Eye-gaze, attention and strategy in visual working memory

Eye-gaze, attention and strategy in visual working memory.

Edmond Stewart

A thesis in fulfilment of the requirements for the degree

of Doctor of Philosophy

School of Psychology

Faculty of Science

May 2021

Thesis Title and Abstract | Declarations | Inclusion of Publications Statement | Corrected Thesis and Responses

Thesis Title

Eye gaze, attention and strategy in visual working memory

Thesis Abstract

In the field of visual working memory (VWM) there exist a number of competing models that attempt to describe the capacity and storage features of memory. However, these models lack an explanation of how items are encoded. This presents a significant problem in the field, as it is currently difficult to distinguish between competing models such as the variable precision model (Van den Berg, Shin, Chou, George, & Ma, 2012) and the slots-plus-averaging model (Zhang and Luck, 2008). Given that these models have distinct theoretical accounts of memory, the lack of difference between the models' behaviour creates a problem for their explanatory power. To create a point of difference in these theoretical accounts, we wanted to investigate the encoding state of VWM to see if how attention was distributed during a task. We used eye gaze as a proxy for attention to investigate connections between encoding and performance in VWM tasks. In standard tasks we found little connection between when and if the target was fixated and task performance. Instead, participants most frequently focus the centre of the screen and appear to learn to decrease their eye movements across the course of the experiment. We then switched to investigating a gaze contingent paradigm and saw not only a connection between fixation of the target and task performance, but most prominently an effect of recency. While some of our findings were more suggestive of the variable precision account compared to the slots-plus-averaging account, we conclude that the prominence of the recency effect is most in line with interference accounts of VWM (Oberauer & Lin, 2017). As well as our investigation into attention and performance, we also investigated task strategy. In our standard tasks, we saw a preference towards covert strategies that made few fixations however there were participants in these tasks that made consistent item fixations. This demonstrated individual differences in task approach. We found some consistent search strategies such as participants generally favouring an anticlockwise pattern when exploring study arrays that were presented on a circle. Given the individual differences we found, we recommend that strategic approach be investigated further to create a more complete understanding of VWM. In general, we caution the use of eye tracking in standard VWM tasks and believe that gaze-contingent designs should be utilised to explore the relationship between attention and VWM.

Thesis Title and Abstract | Declarations | Inclusion of Publications Statement | Corrected Thesis and Responses

ORIGINALITY STATEMENT

☑ I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

COPYRIGHT STATEMENT

AUTHENTICITY STATEMENT

☑ I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis.

Thesis submission for the degree of Doctor of Philosophy

Thesis Title and Abstract | Declarations | **Inclusion of Publications Statement** | Corrected Thesis and Responses

UNSW is supportive of candidates publishing their research results during their candidature as detailed in the UNSW Thesis Examination Procedure.

Publications can be used in the candidate's thesis in lieu of a Chapter provided:

- The candidate contributed **greater than 50%** of the content in the publication and are the "primary author", i.e. they were responsible primarily for the planning, execution and preparation of the work for publication.
- The candidate has obtained approval to include the publication in their thesis in lieu of a Chapter from their Supervisor and Postgraduate Coordinator.
- The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis.

☑ The candidate has declared that **their thesis contains no publications, either published or submitted for publication**.

Candidate's Declaration

I declare that I have complied with the Thesis Examination Procedure.

Acknowledgements

First, thank you to my primary supervisor, Chris Donkin. Anything I write here can't fully acknowledge the level of support you have given me throughout this process. From basic programming guidance to our more philosophical conversations to your reading and re-reading this thesis, I can't thank you enough. I'm not how many people would say they would want to be friends with their supervisor post-submission, but I am one of them.

Thank you to my co-supervisor, Mike Le Pelley for his generous assistance with my many questions about MATLAB and eye tracking. If it weren't for your clear explanations, I would probably still be trying to code my experiments.

Thank you to Poppy Watson for enthusiastically updating my experiments for the Tobii Pro Spectrum and simultaneously offering your debugging assistance.

Thank you to my friends, housemates and family who have cooked for me, kept me sane and put up with me throughout. In particular, Sarah Wagstaff and Gareth Kindler, for their patient ears and kind words when I needed them most.

Thank you to the Cognition Lab at UNSW who provided much needed friendly distraction at the best of times. A special thank you to Jeremy Ngo who is, undoubtedly, the best Marth of all time.

Abstract

In the field of visual working memory (VWM) there exist a number of competing models that attempt to describe the capacity and storage features of memory. However, these models lack an explanation of how items are encoded. This presents a significant problem in the field, as it is currently difficult to distinguish between competing models such as the variable precision model (Van den Berg, Shin, Chou, George, & Ma, 2012) and the slots-plus-averaging model (Zhang and Luck, 2008). Given that these models have distinct theoretical accounts of memory, the lack of difference between the models' behaviour creates a problem for their explanatory power. To create a point of difference in these theoretical accounts, we wanted to investigate the encoding state of VWM to see if how attention was distributed during a task. We used eye gaze as a proxy for attention to investigate connections between encoding and performance in VWM tasks. In standard tasks we found little connection between when and if the target was fixated and task performance. Instead, participants most frequently focus the centre of the screen and appear to learn to decrease their eye movements across the course of the experiment. We then switched to investigating a gaze contingent paradigm and saw not only a connection between fixation of the target and task performance, but most prominently an effect of recency. While some of our findings were more suggestive of the variable precision account compared to the slots-plus-averaging account, we conclude that the prominence of the recency effect is most in line with interference accounts of VWM (Oberauer & Lin, 2017). As well as our investigation into attention and performance, we also investigated task strategy. In our standard tasks, we saw a preference towards covert strategies that made few fixations however there were participants in these tasks that made consistent item fixations. This demonstrated individual differences in task approach. We found some consistent search strategies such as participants generally favouring an

anticlockwise pattern when exploring study arrays that were presented on a circle. Given the individual differences we found, we recommend that strategic approach be investigated further to create a more complete understanding of VWM. In general, we caution the use of eye tracking in standard VWM tasks and believe that gaze-contingent designs should be utilised to explore the relationship between attention and VWM.

Table of contents

Introduction

In recent years, it has become popular to try to understand visual working memory (VWM) through the use of computational models. These models attempt to address fundamental questions in VWM. For example, whether VWM has a strict capacity limit and how likely a stimulus is to be remembered. In short, there are two dominant accounts of behavioural data in VWM, the slots plus averaging model (Zhang & Luck, 2008) and the variable-precision model (Van den Berg et al., 2012). While these models account for the data well, they also mimic each other. This creates a significant problem. Without distinct models it is impossible to determine which gives a more accurate account of VWM. At the start of this thesis, we set out to attempt to provide insight into this issue by using eye-gaze data to examine the encoding process in VWM, and the role of strategy in VWM tasks. Exploring the encoding process goes beyond the current models, which account for what happens once things are in (or out of) memory. Through this analysis, we hope to gain more insight into VWM and we attempt to resolve the question of whether VWM is better described as a set of discrete slots or as a flexible resource. In our exploration of strategy, we consider both group-level and individual preferences for a variety of VWM tasks. In this introduction we provide a description of main types of models in VWM and then discuss the issue of mimicry. We then discuss the use of eye tracking as a method of measuring the encoding process, and its relation to attention. Finally, we explore the literature on strategy and attention in visual tasks and conclude this section with how we will address our two lines of inquiry: attention and strategy in VWM.

Models in VWM

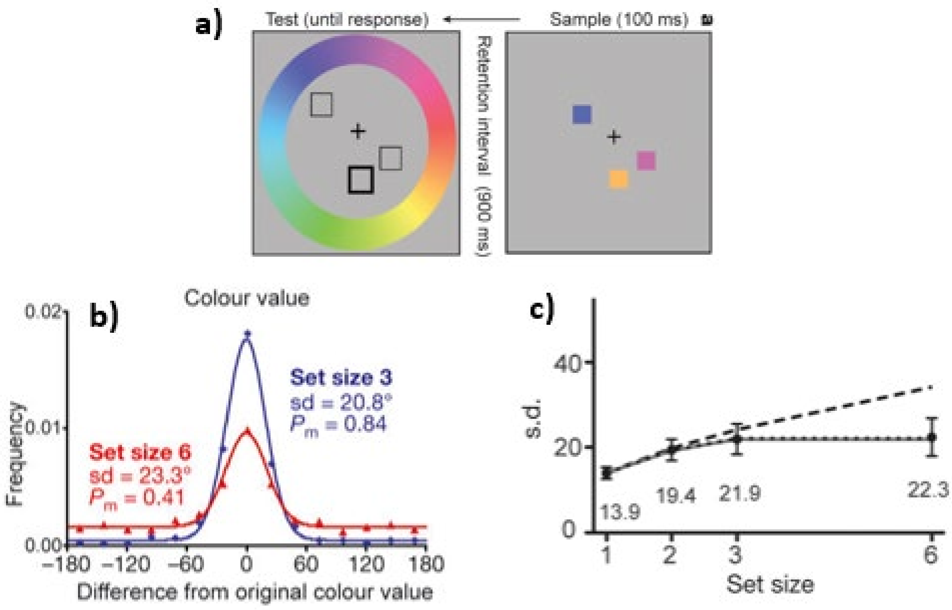


Figure 0.1. a) The colour recall task used by Zhang and Luck (2008). Participants were presented with an array of stimuli and on test had to reproduce the colour of a square at a probed location using a colour wheel. b) the frequency distribution of responses as difference from actual colour value (in degrees). Frequency of accurate response (no difference = 0) decreases with set size as does the probability of an item being in memory (P_m). c) standard deviation of responses as set size increase for actual data (dotted line), as predicted by the slots model (grey line) and as predicted by the slots model (dashed line).

In VWM research, there are two main types of model that have been used to account for its capacity. They are known broadly as slots-based and resources-based models. To contextualise the models, they are based on VWM tasks like the colour recall task (Wilken & Ma, 2004). In this task a participant is presented with a set of items such as coloured squares to remember, and after a short study interval one particular square in the display is indicated by highlighting the location where it was presented during study. The participant is asked to reproduce the colour of the highlighted square by selecting it on a colour wheel. An example, taken from Zhang and Luck (2008) of this sequence in shown in Figure 0.1a and the frequency and accuracy of responses as a function of set size is shown in Figure 0.1b.

Slots-based models

Slots-based models propose that memory behaves like a finite set of slots, each of which is able to hold one item. The slots-based model proposed by Luck and Vogel (1997) is the prototypical account of this type. A schematic for this account is displayed in Figure 0.2. Whether an item is remembered depends on whether it is stored in one of the available memory slots. If the number of items to be remembered is less than or equal to the number of slots in memory then all items in the array will be remembered with high precision. However, if there are more items than memory slots, any items not stored will have no information retained about them. Therefore, if asked about an item that is not in memory, a slots-based account assumes that person will have no information at all to inform their choice and will be forced to guess. This is described as an all-or-none account (Zhang & Luck, 2008) – memories are either very accurate or absent entirely.

Slots-based models may originate from Miller (1956), who found that across a variety of stimuli, information processing appears to be limited in capacity holding around 7 items at a time. However, the more recent and VWM specific description of the slots model comes from Luck and Vogel (1997) who found that memory performance in a change-detection task began to decline with arrays larger than 3-4 items. Luck and Vogel (1997) presented participants with a study array and then following a brief interval, a second array that was either identical to the first or was changed by one feature. Participants were required to indicate whether or not a change occurred between the study and test arrays. Study arrays varied in size from 1-12 items and also varied in stimulus complexity (defined as the number of features in a stimulus). Luck and Vogel (1997) found that only set size effected the task and that regardless of complexity, participants had good memory for 3-4 items. This led the authors to conclude that VWM has a strict capacity limit of 3-4 items, but these items are remembered with a high amount of precision and the representations can accurately store more complex as well as simple items.

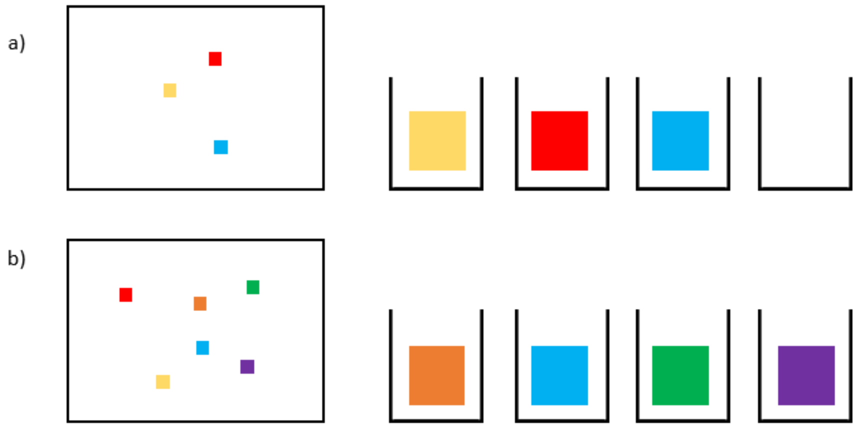


Figure 0.2. The standard slots model (Luck & Vogel, 1997). VWM is described as being a series of 3 to four slots each capable of holding one item. If an item is in a slot it is recalled perfectly. If the number of items displayed are less than or equal to the number of slots, as in a) the whole display can be retained. When asked at test about the location of any coloured square, it will be recalled correctly. If the number of items in the display are greater than the number of slots as in b), some items will not be in memory at all (the yellow and red). In the case of b), if the orange, blue, green or purple squares are probed, recall will be prefect but if the yellow or red squares are probed, the model or participant must guess the location.

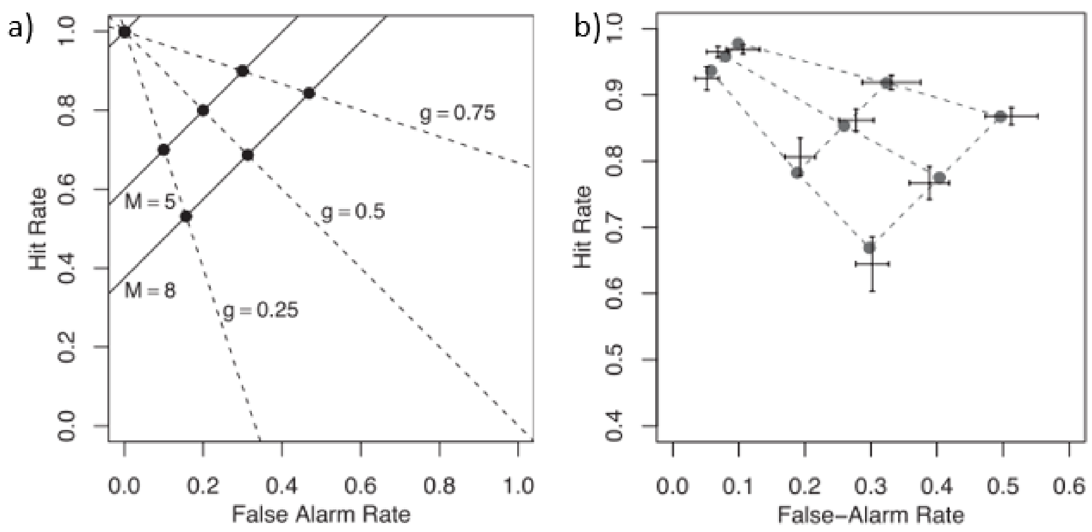


Figure 0.3. ROC figures from Rouder et al. (2008) showing evidence in favour of a slots account of VWM. a) shows predictions about ROC based on the slots model while b) shows the empirical data (error bars) as well as the predictions (dashed lines). Copyright (2008) National Academy of Sciences.

Zhang and Luck (2008) expanded on the basic model to create their slots-plus-averaging model that makes the additional assumption that when the observer has more slots than items to remember, then items are stored in multiple slots. Luck and Vogel's (1997) account makes the surprising prediction that, regardless of the number of items to be remembered – often referred to as set size - any item that makes it into memory will be recalled accurately. For example, an item that is encoded in a 3-item display will be remembered and recalled with the same precision as an item encoded in a 6-item display. The implication from this is that the precision of memory recall should be constant regardless of set size when the item is in memory. Though this pattern was found to hold true for set sizes of three and larger, Zhang and Luck (2008) did find that the precision of memory-based recall was better for small set sizes (as seen in Figure 0.1c). This gave the impetus to include an averaging component to the model, which made it so that when there are more slots available that items to remember, item details can be stored in multiple slots. The information in multiple slots can then be combined to produce a more accurate response, thus leading to better performance when set sizes are small.

Another variation of the slots model was developed by J. N. Rouder et al. (2008) who demonstrated the need for an attention parameter in the model to account for errors made when set sizes were small. Empirically, even when set sizes are 1 or 2 items only, participants will still occasionally make errors in a change-detection task. This is not explained by the base slots model as it assumes that if there is space in memory for an item to be encoded, it will be. Rouder et al. (2008) believed this discrepancy could be explained by adding an attention parameter that accounted for the fact that sometimes participants are distracted and fail to encode items even when they have memory capacity to do so. Rouder et al. (2008) found their slots model with this additional parameter out performed slots-based models without this parameter as well as models with a variable capacity limit. In their task the researchers varied set size as well as the probability of whether a trial was a 'change' or 'same' trial. By varying this probability, the hit rate and false alarm rates were made to vary within each set size. Given a slots model with a fixed capacity and when an item is not remembered, guessing is made, Rouder et al. (2008) were able to make strong predictions about receiver operating characteristics (ROCs). Shown in Figure 0.3a are the ROC predictions that are made by the slots model. With a fixed memory capacity, hits and false alarm rates are predicted to increase with a slope of 1 when set size is held constant (solid lines, Figure 0.3a). Additionally, hit rates decrease while false alarms increase with set size when the probability of a 'change' trial, g, is held constant (dashed lines, Figure 0.3a). The slope of this line is predicted to be equal to $1-1/g$. Rouder et al. (2008) found that by introducing an attention error into the slot model, shown by the dashed lines in Figure 0.3b, a slots-based model could be made consistent with the empirical findings (error bars, Figure 0.3b). Because of this, Rouder et al. (2008) conclude that their slots-based model of VWM to be the most compatible with change-detection data.

Resources-based models

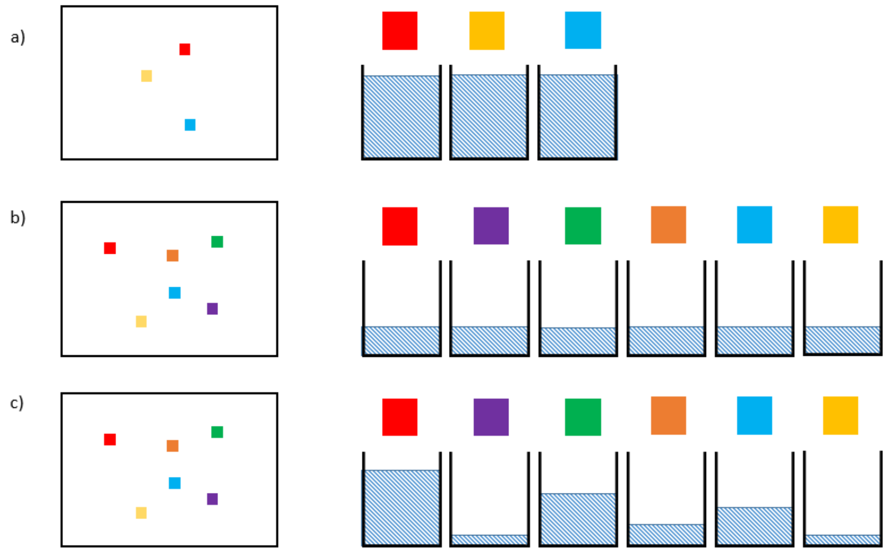


Figure 0.4. A representation of the resources model (Frick, 1988). Rather than having a set number of slots, the resources model states that memory is flexible in the number of items it can contain but has a set amount of a memory resource to distribute between items. With a small number of items, as in a), each item can receive a large amount of memory which leads to accurate memories for smaller set sizes. With larger set sizes, like in b) all items are still represented in memory, but there is less of the memory resource to go around.

The resources-based model, on the other hand, conceptualises memory more flexibly than a slots model. Rather than whole items stored near-perfectly in memory, a resources-based model describes memory as a continuous resource that is divided up and allocated to different items. This mnemonic resource determines the quality or precision of the resultant memories for those items. The more memory resource an item is allocated, the more precise the memory. According to the standard resources model (Frick, 1988), memory is divided equally between all items in the display. Since the amount of memory resource is constant, the more objects there are in a display, the less memory each item is allocated - as seen in Figure 0.4. Unlike the slots model, all items are remembered, but because the amount of mnemonic resource available is fixed, it must be divided up more finely as the number of items increases, which results in less accurate memories. Frick (1988) proposed this model as an alternative to the account proposed by Miller (1956). Miller (1956) described memory as limited in capacity but items that were represented in memory were represented with equally high precision. In his review, Frick (1988) proposed the idea of a mnemonic resource, citing evidence from Palmer (1988) that representations for two items were less precise than for one item alone – something that was not compatible with a slots-based account. Palmer's (1988) experiment tested participant's recognition of line length. Either one or two lines were presented for 100ms and recognition of one line was tested at a time. Palmer found a small decrease in accuracy of response for two lines rather than one. Given that the capacity of memory under the slots account was likely more than two items, Frick (1988) suggested that this account could be insufficient in describing VWM (though we know now that this result would be explained by the slots-plus-averaging model).

Resources based accounts have also been described that are based on signal-detection theory, such as that described by Wilken and Ma (2004). These researchers were critical of the notion of the high precision memories described in the slots account. This is because they believe encoding without considerable noise to be implausible. Furthermore, assuming such noise in encoding gives rise to a capacity limit, without the need to define a capacity limit within the model. Wilken and Ma's (2004) experiment consisted of a series of change-detection tasks using stimuli that varied by colour, orientation or spatial frequency. In some of their experiments, the number of targets was allowed to vary ($T = 1, 2, 3$, or 4) – that is the number of items which were able to change in a change trial varied. ROC curves were used to compare the prediction of a slots-based account with signal-detection accounts that assumed constant noise or variable noise. Based on the results, Wilken and Ma (2004) found that the signal-detection models that were defined to detect noise changes between the study and test arrays provided the best account of the data.

Additional variations of the resources-based models exist that allow the model to distribute its resources more flexibly, or models that favour selecting a few items to focus most of their memory on. Bays and Husain (2008) provided evidence in favour of such an account in a study they ran using location and orientation recall tasks. In their experiments, participants were required to reproduce the original location or orientation of a probed item they saw during the study array. There was also another condition in which participants were either required to make a saccade – that is to move their eye-gaze – to an item of a pre-specified colour or not. It was found that memory performance, specifically the precision of the participants recall responses, was improved when the probed item was the target of the saccade. Bays and Husain (2008) suggested that this was evidence of eye-gaze and attention altering the distribution of a memory resource, indicating VWM might have a flexible, uneven distribution of resources.

A resources model that is in line with this flexible allocation account of mnemonic resource comes from Smith, Lilburn, Corbett, Sewell, and Kyllingsbæk (2016). These authors proposed that memory is distributed amongst items according to attention allocated to each item. Following a series of phase discrimination experiments, in which participants had to judge which side of a half black, half white circle was black, Smith et al. (2016) found that their results were best accounted for by a model which allocated the majority of its mnemonic resource to one item and the remaining attention split between the other items in the display.

However, this study was only tested set sizes of n = 1, 2, 3 or 4, specifically to avoid complications of potentially overrunning capacity. Therefore, while their experiment appeared to be evidence in favour of a resources approach, the lack of data from larger set sizes means that a variant of the slots plus averaging model may plausibly also account for their results.

The variable-precision model

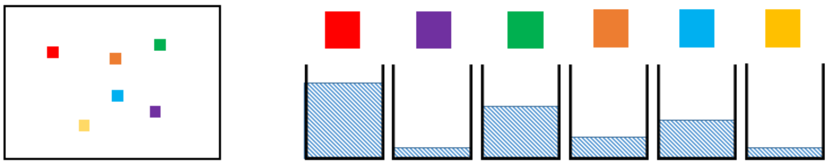


Figure 0.5. A representation of the distribution of memory in the more successful variable-precision model (Van den Berg et al., 2012) is not uniform but variable based on attention or noise.

In terms of recall tasks, the slots-plus-averaging model by Zhang and Luck (2008) arguably gave the best account of the data for some time. As shown in Figure 0.1, their results demonstrated a decreased probability of an item being in memory with higher set sizes (Figure 0.1b) as well as a response precision that, while higher for smaller set sizes, is generally consistent (Figure 0.1c). Both of these findings are more consistent with a slots account than the resources accounts provided up until then. As such, the slots-plus-averaging was proposed as the best description of VWM.

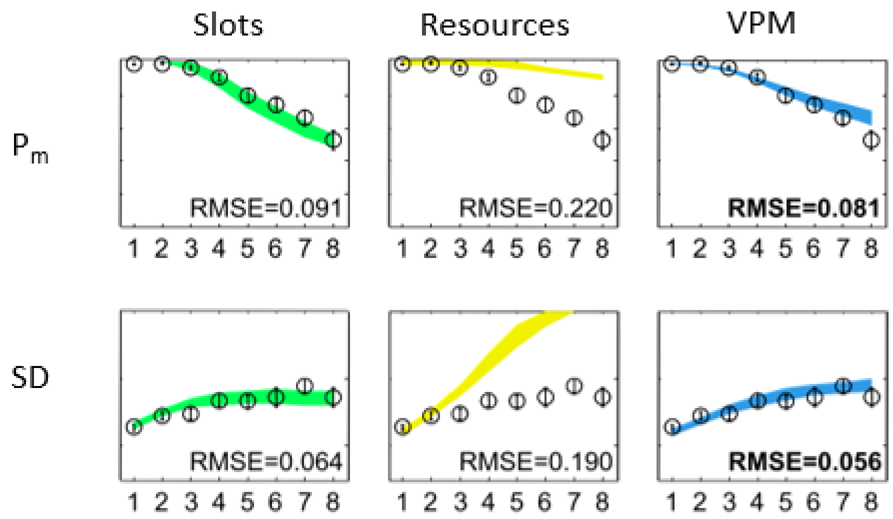


Figure 0.6. A comparison of the slots, standard resources and variable-precision models (Van den Berg, Shin, Chou, George, & Ma, 2012).The predictions of each model for probability of an item being in memory (P_m) and the standard deviation of response (*SD*) are depicted in the coloured lines. The experimental data is given by the open circles. The root mean-square error (RMSE) indicates than the model with least error is the variable-precision model.

More recently, this finding was challenged by Van den Berg et al. (2012), who developed a resources-based model they named the variable-precision model. Similar to the model from Smith et al. (2016), the variable-precision model assumed that the mnemonic resource could be distributed unequally between items in memory. However, unlike Smith et al. (2016), the restrictions on how precision was allocated were less strict. That is to say, different items in a study set could be remembered with different levels of precision. Figure 0.5 gives an example of how memory may be distributed under this model. Van den Berg et al. (2012) compared the standard resources, slots-plus-averaging and variable-precision model and found that variable-precision model gave a better account of the data from recall tasks. Figure 0.6 shows these results in which the variable-precision model is better able to account for data from a typical recall VWM task. Furthermore, in a large reanalysis of multiple data sets from multiple testing sites, Van den Berg, Awh, and Ma (2014) compared a host of computational models of VWM including the variable-precision model. Overall, the variable-precision models provided the best account of VWM data.

One issue with the variable-precision model is that its memory-allocation process is extremely flexible. As such, it can produce many memory states that closely resemble what is predicted by a slot-based model. For example, it is possible that in a display of six items, the mnemonic resource could be allocated roughly equally between four items in the display with practically no resource allocated to the other two items. These very low precision memories would have so little information attached to them that if these items are probed, the predictions of the model are indistinguishable from guessing. Such a memory state is indistinguishable from a slots model, in which the items in slots are remembered well but those that are not in slots are not remembered at all. What complicates matters is that despite making the same predictions for responses, the interpretations are very different. The variable-precision model states that wildly incorrect responses are caused by extremely low precision memories, while a slots model says that such responses are not based on memory.

While the variable-precision model can mimic the responses of the slots models, it is clear that the theory behind these models is quite at odds. However, at our current level of understanding it is not possible to distinguish between these models.

Solving the mimicry problem

Model mimicry presents a large problem with the current status of VWM research. While the variable-precision model is the best performing model, if it cannot be distinguished from a slots-based account, the explanatory power of the model is compromised. It is difficult to generate unique, testable predictions from models when one model can always mimic the other.

Here, we pursue a potential solution to the problem by focusing on aspects of VWM that are currently overlooked. Existing models are similar in that they describe the capacity of memory, the probability of items being in memory, and the precision of such memories. Each of these factors, while important to include in an account of the VWM process, are at a computational level (Marr, 1982). That is to say, they attempt to discriminate between different models of VWM by interrogating the consequences of different types of memory content. These models do not attempt to characterise the processes that govern VWM, such as encoding, maintenance, or retrieval. Unfortunately, discriminating between such computational-level models has reached an impasse, since models with vastly different kinds of memory representations can produce the same observed behaviour.

Creating a model that takes into account the process of how items get into memory would provide a new level of explanation that has seen limited exploration in the VWM literature so far. However there have been some notable examples. Smith and Ratcliff (2009) used response times from a stimulus detection task to inform their theory that linked visual encoding, spatial attention and VWM. In this theory, a stimulus is first processed as a sensory response that is then encoded into VWM with the aid of spatial attention. The process of an item being encoded into memory is dependent on time and is made more efficient with the assistance of spatial attention. This account provides a theoretical link between attention and VWM. The paper by Smith et al. (2016) is a spiritual successor to this work by Smith and Ratcliff (2009). However, neither theory has been tested (by design) in displays that exceed the purported item limit of VWM. Therefore, the extent to which these theories could be used to describe larger displays is unclear.

The only computational model that attempts to account for the encoding process in VWM tasks (to the best of our knowledge) is the previously mentioned attention-weighted sample size model by Smith et al. (2016). This model describes attention as a guiding process in VWM, which determines how mnemonic resources are allocated in the display. Specifically, one of the displayed items receives the largest portion of attention and thus memory resources, while the remaining attention is divided equally among the other items in the display. This account is an encouraging link between attention and VWM. However, to avoid complications with capacity limits, Smith et al. (2016) limited their study to set sizes of n = 1, 2, 3 or 4. Naturally, this means that the model is untested on larger set sizes. It is unclear, for example, whether a slot-like item limit of four items would need to be implemented in their model to explain behaviour in larger set sizes. In this thesis, we hoped to be able to extend process accounts such as Smith et al.'s to larger display sizes.

By furthering our understanding of the encoding process in VWM, we could work to resolve whether VWM is best categorised by a slots or resources model. This is because under these different models, different patterns of encoding and attention distribution might be expected. For example, within a slots model there is a hard limit on the capacity of memory. Perhaps this hard capacity limit is reflected in how items are encoded. That is to say, within a trial, an individual may only attend and encode 3 or 4 items. By contrast, a resources model (such as the variable-precision model) suggests all items in the display are encoded, however precision for items is reduced with the more items to be remembered. If a resources account were true, we might expect an individual to attempt to attend and encode all items in the display, as opposed to focusing on a subset as we might expect under the slots account. Thus, by examining how attention is distributed in VWM experiments, it may be possible to develop a processed based account of VWM which would lead to the resolution of the current model mimicry problem. Gaining such an understanding of attention in VWM is the first motivation for this thesis.

