considered strong evidence in the given direction (Wetzels et al., 2011). Bayesian repeated-measures ANOVA and Bayesian pairwise t-tests were conducted using JASP (JASP Team, 2023) using default priors. Bayes factor in favour of the null hypothesis is given in italics for clarity.

Follow-up analysis methods for Bayesian individual difference model comparison and temporal patterns on precision are discussed below.

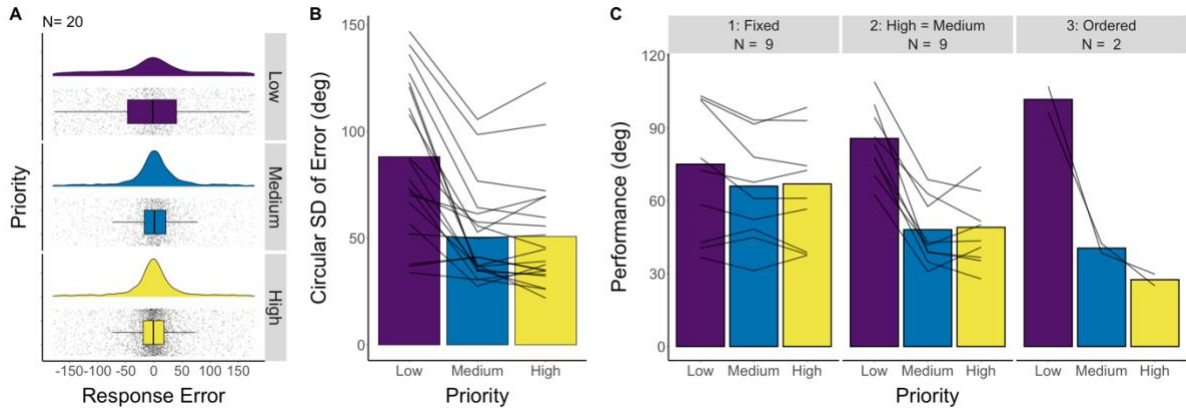
Results

Experiments 1a and 1b

The purpose of each experiment was to examine whether more than two goal-

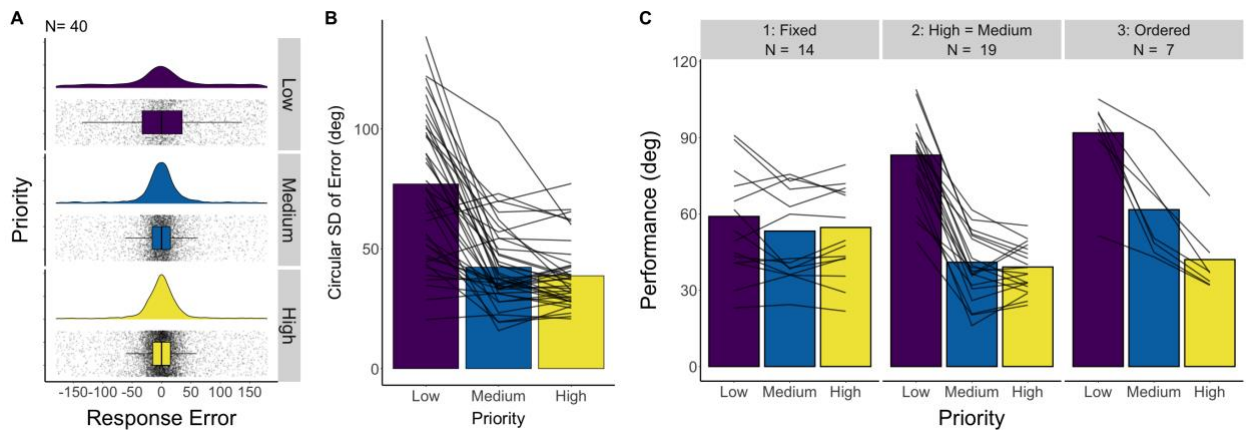
driven attention levels can guide attention by having two explicit attentional cues to high and medium priority items, while the uncued items were not to be ignored and were still test candidates. The performance measure of circular SD was compared across conditions. In Experiments 1a and 1b, repeated-measures ANOVAs between priority condition (high vs. medium vs. low) revealed significant main effects (both $BF_{10} > 1000$), showing participants differentially prioritized items given the cue. Planned comparisons showed strong evidence that uncued, low-priority items (1a SD = 88.14°; 1b SD = 77.04°) were reported with worse precision than medium-priority items (1a SD = 50.73°; $BF_{10} = 695.70$; 1b SD = 42.33°; $BF_{10} > 1000$) and high-priority items (1a SD = 50.46° and 1b SD = 38.76°; both $BF_{10} > 1000$), demonstrating that participants were able to prioritize cued items over uncued items. Although Experiments 1a and 1b were run concurrently with similar designs, Experiment 1b had a longer study presentation time which may have made it easier for participants to use the cues. However, there was no difference found between the medium and high priority items: in Experiment 1a there was substantial evidence of no difference between the medium- and high-priority items $BF_{01} = 4.27$; while in Experiment 1b there was no evidence to support either a difference or no difference between the conditions ($BF_{10} = 1.24$, in a one-tailed test). That there was no difference in response precision suggests there was no difference in the allocation of attention between the two higher priority conditions. See Figures 2-2 A-B and 2-3 A-B.

Figure 0-2
Results of Experiment 1a



Note. **A.** Distribution of error in degrees by priority conditions. Boxplots overlaid on all data points. **B.** Graph of the circular standard deviation of error by priority. **C.** Performance by priority grouped by the best fit strategy according to the Bayesian model comparison.

Figure 0-3
Results of Experiment 1b

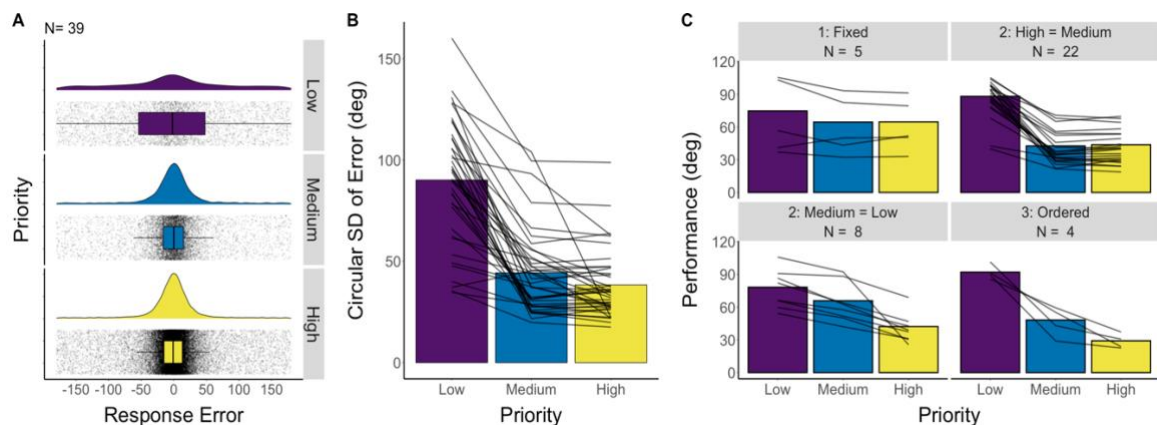


Note. **A.** Distribution of error in degrees by priority conditions. Boxplots overlaid on all data points. **B.** Graph of the circular standard deviation of error by priority. **C.** Performance by priority grouped by the best fit strategy according to the Bayesian model

Experiment 2

In Experiments 1a and 1b, there was no evidence the high- and medium-priority conditions were differentiated. This may have been because there was not a large enough difference in the probe probabilities to incentivize this strategy. Thus, in Experiment 2, the difference between the high and medium conditions was increased (70%, 20%, 5%). As before, there was strong evidence for the effect of priority condition ($BF_{10} > 1000$), showing participants differentially prioritized items. Planned comparisons showed strong evidence that the low-priority items ($SD = 90.15^\circ$) were reported with worse precision than medium-priority items ($SD = 44.36^\circ$, $BF_{10} > 1000$) and high-priority items ($SD = 38.40^\circ$, $BF_{10} > 1000$), demonstrating that participants were able to prioritize cued items over uncued items. In contrast from the previous two experiments, there was anecdotal evidence of a difference between the medium- and high-priority items ($BF_{10} = 2.255$). See Figure 2-4 A and B. Although weak, this provides some evidence that three priority

Figure 0-4
Results of Experiment 2



Note. **A.** Distribution of error in degrees by priority conditions. Boxplots overlaid on all data points. **B.** Graph of the circular standard deviation of error by priority. **C.** Performance grouped by the best fit strategy according to the Bayesian model comparison.

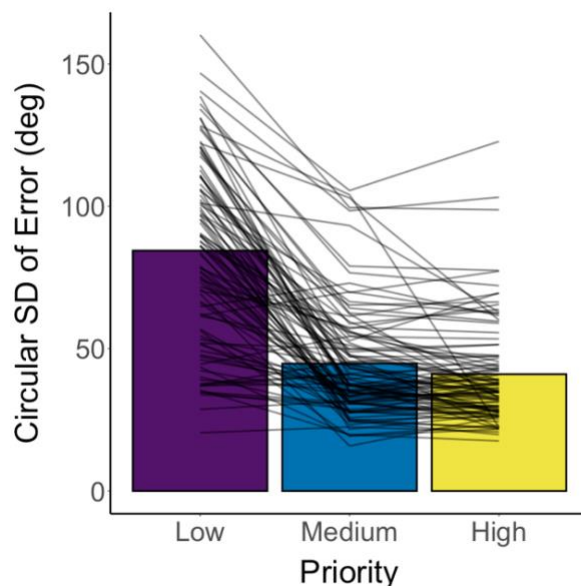
levels could be used to guide attention and memory resources when the priority levels are substantially different in a simultaneous cuing task.

Combined Experiment Data Analyses

Due to the conceptual and methodological similarity of the experiments, the data of all three experiments were analysed with a Bayesian repeated-measure ANOVA with priority condition and experiment as predictor variables. Consistent with the prediction that cue priority would correlate with response precision, the best fitting model was one with only priority condition as a predictor ($BF_{10} > 1000$). All levels of priority are significantly different from each other in the combined sample: low ($SD = 84.44^\circ$) vs. medium ($SD = 44.77^\circ$) ($BF_{10} > 1000$); low vs. high ($SD = 41.03^\circ$) ($BF_{10} > 1000$); medium vs. high ($BF_{10} = 6.702$). See Figure 2-5. This analysis provides evidence that

Figure 0-5

Circular standard deviation of error from combined data by priority condition



Note. Combined samples $N = 99$. All priority conditions are significantly different: low vs. medium ($BF_{10} > 1000$); low vs. high ($BF_{10} > 1000$); medium vs. high ($BF_{10} = 6.702$)

three levels of priority may be maintained to guide attention and memory, but with the caveat that it is a smaller effect size than we had expected based on results from studies with two levels of priority.

Temporal Analysis

Although the effect is small, the results of the three studies combined provide evidence that it is possible to maintain three levels of attentional priority. One possible explanation for the modest effect may be that performance on the task changes over time through practice effects or fatigue. We may expect an effect of fatigue showing time on task leads to reduced accuracy and it may interact with item priority showing accuracy to become more similar over time (converging performance). Alternatively, if time on task interacts with priority conditions to increase the difference in performance (diverging performance) it may indicate learning to effectively use the cues.

Method

To examine whether performance changed over time, we examined absolute response error using mixed-effect models with fixed effects for the test variables of trial and priority condition, and random intercepts for each participant. Trial number was centered and scaled such that estimates indicate change per 100 trials (over the 400 trials of Experiments 1a and 1b, and the 1000 trials of Experiment 2). Priority was treated as a categorical variable for all three experiments. Mixed-effect models were constructed in R using the *lme4* package (Bates et al., 2015), and ANOVA (type 2) from the *car* package (Fox & Weisberg, 2019) was used to test significance of fixed-effects and interactions. Where the interaction was not significant, a model constructed without the interaction was used to estimate and test first-order terms.

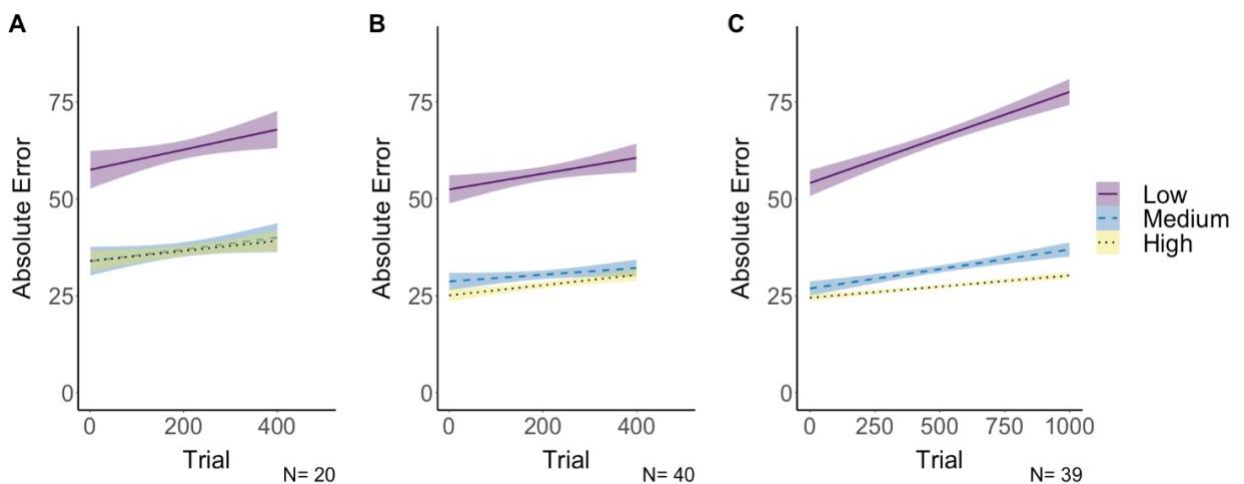
Results

For Experiments 1a and 1b, this analysis showed there were significant main effects of priority condition, (1a: $X^2(2) = 547.07, p < .001$; 1b: $X^2(2) = 1397.28, p < .0001$), and trial, (1a: $X^2(1) = 16.06, p < .001$; 1b: $X^2(1) = 21.06, p < .0001$), but no interaction, (1a: $X^2(2) = 1.69, p = .429$; 1b: $X^2(2) = 1.76, p = .413$). Meaning that over 100 trials, participants in 1a increased their absolute error by 1.67° ($se = 0.42$), and participants on average were 6.68° less accurate at the end of the task relative to the start, $t = 4.01, p < .001$. See Figure 2-6 A. Similarly, participants in Experiment 1b increased their absolute error by 1.24° per 100 trials ($se = 0.27$), meaning participants on average were 4.96° less accurate at the end of the task relative to the start, $t = 4.59, p < .001$. See Figure 2-6 B.

For Experiment 2, again there were significant main effects of priority condition, $X^2(2) = 3613.96, p < .001$, and trial $X^2(1) = 167.62, p < .001$. Across all priority

Figure 2-6

Absolute error for each priority condition by trial number (time)



Note. 95% confidence intervals shown. **A.** Experiment 1a; priority conditions: 50%, 25%, 12.5% probe probabilities **B.** Experiment 1b; priority conditions: 50%, 30%, 10% probe probabilities. **C.** Experiment 2; priority conditions: 70%, 20%, 5% probe probabilities.

conditions, absolute error increased by 1.33° per 100 trials ($se = 0.09$), $t = 15.07$, $p < .001$, meaning participants on average were 13.3° less accurate by the end of the 1000 trials. However, there was also a significant interaction, $X^2(2) = 69.17$, $p < .001$. The accuracy difference between medium- and high-priority conditions (4.54° at trial 500) was estimated to increase at a rate of 0.47° ($se = 0.17$) for every 100 trials, $t = 2.83$, $p < .001$. The accuracy difference between low- and high-priority conditions (38.43° at trial 500) was estimated to increase at a rate of 1.80° ($se = 0.22$) for every 100 trials, $t = 8.13$, $p < .001$. The pattern of the interaction suggests that fatigue effects were more pronounced in the lower priority conditions than the highest priority condition. See Figure 2-6 C.

Bayesian Individual Differences Analysis

Individual differences in flexible allocation ability may be responsible for the current results: It may be that some participants are using three distinct priority levels, while others are not, thus obscuring the expected differentiation of response precision at the group level in Experiments 1a and 1b particularly. If participants are able to use three distinct levels of attentional prioritization, it is expected that behaviour would be organized into three separable response bins based on increasing error with decreasing priority; we consider this to be *ordered* response type. However, three alternative attentional template strategies are considered: *fixed*, meaning error across the three levels of priority was equal (i.e., did not use cue information); *high = medium*, meaning the high- and medium-priority conditions were equal but had less error than the low priority; or *medium = low*, meaning the medium- and low-priority conditions were equal but had greater error than the high-priority condition; see Table 2-1 for summary. Of note, the

strategy of grouping *high = medium* could also be considered a strategy of grouping the cued items and prioritizing them over the uncued items; however, the alternative *medium = low*, does not share this feature.

Table 0-1
Summary of Response Strategy Model Parameters

Model name	Condition Groupings of Response error (SD)	Corresponding number of attention levels
Fixed	High = Medium = Low	One
High = Medium	High = Medium > Low	Two
Medium = Low	High > Medium = Low	Two
Ordered	High > Medium > Low	Three

Method

To investigate different patterns of behavior exhibited by participants, Bayesian model selection was run following the methods of Dowd et al. (2015): testing the best model fit between four specified models (described above, summarized in Table 2-1) for each participant. All models were fit to the SD error parameter calculated using MemToolbox (Suchow et al., 2013). As described by Dowd et al. (2015), the best fitting model for each participant was determined using deviance information criterion (DIC), a relative measure of model fit calculated using the MemToolbox (Suchow et al., 2013). This analysis was completed using MATLAB R2017a.

Results

Each participant's data was first fit to all 4 models reflecting possible patterns of response behavior, and the best fit model was chosen: *fixed*, *ordered*, *high = medium*, and *medium = low*. Of the 99 participants in the three experiments only 13 show evidence of using 3 attention levels by best fitting an *ordered* model; 2, 7, and 4 people in each experiment, respectively. In contrast, 58 participants showed evidence of two attention

levels; these were 9, 19, and 22 participants best fit by a *high = medium* model in each experiment, respectively, as well as 8 participants in Experiment 2 best fit by the alternative *low = medium* model. Finally, 28 participants were best fit by a *fixed* model, suggesting they used only one attention level. In other words, these participants were not using the cues at all. In summary, the majority of participants only maintained two or fewer attention levels, independent of experiment. However, these results also indicate that using three attention levels is not impossible, but not the most common pattern of behavior observed. See Figures 2-2 C, 2-3 C, and 2-4 C showing the breakdown of best fit models in each experiment.

Discussion

Using probabilistic cueing paradigms, research has demonstrated that VSTM can be conceived as a continuous resource that is flexibly allocated according to the attentional priority of an item (Dube et al., 2017; Emrich et al., 2017; Huynh Cong & Kerzel, 2022; Klyszejko et al., 2014). When items are indicated as more likely to be tested, they are prioritized in memory resulting in more precise memory reports. The current studies examined whether three levels of priority could guide attentional allocation through three levels of task relevance. Consistent with previous studies, the data support evidence that memory resources can be flexibly allocated: Across all three experiments, there was increased precision for the high-priority cued items compared to the low-priority uncued items. Despite that, across all three experiments there was only moderate evidence that three levels of priority were readily used to guide attention when simultaneously cued. In general, we observed that performance was equivalent for the high- and medium-priority items suggesting that participants did not distinguish between

these two conditions. However, when all the experiments were combined, there was moderate evidence that the three goal-directed priority levels were used to guide attention. This may suggest that each of the studies were under-powered on their own, and that the effect was smaller than anticipated based on the strength of the correlation between priority and precision when only two attention levels are needed.

The observation that only a small percentage of participants seemed to prioritize VSTM contents with three levels of priority using a simultaneous cue is surprising in light of work by Allen and Ueno (2018) and Yoo et al. (2018), both showing that three priority levels were effectively engaged when information was given by a pre-cue. In the study by Yoo et al. (2018), participants were pre-cued using coloured wedges with proportional radial size to indicate the relative priority of each quadrant, with cuing probabilities of 60%, 30%, 10%, and 0%. Their results showed three distinct levels of response error which matched the behavioral relevance of the item. Experiment 4 in Allen and Ueno (2018), used distinct levels of reward, assigning each position on the display an ordinal reward level by a pre-cue, and found correspondingly graded accuracy. Thus, it may be that preparatory orienting from pre-cues is necessary for three or more levels of priority while simultaneous cues allow for only two levels of priority to readily guide attention even when viewing time is extended to 1200 ms, as in Experiments 1b and 2. Indeed it has been long found that preparedness for the stimulus, or some part of a stimulus set, greatly improves accuracy of memory report compared to simultaneous and retro-active instruction (DiPuma et al., 2023; Sperling, 1960). Considering this, it is plausible that mechanisms of memory encoding are primed or off-loaded when given a pre-cue versus a simultaneous cue.

Interestingly, when attentional priority is manipulated through physical saliency rather than goal directed task relevance, there is a clear parametric relationship between the priority and task performance not observed in the current data. Constant and Liesefeld (2021) found that when a target bar was made more salient in a crowded display by increasing the tilt from 12° to 45° recall error of the feature colour decreased (i.e., better performance for higher saliency). Moreover, the resultant graded attentional prioritization from saliency was nearly impossible to overcome with goal directed incentives of reward or task relevance in a follow-up study, with only weak evidence found at very long presentation times of 3000ms (Constant & Liesefeld, 2023). This evidence aligns with the current study suggesting that top-down control of attentional prioritization is not a simple process, but not an impossible task either. Although the most common strategy used in the current study was not a strong match to our prediction, participants may have used a strategy that they believed would optimize their performance for the lowest effort; However, what “optimized performance” means to each participant may have differed from our expectation – for instance *minimizing error* instead of *maximizing precision* through the use of cues (van den Berg & Ma, 2018). Because we did not reward participants for better precision on high-priority trials, it is reasonable that participants chose to minimize error overall by more evenly distributing resources among the memory items. This could have been done by grouping multiple priority conditions or even equating all items as in the *fixed* response error strategy. Yoo et al. (2018) found that participants were inclined to use a *minimize error* strategy even when they would have received a greater reward by maximizing precision on high-priority trials. In our experiments, performance may not have reflected the three different priority conditions

because participants were not trying to perform better on high-priority trials than on medium-priority trials, instead trying to perform “well enough” on the majority of trials. For example, the interaction in the temporal analysis in Experiment 2 could be interpreted as participants attempting to minimize error by protecting high-priority items at the cost of low-priority items. It should be noted that in Experiment 2 the tested item was either a high- or medium-priority item on 90% of trials, meaning participants would only have the feeling of catastrophic memory errors on 10% of trials if they ignored the low-priority items.

Temporal Analysis

It is possible that participants were using all three priority levels but inefficiently or inconsistently. For instance, participants may have learned to use the cues over time, or they may have stopped using them as time-on-task continued. To test these alternatives, we investigated response error over the duration of the task using a mixed-effect model with random effects of participant and fixed effects of condition and time (trial number). In each experiment performance decreased over the duration of the task such that the SD of response error was between 4.96° and 13.3° greater by the end of the task. Only in Experiment 2 was there an interaction between priority condition and trial number suggesting that participants treat the cues differently over time. Specifically, the performance decrease was more prominent in the low- and medium-priority conditions than in the high-priority condition. While it may be that participants learned to use three different cues over time, it could also be that the relatively higher probe probability for the high-priority items in Experiment 2 incentivized participants to protect them from the effects of fatigue (i.e., rate of increasing error).

We argue the findings from the temporal analysis support flexible allocation theories. One conceptualization of the impact of fatigue may be that fatigue decreases the total amount of memory resources available as the task goes on. In a flexible and continuous framework, we suggest that participants will allocate fewer resources to low-priority items to better maintain performance of higher priority items rather than equally diminishing performance on all items (which could be non-flexible continuous resource allocation). If resources were neither continuous nor flexible, and instead were all-or-none per item, an increase in fatigue would result in a sharp drop off in performance as participants lost discrete resources to allocate to lower priority items, i.e., effectively lowering maximum working memory capacity. In contrast, participants were not only able to maintain the allocation of resources over time, but the differences also became exaggerated as resources became depleted due to fatigue. Thus, while the current task was not specifically designed for this investigation or with this hypothesis in mind; future studies could purposefully investigate differential resilience of memory performance under flexible task demands.

Individual Differences Analysis

We also examined the possibility that not all participants were able (or attempted) to employ three attention levels by investigating individual differences in the patterns of response errors across conditions. Bayesian model selection was used to determine which of four models designed to reflect potential patterns of responses was the best fit for each participant. Overall, most participants (58 of 99) were best fit by one of the models reflecting two attention levels. By contrast, 13 participants across all experiments were best fit by the ordered response model reflecting the use of three attentional cues.

Interestingly, these participants were not predominantly from Experiment 2 as may be expected from the initial analyses, which had the strongest evidence of three priority levels. Four participants from Experiment 2 were best fit by this model, suggesting that the moderate evidence in the group analysis may instead reflect an averaging of the other response patterns (i.e., *low = medium, medium = high*). Thus, while making the priority levels dramatically different did influence how participants used the cues, it may not have had the intended effect of motivating a three attention level strategy throughout the task and may instead have motivated alternative strategies that utilize two attention levels. Regardless, while we did find evidence that it is possible for some participants to use three attention levels defined by task-relevant priority conditions, it was not the primary response pattern observed in any of the three experiments.

While most participants used cue information when allocating memory resources, some participants appeared to ignore the cues altogether, as evidenced by a fixed pattern of response error. It can be noted that fixed response errors were apparent in two forms: all low error responses (high precision) or all high error responses (poor precision). Thus, it seems some participants did not use the priority cues because they felt they were unnecessary to maximize their performance, whereas other participants did not use the priority cues because they lacked the intention or ability to maximize their performance. This phenomenon may be explained by the recent work from Irons and Leber (2020) which suggest that attentional control strategy is an individual difference that can be predicted by how effortful the person finds the optimal strategy. Indeed, when considering the observed increases in response error over time, it might be assumed that the optimal strategy was quite effortful. Response error tended to increase over the

duration of the task, particularly for low-priority items, suggesting it may be more effortful to maintain low-priority items; perhaps this pattern extends to maintaining multiple attention levels. Thus, although participants were instructed to use all three levels of priority, there are reasonable strategic approaches to the task that only require two-levels of prioritization, which may reflect individual differences in abilities or strategies.

Limitations

Although Ma (2018) argued that raw error is the best measure when comparing across manipulations, it is possible that it does not capture all aspects of VSTM performance. Indeed, models have been devised to capture several theoretical aspects of memory performance, such as binding errors (Bays et al., 2009; Swan & Wyble, 2014), categorical responses (Bae et al., 2015; Hardman et al., 2017), resource rationality (van den Berg & Ma, 2018) and perceptual similarity (Schurgin et al., 2020). While my analysis may not model all nuances of resource allocation, most or all of those theorized aspects of responses are captured by a measure of raw error. Although the model that best characterizes memory performance may change over time, the model that best accounts for behavior clearly needs to capture both the flexible control of resources (Dube et al., 2017; Emrich et al., 2017; Huynh Cong & Kerzel, 2022; Klyszejko et al., 2014), as well as the limitations therein. Moreover, the results of the present study suggest that individual differences are an important driver in performance that need to be considered to accurately understand both what is possible and what is readily done by participants during VSTM tasks.

The goal of the current study was to use three levels of goal-directed attentional priority to direct the flexible allocation of VSTM resources; however, in each of the experiments the low priority condition was always uncued; whereas, the high and medium conditions both were cued. The general finding that participants tended to group the high and medium conditions together could be the result of grouping the cued versus the uncued objects in memory. Specifically, the result of the individual differences analysis found that 51 of 99 participants used a two attention level strategy (*high = medium*) that is indistinguishable from a cued versus uncued strategy. While this limitation does not take away from the finding that some participants were able to flexibly allocate VSTM resources to three levels, it is possible that more participants would have used a three attention level strategy if all of the objects had been cued. A future study could address this limitation by using cues that indicate the low priority items as well.

Conclusion

Together, these results add to the body of literature demonstrating that VSTM resources can be flexibly allocated via attention manipulations and further that multiple attention levels can guide attention and thus the distribution of memory resources. Specifically, it was found that although resources can be flexibly allocated, it is not readily done with more than two goal-driven attention levels when simultaneously cued unlike in pre-cue studies. Our results also reveal individual differences in approach or cognitive flexibility in response to multiple cue conditions. Thus, a full understanding of VSTM may require the examination of individual differences in flexible allocation of VSTM resources.

References

- Adam, K. C. S., & Vogel, E. K. (2018). Improvements to visual working memory performance with practice and feedback. *PLOS ONE*, *13*(8), e0203279. doi.org/10.1371/journal.pone.0203279
- Agostinelli, C., & Lund, U. (2023). *R package "circular": Circular Statistics (0.5-0)* [R]. cran.r-project.org/package=circular
- Allen, R. J., & Ueno, T. (2018). Multiple high-reward items can be prioritized in working memory but with greater vulnerability to interference. *Attention, Perception, & Psychophysics*, *80*(7), 1731–1743. doi.org/10.3758/s13414-018-1543-6
- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2021). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*(63). doi.org/10.12688/wellcomeopenres.15191.1
- 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. doi.org/10.1016/j.tics.2012.06.010
- Bae, G.-Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, *144*(4), 744–763. doi.org/10.1037/xge0000076
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi.org/10.18637/jss.v067.i01

- Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, 9(10):7, 1–11. doi.org/10.1167/9.10.7
- Constant, M., & Liesefeld, H. R. (2021). Massive Effects of Saliency on Information Processing in Visual Working Memory. *Psychological Science*, 32(5), 682–691.
- Constant, M., & Liesefeld, H. R. (2023). Effects of salience are long-lived and stubborn. *Journal of Experimental Psychology: General*, 152(9), 2685–2694. doi.org/10.1037/xge0001420
- Emrich, S. M., & Busseri, M. A. (2015). Re-evaluating the relationships among filtering activity, unnecessary storage, and visual working memory capacity. *Cognitive, Affective, & Behavioral Neuroscience*, 15(3), 589–597. doi.org/10.3758/s13415-015-0341-z
- DiPuma, A., Lockhart, H. A., Emrich, S. M., & Ester, E. F. (2023). Retrospective cue benefits in visual working memory are limited to a single location at a time. *Attention, Perception, & Psychophysics*. doi.org/10.3758/s13414-023-02661-0
- Dowd, E. W., Kiyonaga, A., Beck, J. M., & Egner, T. (2015). Quality and accessibility of visual working memory during cognitive control of attentional guidance: A Bayesian model comparison approach. *Visual Cognition*, 23(3), 337–356. doi.org/10.1080/13506285.2014.1003631
- Dube, B., Emrich, S. M., & Al-Aidroos, N. (2017). More than a filter: Feature-based attention regulates the distribution of visual working memory resources. *Journal of Experimental Psychology: Human Perception and Performance*, 43(10), 1843–1854. doi.org/10.1037/xhp0000428

- 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.
doi.org/10.1037/xhp0000398
- Fougnie, D., Cormiea, S. M., Kanabar, A., & Alvarez, G. A. (2016). Strategic trade-offs between quantity and quality in working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 42(8), 1231–1240
doi.org/10.1037/xhp0000211
- Fox, J., & Weisberg, S. (2019). *An {R} Companion to applied regression* (3rd ed.). Sage.
- Griffin, I. C., & Nobre, A. C. (2003). Orienting attention to locations in internal representations. *Journal of Cognitive Neuroscience*, 15(8), 1176–1194.
doi.org/10.1162/089892903322598139
- Hardman, K. O., Vergauwe, E., & Ricker, T. J. (2017). Categorical working memory representations are used in delayed estimation of continuous colours. *Journal of Experimental Psychology: Human Perception and Performance*, 43(1), 30–54.
doi.org/10.1037/xhp0000290
- Huynh Cong, S., & Kerzel, D. (2022). The allocation of working memory resources determines the efficiency of attentional templates in single- and dual-target search. *Journal of Experimental Psychology: General*, 151(12), 2977–2989.
doi.org/10.1037/xge0001239
- Irons, J. L., & Leber, A. B. (2020). Developing an individual profile of attentional control strategy. *Current Directions in Psychological Science*, 29(4), 364–371.
doi.org/10.1177/0963721420924018

- JASP Team. (2023). JASP (Version 0.17). Retrieved from jasp-stats.org/
- Klyszejko, Z., Rahmati, M., & Curtis, C. E. (2014). Attentional priority determines working memory precision. *Vision Research*, 105, 70–76.
doi.org/10.1016/j.visres.2014.09.002
- 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. doi.org/10.1016/j.tics.2013.06.006
- Ma, W. J. (2018). Problematic usage of the Zhang and Luck mixture model.
doi.org/10.1101/268961
- MacDonald, K. J., Lockhart, H. A., Storace, A. C., Emrich, S. M., & Cote, K. A. (2018). A daytime nap enhances visual working memory performance and alters event-related delay activity. *Cognitive, Affective, & Behavioral Neuroscience*, 18(6), 1105–1120. doi.org/10.3758/s13415-018-0625-1
- Ophir, E., Nass, C., & Wagner, A. D. (2009). Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences*, 106(37), 15583–15587.
doi.org/10.1073/pnas.0903620106
- Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, 2. doi.org/10.3389/neuro.11.010.2008
- Salahub, C., Lockhart, H. A., Dube, B., Al-Aidroos, N., & Emrich, S. M. (2019). Electrophysiological correlates of the flexible allocation of visual working memory resources. *Scientific Reports*, 9, 19428.
doi.org/10.1038/s41598-019-55948-4

- Schurgin, M.W., Wixted, J.T. & Brady, T.F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. *Nature Human Behavior*
doi.org/10.1038/s41562-020-00938-0
- Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Attention, Perception, & Psychophysics*, 78(7), 1839–1860. doi.org/10.3758/s13414-016-1108-5
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74(11), 1–29.
doi.org/10.1037/h0093759
- Suchow, J. W., Brady, T. F., Fougny, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision*, 13(10), 9–9.
doi.org/10.1167/13.10.9
- Swan, G., & Wyble, B. (2014). The binding pool: A model of shared neural resources for distinct items in visual working memory. *Attention, Perception, & Psychophysics*, 76(7), 2136–2157. doi.org/10.3758/s13414-014-0633-3
- van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, 121(1), 124–149. doi.org/10.1037/a0035234
- van den Berg, R., & Ma, W. J. (2018). A resource-rational theory of set size effects in human visual working memory. *Elife*, 7. doi.org/10.7554/eLife.34963
- Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science*, 6(3), 291–298.
doi.org/10.1177/1745691611406923

- Unsworth, N., Miller, A. L., & Robison, M. K. (2021). Are individual differences in attention control related to working memory capacity? A latent variable mega-analysis. *Journal of Experimental Psychology: General*, 150(7), 1332–1357. doi.org/10.1037/xge0001000
- Yoo, A. H., Klyszejko, Z., Curtis, C. E., & Ma, W. J. (2018). Strategic allocation of working memory resource. *Scientific Reports*, 8, 16162. doi.org/10.1038/s41598-018-34282-1
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453, 233–235. doi.org/10.1038/nature06860

Appendix 2A: Supplementary analysis: Absolute error.

Results – absolute error

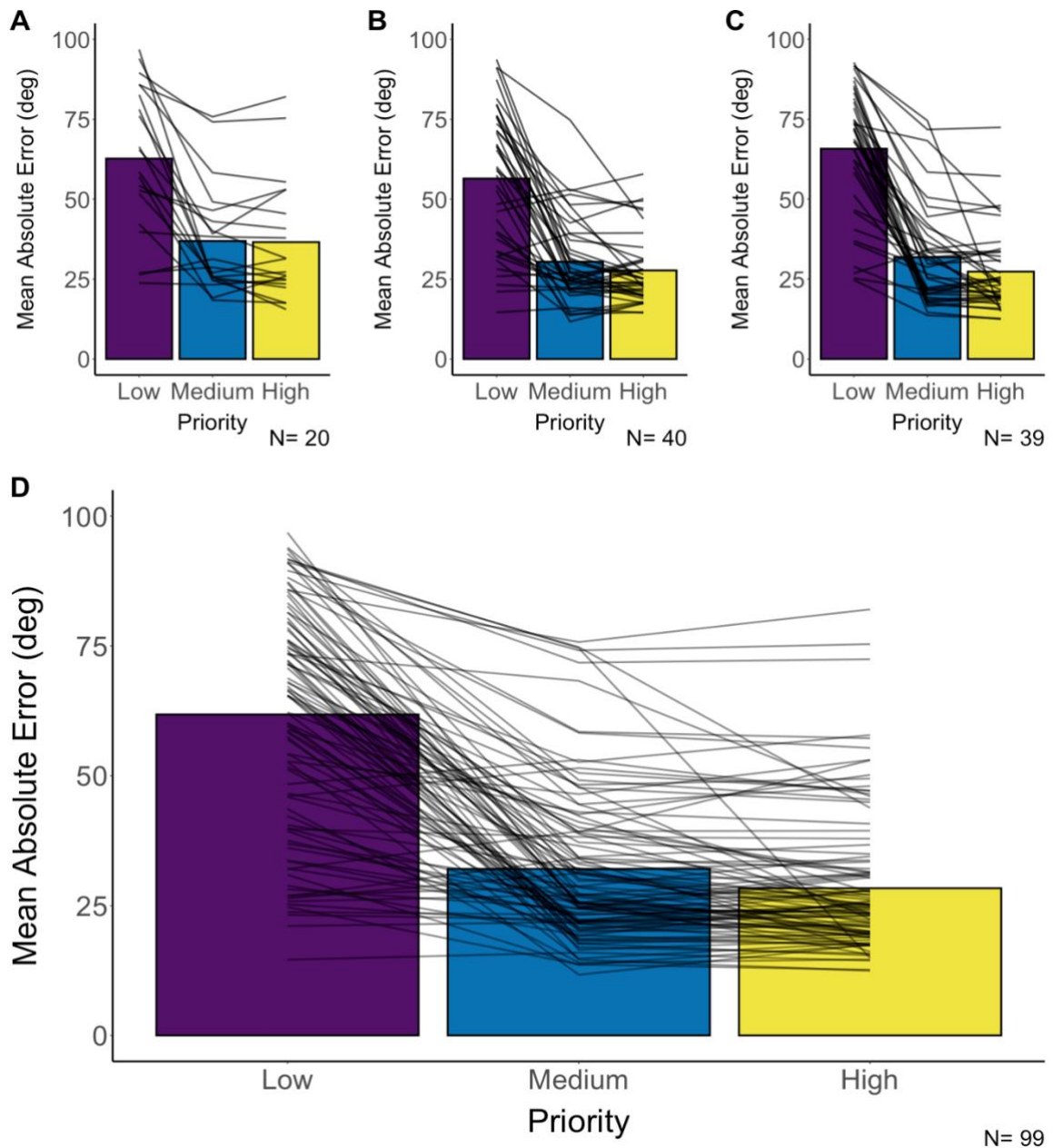
A supplementary analysis using the performance measure of absolute error was prepared repeating the same tests as presented in the manuscript with circular standard deviation.

Experiments 1a and 1b

The performance measure of mean absolute error was compared across priority conditions. In Experiments 1a and 1b, one-way ANOVAs on priority condition (high vs. medium vs. low) revealed significant main effects (both $BF_{10} > 1000$), showing participants differentially prioritized items. Planned comparisons showed strong evidence that uncued, low-priority items (1a $M = 62.69^\circ$; 1b $M = 56.44^\circ$) were reported with worse precision than medium-priority items (1a $M = 36.56^\circ$; $BF_{10} = 746.14$; 1b $M = 30.42^\circ$; $BF_{10} > 1000$) and high-priority items (1a $M = 36.91^\circ$; $BF_{10} = 306.62$; 1b $M = 27.70^\circ$; $BF_{10} > 1000$), demonstrating that participants were able to prioritize cued items over uncued items. However, there was no difference found between the medium- and high-priority items: in Experiment 1a there was substantial evidence of no difference $BF_{01} = 4.16$; while in Experiment 1b there was no evidence to support either a difference or no difference between the priority conditions ($BF = 1.27$). No difference in response accuracy suggests there was no difference in the allocation of attention between the two higher priority conditions. See Figure 2-A1 B.

Experiment 2

As in Experiments 1a and 1b, there was strong evidence for the effect of priority conditions ($BF_{10} > 1000$), showing participants differentially prioritized items. Planned comparisons showed strong evidence that the low-priority items ($M = 65.75^\circ$) were

Figure 2-A1*Mean absolute error for all experiments*

Note. **A.** Experiment 1a; priority conditions had the probe probabilities of 50%, 25%, and 12.5%. **B.** Experiment 1b; priority conditions were 50%, 30%, and 10% probe probabilities. **C.** Experiment 2 the high priority condition was probed more often, priority conditions were 70%, 20%, and 5% probe probabilities. **D.** All participant data combined.

reported with worse precision than medium-priority items ($M = 31.87^\circ$, $BF_{10} > 1000$) and high-priority items ($M = 27.36^\circ$, $BF_{10} > 1000$), demonstrating that participants were able to prioritize cued items over uncued items. In contrast from the previous two experiments, there was anecdotal evidence of a difference between the medium- and high-priority items ($BF_{10} = 2.576$). See Figure 2-A1 C. Although weak, this provides some evidence that three priority levels could be used to guide attention and memory resources when the priority levels are substantially different in a simultaneous cuing task.

Combined Experiment Data Analyses

Due to the conceptual and methodological similarity of the experiments, the results of all three experiments were analysed with a Bayesian repeated-measure ANOVA with priority condition and experiment as predictor variables. Consistent with the prediction that cue priority would correlate with response precision, the best fitting model was one with only priority condition as a predictor ($BF_{10} > 1000$). All levels of priority are significantly different from each other in the combined sample: low ($M = 61.37^\circ$) vs medium ($M = 32.30^\circ$) ($BF_{10} > 1000$); low vs high ($M = 29.36^\circ$) ($BF_{10} > 1000$); medium vs high ($BF_{10} = 12.16$). See Figure 2-A1 D. This analysis provides evidence that three levels of priority may be maintained to guide attention and memory.

Chapter 3: In-memory items are not equal: multiple item report reveals differential memory precision and not proportional memory access for prioritized items.

Acknowledgments:

This project was funded by an NSERC Discovery Grant awarded to SME [435945], and NSERC Postgraduate Scholarship Doctoral Program awarded to HAL. We would like to thank Tom Nelson for programming Experiment 1.

Abstract

Visual short-term memory resources can be flexibly allocated according to an item's priority. However, because only a single item is typically measured on each trial, it is not clear whether participants are truly allocating memory resources flexibly within each trial according to the attentional priority of each item or whether items are probabilistically encoded into limited memory slots. To test this question, participants reported two items on each trial from a sample containing two levels of attentional priority. Task conditions were created based on which items were probed at test: both high-priority (high-high), one each of a high- and low-priority (high-low), or two low-priority (low-low). Against the prediction of probabilistic encoding, there was a greater difference in the report of two high-priority responses (high-high) than two low-priority responses (low-low). The first response of high-low was also distinct from the first response of high-high, falling between the performance of the two high-high responses. This suggests that the memory fidelity of two high-priority items were not equivalent even before the first report. This result is compatible with a flexible and variable allocation of memory resources. A second theoretically driven analysis compared the observed proportion of "in-memory" first responses from the low-low test condition to the predicted proportion that should have been in-memory based on each individual's estimated capacity. Participants' observed data had significantly fewer first low-priority responses in memory than predicted by probabilistic encoding. These results are consistent with a model of flexible allocation of continuous memory resources rather than probabilistic encoding of a discrete memory resource. That is, the priority manipulation appears to change the precision of the memory representation and not the likelihood the item was in memory.

Introduction

Visual short-term memory (VSTM) is the cognitive process that supports the on-line storage of visual information and is characterized by a strict capacity limit on the resources used to maintain information. The precise nature of how this resource is limited is debated between two major families of theoretic cognitive models: continuous and discrete resource models. Continuous resource models propose a limited but infinitely divisible pool of memory resources that are distributed to to-be remembered items (Bays et al., 2009). In contrast, discrete resource models characterize memory as having a limited number of storage “slots”, in which individual items are stored (Awh et al., 2007; Luck & Vogel, 1997).

To test models of VSTM, the majority of studies use procedures in which all items are equally relevant at test. However, this method yokes resource allocation to memory load, where both models make similar predictions. If instead the relative priority of to-be-remembered items are manipulated, thereby requiring flexible allocation of VSTM resources, it is possible to separate out the effects of memory load (i.e., the number of items to be stored) and resource allocation (i.e., how much resources an item receives). That is, if VSTM resources are discrete, each item in the display should receive a similar proportion of resources (i.e., a “slot” of memory), up to memory capacity; alternatively, if resources are continuous and can be flexibly distributed, each item should be allocated a proportion of memory resources determined by the item’s relative priority. Previous studies (Dube et al., 2017; Emrich et al., 2017; Klyszejko et al., 2014; Salahub et al., 2019; Yoo et al., 2018) have demonstrated that the precision of VSTM recall, as measured with continuous report tasks, is strongly predicted by attentional prioritization,

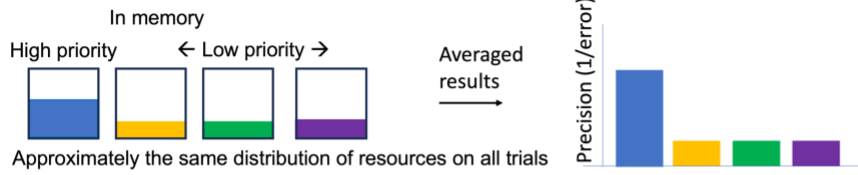
consistent with the predictions of flexible allocation of a continuous memory resource. Relative priority was manipulated by varying the task-relevance of cued items either by making some items more likely to be tested (Dube et al., 2017; Emrich et al., 2017; Salahub et al., 2019; Yoo et al., 2018) by varying the reward associated with a cue (Klyszejko et al., 2014), or the frequency that an attentional template was used in single and dual-target visual search (Huynh Cong & Kerzel, 2022).

While the observed flexibility of resource allocation is consistent with a continuous resource model, it is possible this is an artifact of averaging across trials that include both in-memory responses and guesses, rather than reflecting resource allocation *per se*. This alternative, consistent with discrete resources, proposes that a small number of items are held in discrete slots on a given trial, while others are excluded entirely; critically, the choice of which items are excluded is determined by the probability that each item will be probed (i.e., determined by the cues). With equal precision discrete memory stores, low-priority items that are ‘in-memory’ would be reported with the same precision as high-priority items that are ‘in-memory’; however low-priority items are less likely to be encoded in memory, resulting in a greater number of guess responses. Critically, when averaged across trials, the resulting precision would correspond with the priority set by the attentional cues but would represent a weighted averaged of in-memory and not in-memory (guessed) responses, where the proportion of trials the item was in-memory is determined by the attentional cue. In contrast, flexible allocation proposes that memory resources are directly assigned in proportion to the item’s priority. This averaged data would look identical to probabilistic encoding, see Figure 3-1.

Figure 3-1
Flexible Allocation versus Probabilistic Encoding

A Continuous Flexible Allocation

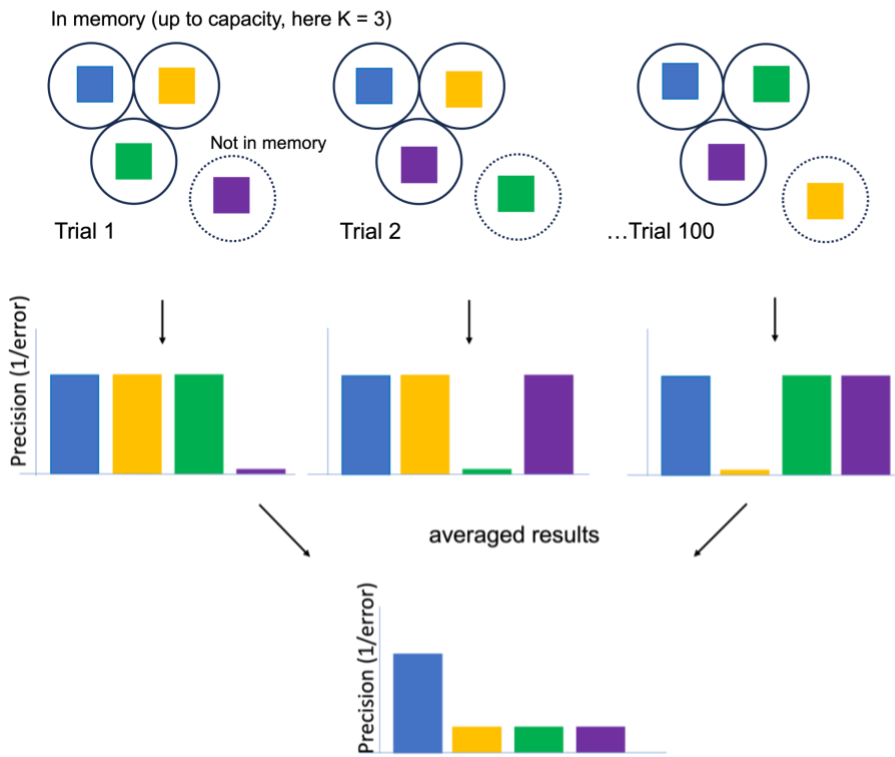
Items receive a portion of memory resources according to attentional priority



High priority (blue) is reported with greater precision than low priority items because it was stored with high fidelity.