Eye-gaze and attention

Eye-gaze data is often used as a proxy measure of attention. This is because eye-gaze and attention most often occur together: typically people will fixate their eyes on that to which they are paying attention (Posner, 1980). Indeed, as detailed vision only occurs at the fovea, fixation is required to receive high-quality information about the stimulus (Livingstone & Hubel, 1988). When a fixation is made via a saccadic eye movement to a target location, it is always linked with an attention shift (e.g. Deubel & Schneider, 1996; Hoffman & Subramaniam, 1995). While it is possible to attend an item in one's peripheral vision, which is to say to use covert attention, it is reportedly uncommon for people to do this without being prompted (Findlay & Gilchrist, 2001; Rayner, 2009). This is clearly demonstrated in numerous tasks where stimuli that capture attention (due to salience in colour, onset , etc.) draw eye-gaze even when this is detrimental to the task (e.g. Le Pelley, Pearson, Griffiths, & Beesley, 2015). As a result, eye-gaze is usually considered a good measure of attention.

In regards to VWM, eye-gaze has been used to demonstrate the link between attention and memory (Fougnie, 2008). For example, a number of studies examine the number and sequence of fixations and how this affects memory performance. Zelinsky and Loschky (2005) used an alternative-forced choice task with naturalistic items. Participants were presented with a study array of 9 items in each trial either simultaneously (experiment 1) or sequentially (experiment 2). In each trial, participants were allowed to freely view the scene while their eye movements were tracked. At some point in each trial, the participant would fixate on the target (randomly selected on each trial) and the number of fixations the participant made after looking at the target was monitored. When the number of fixations made after viewing the target reached a certain criterion (1 – 7), the study array was removed from view and participants were then probed on their memory of the target. The authors showed that the proportion of correct responses was greatest when there were fewer fixations made between viewing the target and test (Zelinsky & Loschky, 2005). This suggests a recency effect such that the more recently a target is sighted (relative to test) the more likely it is to be remembered. Similarly, Irwin and Zelinsky (2002) had an array of natural objects. Participants performed a partial recall task in which they were allowed to explore the display for n = 1, 3, 6, 9, 15 fixations before being tested. The authors found that if the target item was one of the last three items fixated upon, then participants were more likely to recall the item correctly. Again, this suggests an effect of recency in eye-gaze.

A recency effect was also be shown by Nosofsky and Donkin (2016),wherein to-be-remembered items were presented sequentially. In sequential presentation, participants are not free to choose the order in which they look at items. Instead, stimuli are presented to them one at a time, creating an order in which the items must be viewed. While Nosofsky and Donkin (2016) did not use eye tracking, the sequential presentation provides us with a definitive order in which items were seen. By finding a recency effect in these conditions, we see such effects are present even in the absence of free moving gaze. When taken with the eye-gaze studies, this would suggest that the order in which objects are presented has an impact on VWM, and suggests that eye-gaze may be a useful means of understanding the role of attention in VWM.

There are some studies, however, that indicate that eye movement can interfere with spatial memory of items. For example, Lange, Starzynski, and Engbert (2012) found interference between eye movement and spatial memory in their task that compared spatial and verbal memory. The task involved a serial presentation of digits (random selection of 0 –

9, excluding 5), each presented within a spatial grid along with a distractor. The distractor was either of low similarity to the study items (a green triangle) or of high similarity ('5'). At test, participants either recalled the digits presented during study (verbal) or the location of items during study (spatial). Both high and low similarity distractors captured attention (measured here with eye gaze). However, while the only the high-similarity distractor decreased performance in the verbal task, both the high and low similarity distractors reduced performance in the spatial task. The authors believed that the increased impact of the distractor on spatial memory is due to spatial working memory sharing a cognitive representation with eye movements.

In line with this, Lange and Engbert (2013) found that participants are more likely to fixate their gaze on verbal as opposed to spatial stimuli in their working memory task. When participants were encouraged to make more fixations on spatial items, memory performance decreased. The authors believed that, again this result was due to a spatial representation shared by items. However, Czoschke, Henschke & Lange (2019) also found that fixations were less common on spatial compared to verbal items, this did not impact performance. The authors found this effect to be consistent regardless of whether the presentation type was sequential or simultaneous or whether the task allowed free recall of items, or items had to be recalled in a cued order. Taken together, it is unclear as to whether these studies suggest that eye movements interfere with memory or not. However, they do suggest that perhaps the connection between eye gaze and memory may not be as strong or as clear as could be assumed intuitively.

Surprisingly, there are not many studies that show connections to performance on memory tasks and measures such as time fixating on the target. Furthermore, there are studies which utilise eye-gaze in combination with tasks such as change-detection and partial report, it is less common in the study of recall tasks. This is of particular interest because it is these continuous-report recall tasks that are most often used for informing models of the capacity of visual working memory. As a result, the lack of eye-gaze data in these tasks mean we cannot integrate attentional aspects into computational models of VWM with the current available information.

Strategy in visual tasks

When discussing attention and encoding, it seems unavoidable that we discuss of strategy, especially given the long-lasting debates regarding top-down and bottom-up attention (e.g., Connor, Egeth, & Yantis, 2004; Egeth & Yantis, 1997). There do exist literatures on the strategies that participants employ when approaching VWM tasks, and these include consideration of which of these approaches yields the best performance (Cusack, Lehmann, Veldsman, & Mitchell, 2009; Vogel, Woodman, & Luck, 2005; Gazzaley & Nobre, 2012). These debates are similar to those in related fields, such as visual search (Boot, Kramer, Becic, Wiegmann, & Kubose, 2006; Over, Hooge, Vlaskamp, & Erkelens, 2007; Boot, Becic, & Kramer, 2009). Strategy is closely linked to attention, as it is seen as part of the link between the stimuli, encoding and task performance.

Within the domain of VWM, the study of strategy has largely consisted of behavioural methods. Vogel et al. (2005), for example, used a dual task to see how people deploy their attention. Specifically, they asked what people pay attention to and what they ignore. Generally, it was found that some people were able to filter out stimuli better than others and this impacted their effective memory capacity. In this study, strategy was said to be deployed based on participant preference – something that the authors attribute to IQ, a result that was also described by Cusack et al. (2009).

Cusack et al. (2009), compared performance in whole-report and change-detection VWM tasks. In both conditions, participants were shown an array of letters or colours to commit to memory. At test, participants were asked to either determine if the item at the cued location had changed since study, or were asked to recall all the items in the study array. Cusack et al. (2009) found that participants showed better memory performance in the whole-report condition, particularly for larger set sizes. Participants showed worse performance for the change-detection task, especially as the size of the study array increased. The authors found that participants' nonverbal IQ did not correlate with their estimate for VWM capacity, but it did correlate with the decreased performance in the change-detection task at higher set sizes. Cusack et al. (2009) interpreted this indicative of higher nonverbal IQ participants being aware of their capacity limits and adjusting their strategy to the task.

According to Vogel et al. (2005) and Cusack et al. (2009), one could assume VWM strategy depends on inherent features of the individual. However, this notion was challenged by Bengson and Luck (2016) in their study that manipulated task instruction. In a change-detection task, participants were either told to "do your best" (control); to try to remember the entire display even if there were a lot of items (remember all) or if there were too many items for them to remember, to just try to remember a subset (remember subset). It was found that participants were able to change their strategy based on the instructions given. Other studies have been able to demonstrate increased performance in VWM with training, suggesting that performance in these tasks is less stable and more related to attentional control (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Cowan & Morey, 2006). However this idea still remains controversial as other studies, such as Zhang and Luck (2011), were unable to encourage participants to make a quality/quantity trade off (that is, to move between slots-like and resources-like encoding).

Despite the ongoing controversy, Bengson and Luck (2016) suggest that their results were perhaps due to task differences. Donkin, Kary, Tahir, and Taylor (2016) reports results that support such an account. In their paper, the authors suggest that different task features, such as the presence of cues, the global or local nature of the task, and the predictability of the presentation of different set sizes may be factors that sway whether people behave as if memory is discrete and slots-like, or more fluid and resources-like. However, this issue is still open to more research. As such, determining the nature of strategy and its flexibility within VWM tasks is a key interest in this thesis.

Beyond the flexibility of strategy, there is also the question of optimal strategy. As well as finding that participants were able to respond to instruction, Bengson and Luck (2016) found that encouraging participants in the remember-all condition had better estimates of capacity compared to participants in the other conditions. This is in contrast to these results from Cusack et al. (2009) who found the opposite. With this large discrepancy, it is of interest to investigate which strategies are the most successful. In addition, both of the above studies used change-detection tasks and as there is already the suggestion that strategy is sensitive to task design, we are interested in investigating strategy performance with other task types.

Overall, there is evidence that suggests variations in VWM strategy between individuals. However, as the varying results of these studies suggest, this appears to be task sensitive. Of importance to our current project, it should be noted that few of these studies use a recall task meaning that the nature of strategy in these types of tasks is under explored. Additionally, while attention and strategy are discussed in VWM literature, few attempts seem to have been made to measure such things more directly – e.g., by using eye-gaze data. By contrast, study of strategy in visual search more often involves the use of eye-gaze data (e.g. Over et al., 2007; Boot et al., 2006). Indeed, when strategy is studied in the context of VWM, such studies often couple the typical aspects of VWM tasks with features from visual search studies. Woodman and Luck (2007) for example, use a dual task with a visual search and VWM component. They showed that when performing the search task, participants could actively resist fixating on distractors in the search task when they were the items they were told to remember. Similarly, a study in visual search by Dickinson and Zelinsky (2005) showed that visual search is guided by memory as participants performed equally well on search tasks in which items they had fixated on were marked compared to when they were not. Both experiments provide examples of memory guiding attention, which is a helpful link, but it does seem as if the VWM strategy literature is lacking in terms of the use of eye-gaze data.

The current project

In this thesis we aim to address these issues with two lines of enquiry. Firstly, we explore the encoding process in VWM. On a theoretical level, we believe the current models may diverge in their ability to explain different patterns of attention and encoding. Therefore, by examining this process we might be able to distinguish between VWM models and resolve the mimicry issue. To do this, we ran a series of recall tasks with eye tracking. Our aim was to use eye-gaze data as a proxy for attention and thus potentially find a link between encoding processes and the likelihood/precision of an item being in memory. Our hope was that eye-gaze data could be used to formulate or add to an existing model of VWM.

Secondly, we explored strategy use in VWM tasks. From previous studies, we can see that strategy can vary between individuals (Cusack et al. 2009, Vogel et al. 2006) and also with instruction (Bengson & Luck, 2016). Initially, we were particularly interested in factors that could cause a shift in strategy within a single task. Our reasoning for this is that it is possible that the mimicry seen between current models reflects the true nature of VWM. That is to say that people can switch between slots and resources strategies depending on the nature of the task. Therefore, the variable-precision model (Van den Berg et al., 2012) may in fact be the most accurate account of memory as within it, memory can appear to function as a set of discrete slots and as a flexible resource. Some evidence for strategy switching has been reported in other tasks (Bengson & Luck, 2016; Donkin et al., 2016) providing some support for this account. Our aim is to try to first demonstrate a switch in strategy by manipulating features of the task. Beyond strategy switching, we were interested in examining the variety of individual strategies for encoding within a task as well as differences between different VWM tasks.

Chapter 1: The effect of blocked set sizes on attention and strategy in an orientation recall task.

Beginning our investigation into eye-gaze, attention and strategy in VWM we found two questions to be particularly interesting. First, we wanted to know how participants allocated their attention in a standard VWM task. Second, we hoped to manipulate an aspect of the task so that participates would alter their attention strategy. This second aim, in particular, would help us establish the plausibility of using eye-gaze data to study encoding in VWM.

Due to the connection between eye-gaze, attention, and memory, we believed that we would be able to address our first question by measuring the connections between the stimuli participants fixate their gaze on and the stimuli they subsequently remembered. For example, it seemed likely that looking at a stimulus for longer would increase the chance of that stimulus being remembered. While this result may appear relatively benign, it could have potentially significant implications. If participants look at every item in a trial and trial times are of equal length, then naturally participants would have higher fixation durations on individual stimuli in a 3-item compared to a 6-item trial. Thus, if this were the case, such a finding could add an explanation as to how it is that participants perform better in trials with smaller set sizes. Therefore, by investigating how participants allocated their attention, we were hoping to expand on our knowledge of how stimuli are encoded in these tasks.

To address our second question, we wanted to determine if attention strategy could be altered such that, in general, participants attempt a different encoding strategy depending on changes in the task. Previous research has shown support for both slot-like models (e.g. Zhang and Luck, 2008) and for resource-like models (e.g. Alvarez & Cavanagh, 2004) within

VWM tasks. As suggested in the Introduction to this thesis, it is possible that because participants are capable of both modes of memory, and that participants might be encouraged into employing strategies that are more slots-like or resources-like depending on the nature of the task. This account is consistent with the variable-precision model, which gives the best account of performance in VWM tasks at present, in the sense that memory is flexible enough to encompass both slots-like and resources-like strategies (Van den Berg et al., 2012). We believe that differences in latent memory variables, such as the probability of an item being in memory or the precision of a memory, would likely be at least partly determined by the way in which they are encoded. Therefore, given the connection between eye-gaze, attention and encoding, with an appropriate task manipulation, we expected to see a change in general eye-gaze behaviour with a change in task demands.

Manipulating strategy switching

Donkin et al. (2016) suggested that participants may be able to switch their memory strategy depending on features of the task. Specifically, if people know how many items they will be presented with, they are more likely to use resource-like memory compared to if they do not know the set size of the next trial. One explanation for this is that if participants know the set size of the next trial, they can prepare for a trial and divide their attention more equally between all the items. Having memory resources more evenly distributed between all items is more characteristic of a resources model as opposed to a slots model, wherein a subset of items are encoded with high precision.

In their study, Donkin et al. (2016) compared slots-based and resources-based models on data from four old experiments and two new experiments. Of these two new experiments, one contained an “unblocked” set size in which participants were randomly presented with n = 2, 4, 6, or 8 set size trials. The other experiment contained “blocked” set sizes in which each block contained set sizes of only n = 2, 4, 6, or 8. Compared to the unblocked set-size experiments, participants in the blocked experiment appeared more likely to use resource-like encoding (Figure 1.1). By contrast, in experiments with unblocked set size, participants consistently gave results that were better accounted for by a slot-like encoding.

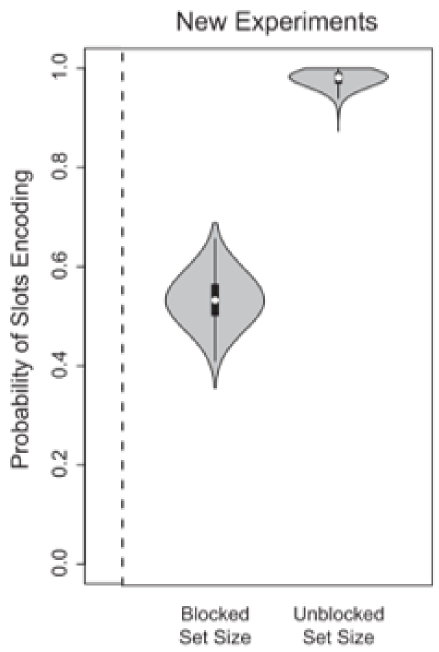


Figure 1.1. Results from Donkin et al. (2016) depicting the likelihood that participants used slots-like compared to resources-like encoding in the blocked and unblocked (new) experiments.

The suggestion made by the authors is that VWM may be more flexibly applied than previously thought. Perhaps if people know the number of items presented on a trial, they will attempt to remember all items instead of focusing on a few. This would increase the chance of an item being in memory, but could lower precision on blocked trials relative to unblocked trials and thus following a more resources-like pattern. While the behavioural data are seemingly consistent with this suggestion, there is currently no attentional data to back up this claim. Running blocked and unblocked experiments with eye tracking would provide more data about where people are allocating their attention during trials. It would therefore be possible to determine if eye movements differ between blocked and unblocked trials, which would demonstrate that overt attention is (at least partly) responsible for the change in strategy used in these tasks.

Experiment 1: Standard VWM with blocking and eye-gaze

In this first experiment, our aim was to begin exploring how items are encoded in VWM by further testing Donkin et al.'s (2016) claims that participants may change their memory strategy depending on whether the set size used was blocked or unblocked. The task used was an adaptation of the colour recall task used by Zhang and Luck (2008). Eye tracking was used in conjunction with this task to attempt to elicit a connection between eye-gaze and the probability and precision of an item being in memory. To test Donkin et al.'s (2016) claims, there was a between subjects condition where participants experienced either a blocked or unblocked version of the task.

While the general task was similar to Zhang and Luck's (2008), the stimuli were altered with the hope of making them more complex. The colour-square stimuli used in the standard experiment are very simple to encode (Eng, Chen, & Jiang, 2005). As a result, a participant may be able to encode items quickly without having to attend to them for very long. It was thought that making stimuli that are more complex would encourage longer fixations and thus yield more heterogeneity in behaviour across trials. The benefit of using coloured squares, however, is that they are easy to reproduce by means of a colour wheel. For the current task, more complex stimuli were desirable, but we also wanted to have stimuli that could be reproduced to form a continuous scale of answers.

For this experiment we used a “ring” set of stimuli. Shown in Figure 1.2a, these stimuli consisted of a coloured ring with a “bead” placed randomly on the circumference of the ring. For the recall task, participants were asked to place the bead on the given ring in the location it appeared during study (Figure 1.2b). These stimuli were (supposed to be) visually more complex while still being able to be reproduced with a continuous response choice.

For this experiment it was predicted that 1) similar to Donkin et al.’s (2016) results, we would find an increased probability of resources-like responding in the blocked condition of this experiment. We would see this as an increase probability of an item being in memory and a decrease in precision, relative to the unblocked condition. Specifically, the difference between the blocked and unblocked conditions would be noticeable for set size 6 and not set size 3, because most existing theories imply that people can easily encode all items in set size 3, but begin to diverge for set size 6. 2) Accompanying this observed change in attentional strategy in the blocked condition, we expect to see more fixations and a shorter average duration of fixations in this condition compared to the unblocked condition. We believe this because of the previous evidence which suggests participants have a more resources-like approach to the task when set sizes are blocked (Donkin et al. 2016) and that a resources account could be categorised by a spreading of attention across all items in the display, rather than focusing on a few items for longer.

Method

Participants. 40 participants were recruited from the paid UNSW sign up system SONA to complete a single one-hour session. Participants were screened during sign up for the study to agree that they had normal or corrected to normal vision. Participants were paid

$15 for their time in exchange for participating. Approval for this study was obtained from the UNSW Human Research Ethics Approval Panel – C (HREAP-C).

Apparatus. A Tobii TX300 eye-tracker, with 300 Hz temporal and 0.15° spatial resolution, mounted on a 23-inch. widescreen monitor (1,920 x 1,080 resolution, refresh rate 60 Hz) was used. Participants' heads were positioned in a chinrest 60 cm from the screen.

Stimuli. The stimuli (Figure 1.2a) were coloured rings with a small, filled circle placed somewhere on the circumference of the ring (J. Rouder, Thiele, & Cowan, 2014; Ricker & Hardman, 2017). The stimuli could be one of eight distinct colours (red, yellow, green, cyan, blue, magenta, brown or salmon pink) and were presented on a grey background. Each stimulus had a ring with a fixed diameter of 120 pixels (visual angle = 3.03°) and a bead with a fixed diameter of 20 pixels (visual angle = 0.51°). The ring had a thickness of 2 pixels (visual angle = 0.05°). Stimuli were presented at random locations on the circumference of an invisible circle with diameter 600 pixels (visual angle = 15.1°). As such, each item was equidistant from the centre of the screen. Items were also restricted such that the angle between them (as measured from the centre of the invisible circle) was equal for all items. The location of the bead on the ring was randomised for each item on each trial. To cue the position of the target to be reproduced, the stimuli would be presented without the bead – i.e. just a coloured ring – in the location in had been presented in the display.

Design. Experiment 1 was a recall task, very similar to Zhang and Luck's (2008) design. The experiment contained 420 trials divided into 14 blocks of 30 trials. There were either N = 3 or 6 items in the study array on each trial. In the unblocked condition this was randomised within each block, with an equal amount of 3 or 6 item displays in each. In the blocked condition the set size was blocked such that the first half of the experiment consisted of trials of all one set size and the second half consisted of only the other set size. The order in which set sizes were presented was randomised across participants.

Procedure. A fixation cross was presented for 500ms at the start of each trial. This was followed by a black screen for 400ms. The study array of *N* gems would then be presented for 1000ms. This was followed by a mask that lasted for 200ms and then a blank screen for 500ms. The participant was then presented with one of the rings they saw in the trial (same location and colour) and was asked to place a bead on the ring at the location it appeared on during the experiment. The participant used the mouse to indicate where they believed the bead was and used the spacebar to confirm their selection. Participants were then given feedback on their selection lasting 1000ms. Their deviation from the "true" bead location was given in degrees as well as some written feedback ("OUTSTANDING!" for deviations less than 10°, "Very good!" between 10° and 20°, "Good" between 20° and 35°, "OK" between 35° and 45° and deviations greater than 45° were labelled with "Poor"). The trial sequence is depicted in Figure 1.2b. After each block, participants were given a break in which they rested for a minimum of 20 second before continuing.

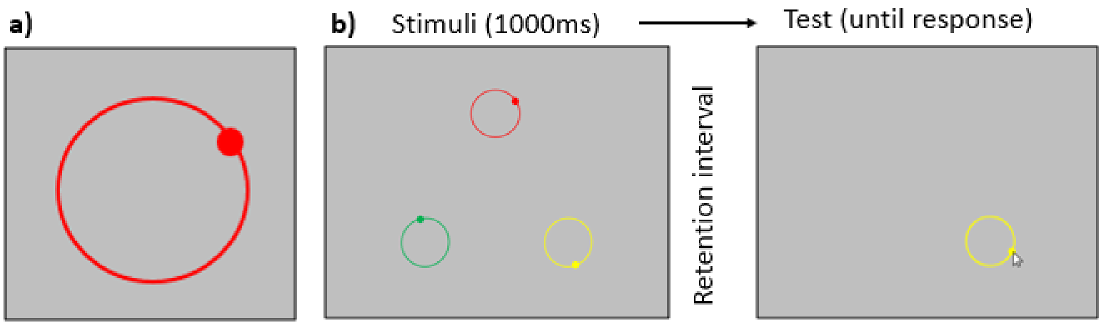


Figure 1.2. a) An example of the bead and ring stimuli used in experiment 1. Stimuli used varied in colour and in location of the bead on the ring. b) The sequence of events in a trial. 3 or 6 differently colour stimuli were presented on a grey background for 1000ms followed by a retention interval (mask then blank screen) of 700ms. Participants were

Modelling procedure. We applied a model to the data to allow us to compare the probability of an item being in memory (P_m) and the precision of memories (*Prec*) between the blocked and unblocked conditions. The model was a Bayesian hierarchical version of the Zhang and Luck (2008) mixture model. The model assumed that the deviation between each response made in the experiment and the correct response either came from memory or from a separate guessing process. Responses based on memory were associated with Von Mises distributions (which are normal distributions, wrapped around a circle) with a mean that was centred around the correct response and a precision that varied depending on condition (blocked and unblocked), set size (3 and 6) and individual participant. Responses based on guessing were uniformly distributed around the circle for all conditions and all participants. The model allocated responses to either memory or guessing process by taking a value from a Bernoulli distribution with a probability of using memory equal to P_m. P_m, like *Prec*, also varied with condition, set size and individual participant. This resulted in four values being calculated for each participant, P_m and *Prec* for set size 3 and set size 6 (remembering that blocked and unblocked conditions are between subjects). Individual-participant level parameters, P_m and *Prec*, were constrained such that they came from their own population-level normal distributions (i.e., one for each parameter in each set size and blocked/unblocked condition). We focus on the eight population-level posterior distributions of P_m and *Prec* in order to compare behaviour across the four conditions of our experiment.

Results

To prepare the data for analysis, trials for which no eye-gaze data was collected were removed (544 trials or 3.24% of trial data). Trials were also removed if they reported more than 10 fixations during the 1000ms presentation window (964 trials, 5.74% of the data) or if the average fixation duration in a trial was less than 100ms (2703 trials, 16.09% of the data). For a fixation to be indicative of encoding or cognitive processing, it is generally thought that it needs to be longer than 100ms (Liversedge et al., 2004; Rayner, Smith, Malcolm, & Henderson, 2009)

Behavioural results. For each trial, the deviation between the participant's answer and the true location of the bead was recorded. As the range of answers varied around the circumference of the circle, the deviation was expressed in radians (3.14 radians = 180 degrees). Figure 1.3 shows the frequency distribution of deviations for set sizes 3 and 6 (green and red lines respectively) for unblocked and blocked set sizes (left and right panels respectively). In both conditions, we saw a similar pattern of responses to results from previous tasks (e.g., Zhang & Luck, 2008) with a large proportion of the responses clustered around the correct response for both set sizes but with more accurate responses for set size 3 (see Figure 0.1b for comparison).

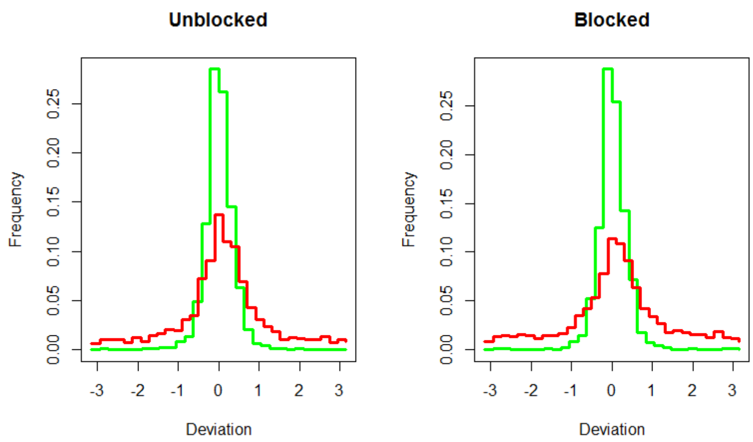


Figure 1.3. The frequency of responses by deviation from actual bead location for the unblocked and blocked conditions. The green line represents set size 3 the red line represents set size 6.

Modelling results. Figure 1.4 shows plots of the population-level posterior distribution for P_m and *Prec* across condition and set size. There was no difference in P_m values for set size 3 between conditions. There was a slight indication of a difference between the unblocked and blocked conditions for set size 6, with smaller P_m values in the blocked compared to the unblocked condition. *Prec* values appear to vary across set size, with higher precision in set size 3 compared to 6. However, there was no observable difference between the blocked and unblocked conditions.

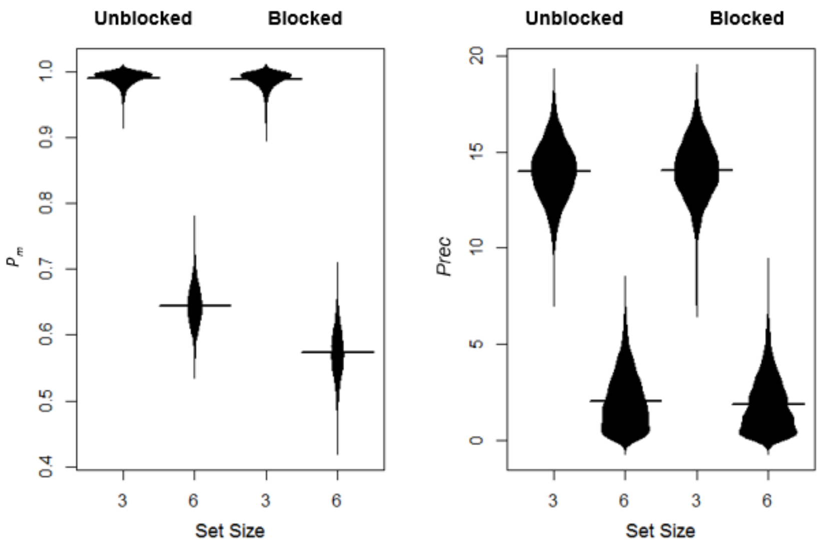


Figure 1.4. Posterior distribution for P_m and *Prec* parameters in blocked and unblocked conditions for both set size 3 and 6. Horizontal lines show the mean of the distributions.

The differences in P_m and precision values between conditions for set size 6 only are presented in Figure 1.5 using smoothed frequency distributions. The difference between the posterior distributions for *Prec* is centred on zero suggesting little difference in precision for set size 6 between conditions. The plot of P_m difference shows higher values for P_m in the unblocked condition compared to the blocked condition. However, this difference is small. Since an appreciable mass of the posterior distribution still surrounds zero, there is relatively little evidence of a difference between the conditions. Furthermore, the difference in P_m trend in the opposite direction of that we expected.

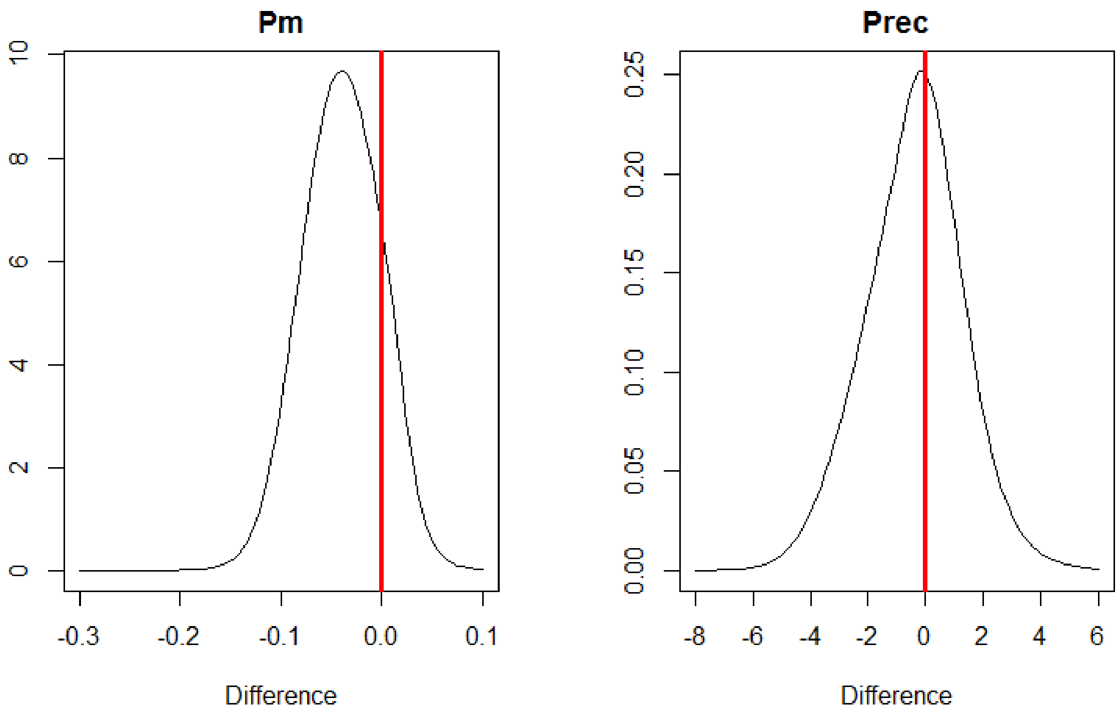


Figure 1.5. Difference in the posterior distributions of P_m and *Prec* between the unblocked and blocked conditions (for set size 6 only).