B Discrete Probabilistic Encoding

Items are probabilistically 'in memory' during each trial according to priority



High priority (blue) is 'in memory' more often than any low priority item

In memory items have high fidelity and are reported with high precision.

High priority (blue) is reported with more consistently high precision than the low priority items.

Note. Averaged results in A and B look identical despite different mechanisms of VSTM resource allocation. **A.** Conceptual diagram of flexible allocation: Items in memory receive amounts of a continuous memory resource according to the item's priority. When responses are averaged, precision reflects the amount of memory resources per item. **B.** Conceptual diagram of probabilistic encoding with discrete memory resources. On a trial-by-trial basis, different items are in-memory or not. On each trial, response precision reflects whether the item was in memory or not. When trials are averaged the results combine in and out-of-memory responses. Since low-priority items were not in memory as often, average responses appear less precise.

Disentangling the mechanism of response precision.

One way to determine whether probabilistic cues affect memory performance according to continuous or discrete models is to have participants make multiple responses on each trial. To disentangle the mechanism of response precision this study will test predictions for two high-priority responses (high-high), two low-priority responses (low-low) and a mixed test condition (high-low or low-high). Since the two alternative models differ in the mechanism that leads to flatter error distributions as item priority decreases (or item load increases), this multi-response format should be able to separate whether prioritized items were more likely to be in memory (probabilistic encoding – discrete VSTM resource models) or memorized with more precision (flexible allocation – continuous VSTM resource models).

Critically, testing multiple items with self-selected response order allows participants to choose to report their better remembered item first. In the case that memory fidelity is equal (i.e., equal resource allocation, or equal ‘in-memory’ status), the selection is effectively random. If responses could be reported simultaneously, it is predicted that responses would have equivalent precision. However, effects of response order are well-documented in multiple report experiments (Oberauer et al., 2018), so instead it is predicted that subsequent responses will be affected by increased time since encoding and first-response interference (Adam et al., 2017; Cowan et al., 2002), resulting in worse response precision. Additionally, besides comparing responses within trials, it is possible to control for response order by comparing the mixed test responses to their comparable priority and order responses in the high-high and low-low conditions. Key predictions of both models are made below.

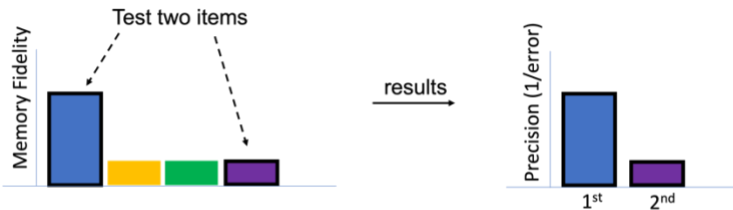
Predictions of flexible allocation on multiple responses' precision

Continuous resource models with flexible allocation predict that the precision of each item is determined by its attentional priority; thus, obtaining reports for items with two different levels of attentional prioritization (high-low/low-high test) should reveal large differences in their response precision (see Figure 3-2 A). In addition, because participants will tend to report the most precisely remembered items first (Adam et al., 2017), high-priority items should be reported first when asked to report both a high- and low-priority item. Importantly, the precision of each response is determined primarily by the proportion of resources it receives, independent of when it was reported. For simplicity, Figures 3-1 and 3-2 depict equal-precision continuous resource allocation within the priority conditions but the property of flexible allocation is not specific to

Figure 3-2

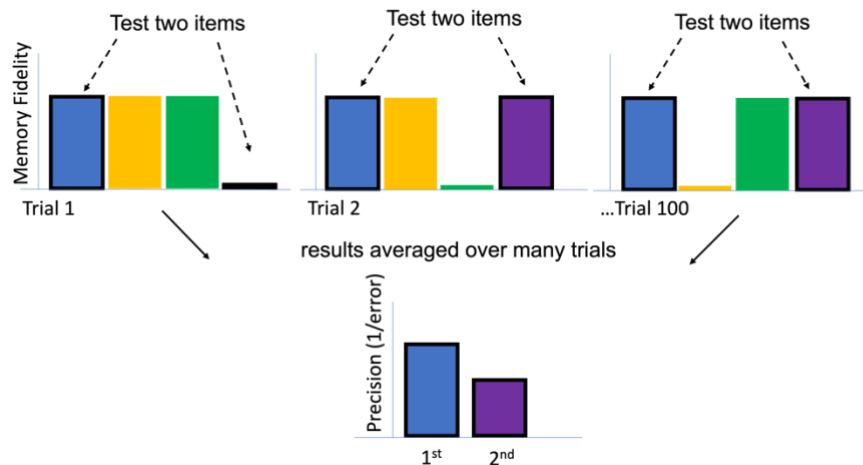
Continued from Figure 3-1, now showing two items tested.

A Continuous Flexible Allocation



Response precision differs from each other according to the item's memory fidelity matching the attentional priority

B Discrete Probabilistic Inclusion



Response precision for the two test items may be equal (in-memory) or extremely different (one in memory, one guessed) on each trial

Averaged responses for test items reflect a mixture of in-memory and out of memory responses (guessing) weighted by the frequency the item was in memory.

equal precision models and is congruent with variable precision models of continuous VSTM resources (e.g., van den Berg et al., 2012). In the current study it is predicted that items of equivalent prioritization should have similar resource allocation.

Predictions of probabilistic encoding on multiple responses' precision

In contrast, the discrete resource model predicts that items in memory are stored with equal precision; thus, the averaged precision reflects the proportion of the time the item was 'in-memory', rather than reflecting differences in within trial resource allocation. In effect, the two responses will reflect different proportions of in-memory and not-in-memory responses where the proportion of in-memory responses is determined first by the prioritization and secondly by the response order. High-priority items should be 'in-memory' on nearly every trial, so long as then number of high-priority items is within the discrete capacity limit; however, low-priority items should only be in-memory if there is sufficient space after storing high-priority items. For example, if high-priority items occupy two of three memory "slots", only one low-priority item could be stored. Thus, responses in high-low test trials reflect a mix of trials in which both items are in-memory and trials in which one item is in memory and the other is out-of-memory.

Probabilistic encoding predicts that when testing two low-priority items the second test item is much less likely to be in-memory, so there is a second source of error due to a larger proportion of guess responses on average in addition to the effects of response order. This is an additional source of response error in the second low-low response that is not predicted by flexible allocation. This should result in greater difference between two low-priority responses than between two high-priority responses, see Figure 3-2 B. In other words, probabilistic encoding predicts a response order by

condition interaction where there is a larger effect of response order in the low-low test condition than in the high-high test condition.

In-memory low-priority responses: an alternative test of the predictions of probabilistic encoding.

As described above, probabilistic encoding proposes that response precision measured by error distributions is a result of a weighted average of in-memory and not-in-memory responses. Thus, instead of comparing the error distributions it is possible to predict the proportion of low-low trials with in-memory responses and compare to the observed proportion of low-low trials with in-memory responses. Probabilistic encoding predicts that in-memory items are all equally precise, but the selection of memory access is determined by attentional goals. This should result in a proportion of trials in which the precision of responses is similar to that of the high-priority items. This would mean that there is a logically determined expected proportion of trials where the tested low-priority items should have been sampled from in-memory items, which can be compared to the observed proportion of trials that are highly precise (thus reflect an in-memory item). By testing multiple low-priority items per trial, participants have a better chance of having at least one of the probed items in memory than if only one of the low-priority items were tested. This analysis is a more direct test of the predictions of probabilistic encoding, by converting the distribution of response precision into a discrete number of in-memory responses.

The current study.

The aim of the current study is to test whether multiple responses can reveal whether items in memory are probabilistically assigned to a limited discrete resource or

flexibly allocated within a continuous resource by attentional prioritisation. The current study was conducted as two experiments that are combined for the current analyses. Experiment 1 used two levels of attentional priority: each high-priority tested on 31.25% of trials, and each of the four low-priority items were tested on 9.38% of trials. This resulted in three conditions, depending on which items were probed on each trial: Both high-priority (high-high), two low-priority (low-low), and one each of high- and low-priority (high-low). In Experiment 2, three levels of attentional priority were assigned; however, as with previous results from our lab (Lockhart et al., 2024), participants tended to group the two cued items into one priority condition, and thus the results from this sample were combined with Experiment 1 (see Appendix 3A for analysis demonstrating that participants combined priority level assignment). In this second experiment, each high-priority item was tested on an average of 32.5% of trials (after combining high- and medium-priority items), and each of the four low-priority items were tested on 8.75% of trials. Importantly, participants were allowed to freely select which of the two test items they would report first, with the hypothesis that first responses were more likely to reflect “in-memory” items, according to discrete models (Adam et al., 2017).

Methods

Participants

Participants were recruited from Brock University using an online research pool or through posters around campus. Participants were offered a choice of research credit for courses or paid at a rate of \$10/hour of participation. All participants reported normal or corrected to normal vision and were tested for normal colour vision. All procedures were approved by the Brock Research Ethics Board.

In Experiment 1 there were twenty participants from Brock University (4 males, 16 females), with a mean age of 20.50 (range 18-26). Participants were given the option to listen to non-lyrical music during the task on their own devices. Four participants chose to listen to music.

In Experiment 2 there were twenty-two participants from Brock University (2 males, 20 females), with mean age 19.9 (range 18 – 33). Two participants did not complete all the trials of their study session and were not analyzed in the analysis that focused on the Experiment 2 sample only (Appendix 3A) but are analyzed here.

Although Experiments 1 and 2 contain the same conditions in name, the exact priority for the conditions are different: high 31.5% – 32.5%, low 8.75%-9.375%. For two participants in Experiment 2 that didn't complete all trials, the resulting ratio of conditions was close to the combined studies, so they were included in the combined sample. Participant A finished 375 of 500 trials resulting in 31.9% high-priority responses and 9.1% low priority response. Participant B finished 346 of 500 trials resulting in 32.6% high-priority and 8.7% low-priority responses.

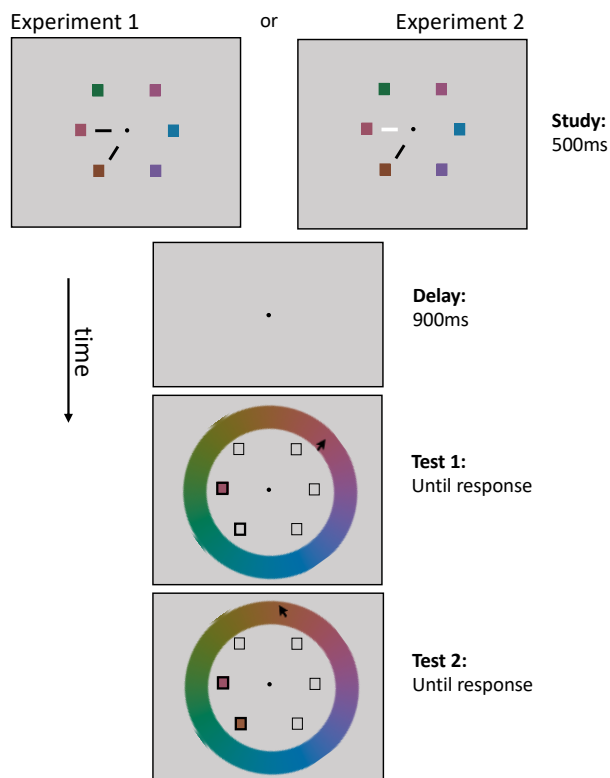
Procedure and Stimuli:

The experiments consisted of a delay-estimation task with two simultaneous spatial cues indicating the high-priority items that were more likely to be probed for test, and with two items probed for test on each trial. There were always six items presented around a central fixation for 500ms; two items were designated high-priority by simultaneously presented spatial cues pointing at the items, and 4 items designated low-priority. In Experiment 1, both the cues were black, but in Experiment 2, one cue was black, and one cue was white to indicate a high and medium priority item (but see

Appendix 3A). After a 900ms delay, each trial concluded with two self-paced responses of the test items in self-selected order. Participants chose which location to report first by clicking inside the square and choosing their response colour with the mouse. The colour inside the chosen tested location changed to reflect the mouse position around the colour wheel (see Figure 3-3). The experiment instructions were written on the screen and read aloud by the experimenter to ensure that all participants received the instructions. All participants completed a session of practice trials before the experimental session and were given the opportunity to ask question or clarify the instructions.

Figure 0-3

Example trial for Experiments 1 and 2.



Note. Study phase in Experiment 1 had two black cues, while in Experiment 2 there was one black and one white cue. Delay and test phases looked the same for Experiments 1 and 2.

In Experiment 1, the test items could have come from three possible combinations: two high-priority items (high-high), two low-priority items (low-low), and finally one item each of high- and low-priority (high-low) to measure memory representations within a trial. The participants were told the frequency of each condition at the start of the sessions in the form “in 50% of trials, both cued items will be probed”, “in 25% of trials, one cued and one uncued item will be probed”. For Experiment 1, participants completed 400 intermixed trials: 200 of high-high (400 high priority reports), 100 trials of high-low (100 each of high- and low-priority reports. This meant that each high-priority item would be tested 31.25% of the time (500/800 reports were high-priority items, but 300/400 trials probed at least one high-priority item), and each low-priority item would be tested 9.375% of the time (300 of 800 reports were low-priority items, but 200 of 400 trials probed at least one low-priority item). These values (31.25% and 9.375%) were not explicitly stated to the participants.

In Experiment 2, the instructions assigned three task-relevant priority levels, however results indicated that participants combined the high and medium conditions making the task the same as in Experiment 1. See Appendix 3A for an analysis demonstrating that high and medium priority were equivalent. Participants completed 500 intermixed trials, there were to be 200 trials of high and medium priority test (equivalent to high-high in Experiment 1), 200 trials of high and low (equivalent to high-low), 50 trials of medium and low (also equivalent to high-low), and 50 trials of two low-priority test (low-low). When combining the high and medium to a unified high-priority condition each item was tested on an average of 32.5% of trials (40% and 25%) and each low-priority item was tested on 8.75% of trials. Participants were told the frequency of each

condition at the start of the experiment.

Additionally, all participants completed a standard whole display change detection experiment with set sizes (N) 2, 4, 6 to measure individual capacity prior to the main experiment. The maximum value was determined to be the individual's capacity.

Capacity (K) was estimated for each set size by Pashler's formula (1988):

$$K = N \times \frac{\text{hit rate} - \text{false alarm rate}}{1 - \text{false alarm rate}} \quad (3.1)$$

Apparatus.

All stimuli were presented using PsychoPy (Peirce, 2008) on a 20" LCD display at a distance of approximately 57 cm. All stimuli were coloured squares of 1° of visual angle with a constant distance from central fixation, the stimuli were equally spaced. Colour stimuli were randomly selected from 360 unique colours from a continuous colour wheel made using CIE L*a*b* colour space with coordinates of a = -6 and b = 14 with a radius of 49, calibrated to the monitor and sampled with a minimum distance of 30°. All responses were made using the mouse by selecting the colour from the colour wheel that best matched their memory.

Analysis of Response Precision.

Response error was determined by circular distance between the response colour and the target colour. The circular standard deviation of error was calculated on an assumed von Mises distribution using the *circular* package (Agostinelli & Lund, 2023). Figures of the error distributions are based on raincloud plots (Allen et al., 2021). Response time was recorded to evaluate response order. Data was analysed using R *stats* package, and JASP 0.17 (JASP Team, 2023) for Bayesian repeated-measures ANOVAs, with post hoc and planned comparisons using Bayesian paired sample t-tests. For

convenience BF_{01} representing evidence in favour of the null hypothesis is given when the result indicates strong evidence of no difference between samples (italicized for clarity); otherwise BF_{10} or $\log(BF_{10})$ is given for evidence in favour of the alternative hypothesis more comparable to p-values in frequentists hypothesis testing. BF_{10} values between 3 and 10 are considered moderate evidence, between 10 and 30 are strong evidence in favour of the alternative, while $\log(BF_{10})$ over 3.4 is used for very strong and over 4.6 is interpreted as decisive evidence in favour of the alternative hypothesis. Note that in Chapter 2 this level of evidence was presented as $BF_{10} > 1000$.

Planned comparisons were made between the high- and low-priority responses from the mixed test condition (high-low) with their counterparts of the same priority conditions to address the two alternative theories. It was expected that response precision would decrease for the second reported item (i.e., response order effects: Adam et al., 2017); thus, comparing the second response on mixed report trials to the second responses from high-high and low-low report conditions isolates the effect of response order. Similarly, comparing the first responses of high-priority items in high-high versus high-low controls for the effect of response order; however, these items differ in one important way: the mixed test condition probed a random high-priority item whereas when both high-priority items are tested it is a truly self-selected response order.

Analysis of discrete resource in-memory responses: observed versus predicted.

A second analysis investigated whether the observed data matched predictions of probabilistic encoding by comparing a predicted proportion of high-resolution in-memory responses from the low-low test condition with the observed proportion of high-resolution responses. In the current experiment, since multiple low-priority items are

probed in this condition, participants have better odds of having *at least one* of the low-priority items in memory and thus giving a high-resolution in-memory response than if only one of the low-priority items were tested as in a typical single probe test.

The theoretic discrete capacity limit determines that only some items are able to be ‘in-memory’ on any given trial, while probabilistic encoding determines which items are granted access based on the person’s priorities (Awh et al., 2012). In the current experiment, the probe probabilities are goal directing cues which should lead to selection of the cued high-priority items over the uncued low-priority items into memory. Over the full experiment, due to this limited capacity, the tested items will be in memory some of the time. The probabilistic encoding hypothesis predicts that the proportion of in-memory response trials would be related to the probe probability of the tested item since the probe probability determines the likelihood that item is in memory. The proportion of in-memory responses should be equal to the probability that an in-memory item was tested – supposing that when an “in-memory” item is tested it must result in a high-resolution “in-memory response”. Thus, by calculating the probability that an in-memory item will be tested, we have a prediction for the expected proportion of in-memory responses that can be compared with the observed proportion of in-memory responses. Importantly, all calculations are tailored to each participant’s performance: the expected proportion of in-memory responses is determined by an individual’s memory capacity, and “in-memory response” is determined by the individual’s response precision since in-memory responses are expected to be high-resolution in discrete resource models.

First, the expected proportion of in-memory responses was determined. For this explanation let the two high-priority items be represented by [X Y], and the four low-priority items be [A B C D].

Step 1: The test of *two low-priority* items results in a set of six possible combinations of the four low-priority items.

(1) Determine all possible test sets for low-low

test set = ${}^4C_2 = 6$:

[AB AC AD BC BD CD]

note. the number of sets any individual item is in equals 3

Step 2: On each trial, each participant selects an integer value of items to be in-memory up to their theoretic capacity (K) from all to-be remembered items. However, since the items are not equal priority, if K is greater than 2 – note that K -estimates measured in a separate change detection task which was greater than 2 for all participants – the in-memory set is a selection of both high-priority items and a selection of ($K-2$) of the four low-priority items.¹

(2) if $K > 2$;

high-priority in-memory set = 2C_2 : [XY]

low-priority in-memory set = ${}^4C_{(K-2)}$: for $K = 2$ to 5:

[], [A B C D], [AB AC AD BC BD CD], [ABC ABD ACD BCD]

Step 3: By taking the probability that the item(s) in memory could be in the test set, we can determine the expected proportion of responses that should be in-memory for each integer capacity. Thus, if capacity was 3, there would be one slot remaining for a low-priority item, which would be tested in 50% of all low-low test trials meaning that 50% of these responses are expected to be in-memory responses (i.e., the expected

¹ This assumption may *underestimate* the expected proportion of trials with at least one low-priority item because it assumes that participants never encoded only 1 high-priority item allowing for another low-priority item.

proportion is 0.5). Table 3-1 summarizes the probability that at least one item in memory is tested, the probability that none of the items in memory are tested and that two of the items in memory are tested for whole number capacity estimates between two and five.

(3) The probability that the item in memory will be tested for each whole number capacity from $K = 2$ to $K = 5$. The identity of the remembered item doesn't matter until test so let the first remembered item be "A", then "B", etc.

a. if $K = 2$,
 $P(I) = 0$

b. if $K = 3$,
 $P(A_{test}) = P(A)$
 $= 0.5$

c. if $K = 4$,
 $P(A_{test} \text{ or } B_{test}) = P(A) + P(B) - P(AB)$
 $= 0.5 + 0.5 - 0.167$
 $= .667$

d. if $K = 5$,
 $P(A_{test} \text{ or } B_{test} \text{ or } C_{test})$
 $= 1$

Table 0-1

Probability that the probed item was in-memory for 'low-low' test trials by capacity estimates.

Memory Scenario:

"____ of the tested items are in memory"

	K-estimate			
	2	3	4	5
None	1	.5	.166	0
At least one	0	.5	.833	1
Two	0	0	.166	.5

Note. It was assumed that $K-2$ low-priority items could be in memory. 'At least one' is the sum of the probability that only 1 of the test items are in memory and the probability that two test items are in memory.

Step 4: The expected proportions of in-memory responses from Table 3-1 were adjusted for each participants' individual change detection estimated capacity since this could be a rational number whereas the theoretic K values in Table 3-1 had to be integer

values. A participant whose K-estimate was 3.05 would have a whole number expected proportion of 0.5 of responses but due to the extra 0.05 capacity, the value is adjusted to 0.52 using the linear difference between the higher ($K = 4$) and lower ($K = 3$) expected proportions. This interpretation of a rational K-estimate is equivalent to suggesting that on 5% of trials, this participant had $K = 4$ capacity and on 95% of trials had a $K = 3$ capacity, rather than assuming that capacity is a fixed value. This results in an expectation that 52% of first responses come from an in-memory item. See Appendix 3B for alternate analysis using the rounded whole number capacity and the rounded down (floor) whole number capacity as predictors instead.

$$(4) (K\text{-estimate} - K_{\text{floor}}) \times (\text{Exp.Prop.}_{\text{upper}} - \text{Exp.Prop.}_{\text{lower}}) + \text{Exp.Prop.}_{\text{lower}} \\ (3.05 - 3) \times (0.833 - 0.5) + 0.5 = 0.52$$

Critically, it was expected that the high-priority items would be in-memory nearly every trial and assumed that an in-memory response would be made every time the tested item was in memory, and according to discrete resource theories, an in-memory response is high-resolution. Thus, I define an in-memory response as one that looks like a high-priority item response. Specifically, I operationally defined this as responses made within the value of one standard deviation for the first response of the high-high test condition for each participant.

Finally, to find the predicted proportion of in-memory high-resolution low-priority first responses, a further adjustment was made based on the individual's performance. Since we had assumed that both high-priority items were always in-memory when calculating the number of slots available to low-priority items, the adjusted expected proportion of in-memory responses was multiplied by the observed proportion of in-memory high-resolution first responses from the high-high test condition. In other

words, since 100% of first responses from the high-high test condition cannot be within one standard deviation, the expected proportion of low-low responses was similarly adjusted. In the example above, if the participant with a capacity of 3.05 only had 80% of their high-high first responses within one standard deviation, the predicted proportion of high-resolution in-memory responses in this condition would now be 41.6% of trials (i.e., 0.8×0.52). See Appendix 3C for analysis with an alternate definition of a high-resolution 'in-memory' response based on removing a uniform guessing distribution.

These analysis steps were also applied to the second responses of the low-priority test condition where even fewer responses are expected to reflect a high-resolution in-memory item. To account for the effects of response order, absolute response error had to be made within one standard deviation of the second response from the high-high test condition instead of the first response. Since it was predicted that the best remembered item would be reported first, the second response was only predicted to be in-memory when the first response had also been in memory, thus the observed proportion of high-resolution in-memory second responses was restricted to trials where the first item had already met the definition of in-memory as well.

Results

Participants were prompted to report the colours of two of the six items presented. The two selected locations tested items from the conditions 1) *high-high* (cuing probability of 31.25% per item in Experiment 1 or 32.5 % per item in experiment 2), 2) *low-low* (9.3% or 8.75 % per item in Experiment 1 and 2 respectively), or 3) mixed test (one high- and one low-priority). Responses were self-ordered. The mixed condition was split by response order. Resulting in a total of four conditions: *high-high*, *high-low*, *low-*

high, and *low-low*. There were no instructions indicating which item participants should report first, but it was anticipated from prior studies that participants would report the better remembered item first (Adam et al., 2017). The mixed test trials were reported in the order *high-low* on approximately 78% of trials (median 84.5%) but this varied by participant between 48 – 99%. Seven participants with fewer than 10 *low-high* mixed test trials were excluded from pairwise analyses of low-high responses.