Eye-gaze results. The unblocked and blocked conditions were compared using the average fixation duration per trial and the average number of fixations for each set size. The mean values for each of these measures, for each condition is plotted in Figure 1.6. On average, more fixations with less fixation duration was seen in the unblocked condition compared to the blocked condition.

We carried out a two-way analysis of variance (ANOVA)[1] of set size and condition on average number of fixations, grouped by participant. There was no significant difference in the average number of fixations between set sizes ($F(1,37) = 0.376$, $p = 0.543$) or between

[1] In this thesis we frequently use ANOVA as part of our statistical analysis. However, this method is not the ideal form of analysis of our data. Time constraints meant that this method was the best to use with the available time, however, as this method is not ideal, results should be interpreted with caution."

the blocked and unblocked conditions ($F(1,37) = 1.531$, $p = 0.224$). A similar analysis conducted for average fixation duration revealed no difference in average duration between the set sizes ($F(1,37) = 0.196$, $p = 0.660$) or between the conditions ($F(1,37) = 1.206$, $p = 0.279$). This suggests no evidence of a difference in the average number of fixations nor in the average fixation duration between trials of difference set size and blocked or unblocked condition.

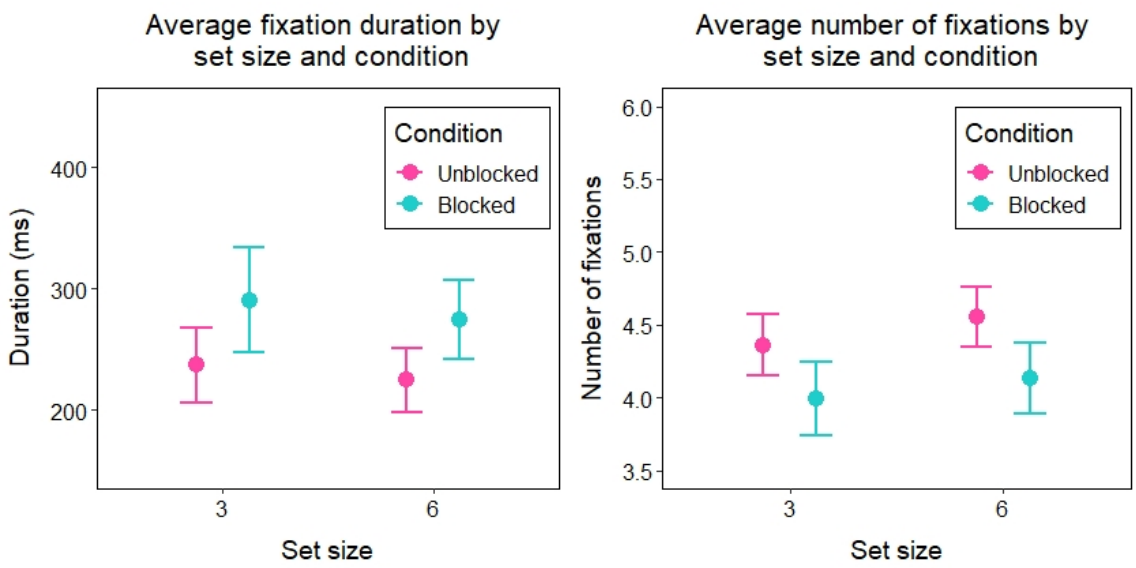


Figure 1.6. Average number of fixations per trial and average fixation duration per trial for the blocked and unblocked conditions for set sizes 3 and 6. Error bars indicate the standard deviations.

As well as the blocked and unblocked conditions, we also examined the effect of fixation on task performance. We examined accuracy as a function of whether participants looked at the target during study or not. Accuracy was expressed in distance (in radians) from the correct answer (i.e. the true location of the bead compared to where participants believed it to be). A plot of the distribution of distance by whether a target was fixated or not is shown in Figure 1.7a for both set size 3 and 6. Examining Figure 1.7a, we found little difference in accuracy as a result of fixation for set size 3, and some benefit of fixation on accuracy in set size 6.

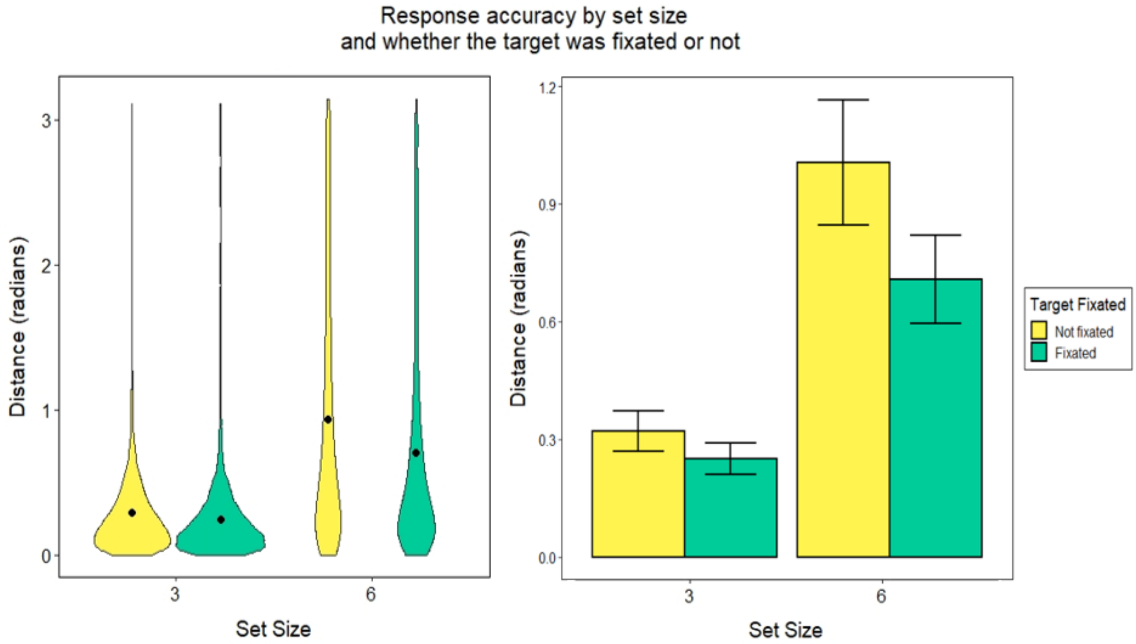


Figure 1.7. a) The distribution of distance (the absolute value of deviation from the correct responses) scores for when the target was fixated during the study interval and when it was not for both set size 3 and 6. b) Mean distance for whether or not the target was fixated and for set size 3 and 6. Error bars are standard error.

Following this observation, we ran a two-way (2x2) ANOVA on absolute value of deviation from the correct response (distance) by set size and whether participants fixated the target or not, grouped by participant. Figure 1.7b presents the means and standard errors of distance for each set size and for whether the target was fixated or not. Responses were significantly more accurate in set size 3 compared to 6 ($F(1,37) = 316.08$, $p < 0.001$) and when the target was fixated compared to when it was not ($F(1,37) = 36.70$, $p < 0.001$). There was also a significant interaction such that the effect of fixation was greater for set size 6

($F(1,37) = 21.72$, $p < 0.001$). This finding, suggests a connection between target fixation and memory performance which in turn is reason to explore the connection between eye-gaze variables and memory performance more thoroughly. We have limited the scope of this chapter to outlining our initial results in this regard, preferring to focus on the issue of manipulating attention. The analysis of the connection between eye-gaze variables and memory is continued in Chapter 3 (page 77).

One aspect of eye-gaze that is worth noting, however, is that many trials within the data were categorised as having no fixations on the target. To examine this further we divided all the fixations across the experiment by the location of the fixation. Fixations were coded as being on the centre of the display, on one of the six items in the display, or at some other location on screen. A fixation was regarded as being at a certain location if it was located within a 160-pixel radius (visual angle = 4°) of the location's centre coordinates. For fixations on items, this area would include the stimulus itself as well as ring 40 pixels (visual angle = 1°) as a margin of error. Across the experiment, the largest proportion of fixations were on the centre of the display as shown in Figure 1.8. Fixations were evenly distributed between the other items in the display.

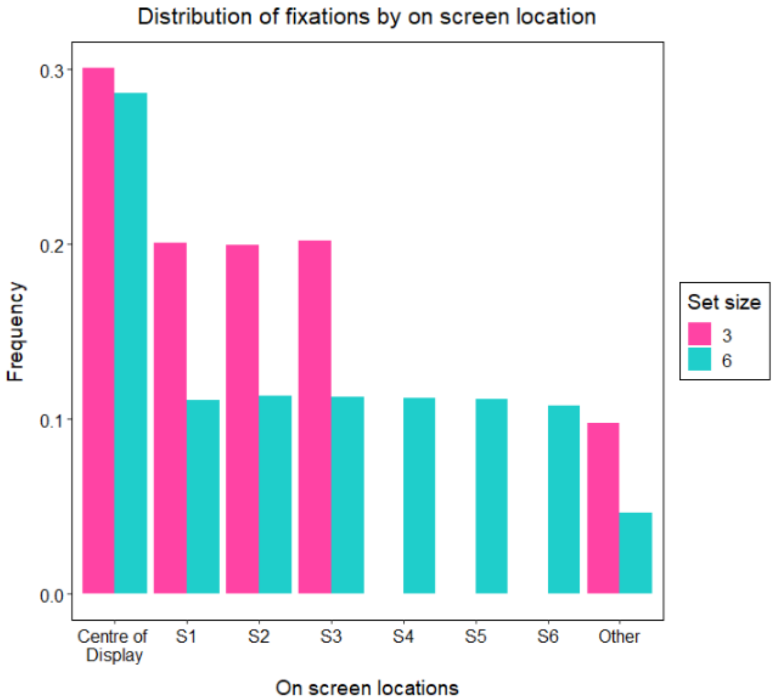


Figure 1.8. What participants fixated on during the experiment as a proportion of the total amount of fixations. The most frequently fixated location was the centre of the display. The other options were the study stimuli. Note, the physical location of items varied between trials. Items 4, 5 and 6 were only presented during 6-item trials.

Discussion

Overall, the behavioural results for the blocked and unblocked conditions were similar. The modelling results indicated similar memory strategies being used in the different conditions. This was indicated by similar values for the probability of an item being in memory and for precision in both conditions. Following from this, there was little difference in the attention patterns demonstrated through eye-gaze. Although there was a suggestion that there was a higher probability of an item being in memory, more eye-gaze fixations, and shorter fixation durations in the unblocked condition. This trend is counter to our prediction that the blocked condition would have a higher probability of items in memory, more fixations, and shorter average fixation duration. However, given the lack of significant behavioural distinction between the blocked and unblocked condition, we are reluctant to attribute any eye-gaze differences as an indication of strategy difference. In regards to this manipulation, there isn't any evidence to suggest a change in strategy between unblocking and blocking conditions.

It is possible that using a recall task instead of a change-detection task may be the reason why we did not observe a blocking effect. In their study, Donkin et al. (2016) investigated whether VWM is more slots-like or resources-like for a variety of task manipulations – blocking and unblocking of set size being one such manipulation. Their study used data from change-detection tasks, for which the majority of manipulations returned modelling results more in line with a slots model account. The implication of this is that slots-like encoding is the preferred memory strategy for change-detection tasks with some manipulations, such as the blocking condition, presenting as an exception by showing a preference towards resources like encoding. This idea that the slots-like encoding is what is used in change-detection tasks is supported by the study from Bengson and Luck (2016). These authors found that they could manipulate participants' strategies in a change-detection task based on the initial instructions given. However, in the condition with neutral instructions (in which participants were told to "do their best") participants tended to show behaviour more in line with a slots account by encoding a subset of the items presented in each trial. This suggests that the default behaviour in a change-detection task is more slots-like. By contrast, our current task showed evidence of not entirely being in line with the slots model. In both conditions, precision varied between set sizes – a pattern that is more parsimonious with a resources account. If our current task is not particularly in line with the slots account then seeing a difference between the blocked and unblocked conditions is less likely than within a change-detection task. This is because there is less room to see a difference made by this blocking manipulation. In order to determine if the design of our task is the cause of the lack of blocking effect, it would be useful to run a change-detection task with eye tracking, to see if the results from Donkin et al. (2016) were able to be reproduced. It is also possible that this effect is limited to change-detection tasks for reasons other than a difference in memory strategy. If this were to be the case, then it demonstrates that this blocking effect is limited due to lack of ability to generalise to other tasks. Running a change-detection task with eye tracking may also provide insight into this question.

As mentioned, the modelling results found in this task were somewhat unusual compared to the standard results. Previous recall tasks, similar to the current experiment show a preference for slots-like encoding (e.g. Zhang & Luck, 2008). However, we found both a change in the probability of an item being in memory with set size and well as a change in precision with set size. As set sizes increases, a slots account predicts that the probability of an item being in memory decreases, while the precision of the remembered items remains constant. By contrast, a resources account predicts that the probability of an item being in memory remains constant, while precision decreased. Our results, therefore, do not reflect one model entirely. It is worth noting that we were able to reproduce the same pattern of parameter estimates using maximum-likelihood estimation, ruling out the possibility that our results were due to the use of a Hierarchical Bayesian model. We also reproduced and tested our code with recovery studies to make sure there were no problem with code, and found no cause for these result within our fitting method.

The decrease in memory precision with set size is incompatible with the slots model. However, why our results differ from very similar recall tasks is unclear. It is possible that the stimuli in our tasks might be the cause for difference. This could be due to emergent stimulus properties that might lead participants to encode the positions of beads relative to one another, for example, rather than relative to each bead's ring. Given that, as discussed above, change-detection appears to be more slots-like by default, perhaps a change-detection task with these current ring and bead stimuli would create a helpful comparison. Another possible change from the standard VWM task is our inclusion of eye tracking. For various reasons (such as calibrating the tracker) participants are made aware of eye tracking. It is possible this could alter participant behaviour, although it is unclear why this would be. While this task is very similar to other recall tasks, there are some differences, such as the stimuli or their placement (in a circle) that may have made some impact on participant behaviour.

Our analyses suggested a preliminary link between fixation on a target item and increased task accuracy. One connection we were interested in exploring was examining the connection between eye-gaze variables and the inferred precision and the probability of an item being in memory. We attempted to include eye-gaze variables into the Bayesian hierarchical model used in our analysis. One such variation defined whether the target was fixated or not a as a factor effect the probability of an item being in memory. However, we were never able to make these models successfully run. While we were not able to make this modelling work here, joint models of eye-gaze and VWM performance should remain of interest to the field.

A reason for our lack of success in modelling eye-gaze appears to come from the nature of the eye-gaze data itself. In examining this data, it was clear that the target was not fixated on many more trials than it was. Participants seem to prefer to fixate the centre of the display more any given stimulus location. This finding gave us insight into our results – the lack of eye movement in general may be a reason for only small improvements in accuracy when the target is fixated. This small change to accuracy may have not been able to properly inform our attempted VWM models that included eye-gaze. Further analysis of eye-gaze data and memory variables from this experiment can be found in Chapter 3, where it is compared more directly with results from Experiment 2. We decided to defer the reporting of some of the results from this experiment to Chapter 3 when we saw similarities between the results of the two experiments. As a result, the discussion of Chapter 3 provides a more integrated summary of the results of this experiment.

Overall, the lack of eye movement impacted our ability to analyse the data from this experiment. Compared to tasks of verbal information, fewer eye movements are observed in spatial tasks (Lange & Engbert, 2013; Czoschke et al. 2019). An explanation for this comes from studies which find that eye movements interfere with spatial memory due to perhaps a shared spatial representation between the systems. (Lange et al., 2012; Lawrence, Myerson, & Abrams, 2004; Lawrence, Myerson, Oonk, & Abrams, 2001; Postle, Idzikowski, Sala, Logie, & Baddeley, 2006). Under this interpretation, the rarity of eye movements is to be expected as participants are keeping their eyes fixated on the centre of the screen to avoid interfering with the spatial representations in their memory.

However, it is possible that the cause of this might be a problem with our task design. It is possible that our use of an orientation recall task, as a departure from the more standard colour recall task may have had an impact on our results. By varying the stimuli in the next experiment, we may determine whether this effect is unique to the stimuli used in the current experiment. As another example, one of the problems with the eye-gaze data is perhaps that there is not enough distinction in where people are looking (their fixation locations) and how long they are looking at stimuli. In this task, we suspect it is possible for participants to encode more than one stimuli in a single fixation, since the relative simplicity of the stimuli permit quick encoding into memory, meaning that participants needed less, or only peripheral attention to encode these items successfully. To remedy this, we planned a second experiment with more spatial separation in items as well as more complex stimuli that we hoped would capture overt attention.

As a result of the blocking/unblocking manipulation not working, we were not able to see any evidence of strategy switching. It is possible that this could also be related to our reasons for not producing good eye-gaze data. If not moving one's eyes was a favourable strategy in this task, then there indeed may be no need to switch behaviour due to a manipulation. Given the original finding of a strategy switch with blocking/unblocking (Donkin et al., 2016) has not been confirmed, we thought it best to look at a more robust behavioural measure that might map on to variations in strategy for our next experiment.

Chapter 2: Attention and serial position effect with small set sizes

One of the main challenges with our first experiment was that we were unable to see a difference in eye-gaze behaviour between the presentation conditions (blocking vs. unblocking). The reason for this is potentially simple: there was no difference to be seen. That is, contrary to the findings of Donkin et al. (2016), there is no difference in strategy or behaviour based on whether set sizes are blocked or not. If this is the case, we would therefore predict there to be no difference in eye-gaze behaviour between these conditions; just as we found. However, by not obtaining to shift in eye-gaze behaviours that we expected between conditions, we were unable to see a connection to eye-gaze and different strategy styles. This means that not only does strategy switching remain to be demonstrated, but more generally, we were unable to produce a robust connection between eye-gaze variables and memory.

While we found a significant effect of fixation on the accuracy of target recall, this effect was not large. This is rather counter-intuitive, as foveating the stimulus has been thought to align with better encoding (Livingstone & Hubel, 1988). Instead, we found participants commonly looking at the centre of the display, rather than on any of the study items. As discussed, features of the display perhaps discouraged eye movements, such as items being relatively close together and to the centre of the screen. As visual acuity is best at the centre of fixation, when items are spread further apart are harder to view together if the eyes are kept still. As such, with more distance between items in the display, participants may be motivated to make more fixations. Stimulus complexity was another such feature that could have minimized the role of eye movements in the experiment. If our stimuli were too simple, they could be encoded quickly, perhaps by using peripheral vision to process them in parallel (Treisman & Gelade, 1980). Increasing stimulus complexity could therefore encourage fixation as participants would need to process items serially.

In our second Experiment, we wanted to address these areas of concern by finding a task that involved more spatially distant and more complex stimuli as well as a task that was already associated with attention effects to maximise the chance of observing eye-gaze effects on encoding and memory.

The serial position effect in Smith et al. (2016)

Smith et al. (2016) aimed to replicate results found by Sewell, Lilburn, and Smith (2014), who demonstrated that for small set sizes (n up to 4) memory performance could be explained by the sample-size model. This model states that VWM draws independent evidence samples of each stimulus. The original version of the model assumes that information about the relevant properties of any presented stimuli is sampled at a constant rate, with attention divided evenly among all items. Given a set duration, a large but fixed number of samples are drawn, making memory in this model classified more as a continuous resource. The appeal of this model lies in its simplicity – it is parameter-free. However, by replicating this experiment with more complex stimuli, Smith et al. (2016) found that the sample size model no longer gave a good account of behaviour.

Sewell et al.'s (2014) task involved orientation judgements. Participants were presented with a set size of n = 1,2,3 or 4 sinusoidal luminance gratings (Gabor patches) and following a retention interval, had to recall the orientation (vertical or horizontal) of one of the items (randomly selected) at study. Smith et al. (2016) altered the task to be a phase discrimination judgement. The stimuli were Gabor patches that were half black, half white (divided vertically) and participants were asked to recall which side was black in a randomly chosen item from the study display. This choice was based on comparisons made by Thornton and Gilden (2007) who showed phase discrimination to be a more demanding, serially processed task compared to orientation judgments which was said to be an easier, more parallel processed task. Smith et al. (2016) decided to test the sample size model with more complex stimuli as the model was claimed to be invariant regarding stimulus complexity. However, the results of their experiment showed that the increased stimulus complexity was associated with results not predicted by the sample size model. If stimuli cannot be encoded in parallel, it appears that attention must be focused on one item at a time. Given that each trial has limited duration, if the encoding time of a stimulus is more than the trial duration divided by the number of items in the display, then necessarily, not all items can be attended equally. Thus, an even distribution of attention between items (as assumed by the sample-size model) seemed less likely if using stimuli that demand more encoding time.

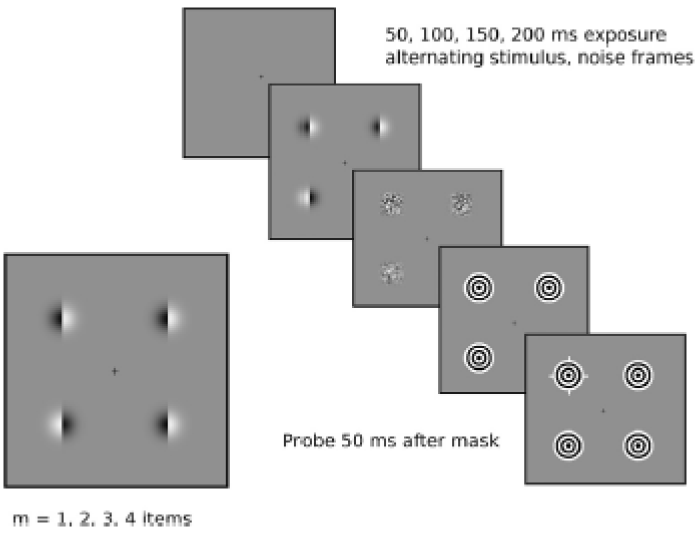


Figure 2.1. The stimulus (left) and the trial sequence (right) as shown in Smith et al., 2016

In their first experiment, Smith et al. (2016) used a simultaneous presentation of the display items in each trial. Like Sewell et al. (2014) they used n = 1, 2, 3 or 4 items in each trial. Each display was presented for 50ms, 100ms, 150ms or 200ms. A trial sequence is shown below in Figure 2.1.

The results show that the sample-size model consistently underpredicted the reduction in accuracy as a result of increasing set size. This was true for both group and individual data. Smith et al. (2016) referred to this as the "excess load effect". Based on this result (as well as others, less relevant for our purposes and so not reported here), Smith et al. (2016) ultimately concluded that the sample size model was incompatible with this phase discrimination task.

To determine the cause of the excess load effect Smith et al. (2016) ran a second experiment, which used the design of the first experiment but with a sequential rather than simultaneous presentation. Like the first experiment, this experiment failed to be accounted for by the sample-size model. This was contrary to findings by Sewell et al. (2014) who found that the sample-size model held for both sequential and simultaneous presentation conditions. Smith et al. (2016) examined task accuracy as a function of serial position and found that task accuracy was much greater for the first item displayed in the sequence compared to the other items in the display. The authors adjusted the sample-size model to take this "serial position effect" into account by including attention weightings. The first item in an m item display presented for t ms received some share of the VWM resources, $c_a(m,t)$. The following items in the display received an equally divided share of the remaining resources, $[1 - c_a(m,t)]/(m-1)$ (Smith et al., 2016). With this adjustment, the model predicted task performance in experiments with both sequential and simultaneously presented items more accurately.

The current experiment

This finding by Smith et al. (2016) is interesting for a number of reasons. Firstly, the authors conclude that the serial position effect is due to one item in the display (the first in a sequential presentation) capturing more attention than other items, with remaining attention being evenly distributed between other stimuli. However, their study makes conclusions about attention by averaging behaviour over the experiment. It is possible that the explanation given by the attention-weighted sample size model could be reflective only of what is occurring on average, but not necessarily during individual trials.

In particular, the attention-weighted sample-size model claims that after one item is attended, the remaining attention is distributed evenly between all other items. Indeed, averaging over trials, one sees that all not-initially-fixated items are remembered roughly equally well. However, such an aggregate pattern could arise for different reasons. For example, consider a four-item display in which a participant attended the first item they saw (as the model suggests), but then spent their remaining attention on just one of the three remaining display items. This leaves two items not attended at all. If a participant repeats this process on each trial, choosing a second item to fixate randomly, then the aggregate behaviour is also in line with what was observed. By having a second measure of attention – participants eye-gaze – we may be able to better assess claims made using the attention-weighted sample size model. Therefore, the first reason we were interested in recreating Smith et al.'s (2016) experiment with eye tracking to determine was to see if the model accurately describes trial-by-trial behaviour.

Additionally, the type of eye-gaze behaviour displayed would distinguish between different accounts of VWM. Smith et al. (2016) classify their attention-weighted sample size model as a resource model of memory. This statement implies that all of the display items are held in working memory and the accuracy of their recall depends on the precision of the memory. The attention weights imply that the more attention given to an item, the more precisely it is remembered. Importantly for this model, it is the first item that is given this benefit. Using eye-gaze as a measure for attention, we believe this model would predict that the degree to which items are remembered is the degree to which they are fixated, with items that are fixated longer being better recalled. To fit with Smith et al.'s (2016) account, we would therefore expect the first item to be fixated the longest with other items being fixated for shorter durations. Relative to each other, the items subsequent to the first item would be fixated on for similar durations. This would describe the pattern of results see by Smith et al. (2016). By contrast, a slots model would predict that rather than subsequent items being remembered less precisely, only a subset of these items is held in memory. As mentioned above, the results from Smith et al. (2016) that imply that attention is equally distributed amongst the items could be an artefact of averaging. The slots model would suggest that this is the case and that beyond the first item, one other item might be remembered. Therefore, by examining the duration and number of items fixated on in a trial we will also be able to distinguish between a slots and resources account.

Finally, the serial position effect displays a strong primacy effect that is associated with attention capture. Our hope here is to use a more reliable effect in order to determine whether eye-gaze maps onto task performance. Primacy effects are a much more robust finding in working memory studies, and are therefore more suitable for this purpose (Oberauer et al., 2018). Additionally, the serial position effect is described by Smith et al. (2016) as being caused by attention capture. This type of phenomena combined with its magnitude suggests that we should see eye movements lining up with the first item presented in sequential displays and this should be associated with better accuracy on the memory task.

Method

Participants. 6 participants were recruited from the University of New South Wales cognition lab and through word of mouth. Participants were paid $20 an hour for their time and each completed 12 hours total (total remuneration of $240). The number of participants was chosen to match that of Smith et al. (2016). Approval for this study was obtained from the UNSW Human Research Ethics Approval Panel – C (HREAP-C).

Apparatus. A Tobii TX300 eye-tracker, with 300 Hz temporal and 0.15° spatial resolution, mounted on a 23-inch. widescreen monitor (1,920 x 1,080 resolution, refresh rate 60 Hz) was used. Participants' heads were positioned in a chinrest 60 cm from the screen.

Stimuli. The stimuli were outlines of pentagons with one side missing as shown in Figure 2.2. There were 5 different variations on the stimuli – all were pentagon outlines but each with a different side missing. The stimuli were 120 x 120 pixels in size (visual angle = 3.03°) and were presented on a grey background. There were four locations stimuli could be presented, which were the centre of each quadrant of the screen. That is, 480 pixels (visual angle = 10.85°) to the left or right of the centre of the screen and either 270 pixels (visual angle = 8.53°) above or below the centre of the screen.

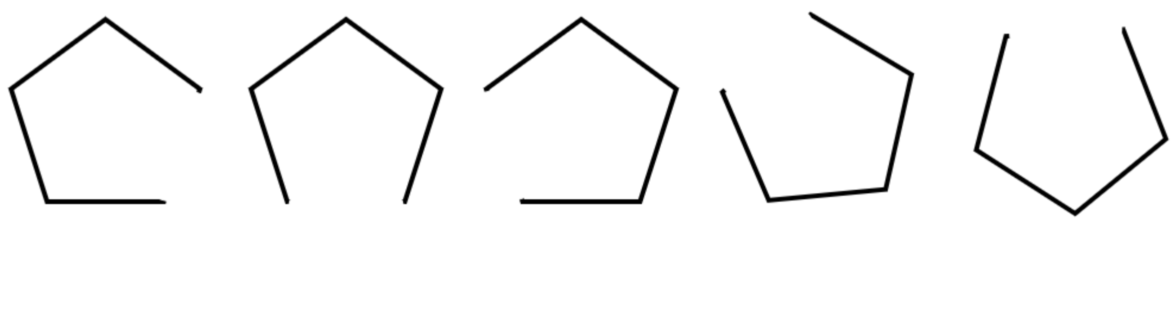


Figure 2.2. The stimuli used in the current experiment. Each was a pentagon of identical size, with a different side missing.

Design. The independent variables in this task were set size, presentation type (sequential or simultaneous) and presentation duration (trial time = 200ms or trial time = 200ms per item). A table showing the different conditions is shown in Table 1. This was a completely within-subjects design with all participants experiencing all conditions. The experiment was conducted over 12 sessions with 4 sessions dedicated to each presentation-style condition. Sessions were blocked in pairs such that participants experienced 2 sessions of the same condition before swapping conditions. Apart from this, the order in which participants experienced the conditions was randomized.

Our presentation durations differed from the ones used by Smith et al. (2016) for three reasons. Firstly, we wanted to minimise numbers of conditions so we did not want to use all the duration conditions used in the original study. Secondly, to maximise eye-gaze data we wanted to choose the longest presentation durations only. Thirdly, we were not entirely clear how presentation duration was timed in the sequential condition. The original study stated that in the simultaneous condition, items were presented for a net trial time of 200ms regardless of the number of items in the display. However, we were not sure whether in the sequential condition items were presented for 200ms per item or if the net presentation time of all items in the trial was 200ms. In order to be most generous with durations to maximise eye-gaze data we chose to present items for 200ms each in the sequential condition. Therefore, the total duration in this condition ranged from 200ms ($n = 1$) to 800ms ($n = 4$). We ran one simultaneous condition to match the timings in the sequential condition (the "time per item" condition). All stimuli were presented on screen together, but the duration of the display varied with the number of items (as in the sequential condition above). Our final condition was the simultaneous condition that matched the original Smith et al. (2016) experiment. Regardless of the number of items in the display, all items were presented together for 200ms (the "net time" condition).

Table 2.1

Condition variations in Experiment 2

Presentation type	Duration type	Set Size	Trial duration
Sequential	200ms per item	1	200ms
		2	400ms
		3	600ms
		4	800ms
Simultaneous	200ms per item	1	200ms
		2	400ms
		3	600ms
		4	800ms
	200ms for whole trial	1	200ms
		2	200ms
		3	200ms
		4	200ms

Set sizes were randomized per block such that each set size was repeated 8 times each block. The target was randomly selected for each trial. Due to the fact that the $n = 1$ trials were identical for each condition, the number of $n = 1$ trials were reduced to make them equally represented overall. All of these condition variations are displayed in Table 2.1.