Response Precision

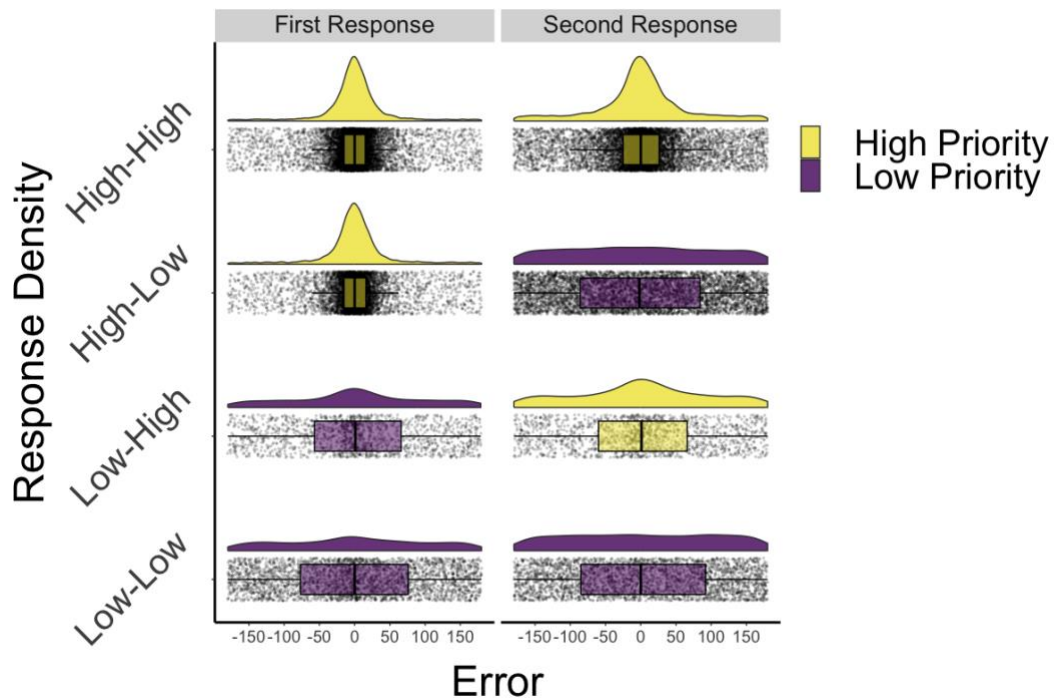
To evaluate how two items are reported under differing priority conditions, a 4x2 Bayesian repeated-measures ANOVA was performed, with experiment as a between-subject factor. There was no effect of experiment, $BF_{01} = 3.605$ in favour of the null, which supports the analysis of the combined samples. There were large main effects of condition, $\log(BF_{10}) = 56.15$, and response order, $\log(BF_{10}) = 19.75$, as well as a large interaction effect $\log(BF_{10}) = 77.18$. The main effect of order suggests that, in general, there was a difference between the first and second responses, consistent with the first-report advantage found by Adam et al. (2017) and other experiments with serial recall or multiple report (e.g., Cowan et al., 2002). The main effect of order remained significant when excluding mixed test trials in a Bayesian 2x2 repeated-measures ANOVA, $\log(BF_{10}) = 15.3$, clarifying that the effect was not driven by the difference in high- versus low-priority responses. See Figure 3-4.

First versus Second Responses: condition by order interactions

I did not make predictions for the magnitude of response order effects in high-high responses for either theory since the effect is not driven by either mechanism. Bayesian paired sample t-tests were used to investigate the condition by order interaction.

Figure 3-4

Response precision for first and second responses by condition.



Note. Distribution of error in degrees by response conditions. Boxplots are overlaid on all data points. Based on Raincloud plots (Allen et al., 2021).

There was decisive evidence of a difference between the first ($M = 39.53$, 95% CI = [35.20 – 43.86]) and second responses ($M = 60.49$, 95% CI = [53.50 – 67.47]) in the high-high test trials $\log(BF_{10}) = 21.76$; a response order effect was predicted by both models as is commonly observed for serial reports (Adam et al., 2017; Cowan et al., 2002).

There was very strong evidence for a difference in the first ($M = 104.50$, 95% CI = [97.79 – 111.22]) and second responses ($M = 119.57$, 95% CI = [114.94– 124.20]) in low-low test trials $\log(BF_{10}) = 4.59$. Probabilistic encoding predicts a second source of error in the second response for the low-low condition in the form of a greater proportion of guess responses. This second source of error should have resulted in a larger magnitude difference between first and second responses in comparison to the difference

between the first and second responses of the high-high test condition which only had the effects of response interference and time since encoding adding to the response error. The Bayesian 2x2 repeated-measures ANOVA found substantial evidence for a difference in the effect of response order for each condition (interaction $\log(\text{BF}_{10}) = 52.34$) but in the opposite direction in that the response order effect was instead larger for the high-high condition than the low-low condition as was predicted by probabilistic encoding.

In the *high-low* mixed test condition there was decisive evidence of a difference between the first (high-priority) ($M = 45.12$, 95% CI = [41.11 – 49.14]) and second (low-priority) responses ($M = 121.40$, 95% CI = [116.70 – 126.09]), $\log(\text{BF}_{10}) = 21.76$ as predicted by both flexible allocation and probabilistic encoding. However, there was evidence that there was no difference for the reverse order (low-high) responses between the first (low-priority) ($M = 89.23$, 95% CI = [81.86 – 96.60]) and second (high-priority) responses ($M = 93.48$, 95% CI = [85.26– 101.70]), $\text{BF}_{01} = 4.15$ in the direction of the null hypothesis (excluding 7 participants with fewer than 10 trials). Specific predictions for low-high response order were not made a priori.

Mixed priority versus same-priority responses

In addition to predictions for response order within the test conditions, each theory has predictions about the mixed test response precision in comparison with the corresponding same-priority test conditions. Firstly, both flexible allocation and probabilistic encoding predict that the precision of the first responses from the *high-low* mixed test and *high-high* test will be equal. Unexpectedly, the precision of first responses (high-low $M = 50.54$ [95CI: 46.36, 54.72] versus high-high $M = 43.98$ [95CI: 39.25, 48.71]) had strong evidence of a difference with greater error when reporting in the

mixed test: $BF_{10} = 36.68$. This result was also found in the Experiment 1 sample alone (Experiment 1, $N = 20$, $BF_{10} = 5.44$; *high-high* first response $M = 39.73$ [95CI: 33.76, 45.69]; *high-low* first response $M = 47.54$ [95CI: 41.68, 53.40]). Experiment 2 had been designed with a high and medium priority item instead of two high priority items. Unlike Experiment 1, there was no evidence either way in the comparison between the high-priority items regardless of the test condition: high-medium compared with the high-low conditions, $BF = 1.438$. Similarly, for the medium-priority item when reported first in medium-high compared with medium-low, $BF = 1.255$. Further analysis of Experiment 2 is found in Appendix 3A.

Flexible allocation predicts the precision of the second responses from *high-low* mixed test and the low-low test conditions will be equal, given that all low-priority items would have similar resources allocated. These two predictions follow that these items have the same priority (i.e., same resource allocation) and same response order effects (response interference and time since encoding). In contrast, probabilistic encoding would predict higher error in the second response in the low-low test condition, because this item is less likely to be in memory. For example, an individual with a capacity of 4, only has a 16.6% chance that both of the tested items were in memory in the low-low test, but a 50% chance that the low-priority item in the high-low test would have been in memory. As predicted by flexible allocation, there was evidence that there was no difference in the second response precision between the mixed low-priority item and low-low test conditions $BF_{01} = 5.12$, in the direction of the null hypothesis.

Probabilistic Encoding: Predicting High-resolution In-memory Responses

Probabilistic encoding predicts that items are encoded in proportion to the priority

of the item, such that when capacity allows, some low-priority items should be encoded into memory. Thus, some proportion of responses from the low-low test condition should have come from in-memory items determined by an individual's K -estimate. Further, in an equal-precision discrete resource model, response precision for any items that were in-memory would be equivalent regardless of item priority.

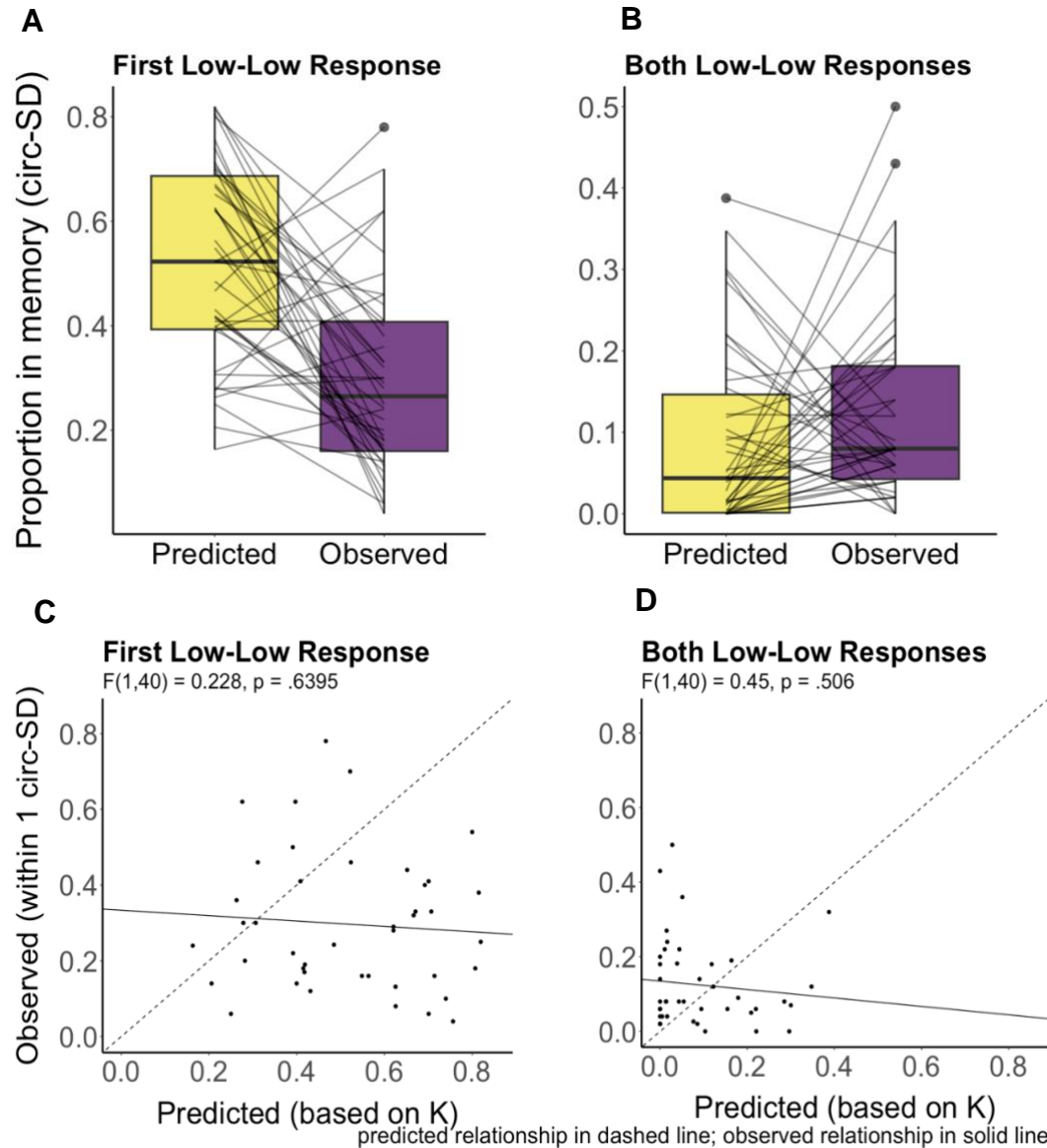
To test these hypotheses, in-memory responses were defined for each participant as responses with the absolute value of error was within the value for the circular standard deviation of the first response error in the high-high test condition. On average 30% of first responses from the low-low test condition were observed to be high-resolution responses (i.e., a proportion of 0.30, range 0.04-0.78). However, the predicted proportion of responses that should have been in-memory according to probabilistic encoding would have been an average of 52% of the time. The difference between the expected proportion and the observed proportion ($M = 0.223$) was analysed in a paired sample t -test and found to be significantly different from zero, $\log(BF_{10}) = 8.542$. See Figure 3-5 A. This suggests that the observed proportion of first responses in the low-low test condition was less than the amount predicted by a proportional encoding model based on the measured individual's capacity and a discrete selection of low-priority items granted access to an equal memory store. Alternative assumptions of capacity estimates and in-memory definition can be reviewed in Appendices 3B and 3C.

Further, investigating the observed proportion of trials where both the first and second responses were considered in-memory, there was an average of 12.4% of trials (i.e., a proportion of 0.124 range 0-0.5). The second response was considered in-memory if it was within one standard deviation of the second response in the high-high test

condition and the first response of that trial had also been considered in memory. An average of 9.1% of trials were predicted to be in memory.

Figure 0-5

Predicted versus observed proportions of low-low in-memory responses.



Note. An “in-memory response” is defined as having absolute response error within 1 circular standard deviation of the high-high condition. **A-B**, boxplots show the distribution, lines connect individual participants. **C-D**, correlation between predicted and observed proportions. Linear model fit shown in solid line. Predicted fit shown in dashed line. **A.** First responses only. Significantly fewer responses were observed than would be predicted by capacity. **B.** Both the first and second responses. There is not a significant difference. **C.** First Responses only. **D.** Both first and second responses.

The difference between the predicted and observed values were tested in a Bayesian paired sample t-test. There was not a significant difference ($M = -0.034$); however, this time there were slightly more observed in-memory responses than predicted, $BF_{01} = 2.737$, suggesting anecdotal evidence for the null hypothesis of no-difference between the observed and predicted, see Figure 3-5 B. Further investigation of the correlation between the predicted and observed value, however, does not produce a positive relationship between the predicted and observed values (First-only: $\beta = -0.071$, $F(1, 40) = 0.228$, $p = .6395$; Both: $\beta = -0.113$, $F(1, 40) = 0.45$, $p = .506$), see Figures 3-5 C and D.

Discussion

The current study sought to investigate a specific alternative explanation for previous findings attributed to flexible allocation of a continuous VSTM resource; namely, whether the results attributed to flexible resource allocation could be instead attributed to probabilistic encoding. Two responses were collected on each trial which may have come from two high-priority items, two low-priority items or a mix of both. In two separate analytic approaches, I have not found support for the probabilistic encoding alternative. When comparing the response precision between high- and low-priority items reported in the same trials, there is clear evidence that the difference in memory priority leads to differences in response precision. Further predictions of resource allocation, such as equivalent response precision for two low-priority responses on the same trial was also found. An alternative approach categorized responses as being an in-memory item or not and compared the proportion of observed in memory responses to a predicted proportion of in-memory responses based on the individual's capacity and response precision. This

analysis found significantly fewer first low-priority responses were from in-memory items than would be predicted if participants were filling the remaining capacity with a random selection of low-priority items.

Response Precision

Although flexible allocation and probabilistic encoding models propose different mechanisms, they sometimes lead to similar predictions. For instance, both models predict that the high- and low-priority item should have different response precision. Flexible allocation makes this prediction because each item is assigned a different parametric amount of a VSTM resource. Probabilistic encoding makes this prediction because high-priority items are in memory more often than low-priority items. Overall, the results best align with a mechanism of flexible allocation, showing that high- and low-priority items are both held in memory, but with predictably distinct fidelity according to their task relevance.

A comparison that had unique predictions for each mechanism was between the two low-low responses. Flexible allocation predicts approximately equal precision within any priority condition. Probabilistic encoding would predict the responses were equivalent if both items were in memory; however, a second low-priority item is unlikely to be in memory. If both high-priority items were stored with a capacity of 4, there is only a 16.7% chance that the second low-priority item tested would also be in memory. Thus, these two low-low responses are predicted to be different from each other on average. Observed was evidence that both low-low responses were equivalent supporting a mechanism of flexible allocation.

Unexpectedly, the precision of the first responses of high-priority items (high-low

versus high-high) were different, which was not predicted by either theory. These items were equivalent in priority and response order so this significant difference must be attributed to another aspect of the test conditions. Although both items were high priority, they may not have been stored equally well. This would suggest there is an effect on memory precision beyond a simple first-report benefit not accounted for by response interference and time since encoding. Although this result was not predicted by flexible allocation it was similarly not predicted by probabilistic encoding since both theories had assumed equal precision for both high-priority items, although for different reasons.

Variable Precision.

The current results suggest that even when participants are prioritizing a small subset of items within the memory array, items with equal priority are not necessarily encoded and maintained with equal resolution. A benefit of testing same-priority (high-high; low-low) and mixed-priority (high-low) conditions is that response error can be compared for the same response order position effectively controlling for many aspects of response order effects. In the current study there is evidence for predicted equivalencies between the second responses of low-low and high-low, but interestingly not first responses of high-high and high-low. This may be because although the first high-priority responses in the high-high and high-low test conditions are equivalent in priority and response order there are notable difference in the ability to select the best remembered item. In high-low test conditions, which of the two high-priority items that was probed was randomly selected. Whereas in high-high condition the test items were the full population of high-priority items, so the best remembered item could always be reported first. The difference between the high-high and high-low first responses suggests that

there was already a difference in the memory stores of high-priority items otherwise these two responses would be equivalent. This evidence aligns with a mechanism for the ‘first report benefit’ being driven by a strategy to report the better remembered item first using meta-knowledge such as the memory strength as opposed to a mechanism where all items were equivalent at the time of first report and only interference or gradual decay reduces the precision of subsequent reports; this evidence is therefore consistent with variable precision models of VSTM (Fougnie et al., 2012; van den Berg et al., 2012). Another VSTM theory that this data may support is the interference model proposed by Oberauer and Lin (2017), whereby one item in memory can be in the focus of attention with higher precision than other items in memory. The current data support the notion that there is variability between items stored in memory rather than equal concurrent memory stores either from continuous or discrete memory resources allocation.

The possibility that two high-priority items have intentionally variable memory quality was the motivation behind assigning three priority levels in Experiment 2. I had predicted that by assigning a middle priority level I would be able to determine which of the two high-priority items would be the better encoded one. However, participants seemingly ignored the attempt to assign which of the high-priority items should be favoured and demonstrated no preference or greater precision for the highest-priority. Despite this there was still clear evidence of variable precision among the high-priority items, it was just not predicted by the assignment of a middle priority. This analysis can be seen in Appendix 3A and replicates the finding in Chapter 2 that three levels of priority are not readily used for resource allocation (Lockhart et al., 2024).

I predicted that both high-priority items would be maintained equally well on all

trials, however, I did not find similar response precision between the two high-priority items even when comparing only the first responses. This could be due to variable precision, but an alternative explanation is that participants only maintained one of the two high-priority items often enough to lead to a lot of variability in the second reported item (i.e., mixing in-memory and not in-memory responses). I had predicted that if participants were only holding three or four items in memory from the set of six that the two high-priority items would be in memory nearly every time. This assumption should result in a more variable average among the low-priority items rather than the high priority items because both high-priority items would be in-memory whereas two low-priority items were likely to come from one in-memory response and one guess response. However, participants may not have used this strategy. Instead, participants could have only remembered one of the high-priority items frequently enough to result in a high number of trials where the second high-priority item was not in memory and the response was essentially a guess. However, if this was the case then I would have expected in Experiment 2 that the highest priority item would have been in memory more often than the medium priority item which was not supported by the response order results or precision in Appendix 3A.

Conversely, I had suggested based on a discrete probabilistic encoding hypothesis that there would have been greater variability in response precision between low-low responses. The first response should have had greater precision because on some of the trials at least one low-priority item should have been in memory by chance and then reported first, as the better remembered item. Whereas significantly fewer trials should have had both low-priority items in memory, the second item would have nearly always

been guessed and thus have low precision. Comparing the first and second responses should have had a greater difference in precision, reflecting greater variability in response error between the first and second responses. However, it was observed that these responses were quite similar in response precision.

Predicted High resolution In-memory Response Proportions

A hypothesis derived from discrete resource models predicts that instead of being encoded into memory with varying levels of precision, each item is probabilistically encoded into memory when there is capacity to do so. Thus, the combined average of responses for low-priority items will include a proportion of responses in which the item was encoded in memory and guesses. This value was estimated as the predicted proportion of in-memory responses for first responses only, and both first and second responses. Comparing observed and predicted proportion of first responses from the two low-priority test condition that were considered “in-memory responses” did not find support for probabilistic encoding of an equal precision discrete memory resource. Specifically, there were significantly fewer in-memory responses than would be predicted by probabilistic encoding.

When considering both first and second responses in the low-low test condition there was weak anecdotal evidence in favour of the null hypothesis, that is that the observed proportion of in-memory responses matched the predicted proportion. Although anecdotal, it could be interpreted that the second low-priority item was, for some reason, easier to remember – as if it could be encoded “for free” when the first item was encoded in memory. Alternatively, this could be a form of strategic guessing, or a gamble that a low-low test trial was “due” and high-priority items were ignored which could explain

this result in a way that is congruent with a mechanism of probabilistic encoding. Further experiments or analysis would be necessary to investigate such participant strategies. It should be noted that this analysis only categorized a second response as being in-memory when the first response was also in-memory. It is expected that the better remembered item will be reported first; however, this is an assumption and the alternative that second items could be in-memory responses when the first item was not tested.

The current analysis made a large assumption that memory capacity is always filled the same way; specifically, always including both high-priority items. However, if instead only one high-priority item was regularly encoded, or sometimes neither were, this would increase the expected proportion of trials that at least one of the tested low-priority items were in memory which would increase the difference between the observed and predicted proportions of in-memory responses. Although the assumption that both high-priority items were always encoded may not be realistic, alternatives that would instead fill capacity with low-priority items are not a better explanation for the observed proportion of in-memory responses because they would have predicted even more in-memory responses than observed.

The meaning of “Not in memory”

Our predictions based on the flexible allocation of a continuous resource hypothesis was based on theories with no true guessing; however, there is evidence that guesses do happen. In disagreement with no-guessing models of continuous resource allocation, a study re-analyzed the data of Emrich et al. (2017) and others and found evidence for guessing for very low-priority items (van den Berg & Ma, 2018). Moreover, Adam et al. (2017) using a randomized probe whole-report procedure observed that some

early responses looked like guesses, consistent with their participant's reported low confidence those responses. Given that in the current study confidence measures were not acquired, it is impossible to assess which trials participants would have self-reported that they were guessing. Regardless of the mechanism of memory allocation, it is probable that aggregated low-priority reports include a significant number of trials for which those items were effectively not in memory, either not encoded or the signal was lost to noise by the time of response. I would argue that there is a distinct difference between memory traces that were never consolidated from their perceptual signals (discrete models), and memory traces that are indistinguishable from noise at the level of the experiment. Although the difference may have little relevance to day-to-day cognition, it is fundamentally distinct in neurologically grounded models of memory processes feeding into basic understanding of core cognition and thus altered states of cognition through disease, plasticity, and individual differences. Conclusion Together, these results add to the body of literature demonstrating that VSTM resources can be flexibly allocated via attentional manipulations. Importantly, it establishes that flexible allocation is a true phenomenon of memory and not an artifact of aggregate responses. Moreover, these data demonstrate that while resource allocation is flexible, memory performance is also variable. Whether this is being able to allocate variable amounts of a memory resource even within a priority condition or protecting the memory stores to differing degrees that results in variable memory resolution at test not explained by an equal VSTM resource.

Although the data did not support the predictions of a mechanism of probabilistic encoding of equal-precision fixed-capacity discrete resource, other varieties of discrete resource models may be able to account for the demonstrated flexibility of allocation and

variability of precision. The novel approach analysing the proportion of in-memory responses attempted to directly observe the predictions of probabilistic encoding rather than its effects on response precision modelling; however, it was necessary to make several assumptions that could be argued or investigated in future studies. Appendices 3B and 3C go through several alternative choices for some of the assumptions used to find the predicted proportion of responses in the two low-priority test condition that should have been from in memory items; however, it was beyond the scope of the paper to design a predictive model for how participants may strategically use their capacity that was not based on the logic that both high-priority items would always be in memory taking one capacity slot each. Whether or not VSTM resources are continuous or discrete, it is apparent that future models and mechanisms of VSTM must consider flexible allocation.

References

- Adam, K. C. S., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology*, *97*, 79–97.
doi.org/10.1016/j.cogpsych.2017.07.001
- Agostinelli, C., & Lund, U. (2023). *R package “circular”: Circular Statistics (0.5-0)* [R].
cran.r-project.org/package=circular
- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2021). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*(63). doi.org/10.12688/wellcomeopenres.15191.1
- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*(7), 622–628.
doi.org/10.1111/j.1467-9280.2007.01949.x
- 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. doi.org/10.1016/j.tics.2012.06.010
- Bays, P. M., Catalao, Raquel. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10):7, 1–11. doi.org/10.1167/9.10.7
- Cowan, N., Saults, J. S., Elliott, E. M., & Moreno, M. V. (2002). Deconfounding serial recall. *Journal of Memory and Language*, *46*(1), 153–177.
doi.org/10.1006/jmla.2001.2805

- Dube, B., Emrich, S. M., & Al-Aidroos, N. (2017). More than a filter: Feature-based attention regulates the distribution of visual working memory resources. *Journal of Experimental Psychology: Human Perception and Performance*, 43(10), 1843–1854. doi.org/10.1037/xhp0000428
- 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. doi.org/10.1037/xhp0000398
- Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, 3, 1229. doi.org/10.1038/ncomms2237
- Hakim, N., deBettencourt, M. T., Awh, E., & Vogel, E. K. (2020). Attention fluctuations impact ongoing maintenance of information in working memory. *Psychonomic Bulletin & Review*, 27(6), 1269–1278. doi.org/10.3758/s13423-020-01790-z
- Huynh Cong, S., & Kerzel, D. (2022). The allocation of working memory resources determines the efficiency of attentional templates in single- and dual-target search. *Journal of Experimental Psychology: General*, 151(12), 2977–2989. doi.org/10.1037/xge0001239
- JASP Team. (2023). JASP (Version 0.17). Retrieved from jasp-stats.org/
- Klyszejko, Z., Rahmati, M., & Curtis, C. E. (2014). Attentional priority determines working memory precision. *Vision Research*, 105, 70–76. doi.org/10.1016/j.visres.2014.09.002

Lockhart, H. A., Dube, B., MacDonald, K. J., Al-Aidroos, N., & Emrich, S. M. (2024).

Limitations on flexible allocation of visual short-term memory resources with multiple levels of goal-directed attentional prioritization. *Attention, Perception, & Psychophysics*, 86(1), 159–170. doi.org/10.3758/s13414-023-02813-2

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

Morey, R. D. (2011). A Bayesian hierarchical model for the measurement of working memory capacity. *Journal of Mathematical Psychology*, 55(1), 8–24. doi.org/10.1016/j.jmp.2010.08.008

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, 144(9), 885–958. doi.org/10.1037/bul0000153

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

Pashler, H. (1988). Familiarity and visual change detection. *Attention, Perception, & Psychophysics*, 44(4), 369–378. doi.org/10.3758/BF03210419

Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, 2. doi.org/10.3389/neuro.11.010.2008

Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin & Review*, 18(2), 324–330. doi.org/10.3758/s13423-011-0055-3

Salahub, C., Lockhart, H. A., Dube, B., Al-Aidroos, N., & Emrich, S. M. (2019).

Electrophysiological correlates of the flexible allocation of visual working memory resources. *Scientific Reports*, *9*, 19428.

doi.org/10.1038/s41598-019-55948-4

van den Berg, R., & Ma, W. J. (2018). A resource-rational theory of set size effects in human visual working memory. *ELife*, *7*, 1-31. doi.org/10.7554/eLife.34963

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.

doi.org/10.1073/pnas.1117465109

Yoo, A. H., Klyszejko, Z., Curtis, C. E., & Ma, W. J. (2018). Strategic allocation of working memory resource. *Scientific Reports*, *8*, 16162.

doi.org/10.1038/s41598-018-34282-1

Appendix 3A: Equivalent precision for high and medium priority items in Experiment 2.

Experiment 2 assigned three priority levels at the study phase of the experiment using a black and white cue to indicate a highest and a medium-priority item in addition to the uncued low-priority items. This was motivated by finding that the first response of high-high in Experiment 1 was better remembered than the high-priority item in the high-low condition. This suggests that items are remembered with variable precision even under equal priority conditions leading to a hypothesis that the choice of the highest priority item could be under goal-directed attentional control. Although with one test item per trial there was no difference in the response precision between high- and medium-priority items (Lockhart et al., 2024), it was predicted that response order (specifically, choice of first response) could reveal a preference for the highest priority item even if the memory precision was approximately equivalent by interpreting the response order as meta-knowledge of the memory strength.

Response Precision

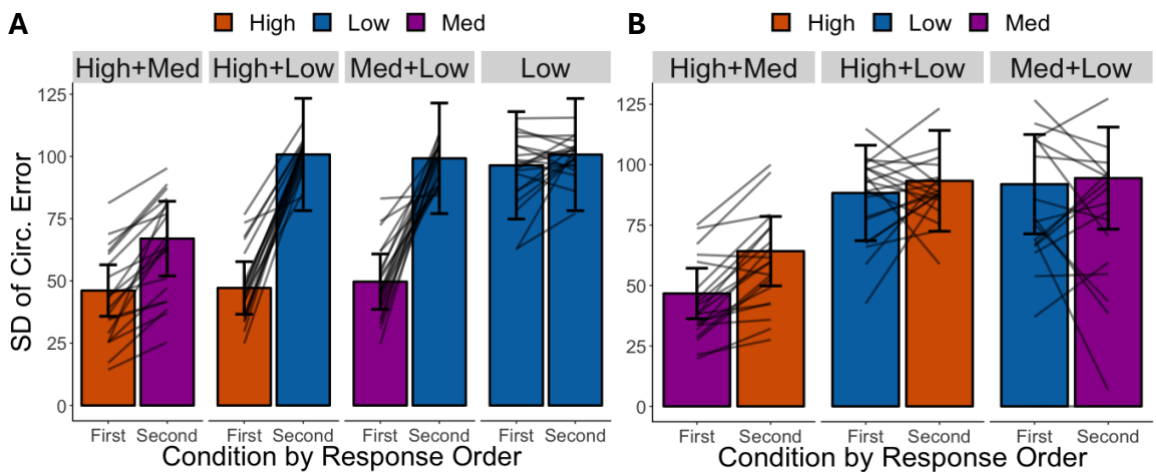
The critical comparison between the highest-priority and medium-priority items did not support the hypothesis that the difference between a preferred and a non-preferred priority item are under goal-directed attentional control either by recall precision or by response order. The circular standard deviation was found for each condition as described in the main text. Planned comparisons were made by Bayesian paired t-tests.

If the medium-priority item was distinct from the highest-priority item, it should be reported with less precision in the same response position. Instead, there was evidence of no difference between the medium- and high-priority items when they were reported

first, $BF_{01} = 4.485$: in the high-medium/medium-high comparison of first responses (high, $M = 39.193$, [95% CI = 32.336, 46.051]; medium, $M = 39.212$, [95% CI = 33.039, 45.385]. Nor when the second item was low-priority, comparing the first responses of high-low and medium-low, $BF_{01} = 2.956$: (high, $M = 42.477$, [95% CI = 37.024, 47.929]; medium, $M = 43.934$ [95% CI = 38.010, 49.858]. Instead, the lack of difference suggests that both high and medium items are equivalent in-memory. See Figure 3-A1 A for high-medium order compared with Figure 3-A1 B for medium-high response order.

Figure 3-A1

Bar graphs for standard deviation of error for Experiment 2 by priority conditions and response order.

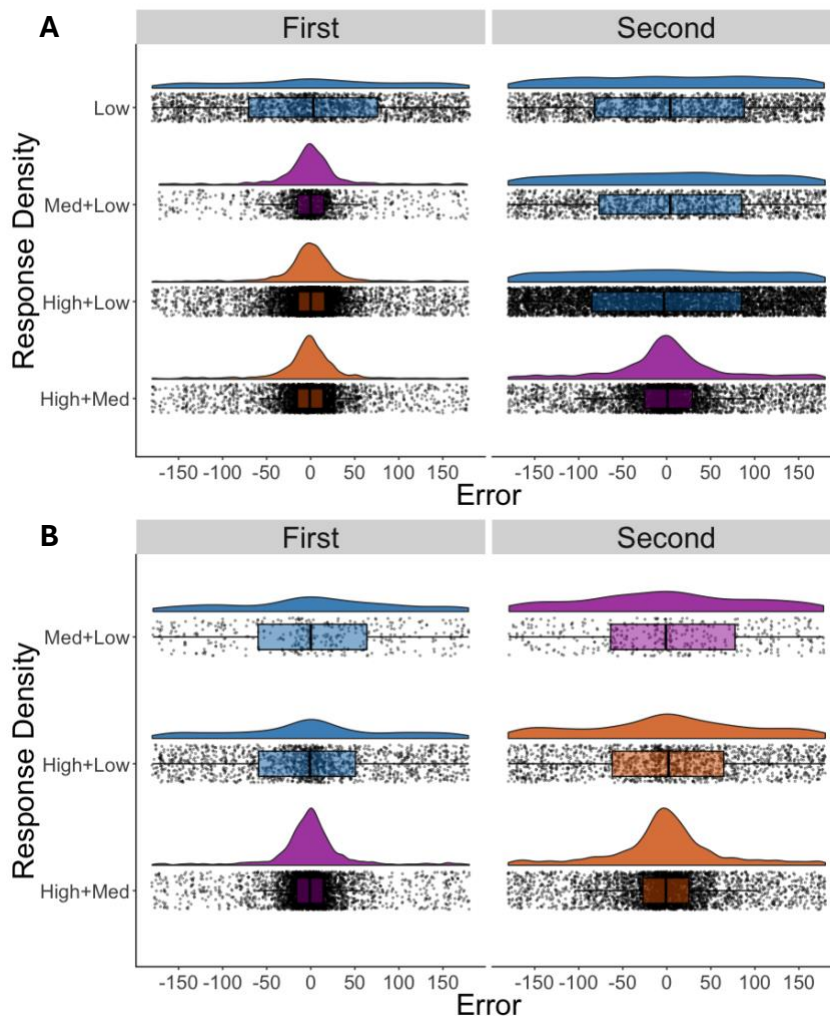


Note. Standard error bars shown. Individual participant data overlaid in grey lines. **A.** Higher priority response made first. **B.** Lower priority response made first.

Within the priority conditions, there was neither evidence for different nor equivalent response precision for the high-priority item regardless of if the second item was medium- or low-priority, $BF = 1.438$. Similarly, for medium-priority item regardless of if the second item was high- or low-priority, $BF = 1.255$. If the memory fidelity was driven entirely by the goal-directed priority level these reports should have been equivalent regardless of the second reported item's priority level.

Overall, for all comparisons of high- and medium-priority items in the first response position there was no evidence of a difference to support the hypothesis that these priority levels were distinctly held in memory. Instead, this result supports the notion that the memory precision of the high-priority and medium-priority item are not distinct from each other as had been found in Chapter 2 (Lockhart et al., 2024). See Figure 3-A2 for response distributions for each priority condition and response order.

Figure 3-A2
Response density distributions of Experiment 2.



Note. Distribution of participants' error in degrees by probe probability conditions. Boxplots overlaid on all data points. Based on Raincloud plots (Allen et al., 2021)
A. Higher priority response made first. **B.** Lower-priority responses made first.

Response Order

A second method to test if the high- and medium-priority items were prioritized as directed is to test the frequency of choosing the higher priority item first. To determine if one selection was made over the other at greater than chance levels the proportion of each choice was compared to chance performance. For each participant the number of trials for each response order was counted and converted into a percent. Chance performance was calculated using the `binom.test()` function in *R stats* package with 0.5 as chance for each condition to find the 95% confidence interval.

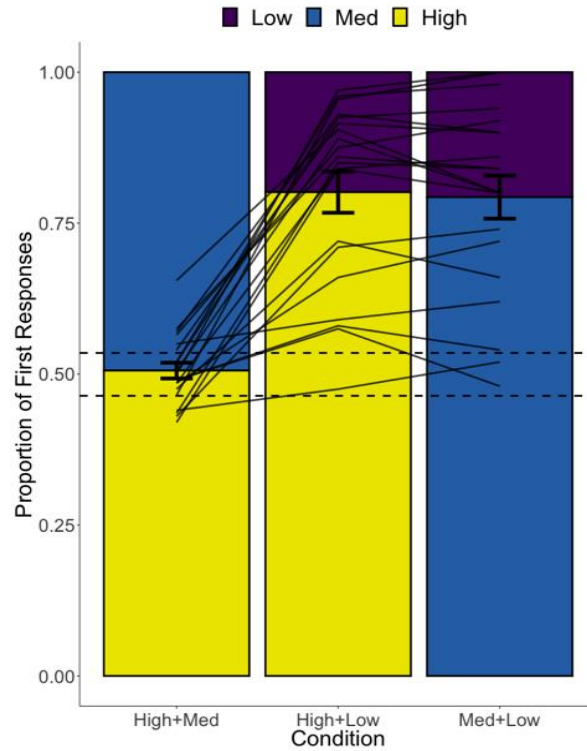
There is no evidence to support that the higher priority item was selected first more often when a high- and medium-priority item were tested together. It was expected that the higher-priority item would be reported first in all conditions including the high-medium condition. Surprisingly, this analysis indicated that participants did not report the highest priority item first at greater than chance levels. The highest priority item was selected first on 50.6% of trials on average. Only 6 of 20 participants selected the high-priority first more often than chance, while another 4 selected the medium first more often than chance (95% confidence upper boundary 53.5%). Conversely, on high-low and medium-low test conditions there is clear evidence that the higher priority item was reported first well above chance levels (high, 80.0%; medium 79.3%; Figure 3-A3).

Thus, in Experiment 2 there is a limitation on the number of goal-directed attentional control sets that can be engaged to allocate memory resources. Specifically, in line with our own past results (Lockhart et al., 2024), there was no evidence that participants allocate memory resources according to the three different priority levels tested, as the high- and medium-priority conditions had the same level of response

precision. This result supports the decision to combine the data with Experiment 1 by treating the medium-priority item as a second high-priority item.

Figure 0-A3

Proportion of first responses made to either priority level by condition



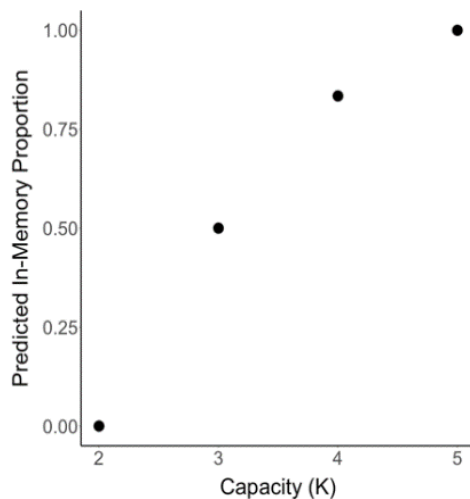
Note. Individual participants shown in semi-transparent lines. Standard error bars shown. 95% Confidence intervals for chance performance (50%) in dashed line.

Appendix 3B: Alternate assumptions for predicting the proportion of in-memory responses.

The expected proportion of in-memory responses was determined logically using the probability that a discrete number of in-memory items would be probed. After storing the two high-priority items, the remaining capacity is used for low-priority items. Although participants may have brought their own biases and strategies to select which low-priority items were in memory this does not impact the expected proportion since selection for test of two of the four low-priority items was randomized. It was only possible to calculate the expected number of items in memory using whole numbers as it is not possible to put half an item in memory. See Figure 3-B1. However, the K-estimates using Pashler's formula (3.1) could be rational numbers which creates an opportunity for differing interpretations of the meaning of fractional portion of capacity estimates in a discrete memory system for a single trial. In the main analysis the choice was made to

Figure 3-B1

Predicted In-memory responses for integer capacity estimate values.



Note. Only whole number capacity values can be used to predict proportion of in-memory responses.

represent this fraction as if on some proportion of trials, participants had an extra capacity slot.

Although the capacity estimates are rational numbers, they are intended to represent a true integer value (Rouder et al., 2011). Thus, an alternative interpretation of the fractional portion of K-estimation could instead be rounded to the nearest whole value for convenience. Using a rounded whole number capacity the difference between predicted and observed proportions was 0.224, $\log(\text{BF}_{10}) = 8.687$, not affecting the interpretation that there is no evidence that the observed proportion of trials with ‘in-memory’ responses matched the predicted proportion of trials.

Another possibility could assume ‘fractional slots’ represented by the rational K-estimates are due to informed guessing and thus over represent the true number of items in memory and the fractional values should be rounded down. This is a similar assumption made for the starting point of the computational models tested by Hakim and colleagues (2020). When using rounded down K-estimates (floor) the difference between the observed and predicted proportions was no longer significant ($M = -0.111$, $\text{BF}_{10} = 1.471$); however, this was merely anecdotal and does not provide evidence that the observed and predicted values matched either.

However, described in depth by Morey (2011), Pashler’s estimate of capacity tends to *underestimate* true capacity due to not accounting for lapses in attention thus rounding up may be crude adjustment that better approaches the true integer capacity. Using rounded up capacity estimates (ceiling), there is a larger difference between observed and predicted $M = 0.336$, $\log(\text{BF}_{10}) = 18.927$, which does not change the interpretation that the observed and predicted values did not match.

Appendix 3C: Alternate definition of “in-memory”

The version of the analysis in the main text has an underlying assumption that responses are never accidentally ‘correct’ which would lead to an over-estimate of in-memory trials if a uniform guessing distribution is assumed. This second definition of in-memory responses instead defines in-memory as all trials that are not guesses. The height of the guess distribution was found by counting the number of responses that were more than 120° away from the correct colour value, and then assuming the guess distribution was uniform. This assumes that no responses on the opposite third of the colour wheel could represent a low-precision memory. The mathematically predicted proportion was again adjusted by the proportion of ‘both high’ responses that met the definition of in-memory.

The mean observed proportion using this alternative definition was 0.22 of responses now categorized as ‘in-memory’. Using the same predicted values with fractional capacity as in the main document the mean difference in the estimated and observed is 0.29, $t(41) = 5.72$, $p < .001$. Similarly using rounded capacity the interpretation does not change; $M = 0.29$, $t(41) = 5.70$, $p < .001$. Using the rounded down capacity, unlike when all low error responses were assumed to be in-memory, is still significantly different from the predicted proportion $M = 0.19$, $t(41) = 3.20$, $p = .003$.

Chapter 4: Recall procedure reveals low-priority objects are stored in memory.

Acknowledgments:

This project was funded by an NSERC Discovery Grant awarded to SME [435945], and Queen Elizabeth II Graduate Scholarship in Science & Technology awarded to HAL. This project was delayed due to limitations on research with human participants during the Covid-19 pandemic.

Abstract

Previous studies supporting continuous resource models of visual short-term memory (VSTM) have tended to use continuous response paradigms, which measure memory precision through characterising error distributions, rather than discrete response paradigms such as change detection, which measure memory accuracy through all-or-none discrimination decisions. The current study used intermixed trials of continuous response and two alternative forced choice (2AFC) to examine whether responses in discrete tasks could be consistent with continuous resource models. In all trials, memory resource allocation was manipulated by varying the probe probability of a cued item compared to the remaining items. In 2AFC trials, the discrimination difficulty was also varied systematically, which allows a psychometric approach to memory accuracy. Performance in the two response styles was highly correlated, confirming that the response style did not test two different aspects of memory. Continuous response precision was predictably graded according to the probe probability condition and best fit by a power-law relationship in line with a continuous resource model. Accuracy on the 2AFC trials was sensitive to the discrimination difficulty and the probe probability. As well, the response time was longer as probe probability declined, and task difficulty increased. Both accuracy and response time patterns of results were consistent with the predictions of noisy memory representations which are produced by the resource allocation in continuous resource models. The current study results aligned with the predictions of continuous resource models, and applications of signal detection models in decision making. This study expands the evidence for continuous resource models beyond the typical paradigm to a discrete response paradigm.

Introduction

Visual short-term memory (VSTM) is the system for keeping visual information ‘in-mind’ for immediate use; however, it is severely capacity limited. The capacity-limited nature of VSTM makes it unique from both iconic memory on the one hand, which is seemingly limitless but fleeting (Sperling, 1960), and consolidated long-term memory on the other, which also seems to have an infinite storage capacity (Standing, 1973). The properties of VSTM resources are currently debated between having a discrete or continuous nature – this debate is often centered around the effect that increasing set-size has on performance. Demonstrating the importance of the set-size effect, a recent fieldwide collaboration by Oberauer and colleagues (2018) agreed that the first benchmark (or fundamental fact of short-term memory) was that as set-size increases there is measurably reduced accuracy, reduced precision, and longer retrieval time. All of these are collectively grouped as set-size effects.

Discrete resource models evolved out of the earliest work characterizing VSTM capacity using change detection. In these studies, as set-size increases, the ability to correctly assess whether an item in a memory array changed relative to the original sample decreases systematically beyond about 3-4 items – clearly demonstrating the set-size effect on accuracy. These findings may suggest that participants are only able to retain 3-4 items in memory (Luck & Vogel, 1997). The discrete resource models characterize memory as having a limited number of storage “slots”, in which discrete items in their own memory slots are stored with equal fidelity. Importantly, if there are more to-be-remembered items than capacity-limiting slots, the extra items would not be in memory, requiring participants to guess. The quantized slot model (i.e., *slot* +

averaging model, Zhang & Luck, 2008), makes the same predictions as the classic slot model when memory demands are at or above capacity, but predicts that items can be stored in multiple slots to improve precision when the set-size is below capacity. Simply put, there are two distinct categories of items at the maintenance stage: in-memory items (up to memory capacity) and not-in-memory items (beyond memory capacity). However, estimating capacity through change detection performance ignores “partial information” in memory in favour of a simplified all-or-none memory store (Pashler, 1988); this conceptualization of memory as all-or-none is incongruent with continuous resource models, which suggest that representations can be imprecise or low resolution. Continuous models do not propose a strict item-based capacity limit; instead, continuous models account for decreasing performance at higher memory loads by suggesting that items receive a smaller proportion of limited memory resources (Bays et al., 2009).

By contrast, contemporary continuous resource models are supported by studies in which memory is assessed by a continuous report paradigm using continuous feature spaces such as orientation or colour. These paradigms measure the fidelity of the memory representations by averaging the trial-by-trial difference between the reported feature and true value of the target item (i.e., error). Early findings using this method reported that the precision (the inverse of error) of an in-memory item decreases as set-size increases (Wilken & Ma, 2004). This led researchers to reconsider the view that items in memory would be equally precise and necessitated the development of continuous resource models (for review see Ma et al., 2014). Continuous resource models account for these findings by suggesting that more neural resources can be allocated to each item when

there are fewer items in memory leading to less noise in the memory signal and thus higher precision responses (Bays, 2014; Bays, 2015).

Both models explain the limited capacity of VSTM: discrete resource models propose the capacity is set by an item-based or quantized unit; continuous resource models propose a limited pool of resources. Discrete resource theories explain the set-size effect on accuracy extremely well (Luck & Vogel, 1997) and have been adapted to explain the set-size effect on precision (Zhang & Luck, 2008). Whereas continuous resource theories explain the set-size effect on precision (when there are more items, each receive a smaller proportion of memory resources), which consequently leads to the set-size effect on accuracy (i.e., poor precision is measured as low accuracy). However, differing practical use of alternative models that align with each theory have been noted by Ma (2018). Specifically, Ma (2018) noted that experiments that fit the variable precision model (a continuous resource model) usually compared multiple alternative models then did not further analyze the parameters; while experiments fitting the discrete resource model (specifically the Zhang & Luck, 2008 *slot + averaging* model) did not test any alternative models. As an illustrative example, Oberauer and colleagues (2018) named “characteristics of errors” broadly as the fourth benchmark, with the characteristic distributions of continuous response tasks relegated to a secondary status within this benchmark noting that this effect was specific to one domain (visual) and one paradigm (continuous response delay-estimation). This distinction reflects that one empirical result is perceived as more important than another. Undervaluing continuous resource models, despite the fits outperforming discrete models, stifles the development of continuous resource theories because it does not facilitate further predictions to be tested.

To further develop our understanding of VSTM resources, it is necessary to equate the importance of the empirical findings central to both theories. In fact, as described in depth in the next section, “set-size effects” are not specific to changes in set-size, rather they reflect changes in resource allocation which are yoked to set-size in typical memory paradigms. Instead, changing the probe probability of an object while keeping set-size the same has demonstrated that resource allocation is responsible for typical set-size effects (Dube et al., 2017; Emrich et al., 2017). The purpose of the current study was to broaden the application of “set-size effects” on memory precision to another paradigm (two alternative forced choice, 2AFC) in VSTM. By having both continuous response and 2AFC trials in the same experiment, the accuracy results can be interpreted within the context of continuous response precision. In addition to analyzing the error distribution, the experimental design allowed for a psychometric approach which compliments the continuous response results and shows how discrete accuracy data can lend support to continuous resource models.