Procedure. Participants were instructed on how to perform the task and experienced 10 practice trials to become acquainted with the task. On subsequent sessions, participants experienced the instructions again for continuity and to remind them of the task. Participants were aware there were different versions of the task but were not explicitly told of the differences in the conditions. However, the practice trials were specific to the different condition types, so following practice participants had experienced the condition they were to complete in that session. On each trial a fixation cross was presented for 500ms, followed by the study array (duration varied with condition) then a mask for 500ms. From the study array, one stimulus was randomly chosen as the target and the location of the target was cued by a red pentagon outline during the test phase. Participants were then asked to indicate which side of that pentagon had been missing when they had seen it in the study array. Participants made their selection by clicking on the missing side with the mouse and confirmed their selection with the space bar. Participants then received feedback as to whether they were correct or incorrect with their selection. Following an inter-trial break of 500ms, the next trial began. There were 448 trials in total, which were divided in 14 blocks. After each block participants were given a break for a minimum of 20 seconds. Once this minimum time had elapsed, participants continued the experiment by pressing the space bar. This procedure was the same for each of the three presentation-duration conditions. The only variation being the duration of each trial (200ms to 800ms in the per-item conditions or 200ms in the total condition) and the presentation type (sequential or simultaneous). A schematic of the trial sequence is shown in Figure 2.3.

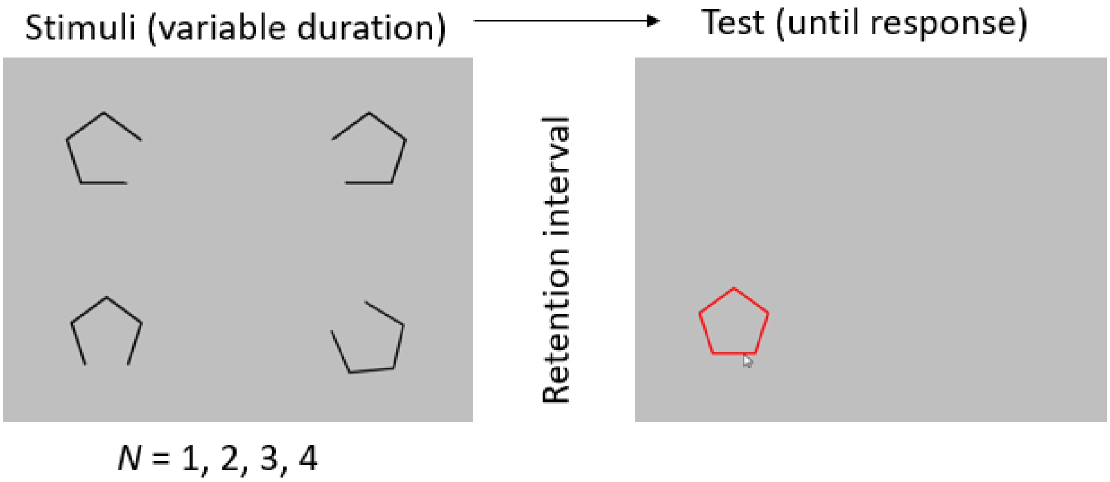


Figure 2.3. An illustration of the trial sequence. The stimuli presentation time was variable based on the condition type and set size (see Table 1). Following a retention interval of 500ms that included a mask, participants needed to select the missing side of the stimuli presented at that location during study (the target).

Results

Across all six participants, 32256 trials were recorded. Trials were excluded from analysis if they did not record any eye-gaze data, or if more than 10 fixations were recorded in any given trial. No trials were removed for having no eye-gaze data, and 803 (2.5%) trials were removed for having more than 10 fixations.

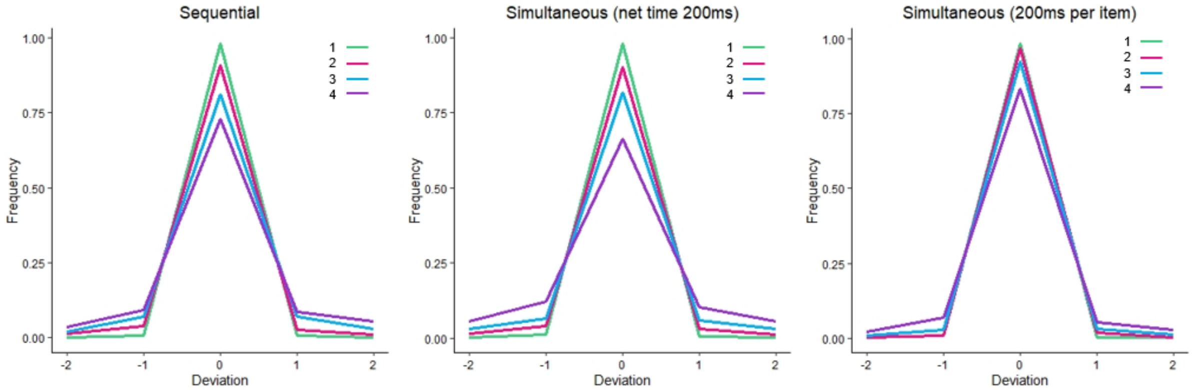


Figure 2.4. Frequency distribution of accuracy of responses in the sequential condition for all set sizes. Accuracy was measured as deviation from the correct response with a deviation of zero indicating the correct response was chosen. As set size increased, task accuracy decreased across all conditions. This effect was most pronounced for the simultaneous net time condition.

Set size effect. Following Smith et al. (2016), we examined the effect of set size on accuracy of response. We measured accuracy as the deviation from the correct response. At test, if the participants correctly identified the missing side in the target pentagon, their deviation was recorded as 0. If the side they selected was directly adjacent to the correct side they were given a score of +1 or- 1 (the sign denoted whether the answer was clockwise of the correct response or not, respectively). Likewise, if the participant selected an incorrect side, not directly adjacent to the correct answer, their response was scored as +2 or -2. The frequency distribution for these deviations, for all set sizes within each condition, are displayed in Figure 2.4.

A repeated measures 4 (set size) x 3 (presentation condition) ANOVA[2] was carried out on average accuracy grouped by participant. Accuracy was coded as 1 for trials in which the target was correctly identified and 0 for trials when it was not. There was a significant main effect of set size ($F(3,15) = 23.1$, $p < 0.001$) demonstrating the standard effect that recall accuracy decreased with set size, as seen in Figure 2.4. There was also a significant main effect of presentation condition ($F(2,10) = 10.27$, $p = 0.004$) on accuracy. A significant interaction was present ($F(6,30) = 17.28$, $p <.001$) suggesting the decrease in accuracy due to increasing set size was on average different depending on the presentation condition. A follow up one-way ANOVA on condition revealed that while the differences in accuracy are small for set size 1 ($F(2,10) = 0.15$, $p \approx 1$), there is a larger difference between the conditions at set size 2 ($F(2,10) = 5.71$, $p = 0.022$), set size 3 ($F(2,10) = 9.39$, $p = 0.005$), and set size 4 ($F(2,10) = 19.9$, $p < 0.001$). As shown in Figure 2.4, the decrease in accuracy with increased set size is most pronounced for the simultaneous (net time) condition and least for the simultaneous (time per item) condition. Pairwise comparisons with a Bonferroni adjustment suggest that for when set size was 2, 3, or 4 there was a significant difference between the simultaneous (net time) and simultaneous (time per item) conditions ($t(5) = -3.83$, $p = 0.012$; $t(5) = -5.62$, $p = 0.002$; $t(5) = -8.01$, $p < 0.001$; for each set size respectively). Unlike the simultaneous (time per item) condition, the sequential condition showed clearer differences in accuracy with set size. From Figure 2.4 the sequential condition appears to be an intermediate point between these conditions. Pairwise comparisons revealed that the sequential condition was significantly different from the time per item condition at set size 3 ($t(5) = -2.81$, $p = 0.037$), and from the net time condition at set size 4 ($t(5) = 3.85$, $p = 0.01$). Overall, these

[2] In this thesis we frequently use ANOVA as part of our statistical analysis. However, this method is not the ideal form of analysis of our data. Time constraints meant that this method was the best to use with the available time, however, as this method is not ideal, results should be interpreted with caution.

results suggest that the simultaneous (net time) and the simultaneous (time per item) conditions differed the most regarding how accuracy varied with set size.

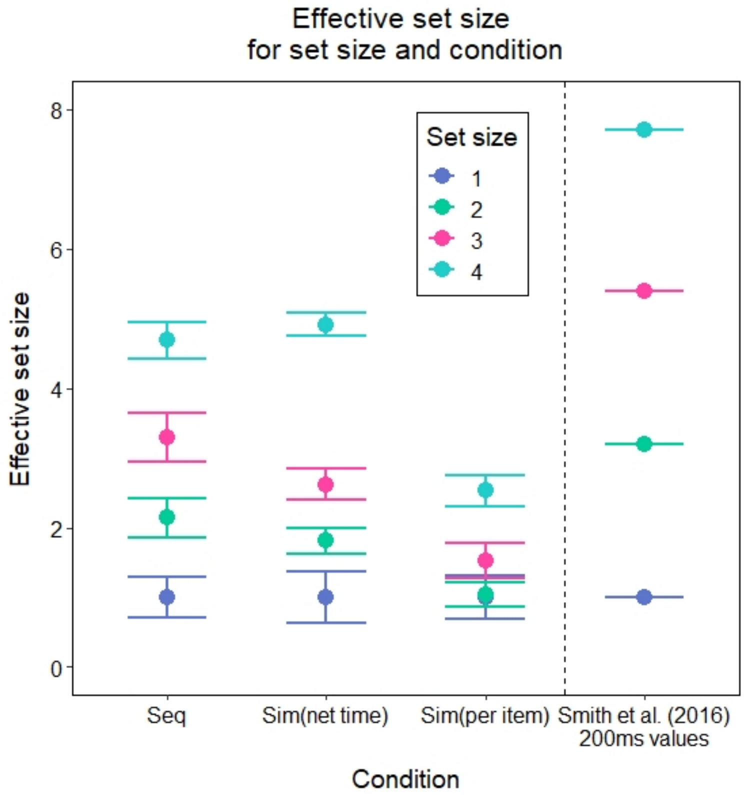


Figure 2.5. Effective set size and actual set size for each presentation condition, both for the current experiment and from Smith et al. (2016). The simultaneous (per item) condition in particular has low effective set size values relative to the actual set size suggesting this condition was easier than the others. The values from Smith et al. (2016) are values for the 200ms duration condition from their Experiment 1. The effective set size values suggest their task was more difficult than the conditions in the current experiment. Error bars indicate standard error.

Following Smith et al. (2016), we calculated the effective set size as predicted by the sample size model (Sewell et al, 2014; Smith et al, 2016). The sample size model predicts that *d'* in a display of *m* items is given by:

$$d'_m = \frac{d'_1}{\sqrt{m}}, \quad m = 2, 3, \ldots$$

By substituting the estimates for d'_m and d'_1 that we calculated from our experimental data, we created an effective display size based on this model.

Figure 2.5 shows a graph of effective compared to actual set size for each presentation condition. For comparison, Figure 2.5 shows the effective set size values obtained by Smith et al. (2016) for each of their set size conditions (to the right of the dashed line). The effective set size in the sequential condition was very close to linear. This suggests that, according to the sample size model, this condition is as difficult as the model would predict. The simultaneous net time condition was consistent with the sequential condition for smaller set sizes, but saw a sharp increase in effective set size when set size n = 4, suggesting this condition was more difficult than predicted at this set size. By contrast, the simultaneous condition shows lower effective size relative to actual set size, suggesting this condition is easier than expected. When compared to Smith et al. (2016) our effective set sizes overall seem lower and, especially in the case of the sequential condition, more in line with the sample size model than the excess load effect demonstrated in their experiments.

A repeated measures 3 (condition) x 4 (set size) ANOVA of effective set size was carried out on the grouped data. There was a main effect of set size ($F(3,15) = 222.183$, $p < 0.001$) as well as a significant interaction of condition and set size ($F(6,30) = 45.32$, $p < 0.001$), but no main effect of condition($F(2,10) = 3.27$, $p = 0.08$). This suggests that while there is no difference in condition average over set size, there are differences in the effects of set size within each condition. Examining Figure 2.5, it is clear that the distribution of effective set size in the simultaneous (pet item) condition is narrower, and tends smaller than the other conditions.

Serial position effect. Our next point of comparison to Smith et al. (2016) was to determine if there were similar serial position effects in our sequential condition as were found previously. To determine the effect of serial position on memory, we looked at the proportion of correct responses depending on set size and serial position of the target. That is, whether the target was the first, second, third or fourth item to be shown within a trial. Figure 2.6 displays the proportion of correct responses depending on serial position and set size for sequential trials, averaged over participants.

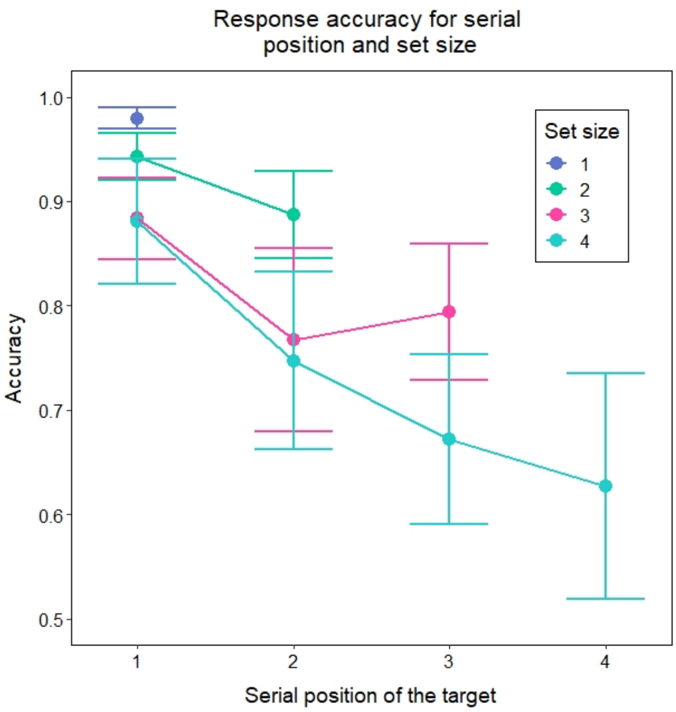


Figure 2.6. Accuracy as a function of set size and serial position, averaged over all participants. Overall, the proportion of correct responses decreases the later the target appears in the trial, suggesting a primacy effect. Error bars indicate standard error.

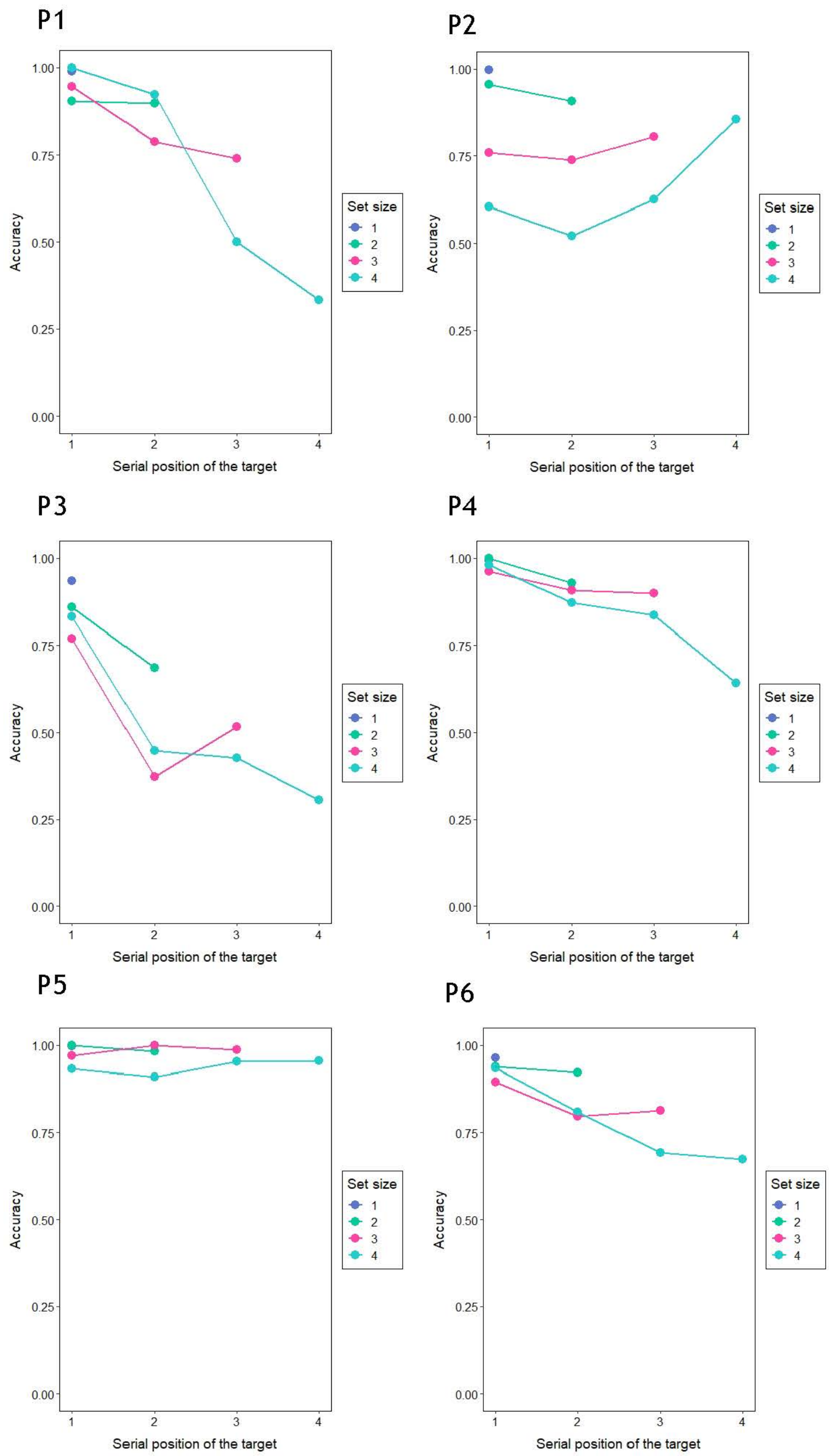


Figure 2.7. Accuracy as a function of set size and serial position for each individual participant. Most participants display the primacy effects shown in the overall result. However, P2 shows a recency as opposed to a primacy effect. P5, shows no effect.

As seen in Figure 2.6, accuracy appeared to be greatest when the target was the first item seen in the trial. The later an item was shown in the trial, the less accurate the responses were. However a 4 (set size) x 4 (serial position) repeated-measures ANOVA revealed no main effect of set size ($F(3,15) = 0.66$, $p = 0.42$) or serial position ($F(3,15) = 0.02$, $p = 0.87$) on accuracy, nor was there a significant interaction ($F(9,45) = 0.61$, $p = 0.44$). We examined the same serial position effect for each individual participant (seen in Figure 2.7) and found that, similar to the group result, there seemed to be a primacy effect such that when the target was seen first in the trial (or closer to the start of the trial) the participants recall accuracy was higher. This was not the case for two of the six participants. P2 showed a primacy and recency effect while P5 showed no effect of serial position.

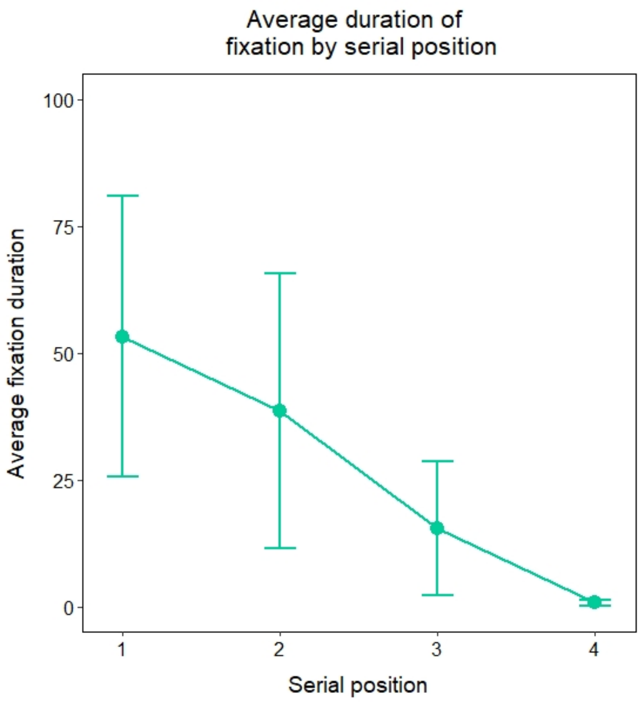


Figure 2.8. Average fixation duration (ms) as a function of serial position (sequential trials, set size $n = 4$ only). On average, the first item shown in the trial it is fixated on longer than later items in the trial. Error bars show standard error.

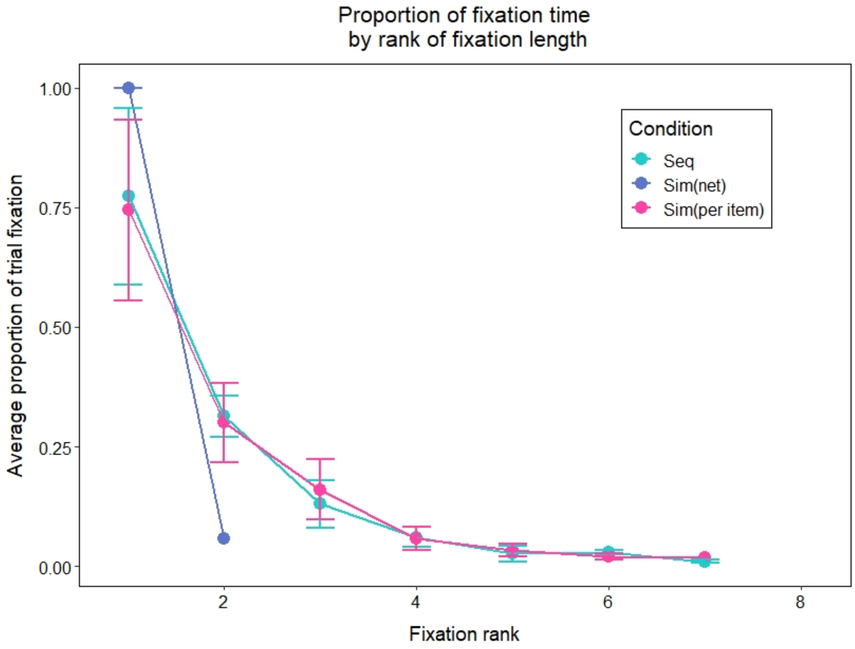


Figure 2.9. The average proportion of fixation time on a stimulus by rank of duration length within that trial (set size 4). The stimulus that is fixated on for the longest duration in the trial is rank 1. The stimulus fixated for the second longest duration is rank two, and so on. In each condition, one item takes up a majority of the total fixation duration on average. Error bars indicate standard error. Note: the proportions do not add to 1 because they are averaged across participants

Eye-gaze data. We examined the eye-gaze data to determine if there was an attention-based effect that mirrored the observed serial position effect. Since we observed a primacy effect in terms of accuracy, according to Smith et al. (2016), we should expect to see the first item shown in a sequential presentation receiving preferential attention, with subsequent items sharing the remaining attention. To do this we examined fixation duration on the study items as our measure of attention. We isolated trials of set size 4 and examined the relationship between serial position and fixation duration. As shown in Figure 2.8 above, fixations were longer for items presented earlier in the trial. This was supported by a one-way

ANOVA of serial position on fixation duration (grouped by participant) which found there were significant differences in average duration across the serial positions ($F(3,15) = 10.45$, $p = 0.004$).

This result suggests a relationship between serial position of item presentation and fixation duration. However, this analysis is limited to only the sequential presentation condition. Smith et al. (2016) found that the distribution of attention predicted by the sample size model held in both the simultaneous and sequential presentation conditions. To examine the per trial distribution as well as to compare the simultaneous conditions to the sequential condition, we examined the distribution of fixations across the trial. Instead of serial position, in the simultaneous conditions we examined fixation duration by rank. That is, we were interested in how much more the stimulus that was fixated longest was looked at compared to the stimulus looked at second longest, and so on. Figure 2.9 shows the distribution of fixation duration across the three conditions for set size 4. Fixation duration is coded as a proportion of time spent looking at any stimulus. We included fixations that were on stimuli that had been fixated before in the trial, which explains why the maximum number of fixations in a trial shown here (7) is greater than the maximum number of items in a trial (4). The sequential and simultaneous time per item conditions were very similar in distribution of fixation. By contrast, in the simultaneous net time condition, a higher proportion of the trial time is spent on a single item, which is perhaps unsurprising given the total stimulus presentation time of 200ms.

To determine if this trend in the simultaneous condition was similar to the effect of serial position seen in the sequential condition, we examined the proportion of trials on which the longest fixation occurred on the first item seen. Overwhelming, the longest fixation occurred on the first stimulus fixated in all conditions (sequential 96.56%, simultaneous net time: 98.9%, simultaneous time per item: 80.65%). While this was true for all conditions, the first item was less likely to be the longest fixated on item in the simultaneous time per item condition.

Our results suggest that on average an attention capture of items in the first serial position for the sequential conditions. The first item is typically looked at the longest with subsequent items looked at for shorter durations (Figure 2.8). This distribution of attention is not unique to the sequential condition, as the distribution of attention is also similarly favoured towards one item in the simultaneous conditions. In all conditions, the item that appears to be fixated longest is the one participants fixate on first, showing a fixation duration preference for primacy.

Trial-by-trial attention. As well as on average, we examined how participants distributed their eye-gaze within each trial. The aim here was to see what proportion of trials showed a pattern of attention that is consistent with that expected from Smith et al. (2016). As mentioned previously, to be consistent with Smith et al.'s (2016) account a majority of the attention in a trial is given to one item (the item in the first serial position in sequential trials) with the remainder of the attention divided between remaining items in the display. Assuming that eye-gaze data is in line with attention, we therefore expect to see participants fixating every item in the display, but with one item given more attention than others[3]. Contrasting to this account, the pattern of eye-gaze data seen at the group level by Smith et al. (2016) may be caused by an alternate distribution of attention. As described in the introduction to this experiment, this group level result could be produced by focusing on only a subset of the items in a trial, rather than necessarily dividing the attention between all of them. In order to distinguish between these accounts, we first examined how often participants fixated on all

[3] Upon reflection, we may also expect to see just one item receiving all of the fixation, with the remaining items receiving peripheral or covert attention.

the items in a trial. Figure 2.10 shows on how many trials participants fixated on 1,2, 3 or 4 unique stimuli for each set size and for the sequential and simultaneous (time per item) conditions. The simultaneous (net time) condition was not included due to fixations on any items being rare in this condition.

As Figure 2.10 shows, it is more common for no stimulus fixations to occur in a trial, regardless of set size. The proportion of item fixations increases with set size, which makes sense because in both the sequential and simultaneous (time per item) conditions, trial time increases with set size. However, most notably, it is exceedingly rare for all the available items in a trial to be fixated. Rather, it appears that participants tend to fail to fixate on one of the items in the display, regardless of set size. It looks as if participants are looking at more items, as there are more to see, but fail to be able to calibrate their eye movements to fixate on all items. At the very least, this suggests that the overt eye-gaze movements are not consistent with the deployment of attention proposed in Smith et al. (2016), where all items, except for the one first fixated, receive equal attention. For such a pattern to be observed in overt attention, all items would need to be fixated, however as we can see, this is not the case.

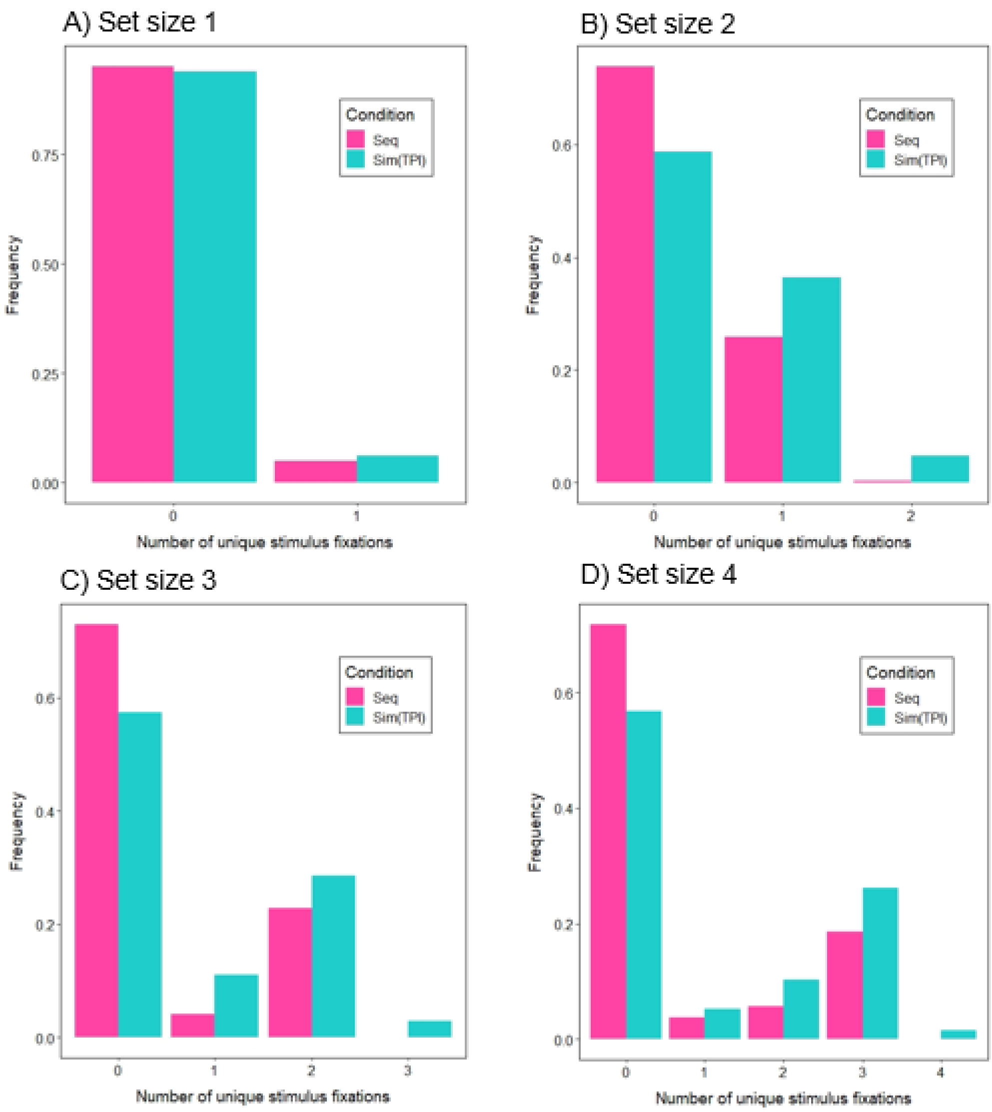


Figure 2.10. Frequency with which participants fixate on 1,2,3,4 or no unique items in the sequential and simultaneous (time per item) conditions. A,B,C and D refer to set sizes 1,2,3 and 4 respectively.

For our current purposes it is relevant to note that fixation data from most of the participants, most of the time does not match with the attentional claims made by Smith et al. (2016). This topic is addressed further in the Discussion. It is also worth noting that by far the most common number of fixations was zero. We will soon discuss that result more.

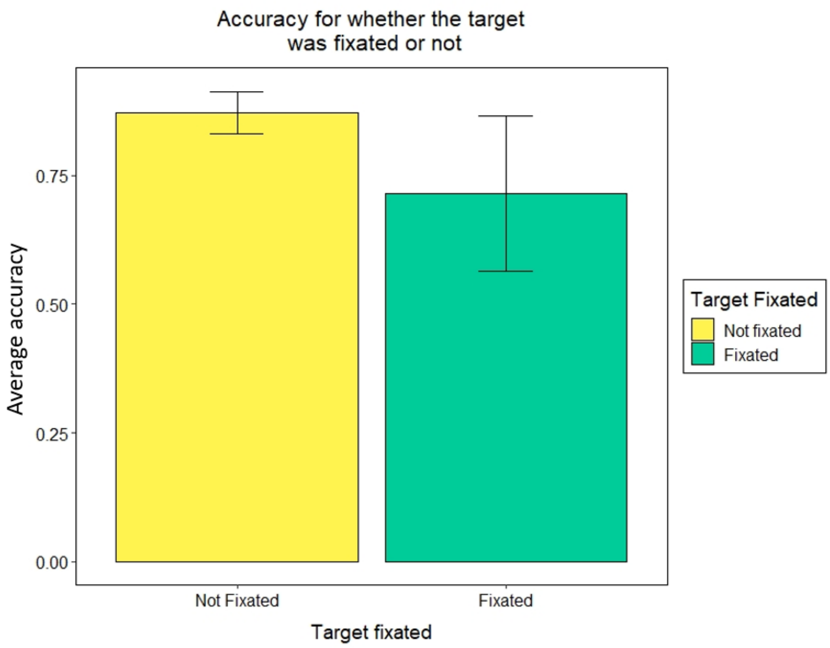


Figure 2.11. The proportion of correct responses for sequential trials by whether the target was fixated or not. Accuracy was similar regardless of whether the target was fixated or not. Error bars show standard deviation.

Fixation, serial position of fixation and accuracy. While we were not able to find an effect of serial position of item presentation on accuracy, we wanted to determine the effect of other eye-gaze variables on accuracy. In line with this, we wanted to examine accuracy depending on whether the target was fixated or not. Average accuracy by whether the target was fixated or not in a trial is shown in Figure 2.11. Using a paired samples t-test we found that the proportion of correct responses (grouped by participant) was not significantly different whether the target was fixated or not ($t(5) = 1.0751$, $p = 0.332$). Given that fixations on the target were on average longer when the target was in the first serial position, and so the relationship to accuracy may be strongest, we ran a second test on just the data for serial position one. Again, accuracy was not significantly changed by fixating on the target or not ($t(5) = 1.2058$, $p = 0.282$). This difference was not significant for set size 4 in the simultaneous time per item condition ($t(5) = 1.245, p = 0.237$) or for the simultaneous net time condition, even though the difference was larger in the latter case ($t(5) = -3.7$, $p = 0.066$).

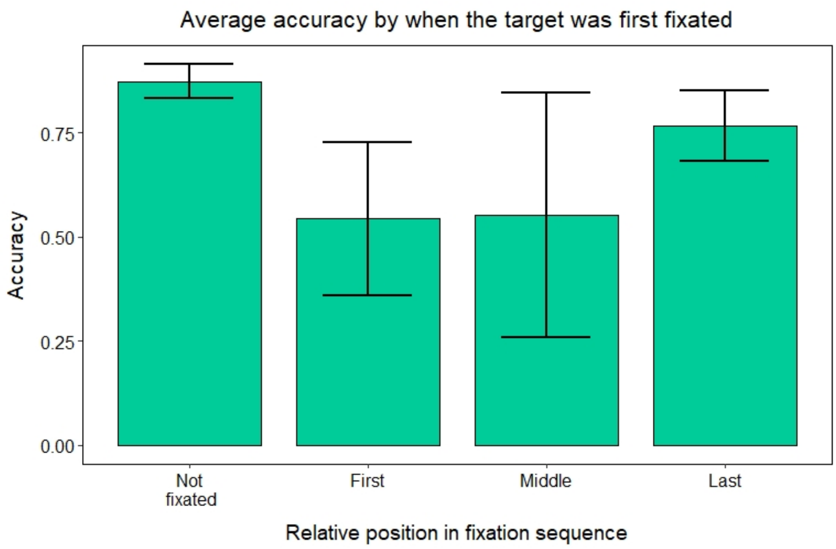


Figure 2.12. Average accuracy of response as a function of when the target was first fixated on during the trial (for sequential trials, $n = 4$ only). Trials were grouped by whether the target was the first item fixated in the trial, the last item fixated or whether the fixation occurred somewhere in the middle of the sequence of fixations. If more than one target fixation was made, only the first instance was used to determine the grouping. Error bars indicate standard error.

While we found no effect of target fixation on accuracy, we wanted to determine if there was any variation in accuracy with the order in which items were fixated – that is, the serial position of fixation. According to Smith et al. (2016), the increase in accuracy found when the target is located in the first serial position is due to the increase in attention. While our distribution of overt attention was not entirely consistent with that expected by Smith et al. (2016), we still found an effect of primacy, with participants showing attention preference to items presented towards the start of the trial. We therefore expected to see this attention effect to be associated with an increase in task performance for when the target was fixated on earlier in the trial.

Unlike the serial position of stimulus presentation, the serial position of fixations varies between trials and individuals. To be explicit: there are always four serial positions of stimulus presentation when $n = 4$ (in the sequential condition), but the number of fixations when $n = 4$ is variable. To make the fixation data points more uniform, we grouped participants fixations based on the when they saw the target relative the other fixations made in that trial. Given the strong effects of primacy reported by Smith et al. (2016) combined with recency findings in other VWM experiments (e.g. Nosofsky & Donkin, 2016), we decided to group trials by whether the target was the first item seen, the last item seen or if it was fixated on at somewhere in the middle of the fixation sequence. We also recorded trials where the target was never fixated. If the target was fixated multiple times in a trial, we grouped this trial based on the first instance of target fixation. This was motivated by Smith et al.'s (2016) finding that primacy effects were prominent in their task. Figure 2.12 displays the average group data of accuracy (proportion correct) given when the target was last seen in a trial, for set size 4 in the sequential condition. However, accuracy does not vary greatly based on when the target is first seen in the trial but tend towards showing an effect of recency rather than primacy.

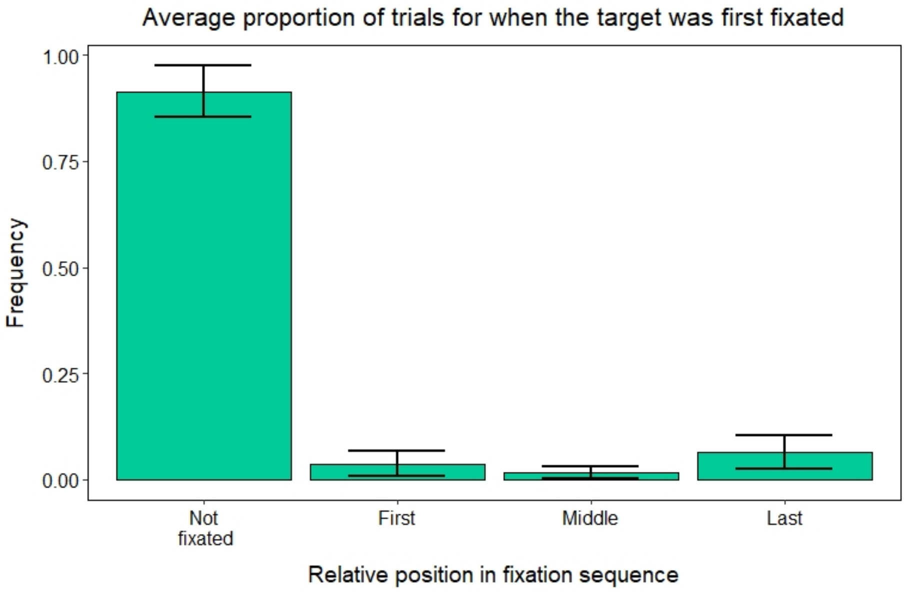


Figure 2.13. The average frequency with which the target was fixated on as a function of when the target was fixated on during trial (for sequential trials, n = 4 only). Overall, fixations on the target were rare. Error bars indicate standard error.

Given that the first fixation is generally the longest, and that accuracy was highest for the first item to be presented, it appears odd that the item seen most recently, if anything, seems to be associated with a more accurate response. To determine why this might be the case we examined the frequency with which the target fell into each of our grouping categories. As Figure 2.13 demonstrates, when the target is fixated, it is most commonly fixated on as the last item in the trial. However, it is most common for the target to not be fixated at all. Trials with no target fixations accounted for 84.7% of the data (sequential condition, n = 4).

Location. Surprised by the proportion of fixations that were of a target item, we examined the locations of all fixations across the experiment and found that for the majority of the task, participants fixated on the centre of the screen. As shown in Figure 2.14, the overwhelming proportion of fixations were recorded at the centre of the screen. This pattern of results was not only consistent across the whole experiment (shown in Figure 2.14) but also for the trials of set size 4 in the sequential condition and across all sequential trials.

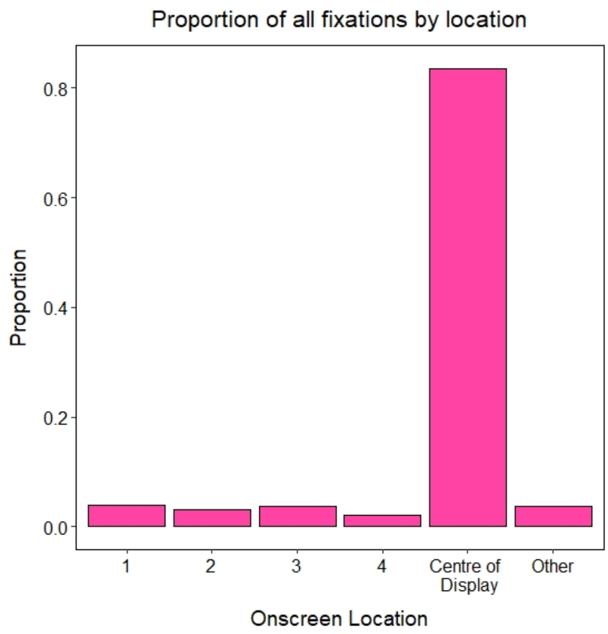


Figure 2.14. The proportion of frequency of fixation by onscreen location. Locations 1, 2, 3, 4 indicate the locations of stimuli. The vast majority of fixations occur at the centre of the display.

Following this overall finding, we were interested in the behaviour of individual participants. To determine if individual participants varied in the number of unique item fixations they made, we examined the distribution of unique stimulus fixations across the course of the experiment. Shown in Figure 2.15, there are large differences in participant's eye-gaze behaviour across the experiment. Quite strikingly, we see that most participants tend to make no fixations across most of the trials. As such, we encourage the reader to take care when considering the results of our eye-gaze analysis thus far.

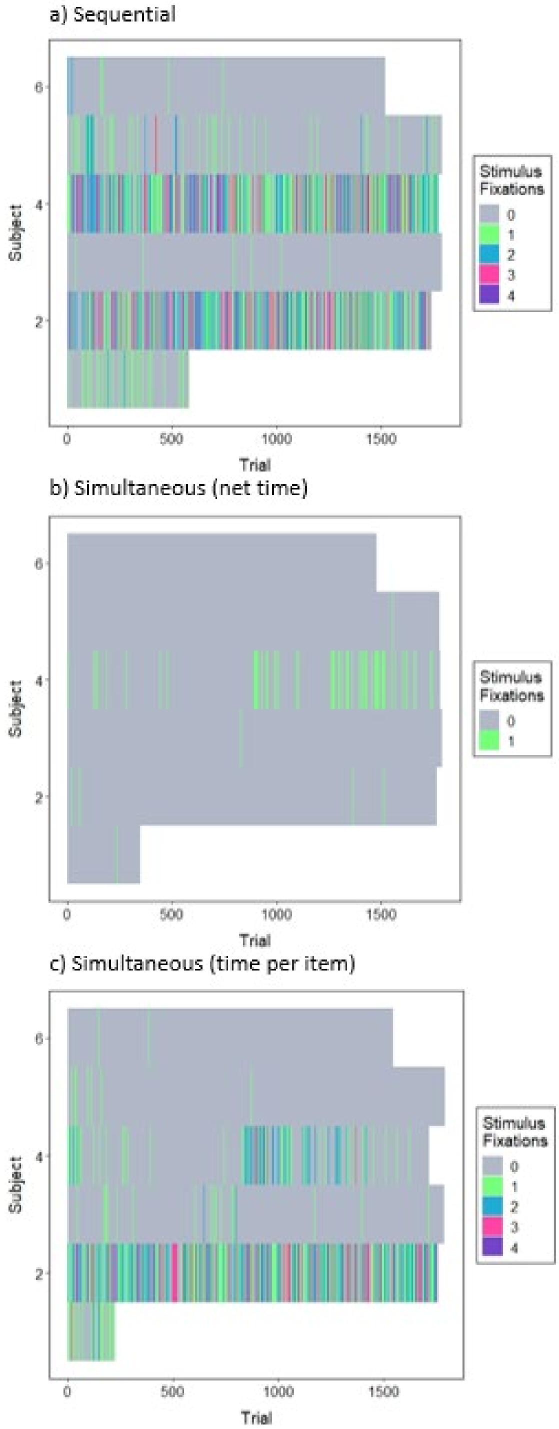


Figure 2.15. Heat maps displaying the number of unique stimulus fixations for each trial, for each subject and in a) the sequential condition, b) the simultaneous (net time) condition and c) the simultaneous (time per item) condition.

Given the lack of fixations on the target, as well as the large differences in quality of individual eye-gaze data, our analysis of the eye-gaze data was greatly limited as it pertained to answering questions responding to Smith et al.'s (2016) work. Given this is the focus of the current chapter, we limited the eye-gaze results to the above for this chapter. Chapter 3 contains a continuation of eye-gaze data from the current experiment (as well as Experiment 1).

Discussion

The central aim of this experiment was to examine eye-gaze behaviour in relation to the attention-based findings reported by Smith et al. (2016). We wanted to replicate the findings of the original study – in particular, the serial-position effect – as well as examining three key points implied by this paper. Firstly, we were interested in determining if the strong serial position effect identified in the previous study was reflected in the eye-gaze data. Secondly, whether the serial position effect described by Smith et al. was consistent on an individual trial-by-trial basis, or if it were an artefact of averaging. Finally, we were interested to examine the nature of the eye-gaze data to determine if there was evidence favouring a slots-based or resources-based account of the VWM performance.

Serial position effect. Our overall result demonstrated a serial position effect somewhat similar to what was found by Smith et al. (2016). However, this effect was less pronounced than what they found. Smith et al. (2016) showed an increase in phase discrimination accuracy when the target was the first item shown in a sequential display, with other items receiving an equal amount of attention. While we showed the same boost performance to the first item over the others, the drop off of performance was more gradual than described previously. It seemed that as serial position increased, the accuracy for each item became less than the previous. This gradual decrease is counter to the previous finding that items subsequent to the first are equally attended (Smith et al., 2016) but did seem to reflect other findings that an increase in set size leads to a decrease in performance for all items (Gorgoraptis, Catalao, Bays, & Husain, 2011). One explanation for this could lie in the stimuli used. In their experiment, Smith et al. (2016) stated that they believed a feature of their observed attention capture effect depended on stimulus complexity. This is due to the understanding that complex stimuli require more attention to encode, and that results from previous experiments with less complex stimuli do not show this result (Sewell et al. 2014). While our choice of stimuli was an attempt to match the difficulty of the phase discrimination task, it seems that our stimuli were not as complex as we expected.

Support for this complexity-based explanation comes from our estimates of *d'*. Compared to Smith et al. (2016) our values for *d'* were increased, which suggests increased discriminability of stimuli in the current experiment compared to the previous. In short, our stimuli were easier to distinguish from one another, meaning the experiment was easier. Comparing set size and effective set size we found that our stimuli produced patterns that were much closer to linear than the previous experiment. This suggests that an excess load effect was less present in our data. This makes our results more in line with the sample size model than the data collected by Smith et al. (2016).

These findings provide support to Smith et al.'s (2016) conclusion that their results are sensitive to stimulus complexity. Our choice of stimuli was based on the results of Thornton and Gilden (2007), who found that these missing-side stimuli evoked slow and serial search. It is possible that having to search for items with a particular side missing is more difficult than encoding the same items into working memory. Such an explanation implies that the 'complexity' of an item may be more task dependent than might be otherwise anticipated. Future research would be helpful to define the boundaries of this effect to better define the type of stimuli needed to produce the kind of strong primacy effects observed in Smith et al. (2016).

Serial position of fixation. In terms of accuracy, we were not able to recreate the same serial position effect as found by Smith et al. (2016). Given this, one would not expect to see the attentional data recreate the original effect either but instead to map on to our own results. In our preliminary analysis we found some evidence of a serial position-based attention capture. We found that on average, participants fixated longest on the first item presented in the trial. Average fixation duration decreased with subsequent items. Interestingly, this seems contrary to the sample-size model, which assumes that attention is spread evenly between all items in the display. Our eye-gaze data allowed to give us some insight beyond the sequential condition and allowed comparison with both simultaneous conditions as well. When examining the distribution of fixations, the simultaneous (time per item) and the sequential condition showed similar results. In both conditions, fixations in a trial were of unequal length with one fixation being clearly longer than the others. Typically, this fixation was also the first fixation made. In line with our behavioural results, and inconsistent with what might be expected from the attention-weighted sample size model of Smith et al. (2016), when ranked by duration, the decrease in duration of fixation was roughly linear. In other words, it does not appear that one item captures attention while others receive an equally small amount – fixations other than the longest fixations are not equal.

In isolation, this result seems to provide some counter evidence that our data, unlike what is shown by the behavioural analysis, may not be strictly compatible with the sample-size model. However, by the end of our results section, we see that this result should not be taken too seriously given the confusing nature of our eye-gaze data. In particular, when we analysed the accuracy of identifying the target by when it was seen in the trial, our results suggested that accuracy should be highest when the target was the last item seen in the trial, which seems to counter the idea that items in the first serial position are responded to more accurately. However, this could be confounded by the fact that more item fixations occur later in the trial.

More importantly, our results made it apparent that fixations on items were however very infrequent in this task. This is made apparent by the distribution of eye-gaze over different locations. Overwhelmingly, fixations occurred at the centre of the screen rather on any of the study items. This lack of finding is exemplified in Figure 2.11 which demonstrates that accuracy is not affected by whether the target is looked at or not. Further investigation and discussion of these eye-gaze results including the timing of fixations is expanded upon in Chapter 3 (page 77).

Individual trial attention. In their original experiment, Smith et al. (2016) found that their results were best described by an attention-weighted sample-size model. The item in the first serial position was given the most of the attention with subsequent items receiving an equal share of the remaining attention. This model was found to hold for both simultaneous and sequential presentation conditions. As mentioned by the authors, this model is most like a resources account (Smith et al. 2016). However, this model describes the distribution of attention on average and doesn't account for differences across individual trials. We sought to determine the degree to which individual fixation patterns vary from trial to trial. To do this we examined the number of fixations on items for each trial.

While we were, again, limited by the lack of item fixations, we found that participants rarely fixated on all the items presented in a trial. For example, a set size of 4 most frequently yielded fixations on three unique items. Considering that not all the items in a trial are looked at, the eye-gaze data here suggests that Smith et al.'s (2016) account that one item receives most of the attention and with the remaining attention split evenly between all remaining items is not observed in overt attention. It could be that attention is divided evenly among items in a covert fashion – a claim that is quite difficult to rule out – but then it is perhaps puzzling why there are multiple overt fixations of additional items beyond the initial item. Again, the answer to this puzzle may lie in the differences between the stimuli we used and the phase-discrimination task. That is, it may be that the phase-discrimination stimuli were especially difficult to encode, and had we used the same stimuli or shorter presentation times, we would have never seen more than one fixation.

What is clear from our eye-gaze analysis, when reviewing individual participants' fixations over the course of the experiment, is that covert attention plays a large role in visual working memory tasks. For example, the majority of our participants rarely fixated on any stimuli, and yet, participants were capable of performing very well on the task. This would suggest that participants are attending the stimuli in order to encode them but doing so covertly – that is, attending without fixating. Overall, this suggests a decoupling of attention and eye-gaze in this task. At the very least, we can say that our eye-gaze data were not completely in line with an attention-weighted account, however, our results are not clearly explained by any alternative because of this disassociation.

This lack of eye movement therefore makes it difficult to make claims about whether the data supports a more slots-like or resources-like approach. Only viewing a subset of items is more in line with what we predicted from a slots account, which is generally what we see when participants do fixate on items. However, it is unclear whether participants are limited in the number of items they visit due to the limited trial duration. It is possible with more trial time participants would be more likely to try to fixate all the items in the display. In order to determine if trial time is a factor in the number of unique items participants fixate, we would need to examine experiments with long study durations. This is partly the reason we cannot say that our data supports a slots account. More importantly, the lack of eye movement that participants display in this task was not accounted for in our predictions of what eye-gaze behaviour would look like under each model.

Limitations. It might be said that differences in stimuli we used might have caused the lack of eye-gaze data, as our stimuli may have been too simple as to require people to fixate. It is possible that more complex stimuli may have required fixation to encode, and this would have altered the outcome of the experiment. While our serial-position effect was not as large as that reported by Smith et al. (2016), it was still present. This suggest that this effect does not rely on overt attention. The lack of informative eye-gaze data found in this experiment seems unlikely to be caused by a feature of the stimuli alone, and this is supported by our results in Experiment 1 which also found a relative lack of eye movement. This similarity occurs despite differences in the tasks in terms of stimuli, set size, and trial duration, and thus it appears that use of covert in these tasks is a common strategy. In Chapter 3, we conduct a further examination of the eye-gaze data from both Experiments 1 and 2 in an attempt to determine the cause of lack of eye movements to get a better understanding of participant behaviour in these types of VWM tasks.

Another consideration of our trial-by-trial results is that our duration times were not similar to Smith et al. (2016). Indeed, our simultaneous (time per item) condition was longer than in the original experiment, but was made to match our sequential condition. Comparing our timings to those in Smith et al. (2016) are complicated by the descriptions of their timings in the sequential condition not being perfectly clear. However, it could be that both their sequential and simultaneous conditions were time matched with our simultaneous (net time) condition. If that is the case, it could explain why our results differ – with more time, participants can apply more attention to more items and this means that the distribution of attention is less likely to be that which was predicted by the sample size model. However, if this is the case our results here demonstrate that participants in the simultaneous (net time) condition only really have time to make one fixation, rarely two. If participants only ever can make two fixations, with the second fixation being very short, when trials are averaged over each other, we may get the result that Smith et al. (2016) find – one item is attended two and the remaining items get equally small amounts of (covert) attention. It is of course possible that covert attention can account for much more of the performance in these tasks, more than can be seen by eye-gaze. However, our results do give a preliminary suggestion that rather than some mechanism of attention distribution, this could simply be due to a lack of time to do anything more than attended one and a bit stimuli per trial.

Conclusion. The current experiment was able to find some evidence of a serial position effect, however not to the extent that was described by Smith et al. (2016). From our measurements of stimulus discriminability, we conclude that our stimuli were perceived as less complex and therefore easier to encode than the stimuli of the previous experiment. As Smith et al. (2016) believe this effect is dependent on a certain degree of complexity, our results seem plausible given our stimuli were perceived as simpler. While our effect of serial position was not as described by Smith et al. (2016), we did see an effect of primacy. We saw a preference for longer fixations on the first item displayed in the sequential condition with fixation duration decreasing as serial position in the display increased. Additionally, the distribution of attention in eye-gaze trials was similar for sequential and simultaneous (time per item) conditions. This suggests, similar to what Smith et al. (2016) found that attention is utilised similarly regardless of presentation type when trial duration is equal. A lack of interpretable eye-gaze data prevents us from making any strong claims about the effects in this experiment. Because it is common for most participants to not fixate on any items in a trial, our data is not incompatible with either of slots or resources accounts of VWM. Instead, given that task performance is similar regardless of whether the target is fixated, it appears that participants are able to complete this task using covert attention. Further examination of these issues is now pursued in Chapter 3. While this experiment did not provide the strong link between task performance and eye-gaze variables that we had hoped, it seems relatively clear that the use of passive eye tracking in such VWM tasks is not advisable.

Chapter 3: Further analysis of the eye-gaze data for Experiment 1 and 2

So far, the results of our eye-gaze analyses have been underwhelming. In Experiment 1, we found some evidence that fixating the target results in more accurate responses. While this difference was not sizeable for 3-item trials, fixation was associated with better memory performance for 6-item trials. However, this difference was perhaps smaller than might be intuitively predicted. Rather, our results suggest that fixation plays a much lesser role in task performance compared to covert attention in this task. As discussed, we suspected that this lack of result for Experiment 1 could be in part due to the lack of any effect of our blocking manipulation, as well as potentially some other features of the task such as arrangement of items in the display. We addressed these issues in our second experiment by following a design by Smith et al. (2016) who demonstrated robust primacy and serial position effects that they attributed to the distribution of attention in the task. In Experiment 2 we did not exactly replicate the findings of Smith et al. (2016), however we did see a preference in attention towards the first item presented in the display in the sequential condition. This was demonstrated by longer fixation durations on items in early serial positions. However, these longer fixations were not associated with an increase in response accuracy. Furthermore, unlike our first experiment, in Experiment 2 we found no connection between fixation and task performance. The explanation behind the results of both experiments appears to be linked to a common cause: generally, participants frequently fixate on centre of the screen in these tasks.

In both Experiment 1 and 2, our analysis of eye-gaze data was restricted by the lack of data resulting from the lack of participant eye movement. This created significant problems for us in trying to examine the connection between attention and memory. In this chapter we detail analyses that we postponed reporting on in the first two chapters of this thesis. Given the lack of data, the results from these analyses should be taken lightly. However, we wanted to include them here to demonstrate the breadth of possible analyses, and to establish our rationale for the next experiments. Firstly, we look at a variety of eye-gaze variables and, using regression, examine a link between these factors and memory performance in both experiments. Secondly, using insights from our second experiment we revisit Experiment 1 to determine any similarities in eye-gaze patterns seen. We then examine the eye-gaze data from both experiments to explore reasons for the lack of eye movements in both experiments. These include the timing of fixations within trials, as well as evidence of learning and individual strategy.

Memory performance and eye-gaze variables

As a first step in attempting to establish the connections between task performance and eye-gaze, we performed a series of linear regressions for each experiment. Our measures for task performance were deviation from the correct response in Experiment 1 and the proportion of correct responses in Experiment 2. For Experiment 1, we also performed regressions to determine a connection between latent working memory variables and eye-gaze. These latent variables came from our estimates for the probability of an item being in memory (P_{mem}) and the precision of an item memory (*Prec*) which were derived using a Bayesian hierarchical mixture model (for details, see Chapter 1). Due to the data from Experiment 2 being 5AFC rather than a continuous report, we did not perform any modelling, as there appears to be very little information in the distribution of errors around the correct response, making it likely that we would learn nothing what we observe in our analysis of accuracy.

Experiment 1

First, we examined the relationship between eye-gaze variables and task accuracy. We grouped the data by subject, taking average target fixation duration and the average number of fixations for each participant and for each set size. We included set size as a variable in some of our analyses since our previous analyses indicated that set size had a large effect on task performance. Furthermore, set size would be potentially confounded with both target fixation duration and the number of stimulus fixations, since, for example, more items in the display may encourage more fixations and shorter fixations. We compared combinations of set size, target-fixation duration and the number of stimulus fixations to see which combination gave the best account of accuracy (i.e., deviation). Table 3.1 shows the complete combination of variables with measures of performance (AIC) for each of these models.

From testing a model of set size alone with accuracy (Model 1), we found that set size alone gave a good account of the variation in accuracy. The model of target fixation duration and accuracy (Model 2) showed that fixation duration on the target was also a significant predictor of accuracy. Model 3, revealed that the number of stimulus fixations was not a significant predictor of accuracy. In Model 4 we introduced set size to Model 2, and found that once set size was introduced, target fixation duration was no longer a significant predictor of task accuracy. Similarly, in Model 5, only set size was a significant predictor of accuracy with the number of stimulus fixations not effecting task performance. Overall, when comparing AIC, it seems as though the best model for accuracy is the one with set size alone.

Table 3.1.

Experiment 1: models predicting accuracy using set size, number of fixations and target fixation duration

	t	*p*	AIC
Model 1			-17.57
Set size	-5.85	<0.001	
Model 2			75.70
Target fixation duration	-4.39	<0.001	
Model 3			83.99
Number of stimulus fixations	-1.06	0.30	
Model 4			11.072
Set size	5.97	<0.001	
Target fixation duration	-1.01	0.32	
Set size x Target fixation duration	1.01	0.32	
Model 5			-2.79
Set size	6.91	<0.001	
Number of stimulus fixations	-0.58	0.56	
Set size x Number of stimulus fixations	0.44	0.66	

As discussed in Chapter 1, the relationship between accuracy and eye-gaze behaviour was relatively small for set size 3. This relationship did appear more pronounced for set size 6, and so we repeated the same analysis on just the set size 6 data. As seen in Table 3.2, neither target fixation duration or the number of stimulus fixations were significant individual predictors of accuracy for set size 6 trials in Experiment 1.

Next, we tried to see if the target fixation duration and the number of stimulus fixations had a better correlation with the latent variable P_{mem}. The results of this analysis can be seen in Table 3.3. As with accuracy, we tested models with only set size (Model 1), only target fixation duration (Model 2), and only number of stimulus fixations (Model 3). Model 1 showed set size to be a significant predictor of P_{mem}. Model 2, showed that target fixation duration was a significant predictor of P_{mem}. By contrast, Model 3 revealed that the number of stimulus fixations did not significantly impact the probability of an item being in memory. As with our analysis of accuracy, we also predicted P_{mem} based on set size and target fixation duration (Model 4), and found that only set size was a significant predictor of P_{mem}. Similarly, in Model 5, which predicted P_{mem} based on set size and number of item fixations, only set size was found to be a significant predictor of the probability of an item being in memory.

Table 3.2.

Experiment 1: models predicting accuracy with number of fixations and target fixation duration for set size 6 only

	t	*p*	AIC
Model 2			32.39
Target fixation duration	0.68	0.49	
Model 3			26.24
Number of stimulus fixations	0.06	0.95	

We also performed a similar analysis of P_{mem} for set size 6 only. Table 3.4 shows that, similar to accuracy, neither target fixation duration nor the number of stimulus fixations are significant individual predictors of the probability of an item being in memory for set size 6.