Flexible allocation of VSTM Resources

Although set-size effects remain a benchmark of models of VSTM, set size tends to be confounded with the proportion of memory resources potentially allocated to each item in memory. For example, whereas a memory load of one item should allow that item to receive 100% of the available memory resources, a load of two would require that resources be split across both items, with each one receiving 50% of available memory resources. In order to test this theory, Emrich et al. (2017), used spatial cues presented during memory encoding that informed participants of the likelihood that a given item would be tested (probe probability). For example, in one condition, six items were

presented but one cued item would be tested on 50% of trials; the precision of memory recall for the cued item was nearly identical to that of trials when two items were cued with 100% probe probability (i.e., when memory resources should be split equally across two items equating to 50% per item). Critically, performance on the uncued trials was similarly predicted by the proportion of resource-allocation, fitted by a power-law function; performance for the 5-uncued items was consistent with each item receiving approximately 10% of the total resources (i.e., $1/5^{\text{th}}$ of the remaining 50%). In other words, probe probability predicts the response precision in line with the proportion of resources that should be allocated to each item demonstrating the same effect on memory precision described by the set-size effect without changes in set-size. The two studies in Emrich et al. (2017) demonstrated this effect across conditions ranging from 9% – 100%, and similar effects have been observed using feature-based cues (Dube et al., 2017; Salahub et al., 2019), reward (Klyszejko et al., 2014), as well as in a memory-guided saccade task (Yoo et al., 2018), and single and dual-target visual search (Huynh Cong & Kerzel, 2022). Thus, studies have demonstrated that (1) memory resources can be flexibly allocated via attention and (2) individual items can receive a small proportion of resources, consistent with a continuous-resource model of VSTM.

Although the evidence to date for the flexible allocation of VSTM resources has tended to come from continuous response tasks, this process (and indeed any model of a cognitive process) should not be dependent on the paradigm being used. Studies investigating the possible continuous nature of VSTM resources have tended to use delay-estimation paradigms, with some exceptions utilizing change detection (Keshvari et al., 2013; Pearson et al., 2014), or change localization (van den Berg et al., 2012). In

studies of change detection and change localization, researchers use a psychometric approach by manipulating the magnitude of the change allowing the researchers to identify properties of memory performance as predicted by VSTM resource models – specifically in regard to “noisy” memory signals. For instance, van den Berg and colleagues (2012) predicted that the magnitude of change would impact memory accuracy to varying degrees in both quantized (discrete) and continuous resource VSTM models, but not in the classic item-limit slot model due to the conceptualization of noisy representations in the previous memory models. In another study, Keshvari and colleagues (2013) varied the magnitude of difference in a change detection task, critically, including close-colour discrimination trials. When large changes were presented, a typical K-estimate of 3.8 was observed. However, if small changes were made, the results suggested that participants had no memory for any items at set-size 6. It should not be concluded that participants had no memory representations under small change conditions; rather it should be acknowledged that the memory signal to noise ratio was not large enough to detect a change. However, by not accounting for the magnitude of comparison, very different conclusions would be drawn from the task.

A similar conclusion can be found in a debate surrounding the reduced capacity of complex items. It is commonly found that change detection tasks using complex item often result in lower capacity estimates than change detection with simple features (e.g., Xu & Chun, 2006). An alternative explanation for this finding is that within-category complex item changes are smaller in magnitude than changes of simple features; when this difference is accounted for by using cross-category changes, similar capacity estimates are measured for both the complex items and simple features (Awh et al.,

2007). Despite this evidence, it is often assumed that at high memory loads, lower accuracy indicates that a discrete capacity has been reached and not that the magnitude change was not large enough to detect. These results highlight that memory resolution as conceptualized as noisy representations influences discrete memory responses, which can be observed when variable magnitude changes are used.

Memory retrieval

Although the competing theories of VSTM resources are most concerned with the accuracy and precision of memory, the set-size effect also impacts the retrieval time (RT) of memory. In addition to accuracy of responses, the RT provides information about the decision processes utilized during VSTM paradigms. A typical model of decision making proposes the accumulation of information up to a decision threshold. The logic from sequential sampling signal detection frameworks, is that the decision time is longer when the information that is sampled is noisier because it takes longer for evidence to reach the threshold (Noorani & Carpenter, 2016; Smith & Ratcliff, 2009). In a robust study of 4,680 change detection trials in three participants, Pearson and colleagues (2014) found that memory quality impacted the rate of evidence accrual and not the decision threshold. Further, manipulations of memory quality (for instance, task-relevance cues) had predictable corresponding changes in RT (Pearson et al., 2014). This approach highlights how continuous resource models of VSTM memory resources fit within broader cognitive processing (decision making) as an application of signal detection. In the current study these findings support a prediction that RT will be longer for low probe probability conditions that are hypothesized to impact memory quality.

The current study

The goal of the current study was to investigate whether evidence of flexible allocation of a continuous VSTM resource could be found in a discrete response task. To test this question, participants were presented a goal-driven attentional prioritization VSTM task, with two response types intermixed: continuous report and two-alternative forced choice. Since the type of response was unknown until test, participants should not have changed any aspects of memory encoding or maintenance in accordance with the continuous or discrete nature of the response types. Critically, the magnitude of difference between the target and lure colour was variable in the 2AFC condition to apply a psychometric approach to measuring the properties of the memory representation (Keshvari et al., 2013). To manipulate goal-directed prioritization, the memory sample included a unique shape that had a higher probe probability than the other two items. The probe probability of this item was either 100%, 80% or 50% of trials per priority condition, leaving the other two items to be probed on the remaining trials in their priority condition. I predicted that the priority manipulation would impact memory resolution with noisier representations of items with lower priority, consistent with continuous resource models (Emrich et al., 2017). Noisier representations (in the low-priority conditions) should result in (1) lower precision on continuous report trials, (2) lower accuracy in the 2AFC trials particularly when the alternate choice is a close-colour, and (3) longer RTs due to slower rate of evidence accrual.

Method

Participants

Participants were recruited from the university population on the online

recruitment platform SONA and through word of mouth. Participants completed an intake form on Qualtrics to ensure that they met the inclusion criteria before scheduling. Thirty-two participants completed the experiment (6 male, 26 female), with an average age of 21.13 (range 18 – 28). Since the experiment included EEG measures (not reported here), one participant was excluded for not disclosing a neurological condition at intake. All remaining participants had normal vision, normal colour vision, no neurological conditions, not currently taking psychoactive medications, and were all right-handed. Additionally, all participants were required to have completed two doses of a federally approved vaccine against Covid-19 and required to wear a medical mask to participate in research on campus at the time of the study. Thus, anyone who was exempt from masking or unvaccinated was also ineligible for the study. All procedures were approved by Brock University's Bioscience Research Ethics Board.

Procedure

The experimental task consisted of two intermixed visual short-term memory tasks: a continuous report task and a 2AFC task. At the start of each trial, participants were presented with three coloured shapes for 250 ms – one circle and two squares – and, following a delay of 1000 ms, were asked to report one of the colours at the test phase. The unique circle distinguished which item was most likely to be probed on each trial. The probability that the circle would be probed was either 100%, 80%, or 50%, creating three task conditions. Thus, in the 80% and 50% circle probe probability conditions, each of the two squares were tested on 10% or 25% of trials respectively, creating five probe probability conditions. Participants were told the probability that the unique circle item would be tested before each block of trials. On the break screens, participants were given

performance feedback of their average absolute response error on the last block of trials.

Response types were intermixed during the blocks: continuous response or two-alternative forced choice. In 100 trials per task condition, the responses were made by selecting with the mouse the colour that best matched their memory (continuous response trials). The colour would appear inside the bolded shape as participants moved the cursor around to make their selection. The response colour was recorded and the difference between the reported colour and the target colour was calculated as the raw error in degrees. In the remaining trials of each condition, participants performed a 2AFC, by selecting with the mouse between two colours presented on either side of the target item. One colour was always the originally presented sample colour, and the second colour was randomly selected between $\pm 15^\circ$ and 60° away from the correct colour. Although the responses were not speeded, there was a maximum time of 5 seconds to encourage the pace of the experiment. Anecdotally, a colour that is 60° away from the target colour may be categorically unique, although still confusable in poor memory conditions, for instance the far-colour trials could have been between blue and green. Far comparisons are expected to be easier than close-colour comparisons, which should be demonstrated in higher accuracy and faster reaction times.

The length of task was designed to maximize the number of trials within the experimental session but anticipated that many participants would not complete all of the trials within the time of the experimental session. As such, the three task conditions were presented in blocks of up to 50 trials at a time, with a random order so that the conditions were maintained regardless of the stopping time. Within the probe probability blocks the cued and uncued trials were pseudo-randomly organized such that probe probability

conditions were upheld if the experiment was stopped after any block, meaning that there would not be a single block of 80% probe probability that was entirely made up of the 10% uncued probe trials. Additionally, trials were constructed to allow 60-65 trials in each of the 6 possible target locations for an EEG analysis not included here. The full experiment contained up to 1665 trials consisting of 390 trials of 100% cued test probe probability (100 continuous recall, 290 2AFC), 495 trials of 80% cued probe probability (80 continuous recall, 316 2AFC; including 99 trials 10% uncued test probe probability, 19 continuous recall and 80 2AFC), and lastly 780 50% cued probe probability (49 continuous, 326 2AFC; including 405 trials of 25% uncued test probe probability of 51 continuous, 354 AFC). An error in the construction of the 50% condition split meant that the actual probe probability of this condition was 48% with the uncued items being probed 26% of the time each. On average 1572 trials were completed by each participant, 19 participants completed all trials, the remaining completed between 1100-1600 with an average of 1426 trials within the allocated time of the experiment. The very low number of 10% probe probability continuous response trials limits the accuracy of the response distribution of this condition (12-19 trials).

Before the practice trials, electrode impedances were checked, and signals were monitored throughout the experiment by a researcher in an adjoining room. All participants completed a session of practice trials before the experimental session which included both types of responses (continuous and 2AFC) and all three attention allocation conditions. Practice session condition order was controlled such that they always learned the 100% condition with the continuous response for the first 5 trials. After the first 5 trials the 2AFC trials were intermixed with the continuous response. Practice trials

continued with the 100% condition, followed by the 80% condition, and then the 50% condition. This order was chosen to facilitate learning of the conditions and response screens. Participants began uninformed of the limited response window but always discovered it during the practice trials, where it was explained verbally. Participants were given the opportunity to ask questions or clarify the instructions. A small number of participants asked to repeat the practice trials. During each experiment, participants were given self-paced breaks a minimum of every 25 trials.

Stimuli and Apparatus

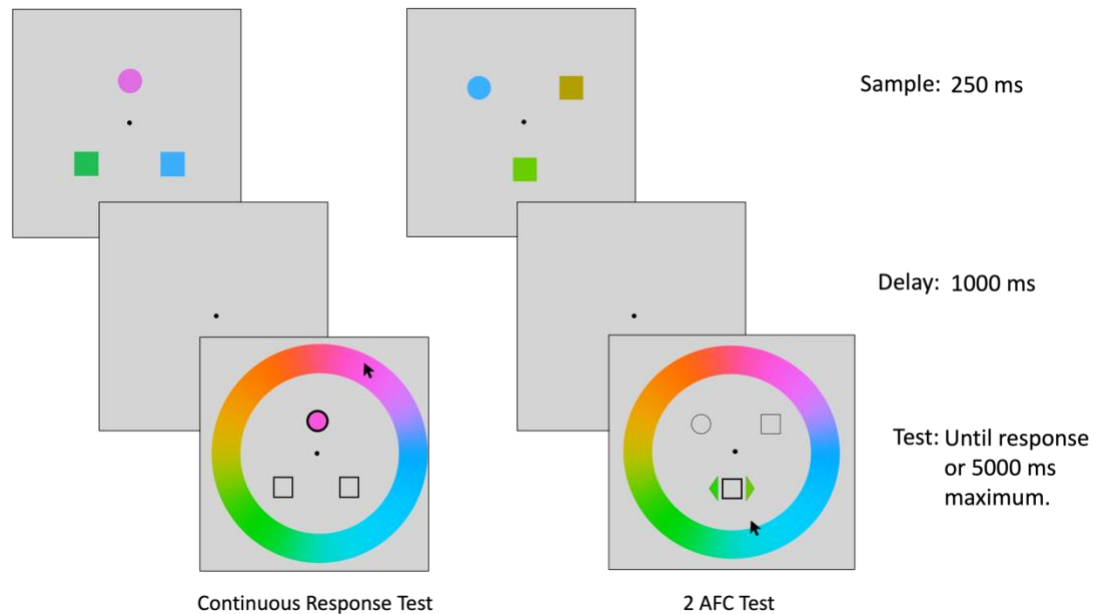
Stimuli were presented using PsychoPy v2021.2.3 (Peirce et al., 2019), on a 20" LCD 2090UXi computer monitor (1600 x 1200 pixels, 60 Hz refresh rate) at a distance of approximately 57 cm. As mentioned, EEG was recorded during the study that is not analyzed here. Recordings were taken using the Biosemi Active-2 system. Participants were fit with 64 channel EEG caps using extended 10-20 sites plus 6 exogenous electrodes placed behind the ears on each mastoid bones, laterally beside the eyes, and below each eye. The signal was referenced online to the common mode sense (CNS) and driven right leg (DRL) electrodes.

Three coloured shapes (one circle, two squares) of 1° of visual angle were evenly spaced around the central fixation during the study phase. The relative spacing was consistent but the shapes could appear among six possible locations thus there were two overall configurations (clock hour-hand positions: [12:00, 4:30, 7:30] or [1:30, 6:00, 10:30]). See Figure 4-1. Their colours were pseudo-randomly selected from 360 unique colours from the CIE $L^*a^*b^*$ colour space calibrated to the monitor and lighting conditions [$L = 70$, $a = -6$, $b = 14$, radius = 49]) with a minimum distance of 30° . A colour

wheel with a radius 7° visual angle was presented during the colour judgment phase of each experiment.

Figure 4-1

Sample trials of continuous response and 2AFC test conditions



Note. The memory sample was displayed for 250 ms, followed by a 1000 ms delay displaying only a fixation point, the test display was either continuous response (left) or a 2AFC (right). The test display was terminated by response or if 5 seconds elapsed. The alternate choice in this trial is 20° away from the probe colour and would be categorized into bin 2.

Analysis

Continuous Response

Error on continuous response trials was calculated as the circular distance between the target colour and response colour. Trials that timed out were excluded from analysis. Response precision was calculated as the circular standard deviation using the *circular* package in *R* (Agostinelli & Lund, 2023). The figures of error distribution are based on raincloud plots showing the shape of the density distribution along with the individual trial data below (Allen et al., 2021). The response precision data are predicted

to fit a power-law relationship with the probe probability (Emrich et al., 2017). Both the response precision and probe probability values were log transformed to be compatible with testing a linear mixed-effect model. The power-law model is compared with the untransformed linear fit. The models were performed using the *lme4* package (Bates et al., 2015), and ANOVA (type 2) from the *car* package (Fox & Weisberg, 2019). The models were each constructed with fixed effect of probe probability and a random intercept for each participant.

Two Alternative Forced Choice

Accuracy

Although the 2AFC responses are discrete in nature, due to the variable magnitude difference of the alternate choice, by taking the circular distance of the response choice colour from the target colour, the response error could be calculated in the same manner as the continuous responses. That is, “hits” always had an error of 0° , whereas “misses” could be between 15° and 60° . From these values, the standard deviation of the error (calculated based on the proportion of incorrect trials to the different alternatives) was calculated on an assumed circular normal distribution.

Although the alternate choice colour was not drawn from the full range of a circle, this data is still circular in nature and therefore modelled with a von Mises distribution. It should be noted the distribution of responses is not normal due to the density at 0° error and the gap between $1-14^\circ$ of error. Thus, the standard deviation should be interpreted with some caution; nevertheless, for comparison to continuous response trials, the circular standard deviation was calculated for each participant in each probe probability condition. See supplementary Figure 4-A1 A in Appendix 4A to see the density

distribution of all 2AFC data by probe probability condition, and Figure 4-A1 B to see the density distribution of absolute error of the miss responses only.

To model the parametric effect of probe probability on response accuracy, the 2AFC data was also split into 9 bins depending on the magnitude of the difference between the alternative colour and the probe colour (discrimination bins). Each bin represented 5° of circular colour deviations, starting with the closest colour comparison of 15-19°, and so on, except for the 9th bin which covered 55-60°. The proportion correct data in 2AFC tasks can be corrected for guessing (Chapter 7, Hautus et al., 2022) using the formula:

$$q = (2 \times \textit{Proportion Correct}) - 1 \quad (4.1)$$

The corrected hit rate for each bin within each probe probability was used for analysis. Trials that timed out were excluded from analysis. The corrected hit rate was used to test whether performance was significantly different from chance (0 for the corrected hit rate) in a one-sample t-test.

As described for the continuous response analysis, to test the effect of probe probability and magnitude difference of alternative choice on 2AFC performance, accuracy was modelled with a linear mixed-effect model with a random intercept for each participant and fixed effects of the predictor variables of discrimination bins and probe probability conditions, as well as the potential interaction. Both probe probability and discrimination bins were treated as continuous variables in the mixed-effect model.

Response Time

Response time data were trimmed using the *trimr* package (Grange, 2022), using an upper criterion of 2.5 SD computed for each participant at each level of probe

probability and discrimination bin (i.e., 10% probe probability and 15-19°, etc.); hit and miss responses were separately trimmed. The effect of probe probability and discrimination bin on response time was tested using the same approach as described above for accuracy, with the one difference that all trials were modelled rather than a summary statistic. This difference allows for the variable number of trials per predictor category (probe probability and discrimination bins) to be reflected in the uncertainty for each participant's model. Since there were no predictions for miss responses, miss RTs are analysed in Appendix 4B.

Comparing performance by response type

To compare the performance between delay-estimation and 2AFC trials, response precision (circular standard deviation) was compared for each participant. The data were fit to a linear mixed-effect model where 2AFC precision was predicted by continuous response precision; however, this resulted in a violation of the model assumptions known as a 'singular' model fit because the variance of the random intercepts was estimated at zero. Consequently, a Bayesian linear mixed-effect-model was fit using the *blme* package (Chung et al., 2013)².

Results

Continuous Report Precision

The purpose of this study was to investigate the flexible allocation hypothesis under a 2AFC paradigm in an experiment which contained both continuous responses and

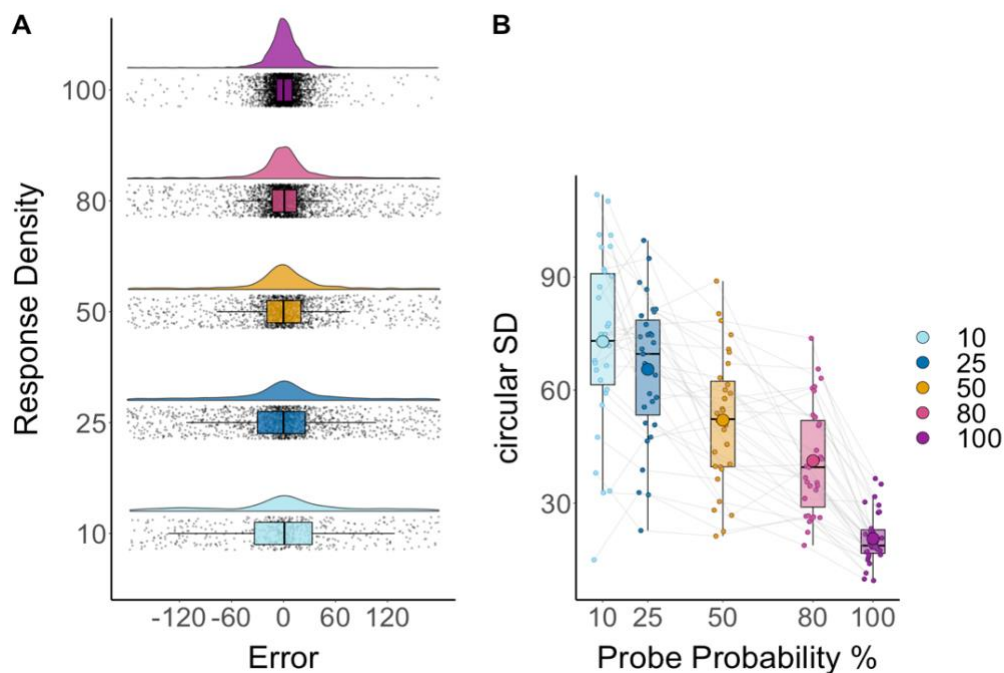
² Singular fit is caused by high co-linearity of a model parameter and can typically be solved by reducing the complexity of the random effects; however, since the only random effect in the current model is the intercept, this is equivalent to fitting a linear model which ignores the within-person structure of the data by treating the intercept as a fixed factor. Instead using the Bayesian linear mixed-effect keeps the random intercept, within-person data structure, by providing a weak prior to the random effect estimate (Chung et al., 2013).

2AFC response types intermixed. To test this question, precision from the continuous response trials was modelled as either a power-law (predicted by continuous resource models) or a linear fit (discrete resource models), as has been established by previous studies (Bays & Husain, 2008; Emrich et al., 2017; van den Berg et al., 2012).

As in previous studies, response error was lower at higher probe probabilities, suggesting greater resource allocation to these items in a flexibly allocated manner (see Figure 4-2 A). To confirm the continuous response data was consistent with predictions of flexible allocation of a continuous resource, a linear model was compared with a power-law model (linear on log-transformed data) (Emrich et al., 2017). As predicted, the

Figure 4-2

Continuous report performance by probe probability conditions



Note. **A.** Response density distributions of the recall error from the continuous report data for each probe probability condition. Boxplots are overlaid on all datapoints. Based on raincloud plots (Allen et al., 2021) **B.** Circular standard deviation (precision) by probe probability conditions for continuous report data.

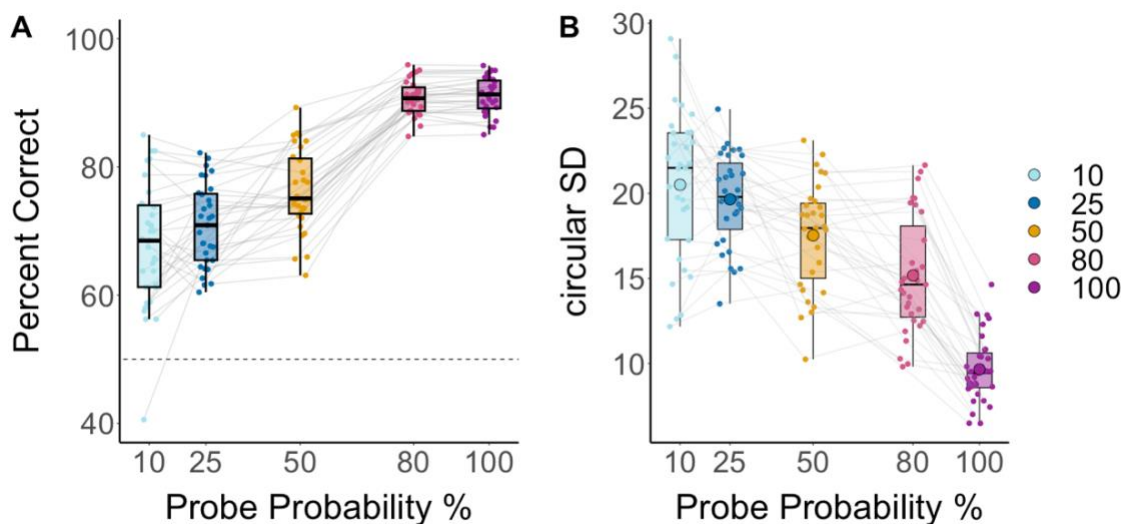
power-law model had a lower AIC value (power-law: -95.31 vs. linear: 1287.89), indicating it was a better fit to the data than the linear model, see Figure 4-2 B.

Secondly, comparisons between the cued and uncued items in the same attention conditions revealed significant difference between response precision in the 80% vs 10% condition $\log(\text{BF}_{10}) = 15.22$, and between the 50% and 25% conditions $\log(\text{BF}_{10}) = 6.55$, indicating the successful use of the feature-based shape cue.

2AFC Precision

First the response accuracy between probe probability conditions was compared. Manipulating probe probability has been demonstrated previously to be comparable to manipulations of set-size (Emrich et al., 2017). Consistent with studies that manipulate set-size, response accuracy (percent correct) decreases when fewer resources were allocated to the item, $F(4, 120) = 187.98, p < .001$, see Figure 4-3 A. Even at the lowest

Figure 4-3
2AFC performance by probe probability conditions



Note. **A.** Percent correct 2AFC performance by probe probability, individual participants data points shown. Chance line shown. **B.** Circular standard deviation (precision) by probe probability for the 2AFC responses, box plot overlaid, individual participants shown; mean performance shown by large circles.

probe probability, responses were 68% accurate. Paired t-tests were used to test for differences between consecutive conditions. There was a significant difference between the 100% and 50% probe probability conditions ($M = 14.8\%$, $t(30) = 14.47$, $p < .001$), which are conceptually similar to set-sizes 1 (100%) and 2 (50%). As well as a significant difference between the 50% and 25% probe probability conditions ($M = 5.4\%$, $t(30) = 4.64$, $p < .001$), which are conceptually similar to set-sizes 2 (50%) and 4 (25%). Finally, there was a significant difference between the 25% and 10% probe probabilities, ($M = 3.0\%$, $t(30) = 2.20$, $p = .036$), which are conceptually comparable to set-sizes 4 and 10. However, there was no significant difference between accuracy in the 100% and 80% probe probability conditions ($M = 0.2\%$, $t(30) = 1.15$, $p = 0.26$), in this case the conceptually similar set sizes 1 (100%) and 1.25 (80%) do not align with a full item step in set size.