Table 3.3

Experiment 1: models predicting probability of an item being in memory using set size, number of fixations and target fixation duration

	t	*p*	AIC
Model 1			-169.88
Set size	-21.28	<0.001	
Model 2			-40.92
Target fixation duration	4.88	<0.001	
Model 3			-29.90
Number of stimulus fixations	-1.40	0.17	
Model 4			-138.33
Set size	-7.32	<0.001	
Target fixation duration	1.12	0.27	
Set size x Target fixation duration	-1.38	0.18	
Model 5			-151.43
Set size	-8.04	<0.001	
Number of stimulus fixations	0.433	0.66	
Set size x Number of stimulus fixations	-0.62	0.54	

We concluded this series of analyses for Experiment 1 by examining whether the target fixation duration or the number of stimulus fixations had an impact on memory precision. Shown in Table 3.5, we tested models with only set size (Model 1), only target fixation duration (Model 2) and only number of stimulus fixations (Model 3). Again, set size and target fixation duration were shown in Model 1 and 2 respectively to be significant individual predictors of precision. However, when each of the eye-gaze variables was combined with set size, neither target fixation duration or the number of stimulus fixations provided a significant contribution in predicting *Prec*. Similarly, predicting Prec for set size 6 only (shown in Table 3.6) revealed that neither target fixation duration nor the number of stimulus fixations are significant individual predictors of precision.

Table 3.4.

Experiment 1: models predicting probability of an item being in memory with number of fixations and target fixation duration for set size 6 only

	t	*p*	AIC
Model 2			-45.94
Target fixation duration	-1.04	0.31	
Model 2			-51.78
Number of stimulus fixations	-0.54	0.5881	

Table 3.5.

Experiment 1: models predicting precision using set size, number of fixations and target fixation duration

Models predicting Precision	*t*	*p*	AIC
Model 1			474.98
Set size	-11.71	<0.001	
Model 2			524.26
Target fixation duration	6.84	<0.001	
Model 3			537.92
Number of stimulus fixations	-0.21	0.84	
Model 4			489.91
Set size	-3.48	0.001	
Target fixation duration	1.20	0.24	
Set size x Target fixation duration	-0.70	0.48	
Model 5			475.06
Set size	-3.85	<0.001	
Number of stimulus fixations	1.47	0.15	
Set size x Number of stimulus fixations	-1.15	0.26	

Overall, set size seems to be the most significant predictor to accuracy, the probability of an item being in memory and the precision of memory, beyond the contribution of any target fixation duration or of the number of stimulus fixations. Given the analysis here, there is no evidence that target fixation duration of the number of stimulus fixations significantly impacts memory performance in Experiment 1.

Table 3.6.

Experiment 1: models predicting precision with number of fixations and target fixation duration for set size 6 only

	t	*p*	AIC
Model 2			201.11
Target fixation duration	-0.49	0.63	
Model 2			194.60
Number of stimulus fixations	0.38	0.71	

Experiment 2

Using the proportion of correct responses as our measure of accuracy, we used regression to determine predictors of task performance. As with Experiment 1, we used set size as a point of comparison for our eye-gaze variables.

Table 3.7 shows the regression modelling results for Experiment 2. We first tested models with single predictors each for set size, target fixation duration and for the number of fixations on stimuli in a trial. Of these single-predictor models, only the model with set size was determined to be a significant predictor. Neither the duration of fixation on the target or the number of fixations on stimuli alone were significant predictors of accuracy in this task. However, as the number of stimuli fixations was close to significant in its single-predictor model, we also ran two predictor model of set size and number of stimuli fixations. In running this model, the significance of the number of stimuli fixations disappeared, with set size as the only significant predictor in the model. For completeness, we also tested the model of set size and target fixation duration, and for that in this model set size was, once again, the only significant predictor. As with Experiment 1, target fixation duration and the number of stimulus fixations do not appear to add a significant contribution to accuracy over that of set size.

Table 3.7

Experiment 2: models predicting accuracy using set size, number of fixations and target fixation duration

	t	*p*	AIC
Model 1			-77.51
Set size	-5.68	<0.001	
Model 2			-50.62
Target fixation duration	-0.06	0.95	
Model 3			-51.76
Number of stimuli fixations	-1.87	0.064	
Model 4			-63.64
Set size	-4.93	<0.001	
Target fixation duration	-0.83	0.41	
Set size x Target fixation duration	0.912	0.36	
Model 5			-58.96
Set size	-4.02	<0.001	
Number of stimuli fixations	-0.29	0.77	
Set size x Number of stimuli fixations	0.366	0.71	

Additional analyses: Experiment 1

Following our analysis of Experiment 2 in Chapter 2, there were two particular findings that we wanted to examine in the data from our first experiment. Firstly, we found that typically one item is focused on the longest with subsequent items receiving less and less fixation time. We wanted to see if this pattern held in Experiment 1 and whether it could be associated with a primacy or recency effect. Additionally, we found large individual differences in the number of fixations per trial which we were curious to compare to the first experiment. Finally, we were interested to see if there was a suggestion of a primacy or recency effect of gaze serial position.

Fixation ranking

We were interested in examining the distribution of fixations across the experiment. In particular, comparing the fixation time given to the longest fixated item compared to the second longest item fixated and so on, as we did in Experiment 2.

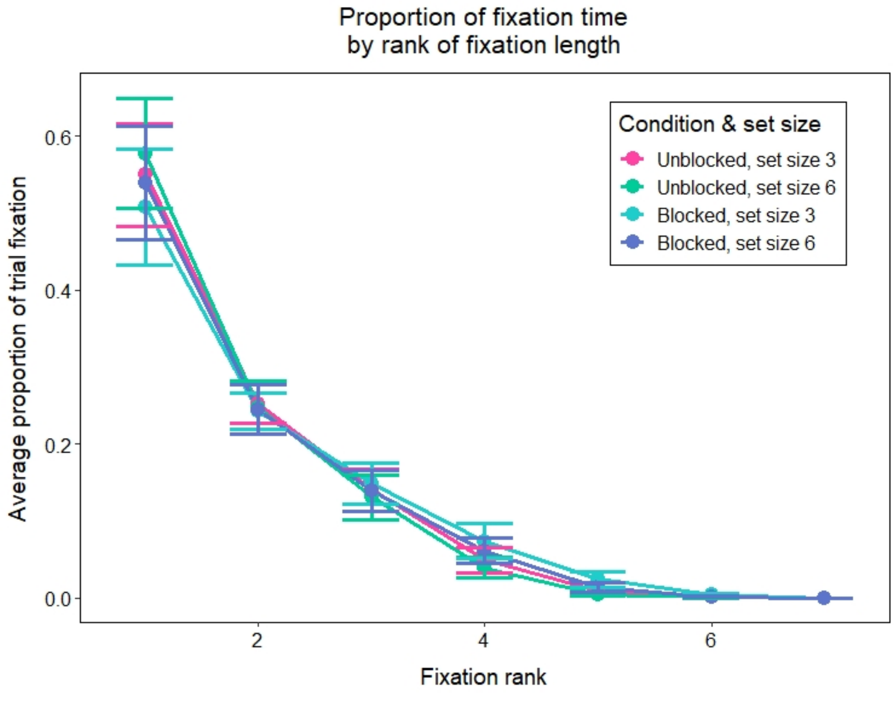


Figure 3.1. Average fixation duration by rank of fixation size and set size. Rather than an even distribution of fixation, one item is fixated on more than other fixations. Error bars are standard error.

As shown in Figure 3.1, rather than an even distribution of fixation time across fixations, on average, one item is fixated more than others with other fixation ranks being allocated less and less time. This pattern is similar to what we found for the Experiment 2 data.

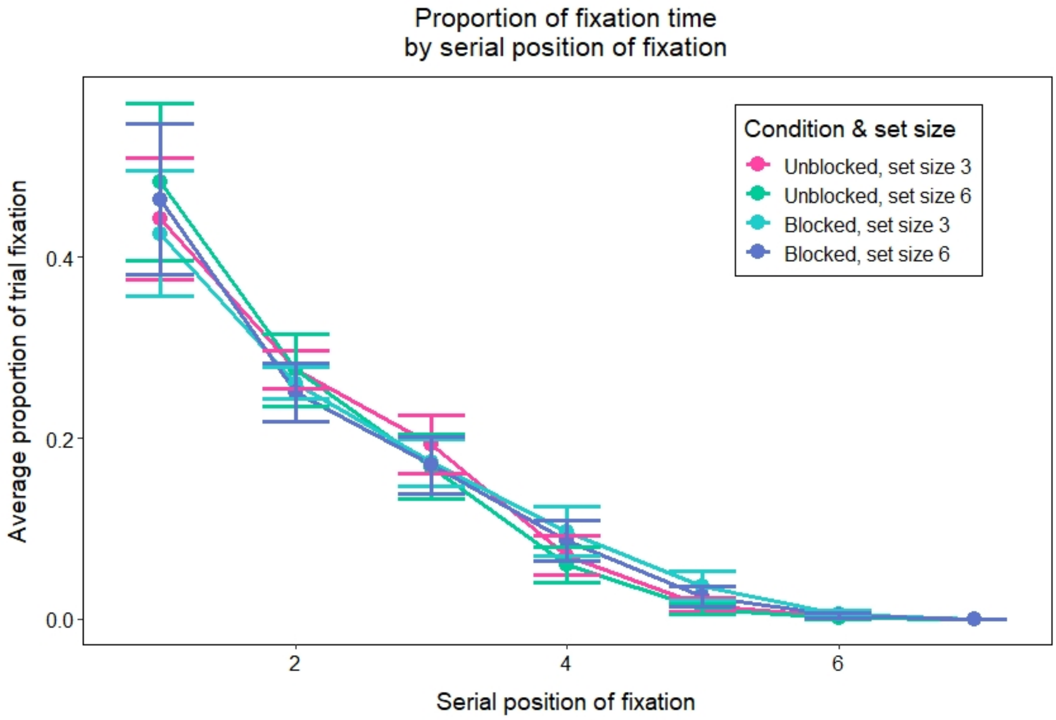


Figure 3.2. Average proportion of fixation by the serial position of fixation for set size 3 and 6. The first fixation made is typically the longest with subsequent fixations becoming shorter. Error bars are standard error.

Following this we also looked at the average duration of the fixation on stimuli by the serial position of fixation. As seen in Figure 3.2 the first fixation on an item is typically the longest with subsequent fixations being shorter. These results are similar to our results from Experiment 2, which show longer durations for earlier serial positions.

It is important to note that Experiment 1 used a simultaneous presentation only. However, similar to the simultaneous (time per item) as well as the sequential conditions in Experiment 2, we see an uneven distribution of attention between trial items with a preference towards dwelling longest on the first fixation made. This suggests a consistent pattern of attention distribution across tasks.

Number of fixations

In Chapter 2 we examined the number of fixations per trial to begin to determine how participants were allocated their attention. In particular, we were curious to see if the number of fixations was more in line with the account put forward by Smith et al. (2016), where participants might be expected to fixate all the items in the display. In addition, we were interested in whether participants in Experiment 1 preferred, like those in Experiment 2, to fixate most often on no items.

Figure 3.3 displays the frequency with which participants make unique item fixations in Experiment 1 for set sizes 3 and 6. For both set sizes, it is most common that participants fixate on 3 unique items in a trial. This suggests that for set size 3, it is most common for all the display items to be seen. By contrast, in set size 6, seeing all of the items in the study array is rare.

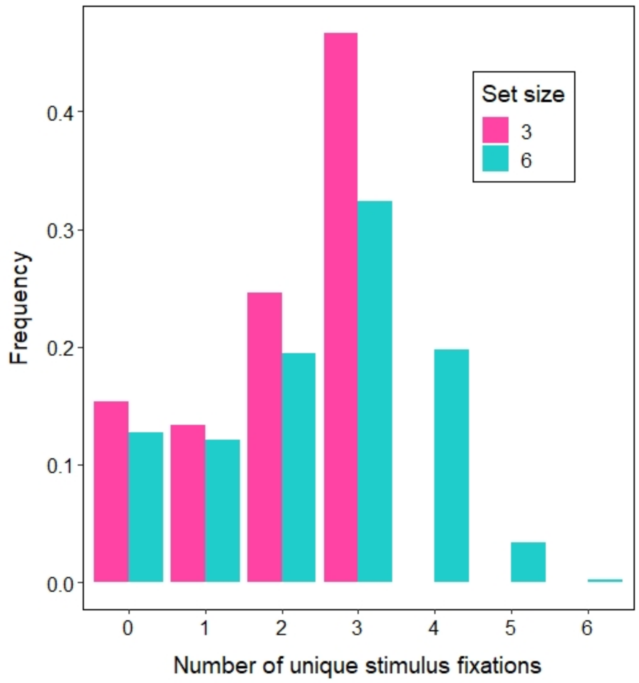


Figure 3.3. Proportion of trials on which a given number of fixations on unique stimuli occurred. For set size 3 and 6, it was most common to fixate on three unique items per trial.

The nature of the eye-gaze data here is more suggestive of a slots model account instead of a resources model. If overt attention was necessary for encoding, then a resources model would imply that participants would fixate on all items in the display, rather than looking at a subset of items (which is more in line with a slots account of attention). Like the results from Experiment 2, however, the number of unique fixations may be limited by the trial time. This is to say that participants might not have enough time to comfortably visit more than 3 items. However, it is important to note that the study duration in this experiment is longer than the maximum duration in Experiment 2 (1000ms compared to 800ms). This increase in study time would, going from our results in Experiment 2, suggest that participants would be able to make more fixations. Indeed, it is not uncommon for participants to make fixations on 4 unique stimuli in set size 6. However, given that the most common number of unique stimuli fixated is 3, it is possible that people simply preferred to fixate on only 3 stimuli (but for longer, presumably). If this was true, this may indicate that participants favour attending a subset of the display rather than all items, which has been shown to be the case in change-detection tasks (Bengson & Luck, 2016). However, this could only be determined in experiments with longer study times, where it is clearer that participants could have comfortably attended all items should they have wished. Additionally, our results also imply that participants must use covert attention in this experiment, since participants do not fixate on all items while being relatively accurate, and there being no clear evidence that fixations are closely related to accuracy. This could mean that participants are attending more than one stimulus for each fixation. To resolve this, in future experiments we would need to design an experiment to tease apart overt and covert attention, something we address more in the Discussion of this chapter.

Serial position of fixation

We were interested to see if when the target was first seen in a trial has an impact of task performance in Experiment 1. To do this we followed the method of grouping fixations described in Chapter 2 (page 65). Task performance was measured by used the absolute value of deviation from the correct response.

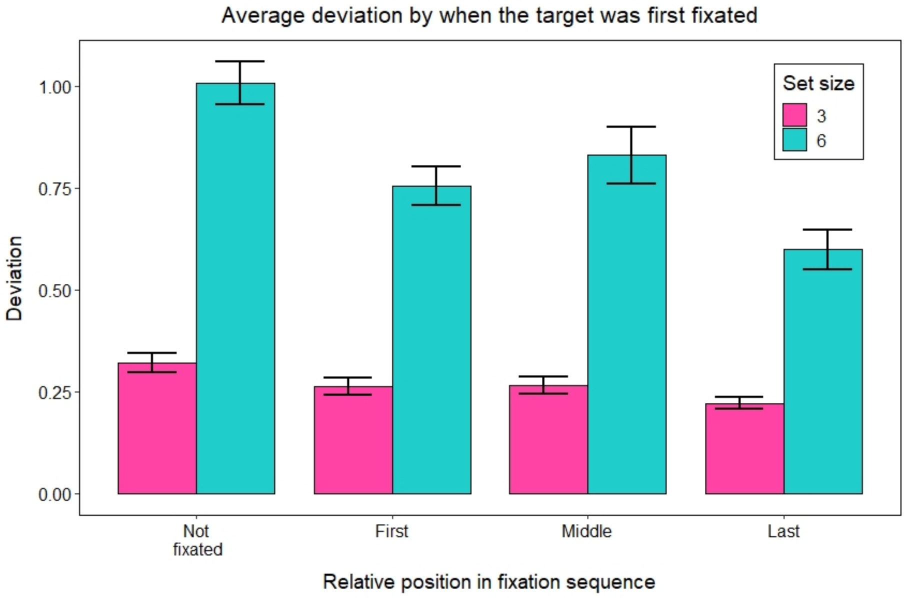


Figure 3.4. Average absolute value of deviation (in radians) by when the target was first fixated in the trial, by set size. Trials are grouped by whether the target was not fixated at all, the first item fixated, the last item fixated or whether it was fixated on somewhere in the middle of the fixation sequence. Deviation appears lower when the target is the most recent item fixated. Error bars are standard error.

As Figure 3.4 depicts, deviation from the correct response did not differ greatly for set size 3. However, there was some suggestion that when the target is the last item fixated on, there is less deviation and therefore more accuracy in responses for set size 6. Figure 3.5, shows the average frequency with which trials fit into each of these groupings. For both set sizes, it was most common for the target not to be fixated at all, however this difference was more pronounced for set size 6. Otherwise, the target seemed equally likely to be the first item seen, last item seen or to be seen somewhere in the middle of the fixation sequence.

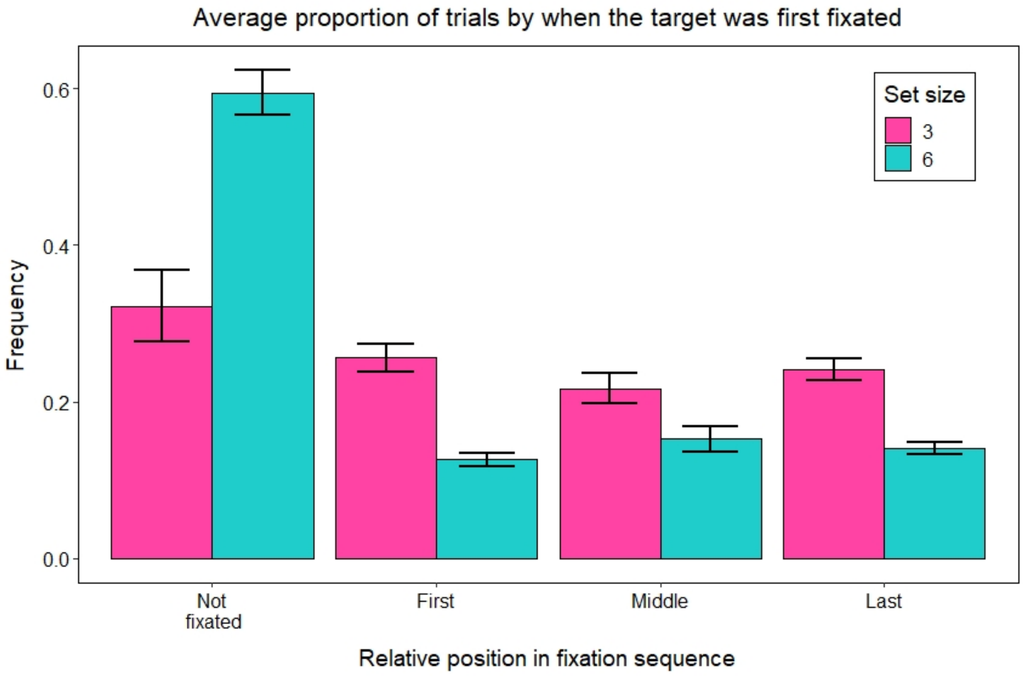


Figure 3.5. Average proportion of trials by when the target was first seen in the trial. Trials are grouped by whether the target was not fixated at all, the first item fixated, the last item fixated or whether it was fixated on somewhere in the middle of the fixation sequence. It is most common that the target is not fixated, however this is more apparent for set size 6. Error bars are standard error.

Like in Chapter 2, if the target was seen more than once, this analysis only recorded the first instance of target fixation. However, running the same analysis but with recording only the most recent target fixation yielded similar results regarding deviation and frequency. While not drastically different, the frequency results of this version of the analysis suggested that it was more likely for the target to be the last item seen, particularly in set size 3. As with Experiment 2 data there is a compounding effect of number of fixations such that with smaller numbers of fixations in a trial, it is more likely the target is to be seen more recently. To disentangle some of these variables, we re-examined the data from both Experiment 1 and 2 in regards to eye-gaze timing.

Delay in fixating stimuli

From the above analyses of Experiment 1 and from the Experiment 2 results presented in Chapter 2, there are some findings that may seem contradictory. Both experiments show a suggestion of serial position effect that would suggest that items that are seen more recently are remembered better. Similarly, however, we have also shown in both experiments that the first fixation made on a study item is usually the longest. As discussed in Chapter 2, in our attempted to normalise serial position of fixation across trials and participants, we over represented more recent responses in this analysis (for details see Chapter 2). As such, we wanted to determine whether it was true that the target was more likely to be looked at later in the trial or whether this was an artifact of the analysis method we used.

A potential reason that this may be the case is that there is some delay between the onset of the trial and when a participant first fixates on an item. While it might be expected that a sudden onset of an item captures attention (Posner, 1980), in the case of trials in which many items occur together, we might expect people to spend some time first deciding where to look, or to covertly scan the scene prior to fixating. To investigate this, we first examined when participants looked at the target, relative to the target's serial position in sequential trials in Experiment 2. We took these findings to see if we could see generalisation in the simultaneous trials in Experiment 2 and in Experiment 1.

Experiment 2

First, we were interested in the lag between when an item was presented and when it was looked at. To do this we only looked at data where one of the stimuli was fixated for the sequential trials only. This accounted for 5077 fixations (of 43652 trials in the sequential condition, 11.6%). For each of these fixations we compared when the fixation occurred relative to when the stimulus that was the subject of fixation was presented. Of the 5077 fixations, 231 of these occurred during the presentation window for that stimulus. A further 71 fixations occurred before the presentation window of a stimulus, while the vast majority of these fixations, 4775, occurred after the stimulus presentation window.

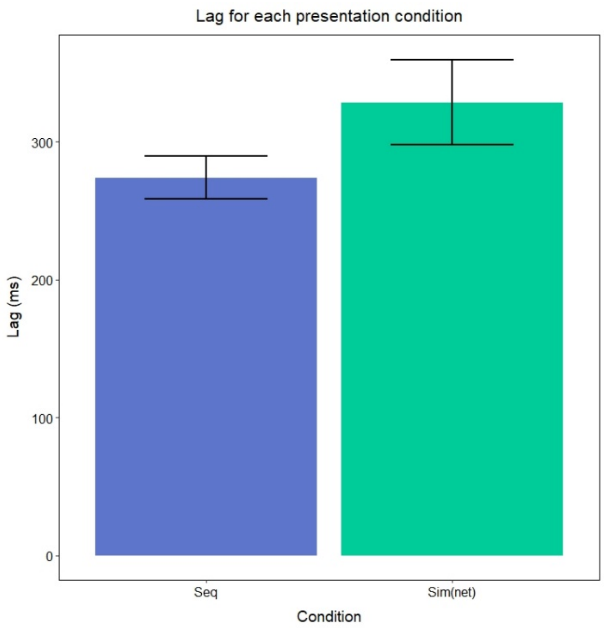


Figure 3.6. Time between start of trials and first item fixated for the simultaneous (per item) and the sequential conditions. There is significantly longer delays in the first item fixation in the simultaneous (per item) condition. Error bars are standard error.

Examining fixations that occurred after the stimulus duration window (what we now call late trials), we found that there was a mean lag between the stimulus being removed from view and the start of the fixation on the stimulus of 101.8ms. If we focus on just set size 4 trials, the late trials had a mean lag of 121.7ms. In summary, this data suggests that by the time participants fixate on an item, it has already been removed from view. Participants are therefore fixating on the location where a stimulus used to be, rather than where a stimulus is. Given that participants still perform this task relatively well (see Chapter 2), it is clear they are capable of using covert attention to do the task. However, the fact that some participants attempted to fixate items also suggests that it is not a completely covert process. Indeed, these results are in line with other studies that have found that eye-gaze is preceded by a shift in attention (e.g. Deubel & Schneider, 1996, Hoffman & Subramaniam, 1995).

Additionally, there does seem to be an effect of condition on the delay of the first stimulus fixation onset. A one-way ANOVA of condition on lag for trials with a least one stimulus fixation revealed significant differences in lag between the conditions ($F(2,5) = 28.57$, $p < 0.001$). However, the results of this ANOVA were driven by the simultaneous (net time) condition, where the total display was shown for only 200ms. On such trials, participants mostly made no fixations, with one stimulus fixation only accounting for 2.3% of all trials. Furthermore, when at least one stimulus was fixated in this condition, it appeared to be the case that the participant happened to be looking at the stimulus location before it appeared onscreen. That is to say, participants sometimes correctly anticipated the location of a stimulus and fixated on that location before it was presented. This resulted in very low lag for the simultaneous (net time) condition. Comparing the remaining conditions (Figure 3.6), lag was significantly larger in the simultaneous (per item) condition compared to the sequential condition ($t(5) = -2.88$, $p = 0.035$). This is perhaps due to the fact that the sudden onset of a single item in the sequential condition means that attention is captured by a single location, which reduces lag time – an effect that would be consistent with findings which supports the idea that attention and eye-gaze is drawn by the sudden onset of stimuli (Yantis & Jonides, 1984). By contrast, in the simultaneous (per item) condition, this exogenous capture of attention, should it occur, must be distributed across many items, and so we see the increased lag in this condition. This may be a result of participants deciding how to allocate their attention, a factor which has been shown to effect performance when participants are presented with more items simultaneously (Palmer, 1994).

Experiment 1

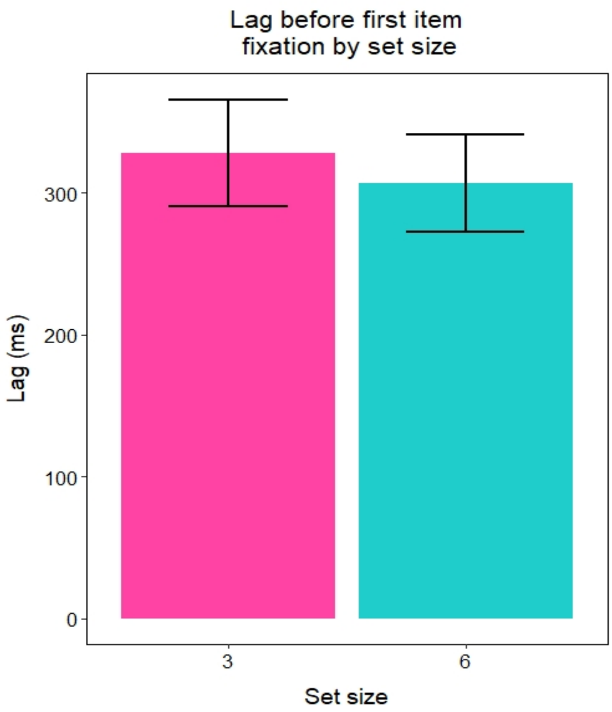


Figure 3.7. Lag in Experiment 1 by set size. There is no significant difference in lag between the set sizes. Error bars are standard error.

Like Experiment 2, we examined lag values in Experiment 1 also. Figure 3.7 shows this comparison between lag and set size. A t-test showed no significant differences in lag between the set sizes ($t(39) = 1.56$, $p = 0.12$). The average lag in this experiment is most similar to that of the simultaneous (time per item) condition in Experiment 2. Given the simultaneous presentation in this experiment, the fact the lag is most similar to the simultaneous (time per item) condition adds to the idea that presentation condition might affect the lag. Interestingly, the lack of an effect of set size potentially suggests an effect of presentation condition beyond that of simply the number of items available at once to be processed (Palmer, 1994).

Comparing both experiments, a limitation in eye-gaze data is revealed. If participants have a consistent delay between trial onset and fixation time on the first stimulus, this goes to explain some of the results. For example, the trend towards more recent fixations being more associated with accuracy in both experiments seems to be related to the fact that more stimulus fixations occur relatively later in the trial. Therefore, our measure of serial position of fixation in both experiments is limited. Additionally, the delay between the trial onset and viewing the stimulus means that the first one or two fixations in a trial will likely be on no item and instead be at the centre of the screen (where participants are encouraged to fixate between trials). This too provides a skew in the fixation data as we see in the distribution of the fixation locations across the experiment. However, these results are interesting in that they suggest that eye-gaze behaviour might be affected by whether the stimuli are presented simultaneously or sequentially. This creates a problem as it seems participants are first processing items using covert attention before making a fixation which creates a divergence of attention and eye-gaze. As such, we are unsure when a participant begins encoding a specific stimulus. This disjunction is something that we attempt to resolve in future experiments within this thesis.

Strategy

So far, we have not been able to demonstrate robust links between eye-gaze and performance in a VWM task. However, we have seen that in both experiments, fixation duration varies with the serial position of presented stimuli as well as the serial position of fixations across the trial. From our analysis of lag in both experiments, we see that the onset of the first stimulus fixation might be affected by whether items are presented sequentially or simultaneously. This suggests that features of the task can alter how attention is distributed in a VWM task. We wanted to examine this phenomenon in more detail to see if we could detect other systematic variations in attention. In particular, we were interested to see if people altered their behaviour over time and if there was evidence of individual strategies within our experiments.

Learning

One way in which attention could vary across an experiment could be over time. As participants become familiar with the task, they may adapt to perform the task more accurately or with better efficiency. We examined this possibility for both experiments.

Experiment 1

To assess learning we first compared the average deviation per block across the whole experiment. For this analysis we restricted our data to the unblocked condition only. This was because the nature of the blocked condition was such that participants would experience one set size for the first half of the experiment and then the second set size for the remainder of the experiment Given that we counterbalance which set size appeared first and given the fact that set size significantly effects deviation, any effect of learning over the experiment may be overshadowed by the set size effects. So, looking at only unblocked trials, we plot deviation averaged over participants, broken down by block (30 trials) for each set size in Figure 3.8. We see that deviation remains consistently low for set size 3 across the experiment, while for set size 6, there appears to be a slight decrease in deviation – an increase in accuracy –

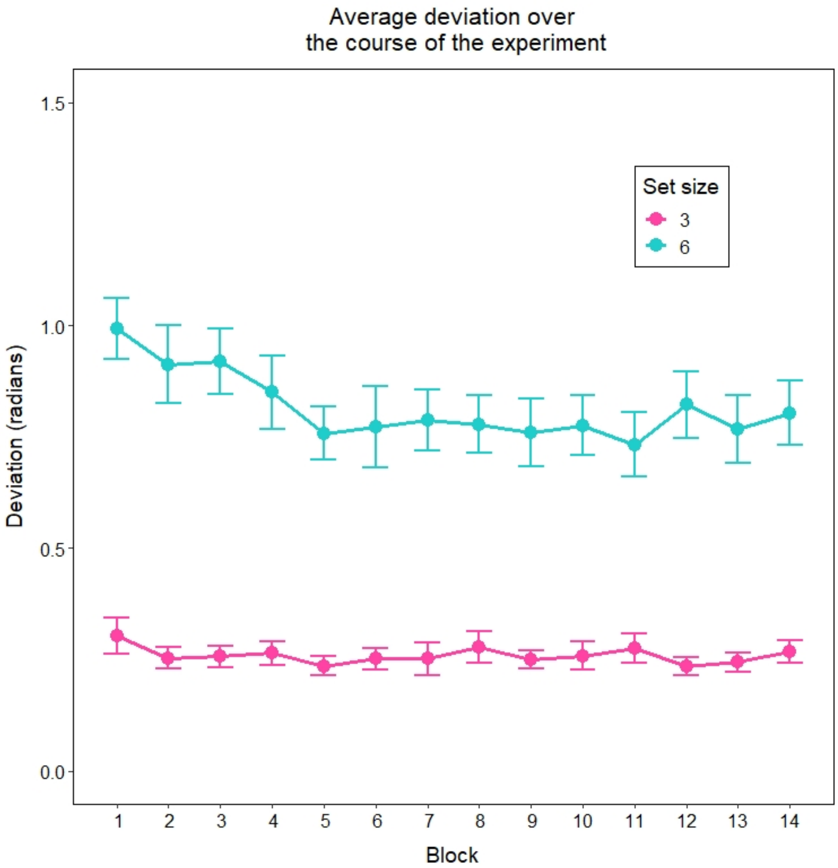


Figure 3.8. Deviation averaged by block over the course of Experiment 1. Unblocked condition only, error bars are standard error.

A two-way 2 (set size) x 14 (block) ANOVA revealed that while there was a main effect of set size ($F(1,19) = 327.15$, $p < 0.001$) there was no main effect of block on deviation ($F(13,247) = 2.67$, $p = 0.10$). There was however, a significant interaction, suggesting that the effect of block is different for the different set sizes ($F(13,247) = 6.48$, $p = 0.01$). A follow up one-way ANOVA on set size 6 only revealed a significant effect of block on deviation ($F(13,247) = 1.78$, $p = 0.047$). A similar one-way ANOVA on the set size 3 data showed no effect of block on deviation ($F(13,247) = 0.86$, $p = 0.59$). Together, this suggests that there is some learning effect for set size 6 in the unblocked condition but not for set size 3.

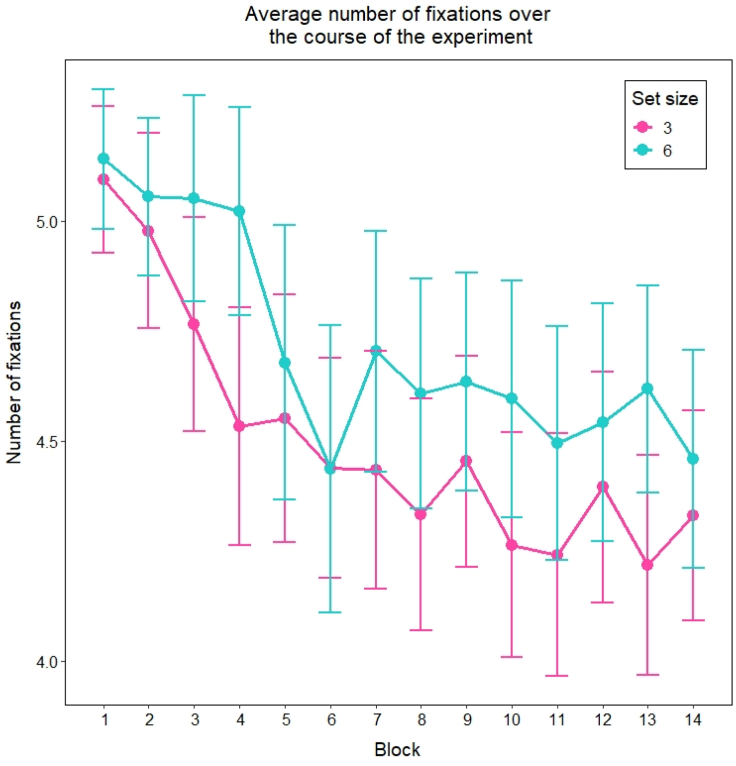


Figure 3.9. Average number of fixations per block. Aggregated over participants. Unblocked condition only, error bars are standard error.

Following our analysis of deviation, we examined the average number of fixations on stimuli, averaged by participant, per block across the experiment. A 2 (set size) x 14 (block) ANOVA revealed no significant main effect of set size on the number of fixations ($F(1,19) = 3.40$, $p = 0.07$), nor was there a significant interaction ($F(13,247) = 0.49$, $p = 0.48$). However, there was a main effect of block on the number of fixations ($F(13,247) = 14.70$, $p < 0.001$). As Figure 3.9 demonstrates, the average number of fixations decreased across the experiment.

These results suggest that over time, participants change their eye-gaze strategy to suit the demands of the task. As seen in Figure 3.10, there does appear to be a trend that participants with lower average deviation (more accurate responses) are also those who made fewer fixations on average. However, the correlation between average deviation and number of fixations for each participant was not significant ($t(18) = 1.70$, $p = 0.11$).

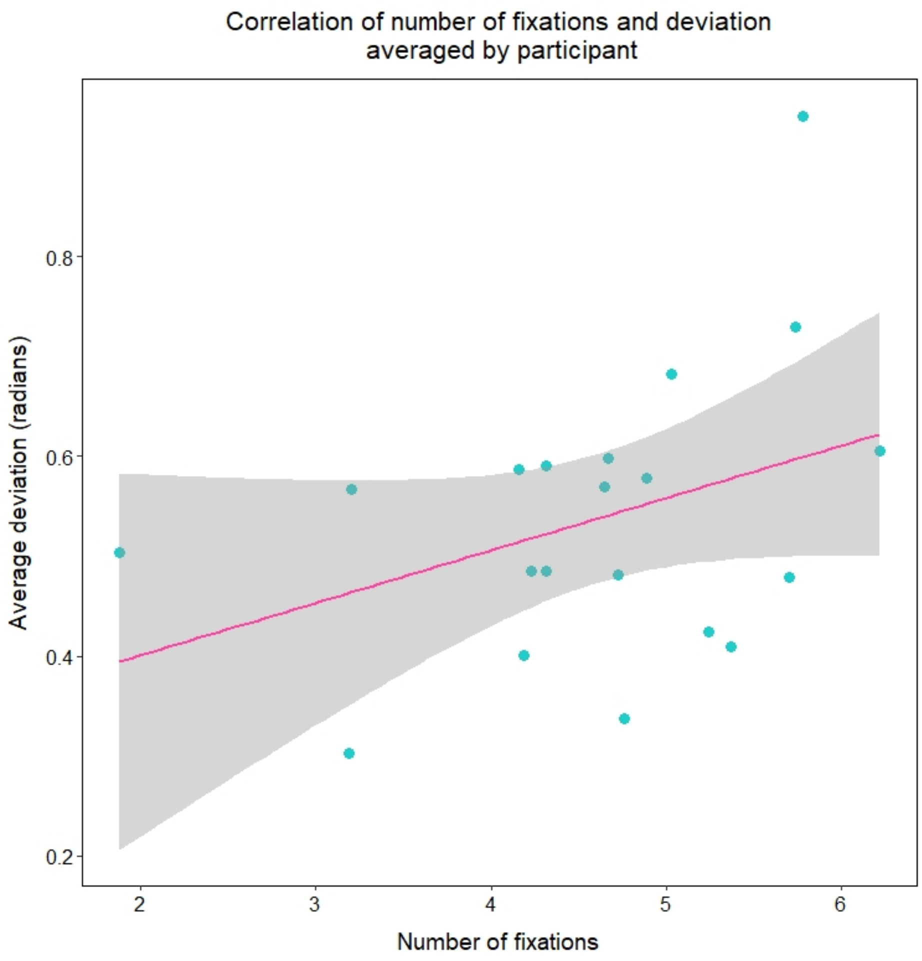


Figure 3.10. The correlation of average deviation and average number of fixations by participant in the unblocked condition only. The pink line is a linear fit with shaded regions showing error.

While the correlation between response accuracy and number of fixations did not reach significance, we do see both decrease over time. One possibility is that there is a direct relationship between these two factors. However, it is also possible that people get better at visual working memory tasks, and they also learn to become more efficient with their deployment of covert attention – requiring less eye movements without compromising accuracy. In either case, we see a change in how attention is used in response to task familiarity. This is somewhat inconsistent with other findings suggesting that participants do not adapt to the task and instead stick with a consistent strategy over time, though our participants seem to have settled into a relatively stable pattern after 150 trials or so (Cusack et al. 2009; Vogel et al., 2005).

Experiment 2

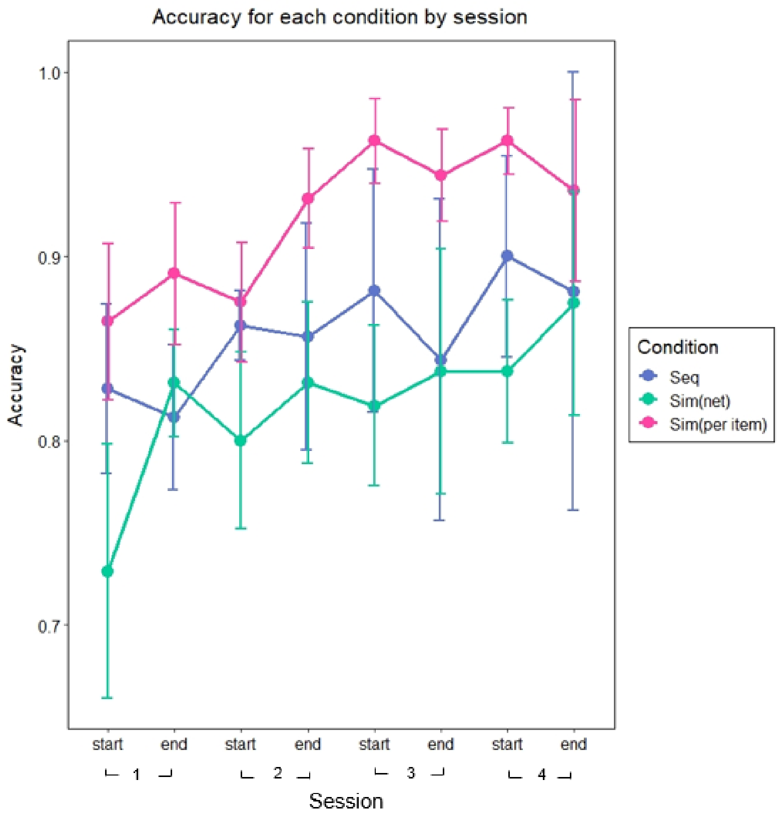


Figure 3.11. The average accuracy over time (trial and session) for each of the 6 participants by block. Error bars are standard error.

To assess learning in Experiment 2, we examined the average deviation and number of fixations for the first and last block (of 32 trials) for each session. Participants completed four sessions 4 in each condition. Figure 3.11 shows average deviation for the beginning and end of each session, separated by condition. Generally, there is a trend towards more accurate responses in all conditions over the course of the experiment. A two-way 3 (condition) x 8 (block) ANOVA found no significant main effect of condition ($F(2,10) = 1.61$, $p = 0.21$), of block ($F(7,35) = 1.53$, $p = 0.22$), nor was there a significant interaction($F(14,70) = 0.34$, $p = 0.57$). These results suggest that there is no clear evidence of learning when it comes to task performance, though the small sample size is unlikely to have yielded considerable statistical power to detect all but the largest effects.

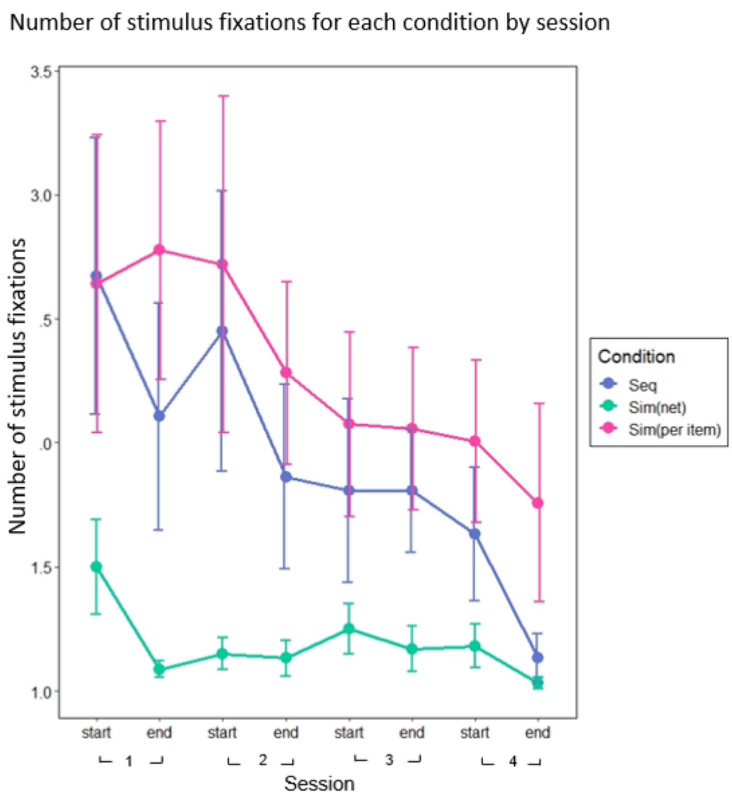


Figure 3.12. The number of fixated stimuli per trial over time (trial and session) for each of the 6 participants. Error bars are standard error.

We also examined the number of stimulus fixations over the course of the experiment. As before, data was averaged for the start and the end of each session, for each condition. Figure 3.12 shows the average number of stimulus fixations for the first and last block of each session. Over the course of the whole experiment, the average number of stimulus fixations per trial tends to decrease.

Given the sparse fixation data in the simultaneous (net time) condition relative to the other conditions, we decide to exclude this condition from the analysis. A 2 (condition) x 8 (block) ANOVA found no significant main effect of condition ($F(1,5) = 0.19$, $p = 0.66$) nor was there a significant interaction ($F(7,35) = 2.93$, $p = 0.09$). There was a significant main effect of block ($F(7,35) = 9.82$, $p = 0.002$). This result suggests that, in the sequential and the simultaneous (time per item) conditions, the number of stimulus fixations decreases over the course of the experiment.

Similar to Experiment 1, the Experiment 2 data shows evidence of a change in how people attend to the experiment over time in that the number of stimulus fixations made decreases. The increase in accuracy over blocks was not significant, though the average accuracy rates do increase. With only six participants, it is difficult to say whether we should expect any increase to replicate. However, it does seem less controversial to say that we would expect participants to become more efficient, by reducing eye movements, over the course of a future experiment.

Covert vs. Overt

In previous visual search tasks, it has been found that participants often have a preference for either a covert or overt search strategy (Boot et al., 2009). For example, in their task with dynamic displays, Boot et al. (2009) found that participants maintained a stable preference for overt or covert use of attention across 4 different tasks. Importantly, this preference persisted even if a particular strategy was not optimal for the task.

While this type of strategy has not been demonstrated directly in the continuous report VWM tasks used in the current project, generally there seems to be a link between the control of attention and VWM (Fougnie, 2008; Vogel et al., 2005). This has also been linked to visual search through experiments such as one by Woodman and Luck (2007), in which participants had to perform a visual search task while holding another item in VWM. It was found that holding the item in VWM made participants better at ignoring the memory target when it appeared in the task as a distractor. We wanted to see if this link between visual search and VWM was carried into covert and overt task strategy.

Experiment 1

Figure 3.13. the average deviation in response depending on strategy. Participants were divided into covert or overt participants based on their average number of fixations. Error bars are standard error.

We categorised participants as overt or covert based on their average number of unique stimulus fixations. Examining the distribution of average unique stimulus fixations, there was a cluster of participants who fixated on more than two unique items per trial and a smaller number who consistently fixated on fewer than 2 items per trial. Therefore, we classified participants who averaged more than two unique item fixations as overt (N = 27) and those with fewer than two as covert (N=13). Figure 3.13 shows a plot of average deviation for overt and covert participants, revealing little difference in accuracy between the two categories. Indeed, this was supported by a two-sided t-test that showed no significant difference between the groups ($t(38) = -0.31$, $p = 0.76$).

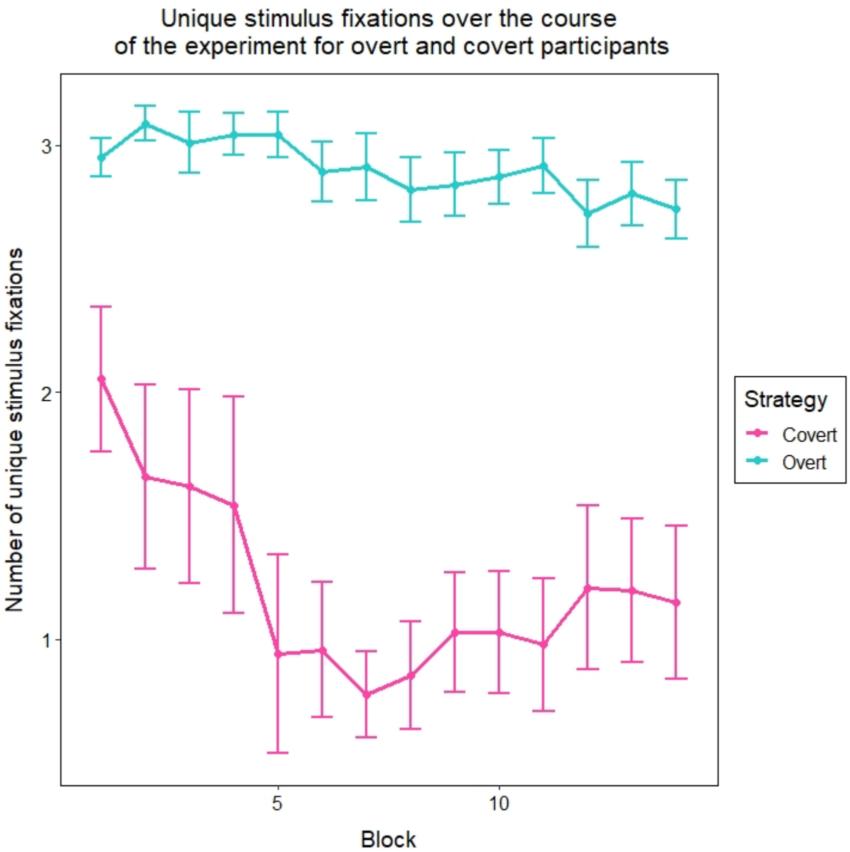


Figure 3.14. the average deviation in response depending on strategy. Participants were divided into covert or overt participants based on their average number of fixations. Error bars are standard error.

While strategy did not have an effect on accuracy, we were curious to see if strategy was stable over time as suggested by Boot et al. (2009). To do this we compared the number of unique stimulus fixations over the course of the experiment for the overt and covert groups. Once again, and for the same reasons as earlier, we only included the unblocked condition in this analysis. Shown in Figure 3.14, the average number of unique stimulus fixations was relatively consistent across blocks in the overt group relative to the covert group.

A two-way ANOVA of 2 (strategy) x 14 (block) showed significant main effects of strategy type ($F(1,18) = 53.95$, $p < 0.001$) and block ($F(13,57) = 7.83$, $p < 0.001$) as well a significant interaction ($F(13,57) = 4.84$, $p = 0.004$). This suggests a different effect of block on the different strategy groups. Of course, the main effect of strategy should be taken with a grain of salt, given that the two groups were created based on their difference on this dependent variable. However, we are interested in this analysis as it speaks to the stability of these strategies. A follow up one-way ANOVA on block for each category revealed that the number of unique stimulus fixations is different over the course of the experiment for both the overt category ($F(13,78) = 4.52$, $p < 0.001$) and the covert category ($F(13,156) = 3.06$, $p < 0.001$). In both the overt and covert groups, participants on average show a reduction in the number of unique stimulus fixations over time. This suggests, somewhat counter to Boot et al. (2009), that individuals do shift their behaviour somewhat over the course of the experiment. However, the decrease in the number of fixations seems larger for the covert participants, who, if anything, appear to increase their reliability on covert fixations.

We also examined individual differences in the number of unique items fixated across Experiment 1. Like our analysis of Experiment 2 in Chapter 2, we found that within individuals there was considerable consistency in terms of the number of unique fixations they made across the experiment. In Figure 3.15, towards the bottom of the figure we see some participants that tend to make no fixations on any of the stimuli. Moving up the plot, we see other participants make consistently few fixations, while some people tended to make many fixations throughout.

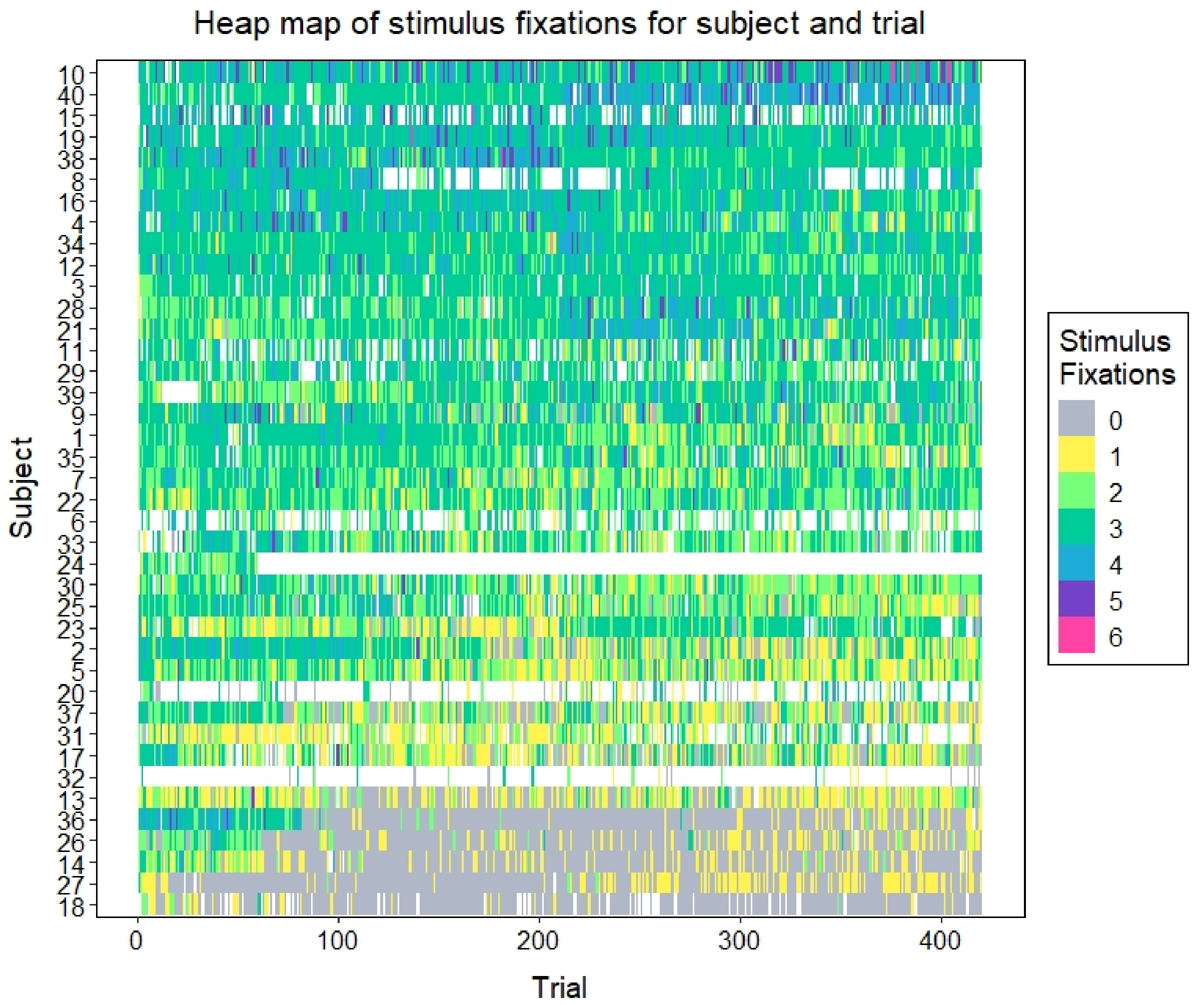


Figure 3.15. Heat maps displaying the number of unique stimulus fixations for each trial, for each subject. Subjects are ordered by there average number of unique stimulus fixations.

Most notable from Figure 3.15 is that there is a wide degree of individual variation as to how participants utilise their attention in a simple VWM task. Clearly, some participants rely strictly on covert attention, while some fixate only a subset of the items in the trial and some prefer to fixate every item in the display. Whether and how such different approaches to overt and covert attention can be mapped onto the various slots or resources accounts of VWM will be discussed later in the thesis.

Discussion

This chapter has been an attempt to make sense of the eye-gaze data from Experiments 1 and 2. Overall, the tendency for participants to rely on covert attention in our tasks mean that our conclusions must be modest, but there are some things we can learn from this analysis to help us guide the design of future experiments.

First, we examined our data using regression to look for eye-gaze variables that would predict latent working memory variables (Experiment 1) and task performance (both Experiment 1 and 2). In both experiments, it was clear that set size was the main significant predictor of task performance beyond either of the eye-gaze variables tested. At the outset of this project, we believed that target fixation duration and the number of items fixated would not only be indicative of task performance, but we also believed that these variables may even help us distinguish between slots and resources accounts of VWM. Given the lack of effect of these variables on task performance, or the latent VWM variables from Experiment 1, our results mean we are unable to make a strong claim using eye-gaze data about how encoding impacts working memory as well as the nature of VWM.

Next, in light of our findings in Experiment 2, we revisited the eye-gaze data in Experiment 1. Like Experiment 2, we found that participants tend to fixate longer on the first stimulus they look at compared to other stimuli. Subsequent stimuli are fixated for progressively shorter amounts of time. The average number of fixations increased between set sizes 3 and 6 demonstrating that participants can adjust their eye movements in an attempt to meet task demands. However, in both set-size conditions it is most common for participants to fixate on three items in both the 3-item and 6-item trials. Similar to Experiment 2, we found that rarely do participants fixate on all the items in the 6-items trials.

In contrast to Experiment 2, we found that individual participants in Experiment 1 were much more likely to make fixations on stimuli. Similar to what was seen in Experiment 2, there were consistent individual differences in the number of fixations participant made per trial. Like Experiment 2, there were some participants that preferred to not make any fixations on the vast majority of trials suggesting that they were utilising covert attention during the task. Even taking into account the relatively small number of participants in Experiment 2, the difference in behaviour may be attributable to the difference in experiment designs. While the trial duration in Experiment 1 was longer, it was not longer by enough to warrant such a stark difference in fixation preference alone (e.g., trials in Experiment 2 with four items had a duration of 800ms in the sequential and simultaneous conditions). Therefore, it may be the case that design features of our second experiment that we thought would increase eye movements actually discouraged them. For example, by having items more spatially distant by placing them in fixed locations on the screen, we hoped participants would be encouraged to fixate more. Instead, it seems as though this may have caused participants to judge seeing all the items in the display as “too difficult” given the time restriction and as such, decided to use covert attention instead. Alternatively, the pentagon stimuli may have facilitated a more ‘holistic’ representation, or made covert attention too effective a strategy. Future work would have to distinguish between such possible explanations.

Our analysis of serial position of fixation in Experiment 1 suggested that accuracy increased the more recently an item was seen during a trial. Such a recency effect of serial position of fixation has been found in other studies (Zelinsky & Loschky, 2005; Irwin & Zelinsky, 2002). This finding contrasts with the similar analysis in Experiment 2 in which there was no clear effect of serial position of fixation. However, given the lack of fixation data in general for Experiment 2, a strong conclusion about the difference here is difficult to draw. The most likely cause would be the difference in trial time allowing for more fixations to be make in Experiment 1 and therefore making accuracy differences for the various serial positions observable. This is expanded upon further in the General Discussion (page 176).

In performing a follow-up analysis, we found that typically there is a lag between the onset of a trial and the first fixation on a stimulus. This was found in both Experiments 1 and 2, where there was typically a 300ms lag between the start of the trial and the first fixation on an item. In addition, data from the sequential-presentation condition in Experiment 2 revealed that often participants often fixated on items when they were no longer present. This led us to speculate that the simultaneous condition might in fact be easier, as the eye-gaze window for when items are onscreen is not as uncertain, and thus participants are more able to make fixations on items while the stimuli are present. Future experiments using sequential presentation of items in combination with eye-gaze measures would need to calibrate presentation times in light of this delay between onset and fixation.

We speculate the tendency for people to delay in their first stimulus fixation because they are to some extent at least relying on their peripheral vision to encode the environment. The lack of fixation data from two experiments suggests this to be true, however it is also supported by the visual search literature that suggests at least some participants prefer covert strategies (Boot et al., 2009; Cusack et al., 2009; Boot et al., 2006;Vogel et al., 2005). We were curious to see if such a distinction could be made in our data and indeed there does appear to be a distinction (in both experiments) between people who tend to fixate on items, and those that rarely move their eyes at all. However, unlike previous experiments (e.g., Bengson & Luck, 2016), we found that people did adapt their strategies to task demands, at least to some extent, over time.

The possibility that participants can use both covert and overt strategies to perform VWM tasks is an interesting one, as our existing studies leave us unsure to what extent participants are using their periphery to aide working memory. From our analysis of covert and overt strategies, we saw that there was an element of participant preference, but that this could be altered over time with the task. The comparison of the sequential and simultaneous (time per item) conditions in Experiment 2 also leads us to an interesting idea that even though the conditions are matched in terms of stimulus numbers and time exposure, the simultaneous (time per item) condition seems to be easier for participants than the sequential condition. This can be seen in our analysis of effective set size (Figure 2.5, Chapter 2) where the effective set sizes for the simultaneous (time per item) condition were lower than the sequential conditions, suggesting this condition is less difficult for participants in general. It is possible that a simultaneous display might aide a remember-all strategy that has been shown to be beneficial in a previous task by Bengson and Luck (2016). In this experiment, the authors note that perhaps the reason why these instructions are beneficial is because they encourage participants to use an approach that incorporates some aspect of statistical learning regarding the whole display (Bays et al. 2011).

Going forward, we want to continue to investigate the connection between eye-gaze and VWM in a design that allows for more robust eye-gaze data. At the same time, we are interested in these preliminary findings about the role over covert strategy and peripheral vision in VWM tasks. In order to create such an experiment, we needed more control in ensuring participants see the stimuli directly (as well as indirectly). For all of the above reasons, we decided to proceed with our investigations using gaze-contingent designs.

Chapter 4: attention and strategy in VWM using a gaze-contingent paradigm

The most obvious shared result across the first two experiments was the lack of insight about visual working memory processes gained from looking at the relationship between eye-gaze data and task performance. Participants appear to be able to complete VWM tasks without utilising overt attention. This seems to be the limiting factor in our investigation as we lack the data to demonstrate a link between eye-gaze, attention and VWM. However, logic dictates that vision must play some type of role VWM, even if participants are reluctant to fixate on individual items in these more standard tasks. Our analysis in Chapter 3 suggested that people may learn to move their eyes less over time to complete the experiment more efficiently. However, while this behaviour is more efficient, it obfuscates the relationship between accuracy and fixation. For this reason, we wanted to continue to investigate the link between VWM and eye-gaze and to try to gain a better understanding of the strategy participants were using in these tasks and why.