Although the alternate colour choice was restricted between 15° and 60° away, the circular standard deviation of response error can still be calculated using both the correct and incorrect responses (see Figure 4-3 B). As with the continuous response data, the continuous resource model predicts that responses will follow a power-law relationship. This was tested in the same manner as described above. Consistent with a flexibly allocated continuous resource the power-law fit was demonstrated to be better than a linear fit by having a lower AIC value (power-law: -242.6, vs. linear: 797.3).

Comparing Precision

The circular standard deviation of response error was compared for all trials of each response type within participant. Figure 4-4 shows the linear correlation between performance on continuous trials and 2AFC trials. The model found that precision on

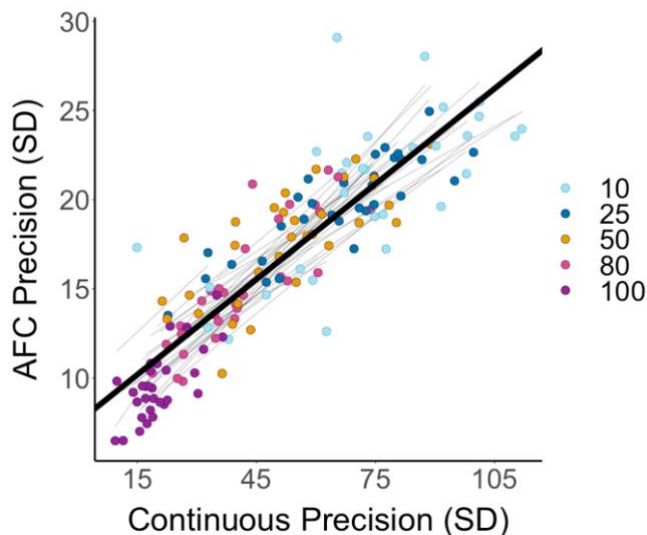
continuous response trials was a significant predictor of precision on 2AFC trials

$F(1, 153) = 548.27, p < .001$. This relationship confirms that participant performance

across the two response types is conserved, such that high performance on 2AFC trials is associated with high performance on continuous response trials.

Figure 4-4

Comparison of circular SD performance between response type



Note. Individual participant linear fits shown in grey lines. Black line shows the linear fit from the Bayesian linear mixed-effect using the fixed effect of continuous precision, and the average intercept for graphical purposes.

Psychometrics of AFC Performance

Continuous resource models draw on the framework of signal detection models, and as such further predictions about response error can be investigated in more detail by analyzing the impact of task difficulty on performance in conjunction with the probe probability manipulation. The 2AFC trials were grouped into 9 discrimination bins based on the magnitude of the difference between the alternate choice colour and the probed target colour. Bin 1 includes the most difficult close-colour comparisons between 15-19°, and bin 9 includes the easiest comparisons of far-colours between 55-60°. The accuracy data was corrected for guessing (Hautus et al., 2022). Linear mixed-effects modeling was

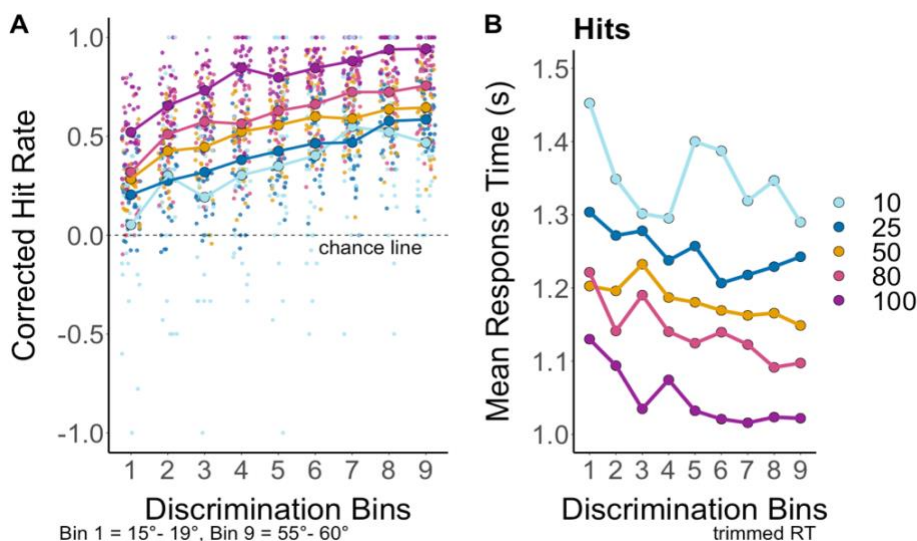
performed to analyse the effects of the attention manipulation (probe probability) and decision difficulty (discrimination bin) on accuracy (corrected hit rate). The model was built in four steps beginning with the random intercept only model. By adding predictors in this way, the unique effect of each predictor is demonstrated. The best fitting model was one that included both bin and probe probability as fixed effect predictors, but not their interaction, see Table 4-1. In the best fitting model, accuracy is predicted by fixed effects of probe probability ($\beta = 0.0046$, $se = 0.0002$, $t = 26.83$, $p < .0001$), and discrimination bin ($\beta = 0.0467$, $se = 0.0022$, $t = 21.05$, $p < 0.001$), such that corrected hit rate increased by 0.37 from the close-colour comparisons (bin 1) to the far-colour comparisons (bin 9). The effect of discrimination bin on response accuracy is not congruent with any discrete model of VSTM resources that does not allow for variability in memory quality (i.e., item based discrete models), but is predicted by all continuous resource models. Furthermore, accuracy increased with greater probe probability, such that the 100% condition had a corrected hit rate 0.23 higher than in the 50% probe probability condition. Figure 4-5 A shows that each predictor affects accuracy without any evident interaction: accuracy increases as the task is easier between discrimination bins 1 and 9 in all probe probability conditions, and accuracy increases from the lowest probe probability (10%) up to the highest (100%) probe probability conditions.

Table 4-1
Summary of mixed-effect models of 2AFC accuracy

	Fixed effects	Random Effect	$X^2(df), p$
1.	Intercept only.	Intercept per Participant	
2.	Accuracy ~ bin	Intercept per Participant	263.2(1), < .001
3.	Accuracy ~ bin + Probe Probability	Intercept per Participant	578.9(1), < .001
4.	Accuracy ~ bin \times Probe Probability	Intercept per Participant	0.303(1), 0.582

One characteristic difference between continuous resource models and discrete resource models is whether there are true guess responses caused by not having the test item in memory or whether there can be very low-resolution memory representations which produce very low precision responses. To assess whether performance in any condition reflected random guessing, the accuracy data was compared to chance performance. Since accuracy data was corrected for guessing before being compared to chance performance, the corrected performance scores are compared to 0 instead of 50%, see Figure 4-5 A. Accuracy in the lowest probe probability condition (10%) at the closest colour discrimination bin (15-19°) was not different from chance; however, performance was better than chance for the next discrimination bin (20-24°), and every discrimination bin thereafter. For all other probe probability conditions at every level of discrimination, accuracy was above chance.

Figure 4-5
2AFC performance measures by task difficulty



Note. **A.** Corrected hit rate by discrimination bins (close-colour Bin 1 = 15-19° difference, far-colour Bin 9 = 55-60° difference) for each probe probability condition; individual participants shown in small circles, mean in large circles. **B.** Trimmed mean response time of hit responses by discrimination bin and probe probability conditions.

Response Time

Response time data relates to the sequential sampling predictions of the signal detection framework: lower probe probability will lead to a noisier representation of the memory item, and then it takes longer to make a decision in noisier conditions. The response time data for hits and misses were separately trimmed with a cut-off of 2.5 SD per participant per experimental conditions (probe probability and discrimination bin), as well as a minimum time of 100ms. These data were fit to a linear mixed-effect model predicting response time from the fixed effects of discrimination bin and probe probability, as well as their interaction with a random intercept for each participant. The full model including the interaction for hit responses was significant:

$RT_{\text{hits}} \sim \text{discriminability} \times \text{probe probability} + (1|\text{participantID})$, ($X^2(1) = 6.44$, $p = .0112$), see Figure 4-5 B. In general, it takes approximately 74.47 ms longer to make a correct response for close-colour comparisons (bin 1) than far-colour comparisons (bin 9) ($\beta = -9.309$ ms, $se = 0.795$, $t = 11.710$, $p < .0001$; estimated at 50% probe probability); and it takes approximately 120 ms longer to make a correct response in the 50% probe probability condition compared with the 100% probe probability condition ($\beta = -2.400$ ms, $se = 0.123$, $t = 19.498$, $p < .0001$; estimated at bin 1 representing a close-colour comparison). Additionally, these effects interact ($\beta = -0.062$ ms, $se = 0.002$, $t = 2.537$, $p = 0.0112$), meaning that the difference in RT between probe probability conditions is greater at close-colour comparisons than at far colour comparisons.

Discussion

The flexible allocation of VSTM resources is well established in studies of continuous response delayed-estimation tasks, but there have been few studies that

directly seek to demonstrate this property outside of continuous response paradigms. The current study analysed both continuous and discrete responses in an intermixed design to assess performance outcomes from discrete responses within the predictions of a continuous resource model. Continuous resource models propose that lower allocation of a flexible continuous resource leads to poorer memory quality (Bay et al., 2009; Fougne et al., 2012; van den Berg et al., 2012). Poor memory quality should result in worse response precision, lower accuracy, and due to the decision-making properties of signal detection model, slower RTs. Thus, in the current study, manipulating resource allocation through probe probability should lead to predictable parametric declines in performance.

Consistent with the continuous resource models, continuous response precision (circular SD) predicted by probe probability was better fit by a power-law model than a linear model. Critically, this same measure in the 2AFC version of the task was also better fit by a power-law model than a linear model. Further, comparing the precision between the continuous response and 2AFC task demonstrates a significant correlation, such that high performance in one response type predicts high performance in the other. This supports the conclusion that the underlying memory representations are the same and do not depend on the type of test used to assess those representations. Together these data provide further support for continuous resource models by demonstrating again that response precision is predicted by probe probability but in a paradigm that has not been used to test this theory before. It is noted however, that the circular SD of the 2AFC data carries the significant caveat that it is not normally distributed.

Further support for the predictions of a continuous resource model comes from the analysis of accuracy across differences in the difficulty of discrimination for each

level of probe probability. By varying the discriminability of the alternate choice, the response decision can be analysed in a psychometric manner according to the principals of signal detection. As expected by flexible models of memory resources, there is a main effect of task difficulty such that the close-colour choices had lower accuracy than the far-colour choices. As highlighted by van den Berg and colleagues (2012) with respect to change localization, the classic item limit discrete resource model does not have any flexibility to accommodate an effect of variable difficulty on accuracy of discrimination because of an assumed noiseless memory representation. Instead, this effect supports a resource model with noisy memory representations.

In addition to the effect of discrimination difficulty, the memory quality is predicted to be proportional to the probe probability as a demonstration of flexible resource allocation. Crucially, probe probability led to graded differences in response accuracy consistent with a parametric effect of signal-to-noise ratio. That is, although closer colour comparisons are more difficult, higher allocation of VSTM resources allows participants to correctly distinguish the target colour from memory which suggests there is a direct effect on the memory quality.

Finally, the response time analysis indicates that greater resource allocation in higher probe probability conditions led to faster reaction times. Further, this interacted with the discriminability difficulty, such that the difference in response time between probe probability conditions was greater at the difficult close-colour trials than at the easier far-colour discrimination trials. This is consistent with the predictions of continuous resource models as applied by the signal detection framework. Signal detection proposes that response time indicates the time needed for evidence to

accumulate to come to a decision between the two alternate choices. Easier decisions should be made faster than more difficult ones when the colours are closer together as demonstrated by the main effect of discrimination difficulty. Secondly, decisions are predicted to be easier (and thus faster) if the memory representation is less noisy which is demonstrated by the main effect of priority condition (probe probability). The model of flexible allocation of continuous resources predicts that greater memory resource allocation results in a higher resolution memory representation, which in turn leads to faster response times. This effect interacted with discrimination difficulty such that the beneficial effect of memory quality had a greater impact on RT for difficult discrimination trials than on easier trials. Similarly, Pearson and colleagues (2014) found that both the memory quality and the discrimination difficulty interacted to affect the slope of evidence accumulation and not the decision threshold or the trial-to-trial variance in the rate of accumulation.

This psychometric approach to the discrete response data further bolsters the predictions of continuous resource models by not requiring the comparison of error distribution parameters. Instead, the accuracy data can be analysed over controlled changes in difficulty. This is an important extension of the application of continuous resource models because it demonstrates support for the predictions of a theoretic continuous resource in a paradigm other than delayed continuous recall. In addition to the evidence from change detection (Keshvari et al., 2013; Pearson et al., 2014), and change localization (van den Berg et al., 2012), there is clear support for continuous resource models regardless of whether memory precision can be directly measured through error distributions or not.

Out-of-memory or low-resolution memories

A secondary debate between VSTM resource models is on the existence of true guesses – that is, trials in which there is no information guiding the response. Some continuous resource models argue that there are no true guess responses, but rather that these presumed guess responses actually reflect low precision memories – although lapses in attention will result in no-information trials (e.g., variable precision models, van den Berg et al., 2012). In contrast, all discrete resource models predict guessing would take place when a probed item was not-in-memory. The corollary of this assumption about guessing is that discrete models suggest there are no true low-resolution memory representations by virtue of having only in-memory (normally distributed) and out-of-memory (uniformly distributed) responses.

A weakness of continuous response paradigms is it is difficult to distinguish a very wide error distribution from a truly uniform distribution which would suggest these responses are random guesses. Often error distributions for very low-priority items are very wide and flat with only some central tendency (e.g., Emrich et al., 2017), which is also demonstrated by the error distributions for the low-priority items in the current study (Figure 4-2 A). Adam and her colleagues (2017) found similarly wide and flat distributions in a whole report task for the fourth to sixth responses and interpreted these distributions as uniform distributions reflecting guessing. Adam et al., (2017) calculated that it would take “over 900 million trials” (p. 94) to distinguish between low precision memory and guessing in a 6-item whole report.

In contrast, alternative forced choice tasks are often used in investigations of long-term memory specifically because they can reveal evidence of “weak” memory through

more simple binary judgements (Brady et al., 2023; Schurgin, 2018). In the current study, the 2AFC responses demonstrated evidence that low precision responses were in-memory responses. The low-priority conditions have wide error distributions in the continuous response trials; however, the same priority conditions in the 2AFC version resulted in above-chance responses. This finding suggests that low-priority items had representations in memory. Although at-chance performance was observed in the lowest probability condition (10%) with the closest colours (15-19°), there was above-chance performance for easier discrimination comparisons. This pattern suggests that there was still in fact a low-priority item in memory, but that the very close target and distractor colour comparisons either did not overcome the noisiness of the representations, or limits of perception. These findings suggest that there was a low-resolution representation rather than no representation in memory, otherwise 2AFC performance would be at chance regardless of the magnitude of difference between the choices.

Limitations

The response distribution of the 2AFC data was modelled on an assumed von Mises distribution to estimate precision through the circular standard deviation just as is done with the continuous response data. However, the data was not normally distributed as assumed by a von Mises distribution because the minimum distance of the alternate choice was not continuous; thus, the measure of distribution from the mean is compromised. If the alternate choice was allowed to be fully continuous between 1-60°, this data would be expected to be normally distributed. However, it was not practical in the current design to include trials where the colour discrimination would be imperceptible to perception and thus not be a test of memory resolution. If only the error

distribution was of interest, future studies could remove the minimum distance of the comparison or set the minimum distance to a measured perceptual threshold of colour difference for each person.

Although a power-law was a better fit to the data, a linear fit also fits the current study data, more so than previous variations of the continuous response paradigm. The data diverged from a perfect power-law relationship in that the performance in the 80% condition was worse than expected, while the performance in the 10% condition was better than expected. These priority conditions are the cued and uncued tests from the same experimental condition so this pattern is what would be expected if a limited resource was split among the to-be-remembered items in a different distribution than the assigned 80/10/10 priority. This instead suggests that more resources were given to the low-priority items at the cost of precision of the high-priority item (Bays & Husain, 2008). There were several differences between the current study and previous versions of probe probability manipulation experiments conducted by our lab. The most obvious difference is having a majority of 2AFC response trials rather than continuous report. Additionally, this experiment included block average feedback on the break screens; perhaps this feedback coupled with the 2AFC response type added incentive to remembering all the items in the 50% and 80% conditions over remembering the most likely item.

Although the results of this study align best with the predictions of continuous resource models, the predictions of any specific continuous resource model were not tested. As well, some results could be explained by flexible variations of discrete resource models which was not ruled out by direct testing. Overall, the data are best

conceptualized as coming from a noisy memory representation which can be improved with greater priority to the item at encoding in a flexible manner. Importantly, these results did not depend on any model estimated parameters, instead comparing conditions on the model-free summary statistics instead (Ma, 2018).

Conclusion

These data demonstrate the flexible allocation of a continuous resource in a discrete response task. Performance was conserved between the continuous response precision and 2AFC precision further supporting the notion that the memory representations, and any other relevant cognitive processes are conserved between the two response types. The data from the intermixed continuous response trials demonstrated the typical pattern that precision could be predicted by the proportion of memory resources fit to a power-law distribution which establishes that the memory processes during the task agreed with the predictions of continuous resource models. Similarly, precision on the 2AFC trials was also predicted by the proportion of memory resources best fit by a power-law distribution. The 2AFC data demonstrate a pattern of both accuracy and response time that would be predicted by differences in noise in the memory signal which is best explained by a flexibly allocated continuous resource rather than an all-or-none discrete resource. As well, above chance performance of the lowest probe probability condition suggests that this condition produces a low-resolution memory which could be confused with guessing in the wide error distributions from continuous responses. This extends the findings in support of the flexible allocation of continuous resources beyond the most commonly used continuous response delayed estimation paradigm and change detection paradigms.

References

- Agostinelli, C., & Lund, U. (2023). *R package “circular”: Circular Statistics (0.5-0)* [R]. cran.r-project.org/package=circular
- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2021). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, 4(63). doi.org/10.12688/wellcomeopenres.15191.1
- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, 18(7), 622–628. doi.org/10.1111/j.1467-9280.2007.01949.x
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi.org/10.18637/jss.v067.i01
- Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience*, 34(10), 3632–3645. doi.org/10.1523/jneurosci.3204-13.2014
- Bays, P. M. (2015). Spikes not slots: Noise in neural populations limits working memory. *Trends in Cognitive Sciences*, 19(8), 431–438. doi.org/10.1016/j.tics.2015.06.004
- Bays, P. M., Catalao, Raquel. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, 9(10), 7.1–7.11. doi.org/10.1167/9.10.7
- Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, 321, 851–854. doi.org/10.1126/science.1158023

- Brady, T. F., Robinson, M. M., Williams, J. R., & Wixted, J. T. (2023). Measuring memory is harder than you think: How to avoid problematic measurement practices in memory research. *Psychonomic Bulletin & Review*, *30*(2), 421–449. doi.org/10.3758/s13423-022-02179-w
- Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., & Liu, J. (2013). A Nondegenerate Penalized Likelihood Estimator for Variance Parameters in Multilevel Models. *Psychometrika*, *78*(4), 685–709. doi.org/10.1007/s11336-013-9328-2
- Dube, B., Emrich, S. M., & Al-Aidroos, N. (2017). More than a filter: Feature-based attention regulates the distribution of visual working memory resources. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(10), 1843–1854. doi.org/10.1037/xhp0000428
- 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. doi.org/10.1037/xhp0000398
- Fox, J., & Weisberg, S. (2019). *An {R} Companion to Applied Regression* (3rd ed.). Sage.
- Grange, J. (2022). *trimr: An implementation of common response time trimming methods*. (1.1.1) [R]. cran.r-project.org/package=trimr
- Huynh Cong, S., & Kerzel, D. (2022). The allocation of working memory resources determines the efficiency of attentional templates in single- and dual-target search. *Journal of Experimental Psychology: General*, *151*(12), 2977–2989. doi.org/10.1037/xge0001239

Keshvari, S., van den Berg, R., & Ma, W. J. (2013). No Evidence for an Item Limit in Change Detection. *PLoS Computational Biology*, *9*(2).

doi.org/10.1371/journal.pcbi.1002927

Klyszejko, Z., Rahmati, M., & Curtis, C. E. (2014). Attentional priority determines working memory precision. *Vision Research*, *105*, 70–76.

doi.org/10.1016/j.visres.2014.09.002

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

Ma, W. J. (2018). *Problematic usage of the Zhang and Luck mixture model*.

doi.org/10.1101/268961

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory.

Nature Neuroscience, *17*(3), 347–356. doi.org/10.1038/nn.3655

Hautus, M. J., Macmillan, N. A., & Creelman, C. D. (2022). *Detection theory: A user's guide* (3rd ed.). Routledge. doi.org/10.4324/9781003203636

Noorani, I., & Carpenter, R. H. S. (2016). The LATER model of reaction time and decision. *Neuroscience & Biobehavioral Reviews*, *64*, 229–251.

doi.org/10.1016/j.neubiorev.2016.02.018

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C.,

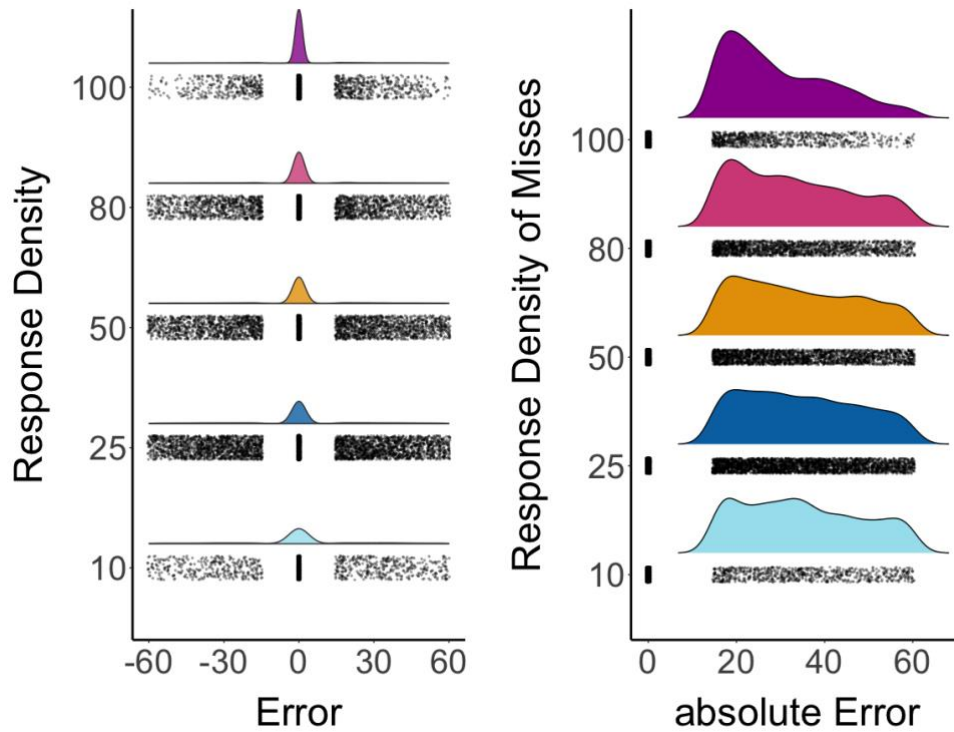
Nee, D. E., Scheppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885–

958. doi.org/10.1037/bul0000153

- Pashler, H. (1988). Familiarity and visual change detection. *Attention, Perception, & Psychophysics*, *44*(4), 369–378. doi.org/10.3758/BF03210419
- Pearson, B., Raskevicius, J., Bays, P. M., Pertzov, Y., & Husain, M. (2014). Working memory retrieval as a decision process. *Journal of Vision*, *14*(2):2, 1–15. doi.org/10.1167/14.2.2
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. doi.org/10.3758/s13428-018-01193-y
- Salahub, C., Lockhart, H. A., Dube, B., Al-Aidroos, N., & Emrich, S. M. (2019). Electrophysiological correlates of the flexible allocation of visual working memory resources. *Scientific Reports*, *9*, 19428. doi.org/10.1038/s41598-019-55948-4
- Schurgin, M. W. (2018). Visual memory, the long and the short of it: A review of visual working memory and long-term memory. *Attention, Perception, & Psychophysics*, *80*(5), 1035–1056. doi.org/10.3758/s13414-018-1522-y
- Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review*, *116*(2), 283–317. doi.org/10.1037/a0015156
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1–29. doi.org/10.1037/h0093759

- Standing, L. (1973). Learning 10000 pictures. *Quarterly Journal of Experimental Psychology*, 25(2), 207–222. doi.org/10.1080/14640747308400340
- van den Berg, R., & Ma, W. J. (2018). A resource-rational theory of set size effects in human visual working memory. *ELife*, 7. doi.org/10.7554/eLife.34963
- 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. doi.org/10.1073/pnas.1117465109
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4(12), 11. doi.org/10.1167/4.12.11
- Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, 440(7080), 91–95. doi.org/10.1038/nature04262
- Yoo, A. H., Klyszejko, Z., Curtis, C. E., & Ma, W. J. (2018). Strategic allocation of working memory resource. *Scientific Reports*, 8, 16162. doi.org/10.1038/s41598-018-34282-1
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453, 233–235. doi.org/10.1038/nature06860

Appendix 4A: Supplementary Figure 2AFC Response Density

Figure 4-A1*2AFC response density showing non-continuity of responses*

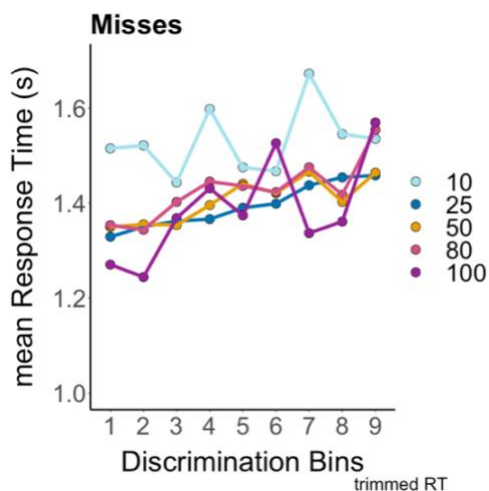
Note. **A.** Response density for All 2AFC responses in each probe probability condition. **B.** Responses density of Miss responses in absolute error, these responses were limited to be between 15-60 degrees.

Appendix 4B: 2AFC Miss RT Results

Miss RT data often behaves differently from hits, for instance, RT trends increase rather than decrease as they do in correct responses as the task gets easier. This was observed presently: as discrimination bins become easier, RTs for misses are longer, see Figure 4-B1. For miss responses it took approximately 114.24 ms longer for far-colour comparisons (bin 9) compared with close-colour comparisons (bin 1) ($\beta = 14.28$ ms, $se = 2.026$, $t = 7.05$, $p < .0001$; estimated at 50% probe probability). The interaction between probe probability and discrimination bin was not significant for the miss responses: $X^2(1) = 2.39$, $p = .122$; the model with only the main effects of probe probability and discrimination bin was significant: $RT_{(hits)} \sim \text{discriminability} + \text{probe probability} + (1|\text{participantID})$: $X^2(1) = 49.58$, $p < .0001$. Similar to correct responses, RTs were longer at lower probe probabilities, but with a smaller difference of only 31.25ms longer at 50% probe probability than at 100% probe probability ($\beta = -0.6249$ ms, $se = 0.1907$, $t = 3.277$, $p = .0011$).

Figure 4-B1

Miss Response times by discrimination bin for probe probability conditions



Note. Trimmed mean response time of missed responses by discrimination bin and probe probability conditions.

Chapter 5: General Discussion

In summary, the current body of work demonstrated the properties and limitations of flexible allocation of a continuous memory resource under a variety of probe probability conditions, response demands, and response styles. In each experiment, the allocation of memory resources was predicted to correspond with the assigned probe probability for each item designating the priority and task relevance. It was previously demonstrated in Emrich et al., (2017) that the probe probability of the item could predict memory performance. This evidence suggested that VSTM resources can be flexibly allocated but left questions about the limitations and properties of a flexible resource that have been addressed in the current studies.

In Chapter 2, the level of control over flexible allocation was tested by assigning three levels of task relevance through two unique cues indicating high and medium levels of priority while the uncued items remained task relevant. It was predicted that response precision should be parametric to the item priority. In general, this predicted pattern was not found, rather participants tended to group the two cued levels of priority. However, analysis of individual strategies showed that a minority of participants were using a three-level allocation strategy, demonstrating that this level of flexibility is possible. The ability to flexibly and finely allocate to three or more levels of priority had been demonstrated when participants were given a pre-cue (Allen & Ueno, 2018; Yoo et al., 2018), but is apparently not the most common strategy when cues are simultaneous to the memory sample as in Chapter 2. This limitation on the control of flexible allocation may suggest that there are mechanisms of memory encoding that are primed or off-loaded when given a pre-cue.

In Chapter 3 two alternative mechanisms of resource allocation – flexible

allocation and probabilistic encoding – were tested using two responses per trial. These mechanisms of resource allocation make different predictions about how items of differing priority are stored simultaneously; consequently, by obtaining responses from two items per trial, several predictions were tested. It was demonstrated that two distinct levels of memory resolution were reported, in agreement with a mechanism of flexible allocation. Further, when multiple responses are made, it is typical for participants to report the better remembered item first. This property made it possible to predict the number of so called “in-memory” responses based on the individual’s estimated memory capacity. This analysis was designed as a more direct test for probabilistic encoding as a potential mechanism of resource allocation. Importantly, the observed number of “in-memory” responses did not match that predicted by probabilistic encoding, again supporting flexible allocation.

Finally, in Chapter 4, using a discrete response paradigm, 2AFC, it was demonstrated that flexible allocation of a continuous memory resource is evident beyond continuous response tasks and supported the underlying signal detection framework of continuous resource models. This study demonstrated evidence that flexible allocation of fewer memory resources results in a noisier memory representation leading to lower accuracy and slower response times as predicted by a signal detection decision-making framework.

Collectively, these studies reveal the properties of a flexible VSTM resource and expand the understanding of VSTM encoding, storage, and recall. These works add to the growing number of studies have demonstrated that continuous resource models, and especially variable precision models, are a better fit to behavioural data (Oberauer, 2023;

van den Berg et al., 2014), and more neurologically plausible (van den Berg & Ma, 2018) than discrete models. Taken together, these data suggest that any model that cannot accommodate a dynamic, flexible resource should be abandoned.

Investigating Response Behaviours

It is clear that VSTM resources are not infinitely flexible. In Chapter 2, I did not find universal utilization of three levels of VSTM resource allocation when given simultaneous memory cues. Even with longer encoding time in Experiment 1b, participants did not readily allocate their VSTM resources differentially to the high- and medium-priority items as expected. Recall error was equivalent for high- and medium-priority items for the majority of participants. However, an alternative explanation could have been that the response measure was not sensitive enough to demonstrate that the highest priority item was prioritized above the medium-priority item. Evidence from the temporal analysis showed that the highest priority item was better protected from performance decline associated with time-on-task fatigue. This may suggest that comparison of mean response precision between conditions is not a sensitive enough measure for some memory processes that differentiate the memory representations of high priority items. Protecting the highest priority item more than the medium priority item suggests that there is some process of resource allocation that did differentiate and rank these two conditions, as expected. Future studies could investigate differences in the change in memory performance over the task in addition to memory recall performance, which may indicate further cognitive processes ongoing during VSTM tasks that are selectively protecting some memories from decay or fatigue but are undetectable in the average.

Another analysis that reveals properties of prioritized resource allocation besides mean performance comparison was reported in Chapter 3, Appendix 3A. Response order effects on performance are well-documented across many VSTM paradigms when multiple responses are recorded as in Chapter 3, so it is not surprising that the first report had better memory performance (Oberauer et al., 2018). However, when the choice of first response is left to the participant, the response order can provide evidence for systemic preferences toward items. In Chapter 3 Experiment 1, it appeared that there was a preferred high-priority item of the two that was better remembered as identified by the self-selected order of memory report. This was the motivation for having three priority levels assigned in Experiment 2. It was predicted that when high- and medium-priority memory items were both probed for recall that even if mean response precision was equal, the response order may indicate that one was actually prioritized over the other. This would suggest that recall error was not sensitive to the prioritization between these conditions, but that there was something that differentiated the two priority conditions in availability as measured by output order. However, the high-priority item was not reported first more often than the medium-priority item (Chapter 3, Appendix 3A). I propose that investigating self-selected response order can provide new information on the relative memory strength of multiple memory stores; however, in the current case, it confirmed that participants were not distinguishing the high- and medium-priority conditions, consistent with Chapter 2.

Predicting High-Threshold Responses

Chapter 3 investigated an alternative mechanism of resource allocation termed probabilistic encoding. Flexible allocation predicts that response precision is

continuously distributed even for low-precision memory representations. Whereas probabilistic encoding proposes that low-priority items are maintained in a high-threshold high precision memory only when there is space in memory capacity to do so. In Chapter 3, participants reported two items on each trial, and it was assumed that the better remembered item would be reported first. Thus, in the low-low test condition (probing two of the four low-priority items) if there was a low-priority item in memory on that trial, it should be reported first. This condition was uniquely useful for this research question since it was more likely that at least one of the test items would be in memory compared with a single probe paradigm or a test condition with only one low-priority item.

This analysis was an exploratory investigation that had been designed to test the predictions of the probabilistic encoding more directly than comparing average response precision. This was the first time to my knowledge that an analysis like this has been used. There is room for improvement on the theoretical assumptions that converted estimated memory capacity scores into trial-by-trial probability of having the item in-memory, but the current iteration is expected to be over-generous to the predictions of probabilistic encoding despite not finding supporting evidence. Specifically, the predicted proportion of in-memory responses was based on a purely rational choice of encoding low-priority items into memory. Realistically, participants are likely to have biased encoding in ways that increase the chance that low-priority items were in memory which would increase the predicted proportion. Rather, the observed proportion was significantly lower than expected thus "better" predictions for the proportion in memory are likely to be more different from the observed value rather than closer. This result

instead suggests that low-priority items are not stored in high-fidelity memory states equivalent to the prioritized items. Future studies could investigate more accurate ways to estimate capacity per trial, or if confidence is measured, an ‘in-memory’ response could be categorized by the participant’s meta-knowledge rather than the response precision.

What Recognition Memory can say about Recall Memory

A benefit of testing memory recall with a continuous measure is that it does not assume a response error is evidence of not having an item in memory, as is the case with the interpretation of K -estimates from discrete VSTM response tasks such as change detection. However, error distributions are unable to differentiate between very low-precision memories and true uniform distribution of random guessing: it would take an unfeasible number of trials to reliably differentiate a flat distribution from a true uniform distribution (Adam et al., 2017). In Chapter 4, this limitation was addressed by using another type of discrete VSTM response task, 2AFC, which measures recognition memory. Recognition memory tasks are commonly used in long-term memory studies to identify weak memory traces (Brady et al., 2023; Schurgin, 2018). During the recognition memory task, participants were better than chance at selecting the target colour even in the most difficult colour discrimination condition at low task relevance (25% probe probability and 15-20° of colour separation); and again, better than chance at all but the most difficult discriminations at very low task relevant items (10% probe probability and 20-25° of colour separation). I argued that evidence of memory for the 10% probe probability condition in all but the smallest discrimination differences was still evidence that these items are represented in memory. This is important evidence for flexible allocation that is only revealed by the recognition paradigm. Evidence from this

recognition memory task supports a model of VSTM with low-precision memory representations that can be lost in a noisy signal. This evidence is in line with signal detection frameworks of VSTM which are in turn supported by plausible neurophysiological mechanisms of neural population coding with noise (Bays, 2014; Bays, 2015).

Signal Detection

Signal detection theory is prevalent across many cognitive and perceptual fields in psychology. In abstract, a signal detection framework theorizes a system with both signal and normally distributed noise around that signal. The strength of the signal over the noise allows the signal to be detected; conversely, the weaker the signal is relative to the noise, the more difficult it will be to detect the signal (e.g., Wixted, 2020). This framework is useful for conceptualizing perceptual and cognitive decisions with uncertainty, as is done across many real-life situations. For instance, many headphones or apps may advertise “sound personalization” which is achieved through a hearing threshold detection test to evaluate the individual’s hearing range, which can be used to apply adjustments to specific frequencies for a better sound experience. This test consists of different frequency beeps given at variable volume and the individual reports when they hear a beep. That is, they are detecting the signal (beep) but may miss tones due to literal noise in the environment, but also due to internal sources of noise, like background neuronal firing. In VSTM, as in the exemplified hearing test and other neural applications of the signal detection framework, a prominent source of noise is thought to come from changes to the firing rates of a population of neurons (Bays 2014; Bays, 2015).

Continuous resource models fit naturally within the context of a memory signal and neural noise: Decreased accuracy with increasing set-size reflects the relative increase in neural noise with higher memory loads (Bays, 2014; Bays 2015; Luck & Vogel, 2013; Ma et al., 2014). At higher memory loads, the population of neurons that are encoding for one item may be competing with the neurons encoding for another item. If there is a limited number of neurons available to encode a memory array, this competition would result in fewer neurons being dedicated to any single memory item and thus less signal per item. Additionally, multiple concurrent memory representations may interfere with each other such that the *signal* for one representation *is noise* for the other. With the additional consideration of flexible allocation, memory load is not the only determining factor for resource allocation, as I have repeatedly shown that resources can be distributed based on attention directing cues. Supporting this idea that attention can direct memory resources, it has been shown that greater attention to the memory sample at encoding (which I propose is the mechanism of flexible resource allocation) results in neural gain at sensory cortex of rhesus monkeys (McAdams & Maunsell, 1999).

If memory representations have a degree of uncertainty, it would also make it more difficult to tell the difference between close colours than far colours. This can be demonstrated by Keshvari et al., (2013) showing that the magnitude of change influences change detection performance such that the capacity estimates would conclude that there were no items in memory at load six. In Chapter 4 it was similarly found that close colour comparisons were more difficult than far colour comparisons. Further, it was demonstrated that the resources allocated to the items (directed by the probe probability cues) influenced the difficulty of the colour comparison such that items with more

predicted VSTM resources allocated to them were easier to discriminate from close colour comparisons than items with less VSTM resources allocated to them. In contrast, perfect memory storage, or uniform high-quality memory storage, would not be affected by the similarity of the target and alternate choice. Thus, Chapter 4 demonstrated that items were held in memory with uncertainty (i.e., noise), and secondly that the probability manipulation was able to affect the signal to noise ratio of the memory representations in a graded manner consistent with a flexibly allocated continuous resource.

Understanding Error Distributions; a Limitation to Interpretation

In delay-estimation paradigms, such as those used in the experiments here, a memory sample is followed by a delay period and a continuous report where the exact colour, orientation or other continuous feature is recalled from memory. Error is calculated as the circular distance between the target and response value. These data have a circular shape meaning the direction of error is not meaningful since the data wraps around where degrees 359 and 0 are next to each other (i.e., $[0, 2\pi)$). Thus, recall error of colour VSTM is represented in a circular normal distribution (the von Mises distribution) centered on the target value as zero. Competing theories of VSTM each make different predictions of the shape of the error distribution, which can be identified through modelling, such as the Zhang and Luck (2008) two-component mixture model, or the three-component mixture model (Bays et al., 2009).

The two-component mixture model proposes that the error distribution can be described by a combination of the circular equivalents of a uniform distribution and a normal distribution (Zhang & Luck, 2008). The circular uniform distribution corresponds

with responses from items not-in-memory that must be randomly guessed, thus take on a uniform shape ($\kappa = 0$). The height of this uniform distribution is taken as the *guess rate*. The circular normal distribution corresponds with in-memory responses that would accurately represent items in memory. This distribution can be thought of as sitting on top of the uniform distribution. The concentration of responses around the target (κ , or circular SD) is reported as the *precision* of memory responses. This distribution theoretically has a fixed width regardless of memory load in equal precision models originally proposed by Zhang and Luck (2008).

A second proposal for decomposing recall error distributions is the three-component mixture model (Bays et al., 2009). In addition to the two components explained above there is a third component that represents non-target errors, a type of binding error where a non-target value (but one that is from the memory sample) is mistakenly reported. These responses appear to be part of the uniform distribution of guessing when compared to the target value; however, if compared to the other feature values in the memory sample they instead look like normally distributed responses (i.e., in-memory responses to the wrong items). Similar to the guess distribution, the value from non-target errors is reported as a proportion of trials where non-target responses are made, or the *non-target rate*. It is argued that these responses are distinct from random guessing as these do not represent out-of-memory responses and do not indicate a capacity limit the way the uniform distribution does. It can be understood that the two-component mixture model is a special case of the three-component mixture model where the non-target rate parameter is fixed to zero (Oberauer et al., 2017).

Further, both of these mixture models can be understood as special cases of the interference measurement model (Oberauer & Lin, 2017; Oberauer et al., 2017). The interference model adds three components accounting for variable noise for each memory representation. When the item is cued for retrieval: (A) all items that are closely related to that cue are retrieved creating “coactivation” interference, (B) all feature values in the memory set are activated, (C) background noise is also activated (Oberauer & Lin, 2017). This model describes not only the shape of the signal in the distributions but the shape of the noise in the distributions. They argue that the addition of these three interference effects best describes the memory performance in VSTM tasks today.

Finally, a hypothesis of variable precision is another variant model that can explain the resources behind error distribution. In contrast to equal precision hypotheses, variable precision proposes that each memory stored has a resolution that is independent of memory load. As with the above models, a response in delay-estimation tasks would be described by the concentration of a von Mises distribution; however, the averaged data is an infinite mixture of von Mises distributions of unique concentration values (Fougnie et al., 2012; van den Berg et al., 2012). There is no component representing guessing in this model because there is no ‘zero-information’ state. According to this theory of VSTM, there is true variability in memory precision both within and between trials that is not a result of fluctuations in attention, or uneven distributions of a memory resource (Fougnie et al., 2012). Rather, variable precision is achieved through random distributions of a continuous memory resource. The interference model similarly allows for variable precision for each item, but the variability is related to variability of the noise in the signal rather than the memory signal (Oberauer & Lin, 2017; Oberauer et al.,

2017). In direct competition, variable precision models tend to provide the best fit to data from numerous studies (Oberauer, 2023; van den Berg et al., 2014; van den Berg et al., 2018). In a recent factorial analysis, Oberauer did not include a direct analysis of the variable precision model but concluded that the best memory model would be one “in which the memory strength of each item varies continuously” (Oberauer, 2023, p. 849)

Given the prevalence of these models, some readers may have wondered why I did not model the error distributions from the delayed estimation tasks using any of the commonly used models in the field today (e.g. Bays et al., 2009; Oberauer et al., 2017; Schurgin et al., 2020; van den Berg et al., 2012; Zhang & Luck, 2008). While the computational components of mixture models are useful to understand insofar as they demonstrate the conceptual mechanisms of the underlying VSTM resource theories, they can limit the interpretation of the data unless model comparison is specifically the goal of the investigation. Oberauer and colleagues (2017) argued that both two- and three-component mixture models are based in theoretically discrete memory states (i.e., in-memory or not in-memory), which is not useful to theories that do not endorse this assumption such as continuously varying memory strength with signal detection models, i.e., continuous resource models. For instance, an investigation found that data simulated by a variable precision model is severely mismatched by the two-component mixture model (*slots + averaging*, Zhang & Luck, 2008). This poor fit especially overestimated the proportion of guesses for low-precision responses (Ma, 2018). Two important details about this particular finding are that at the time of the analysis, variable precision models were reliably the best fitting model for VSTM data (e.g., van den Berg et al., 2014) – suggesting that the real underlying constructs of memory precision were most similar to a

variable precision model – but the most commonly used model to extract summary statistics (memory precision and guess rate) was the two-component mixture model. This is problematic for the study of VSTM using mixture model summary statistics because it means that conclusions made from the model parameters do not necessarily reflect the underlying cognitive processes that were intended. Ma (2018) makes several suggestions for applying these models. Relevant to the current studies is the suggestion to forgo modelling and instead compare conditions using model-free summary statistics such as mean absolute error, circular variance, or circular SD, as was done here.

Conclusion

A unique feature of the current thesis are the various analyses used to investigate the properties of flexible VSTM resource allocation. Since summary statistics were not modelled, the analyses relied on comparisons of response precision in carefully constructed task conditions. In addition, each set of experiments used somewhat unconventional approaches to further explore the rich data available from delay-estimation tasks or supplement that data in the case of Chapter 4 with the 2AFC analysis. Notably in Chapter 2, the pattern of responses was interpreted to reflect a likely allocation strategy used; and a temporal analysis of response precision change over time was considered as well to investigate allocation strategies. Additionally, in Chapter 3, Appendix 3A, a third approach looking at response order investigated potential evidence for a three-level allocation strategy but did not support the common use of three priority levels. Chapter 3 utilized a novel approach to compare the proportion of responses within a certain range (theoretically “in-memory responses”) with a predicted value based on the participants estimated memory capacity. In each case, I think my novel approach to the

data analyses contributes to our understanding of flexible resource allocation, strengthening the evidence from the conventional analysis of response error.

The purpose of these experiments was to refine the understanding of flexible properties of VSTM resources. This property is best aligned with continuous resource models and signal detection frameworks of cognitive processing. The flexible nature of VSTM cannot be left out of models of the nature of VSTM resources and must be investigated to better understand the interactions of VSTM and attention at encoding, maintenance, and retrieval. These presented studies demonstrated a variety of alternative analysis techniques which all supported the flexible resource hypothesis.

References

- Adam, K. C. S., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology*, *97*, 79–97.
doi.org/10.1016/j.cogpsych.2017.07.001
- Allen, R. J., & Ueno, T. (2018). Multiple high-reward items can be prioritized in working memory but with greater vulnerability to interference. *Attention, Perception, & Psychophysics*, *80*(7), 1731–1743. doi.org/10.3758/s13414-018-1543-6
- Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience*, *34*(10), 3632–3645.
doi.org/10.1523/jneurosci.3204-13.2014
- Bays, P. M. (2015). Spikes not slots: Noise in neural populations limits working memory. *Trends in Cognitive Sciences*, *19*(8), 431–438. doi.org/10.1016/j.tics.2015.06.004
- Bays, P. M., Catalao, Raquel. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 7.1–7.11. doi.org/10.1167/9.10.7
- Brady, T. F., Robinson, M. M., Williams, J. R., & Wixted, J. T. (2023). Measuring memory is harder than you think: How to avoid problematic measurement practices in memory research. *Psychonomic Bulletin & Review*, *30*(2), 421–449.
doi.org/10.3758/s13423-022-02179-w
- 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.
doi.org/10.1037/xhp0000398

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, 3, 1229.

doi.org/10.1038/ncomms2237

Keshvari, S., van den Berg, R., & Ma, W. J. (2013). No Evidence for an Item Limit in Change Detection. *PLoS Computational Biology*, 9(2).

doi.org/10.1371/journal.pcbi.1002927

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. doi.org/10.1016/j.tics.2013.06.006

Ma, W. J. (2018). *Problematic usage of the Zhang and Luck mixture model*.

doi.org/10.1101/268961

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

McAdams, C. J., & Maunsell, J. H. R. (1999). Effects of Attention on Orientation-Tuning Functions of Single Neurons in Macaque Cortical Area V4. *The Journal of Neuroscience*, 19(1), 431–441. doi.org/10.1523/JNEUROSCI.19-01-00431.1999

Oberauer, K. (2023). Measurement models for visual working memory—A factorial model comparison. *Psychological Review*, 130(3), 841–852.

doi.org/10.1037/rev0000328

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schwegge, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for

models of short-term and working memory. *Psychological Bulletin*, 144(9), 885–958. doi.org/10.1037/bu0000153

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

Oberauer, K., Stoneking, C., Wabersich, D., & Lin, H.-Y. (2017). Hierarchical Bayesian measurement models for continuous reproduction of visual features from working memory. *Journal of Vision*, 17(5), 11. doi.org/10.1167/17.5.11

Schurgin, M. W. (2018). Visual memory, the long and the short of it: A review of visual working memory and long-term memory. *Attention, Perception, & Psychophysics*, 80(5), 1035–1056. doi.org/10.3758/s13414-018-1522-y

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

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, 121(1), 124–149. doi.org/10.1037/a0035234

van den Berg, R., & Ma, W. J. (2018). A resource-rational theory of set size effects in human visual working memory. *ELife*, 7. doi.org/10.7554/eLife.34963

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.

doi.org/10.1073/pnas.1117465109

Wixted, J. T. (2020). The forgotten history of signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(2), 201–233.

doi.org/10.1037/xlm0000732

Yoo, A. H., Klyszejko, Z., Curtis, C. E., & Ma, W. J. (2018). Strategic allocation of working memory resource. *Scientific Reports*, 8, 16162.

doi.org/10.1038/s41598-018-34282-1

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