Gaze-contingent paradigms

Thus far we have used passive eye tracking as our measure of attention. In such tasks, though the eye tracker records eye position throughout the experiment, where exactly a participant looks has no outcome on anything that happens in the experiment. A gaze-contingent paradigm, on the other hand, may select, restrict, or probe stimuli based on participant fixations. As an example, Zelinsky and Loschky (2005) used a gaze-contingent paradigm in their VWM experiment. In each trial, a predetermined target was selected from the array of study items at random. Participants were allowed to freely view the array of items (which were presented simultaneously) and their gaze was monitored. Once they fixated on the target, the experiment was set to terminate the trial after 1-7 (depending on condition) intervening fixations on other items in the study array. Thus, when the trial ended was contingent upon when the target was fixated by the participant. The advantage of this paradigm over passive eye tracking is that it offers the ability to control what happens in the trial based on the eye movements of the participant or to more closely control what is and is not seen.

The gaze-contingent paradigm has a long history in reading experiments such as those used to determine the span of characters scanned in each fixation during reading (McConkie & Rayner, 1975). More recently, gaze-contingent paradigms have been used with learning experiments to probe or record the specific stimuli that have been attended (Beesley, Pearson, & Le Pelley, 2015), or as a measure of attentional capture (Le Pelley et al., 2015).

In our first two experiments, it seems that participants encoded the study items in each trial, as they were able to perform the task relatively accurately. However, they typically fixated only a subset of the items in the trial, meaning that direct fixation on the target was less common, and thus restricted what we could learn about the links between eye-gaze and task performance. Because the gaze-contingent design can respond to what the participant is fixating on in the moment, we wanted to employ this design to restrict what the participant saw in a trial in a way that would elicit more eye movements than our previous experiments. This could be done by initially covering the study items and requiring the participants to look directly at an item's location before its features would be made available for study. With this design, items could only be seen when they were being looked at directly. Thus, participants would not be able to use their peripheral vision in the same way they did with passive eye tracking.

Design benefits

In our previous experiments, we have seen that participants prefer to fixate the centre of the screen, relying on covert rather than overt attention. A key benefit of this gaze-contingent design is the inability for participants to use their peripheral vision therefore preventing the use of covert attention. As participants must fixate on items to see them (i.e. use overt attention), it seems impossible for this task to not produce a clear link between performance on a VWM task and fixation. However, we use a gaze-contingent task because the nature of the link may be theoretically revealing. Firstly, this task could demonstrate clearer delineations of eye-gaze behaviour in terms of what is expected by slots and resources models. Claims about variable-precision memory models (Van den Berg et al., 2012) are complicated by the potential existence of guessing. One explanation for why guesses occur is because the stimuli are not encoded. Our gaze-contingent task gives us a proxy measure for one way that items may not be encoded (i.e., not fixated). So, though it may be possible that fixated items are not encoded, it may be that slot-like accounts imply that we should not see guess-like responses for fixated items, while that observation would be expected under a resource-like account. Similarly, by forcing the relationship between gaze and encoding, we can ask whether eye-gaze variables predict the accuracy of people's response in a 'continuous' manner. For example, should we find that longer fixations on an item yield more accurate responses, then we have reason to believe that visual working memories are not merely all-or-none (where 'all' usually refers to a fixed, but high precision representation).

In addition to task-related performance, a clear sequence of fixations could assist us in beginning to understand participants' strategies in VWM tasks. While we have provided some analysis and discussion of strategy in our first two experiments, the fact that covert strategies are used frequently limits what we can determine with eye-gaze data. By compelling participants to fixate on specific items we hope that we might start to get an understanding of the overt strategies participants may use in an VWM task, which may help us to speculate about the covert ones. Even within the behavioural strategies that are suggested by certain accounts of VWM (Donkin et al. 2016; Zhang & Luck, 2011), we may expect participants to employ systematic ways of exploring the scene. For example, in the literature of visual search, it is common to see participants use a coarse-to-fine method of exploring a display in which initial fixations are short and far apart while subsequent fixations are longer and more spatially close (Over et al., 2007). Though we do not expect to see this particular pattern, since it seems counterproductive to the current task, such a finding does suggest that we may see systematic exploration patterns.

Finally, it is possible that a gaze-contingent design could disrupt the typical behaviour seen in continuous-report tasks. Evidence from visual search suggests that the number of fixations made in search tasks is optimal for the gradual drop off in visual acuity from the fovea to the periphery of vision, and thus the periphery helps map out scenes (Geisler, Perry, & Najemnik, 2006). Likewise, in studies of context cuing the learning of the displays is disrupted when participants are restricted to viewing a small window of the display (Zang, Jia, Müller, & Shi, 2015). Given these examples, it is possible therefore that restricting periphery information might greatly reduce task performance, since it appears that periphery information is useful in encoding the display. While a decrease in task performance may not seem like a 'benefit' of this design, it would prove highly diagnostic in determining the involvement of peripheral vision in VWM tasks.

The current experiment

To create a comparison between a gaze-contingent experiment and a standard VWM task, we decided that Experiment 3 would be a gaze-contingent replication of Experiment 1. Our experiment was gaze-contingent in that it restricted viewing of the study items. The location of the study items was cued on screen and participants were required to fixate on one of these locations in order to reveal the stimulus they needed to remember. Given that participants tend to utilise peripheral vision where they can in the more standard tasks, it is possible that their approach to a gaze-contingent task may be quite different to what is typically observed. For this reason, we wished to use stimuli and a task design we had used before to create a point of comparison if any such differences were to arise.

We used the same ring and bead stimuli with set sizes 3 and 6 and items would be presented equidistant from each other and from the centre of the screen as they were in the first experiment. We also wanted to include the blocked and unblocked condition that were presented in the first experiment. As a reminder, this was a between-subjects condition in which trials varied by set size randomly (unblocked) or the first half of the experiment contained trials of only one size, with the final half of the experiment containing trials of only the other set size (blocked). It is possible that participants' use of peripheral vision in Experiment 1 prevented us from observing eye-gaze behaviour in line with different strategies for the blocked and unblocked conditions. This experiment, therefore, can be viewed as another attempt to see the anticipated results implied by Donkin et al. (2016), who argued that, for larger set sizes, there would be more stimulus fixations with less average duration in the blocked condition compared to the unblocked condition, since anticipating a large number of study items could prompt more spreading of attention. In Experiment 3, we wanted to use gaze contingency such that the study items could only be seen when they were fixated on directly. Thus, if the blocking and unblocking conditions provoked different attentional responses to the task, we should see a difference in overt looking behaviour.

Following some piloting we found that the gaze-contingent nature of the task was more difficult in that participants needed to exert more effort in order to more actively control their eye movements. Due to this, we increased the trials duration from what it was in Experiment 1 (3000ms compared to 1000ms in the first experiment). Also, though our discussion in Chapter 1 did raise concerns with the simplicity of the stimuli used in Experiment 1, given the increased challenges with the gaze-contingent component, we believed that the rind-and-bead stimuli would provide an appropriate difficultly for this task, and allows for an easier comparison between Experiments 1 and 3.

Method

Participants. 61 first year psychology students from UNSW agreed to participate for course credit. Participants were recruited using the student-specific experiment sign-up system, SONA. Approval for this study was obtained from the UNSW Human Research Ethics Approval Panel – C (HREAP-C).

Apparatus. A Tobii Pro Spectrum 600 mounted on a 23.8-inch widescreen monitor (1,920 x 1,080 resolution, refresh rate 60 Hz) was used for tracking eye movements and displaying the experiment. Participants were seated 40cm away from the screen. A chinrest was used to help participants keep their heads still.

Stimuli. The stimuli were identical to those used in Experiment 1: 120 pixels (visual angle = 3.03°) diameter rings with 2 pixels (visual angle = 0.05°) diameter beads on them (see Figure 1.2, page 28). As before stimuli varied in colour. Stimuli were presented on an invisible circle 600 pixels (visual angle = 15.1°), centred in the middle of the screen, and such that each item was equidistant from the others. (For a full description refer back to Experiment 1, page 27). During each trial, these stimuli were hidden from the participant until the eye tracker registered direct fixation on the item's location. The location of each of the stimuli was shown by a pink circle 160 pixels in diameter (visual angle = 4°), centred on each stimulus (as seen in Figure 4.1). Fixations on the pink circle were counted as fixations on the stimulus. As soon as gaze was detected within the area of a pink circle, the stimulus at that location would be revealed. Similarly, if eye-gaze moved out of the pink circle's area, the pink circle would reappear and the stimulus would be removed from view. Participants were able to freely view the items and were able to revisit items without limit within the duration of the trial.

Design. The design of this experiment was a replication of Experiment 1 with the addition of gaze-contingency. All details, unless otherwise mentioned, are identical to those in Experiment 1. Set size of the stimuli was either 3 or 6 items per trial with an equal amount of each across the whole experiment. As previously, the blocking manipulation varied between subjects. Participants either were in the unblocked condition (set size on any give trial is random) or in the blocked condition (first half of the experiment is one set size, the second half is only the other). Trials were limited to being 3000ms in duration, longer than the 1000ms used in Experiment 1, because pilot tests indicated that durations shorter than 3000ms made it difficult to manage the gaze-contingent aspect of the task. This change resulted in fewer trials compared to Experiment 1: 300 total trials with 10 blocks of 30, down from 420 total trials.

Procedure. Before beginning the main task, participants were given both verbal and written instructions on how to do the task as well as some practice trials to adjust to the experiment. On each trial, participants were shown a fixation cross (500ms), a brief interval of a black screen (400ms) and then the stimulus screen was presented for 3000ms. Each trial could contain either 3 or 6 items presented simultaneously. Each item was obscured from the participant by a pink patch at the location of each item. The pink patches were visible for the entire length of the stimulus presentation window. To view a study item, participants were required to fixate their gaze on a pink patch. When they did so the stimulus would be revealed to them. Participants were asked to try to remember the location of the bead on the ring for each of the stimuli. Once they stopped fixating on an item's location, the pink patch would reappear to obscure the stimulus once again. During the 3000ms presentation window, participants were freely able to "uncover" as many or as few items as they wanted. Participants were able to look at the same item multiple times if they wished. There was also no restriction on how long participants could look at one item. For example, if a participant wished to look at one item for the whole 3000ms, they were able to do so.

After the presentation interval, all items were removed and a mask was displayed for 500ms. A stimulus was randomly selected to be the test item and this item's ring was presented at the location it appeared during study.

The participant then needed to use the mouse to place the bead on the ring where they remembered it to be during study. They then confirmed their selection with the space bar. After answering, participants received feedback for 1000ms. This feedback depended on how accurate their answer was ("OUTSTANDING!" for deviations less than 10°, "Very good!" between 10° and 20°, "Good" between 20° and 35°, "OK" between 35° and 45° and deviations greater than 45° were labelled with "Poor"). The trial sequence is depicted in Figure 4.1. After each block, participants were given a break in which they rested for a minimum of 20 second before continuing.

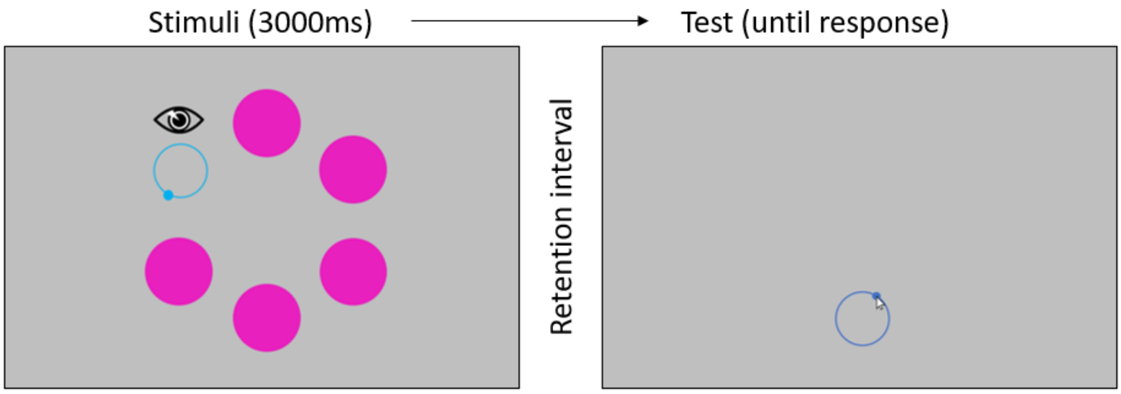


Figure 4.1. The trial sequence for Experiment 3. Participants are presented with a stimulus array that is initially hidden from view. The locations of the stimuli are indicated by the filled pink circles. The image of the eye is representative of the participant looking at the stimulus below it in the diagram. When looked at directly, the pink circle disappears revealing a ring and bead stimulus.

Results

Across all participants, 18300 trials were recorded. Trials were censored with a similar protocol as they were in Experiment 1. We removed trials that recorded more than 30 fixations during the presentation window (958 trials, 5.2%). Note that more fixations were permitted owing to the longer stimulus presentation window in this experiment compared to the first. We also removed trials if they had an average fixation duration less than 100ms (1,358 trials, 7.4%) (Liversedge et al., 2004; Rayner et al., 2009). Trials were also removed if they had no eye-gaze data (471 trials, 2.6%). One trial was removed due to no response from the participant being recorded. This resulted in a total of 2789 trials being removed (15.2% of the data). Compared to Experiment 1, a lower percentage of trials were removed due to the average fixation duration being less than 100ms. This is another indication that this method encouraged participants to change their fixation behaviour to suit the new paradigm.

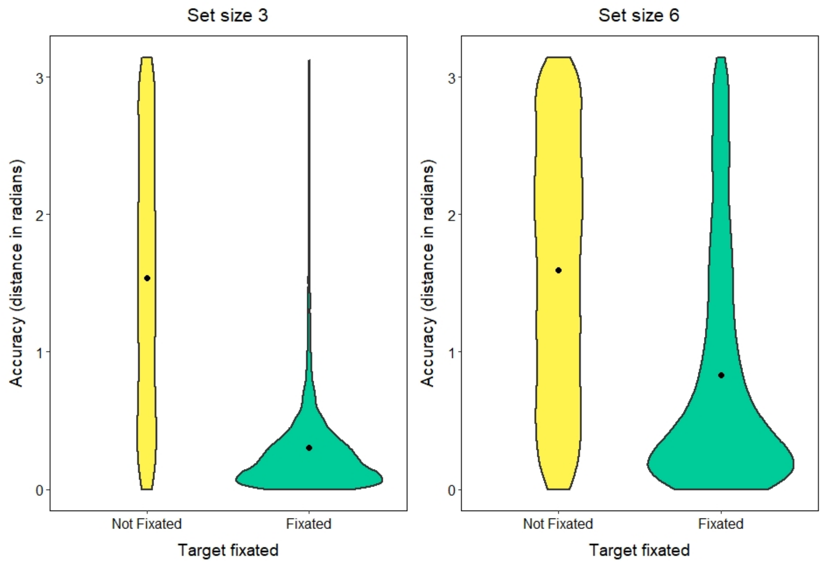


Figure 4.2. The distribution of deviation from the correct response by set size and whether the target was fixated or not. Fixating the target is associated with an increase in accuracy.

Eye-gaze and accuracy. First, we examined the connection between task performance and eye-gaze variables. Task performance was compared by examining the distribution of response accuracy in the task. We analysed accuracy separately for set sizes 3 and 6, as well as for when the target was fixated during the trial and when it was not (Figure 4.2). A two-way, 2 (target fixated) x 2 (set size) ANOVA was carried out on the absolute value of the deviation of responses from the correct response[4]. The results indicated that the response deviation was significantly higher when the target was not fixated compared to

[4] We hoped to use the hierarchical Bayesian mixture model to decompose accuracy into P_{mem} and *Prec* parameters (see Chapter 1). For example, we wrote code for a model that would predict these parameters on the basis of fixation durations. However, all such attempts to fit the model failed to work (i.e., the code did not run). With insufficient time to spend debugging or rewriting the model code, we simply abandoned our attempt to use modelling for this (and the next) chapter. While ANOVA is not an ideal method of comparison here, we decided to use it here given our failed modelling attempt and time restrictions. The reader is advised to interpret with caution.

when it was ($F(1,59) = 330.01$, $p < 0.001$). As must be trivially true in our gaze-contingent task, this result indicates that fixating an item affects accuracy. Set size was found to significantly impact response deviation ($F(1,59) = 19.12$, $p < 0.001$), with lower set size leading to more accurate responses (less deviation). The interaction, though uninteresting, was also significant ($F(1,59) = 13.50$, $p < 0.001$).

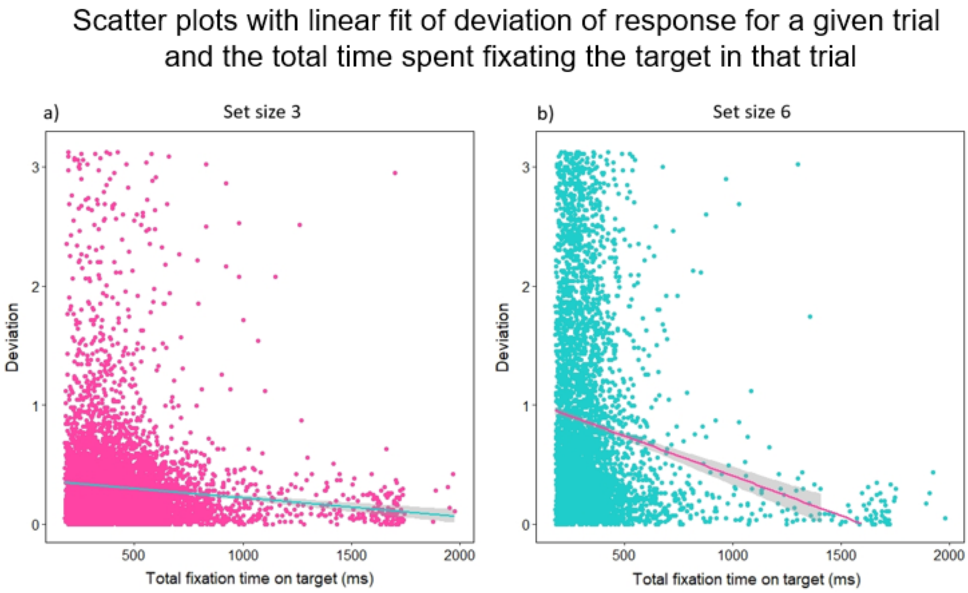


Figure 4.3. Scatter plots of deviation of response for a given trial and the total time spent fixating the target in that trial. a) set size 3 b) set size 6. Generally, more total fixation time on the target is associate with more consistently accurate responses. The line is a linear fit with shaded regions showing error.

In order to understand what was driving the increased performance due to fixation, we then examined the effect of fixation duration on response deviation. Fixation duration was measured as the total amount of time spent looking at the target in one trial. This meant that if participants looked at the target more than once, these durations were added to form the total fixation duration. As demonstrated in Figure 4.3, generally speaking, the longer the total fixation duration on the target the more likely the response was to be accurate. This relationship appeared to hold for both set size 3 (Figure 4.3a) and for set size 6 (Figure 4.3b). This correlation was more pronounced for set size 6 (r = -0.18) than for set size 3 (r = -0.11).

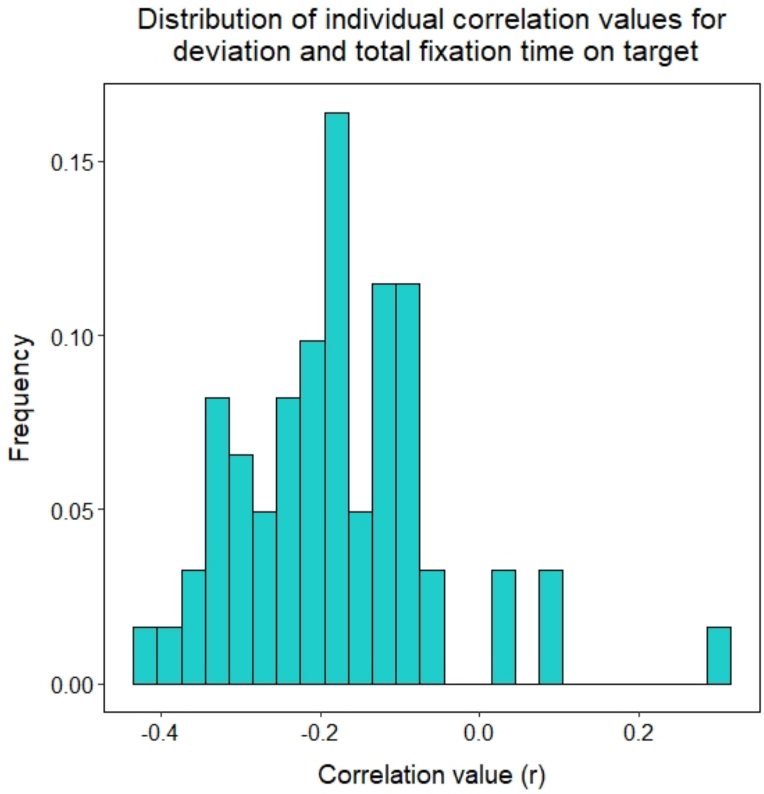


Figure 4.4. A frequency distribution of correlation values (r) for deviation and total fixation time on the target for each participant. The shape of the distribution suggests most participants show a decrease in response deviation with increased total fixation time on the target.

We examined whether this negative correlation between response deviation and total fixation time on the target held for individual participants. We calculated this correlation for each participant, and the resulting distribution of correlation values is displayed in Figure 4.4. The mass of the distribution is centred on a value of r = -0.2, suggesting that for most individuals, response deviation decreased as they spent more total time looking at the target.

Referring again to Figure 4.3, it is interesting to note that when the total fixation time on the target is 600ms or greater, responses appear to be consistently low in deviation (which is to say, more accurate). However, these longer fixations appear less common, with most of the fixations in this task (regardless of set size) appearing to be shorter than 500ms. These shorter fixations seem to be associated with much more variable accuracy in set size 6 (Figure 4.3b). Such inaccurate responding is perhaps indicative of what a slots model would usually classify as "guessing".

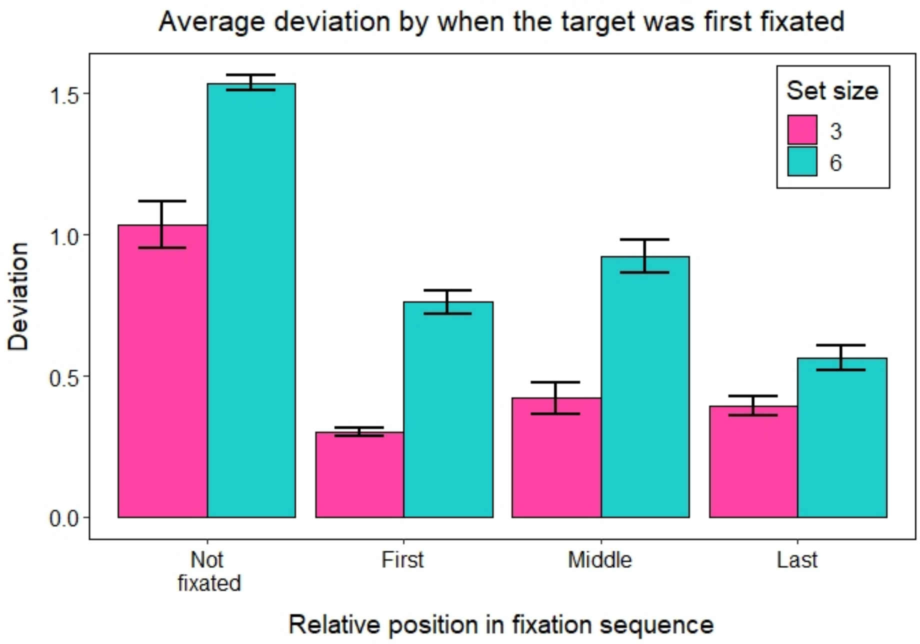


Figure 4.5. Average deviation by when the target was fixated relative to the sequence of fixations a participant made. Fixations on the target were separated by whether they were the first made, the last fixation made or whether the target fixation occurred somewhere in the middle of the sequence of fixations. Error bars show standard error.

We also examined the effect of recency on deviation by looking at accuracy as a function of when the target was last fixated in the trial. The process of this analysis was similar to that which we first used in Chapter 2 (page 65). For each trial we determined when the participant fixated on the target in their sequence of fixations. That is, whether the target was the first item they fixated, the last item they fixated or whether the target fixation was in the middle of the fixation sequence. When there was more than one target fixation in a trial, we made the categorisation based on the first target fixation[5]. Displayed in Figure 4.5, the results indicate that deviation from the correct response is smallest in set size 6 when the target is the last item seen. This suggests a recency effect such that accuracy is improved when the target item is seen closer to test in trials with 6 items. There appears to be little effect of recency for set size 3, but if anything, there appears to be a small primacy effect.

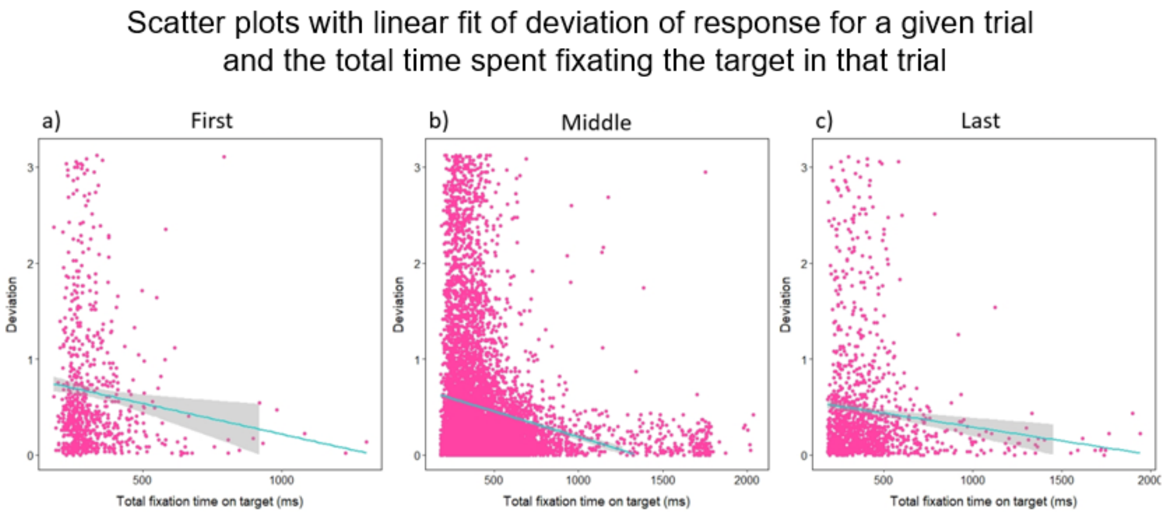


Figure 4.6. Scatter plots of deviation of response for a given trial and the total time spent fixating the target. a) contains data for when the target is the first item fixated on in a trial, b) is for when the target is fixated in the middle of the trial and c) is for when the target is the last item fixated. The blue lines are a linear fit with error bars as shaded regions.

[5] Similar to our previous uses of this analysis, we re-ran the analysis but with preferencing the most recent rather than the first instance of target fixation and found little difference to the pattern of results.

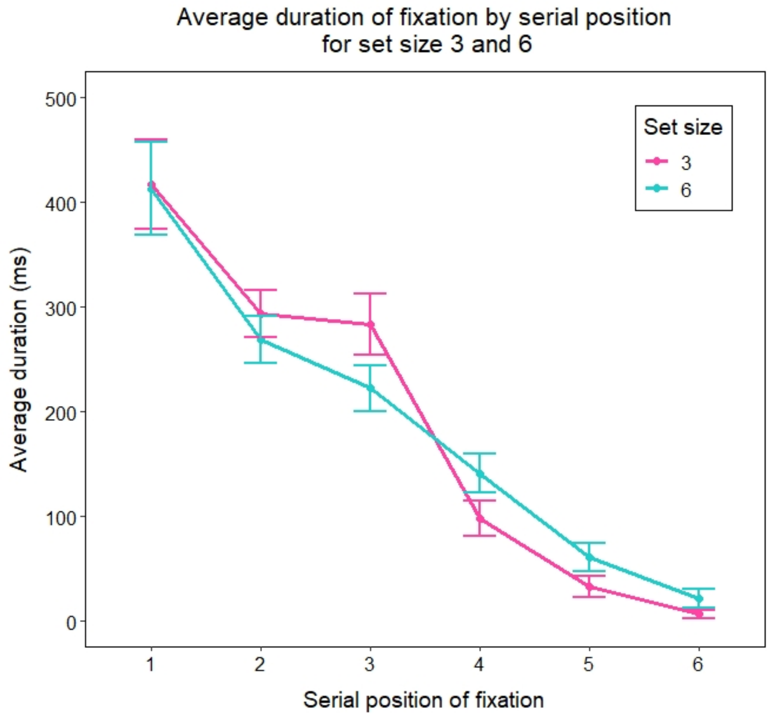


Figure 4.7. Average duration of fixation by the serial position of fixation. The first fixation is the longest with subsequent fixations shorter. Error bars show standard error.

Within our analysis, 'total fixation time on the target' and 'recency' are related measures, because the longer an item is looked at, the later in the trial the last fixation on that item will occur. As an extreme example, consider when the target is the only item that is looked at during a trial. Since our analysis favours recency, this item will be classified as the "most recent" item seen. To determine if these effects were separate, we examined accuracy by total fixation time on the target for each of the three categories bins created in the preceding analysis. The result, shown in Figure 4.6, suggests that the effect of fixation duration occurs at each of our defined points in the fixation sequence. Examining the correlation at each category reveals a negative correlation between total target fixation time and deviation when the target is the first item seen (r = -0.11), the last item seen (r = -0.11) and when it is seen within the middle of the fixation sequence (r = -0.21). This result suggests that total fixation time on the target and recency have an independent effect on accuracy within our task.

Serial position of fixation. Following from our analysis in Chapter 2 and 3, we were interested to see if a similar effect of serial position of fixation was present in the gaze-contingent experiment. Figure 4.7 displays average duration of fixation as a function of serial position of the fixation. The first fixation is, on average, the longest in duration, with subsequent fixations receiving progressively less time for both set sizes. A noticeable difference the set sizes on occurs between fixations 3 and 4. The difference in average duration between these fixations is greater for set size 3. This makes sense as following the third fixation participants would have typically seen all the items in the set size 3 condition and thus subsequent fixations are shorter (and less likely).

Blocking and unblocking. Given the connection between fixation and task performance, we were interested to see if we were able to see the blocking effect that we failed to find in Experiment 1. Figure 4.8 shows the frequency distribution of deviation for set size 3 and 6 in both the blocked and unblocked conditions. The typical differences between set sizes is observed – a peak in frequency around the correct response for both set sizes but with more accurate responses for set size 3. As seen in Figure 4.8, this effect was similar regardless of condition. A 2 (set size) x 2 (blocking condition) ANOVA on average absolute value of deviation revealed a significant effect for set size ($F(1,118) = 285.35$, $p < 0.001$), but no effect of blocking condition on accuracy nor was there an interaction (both ps

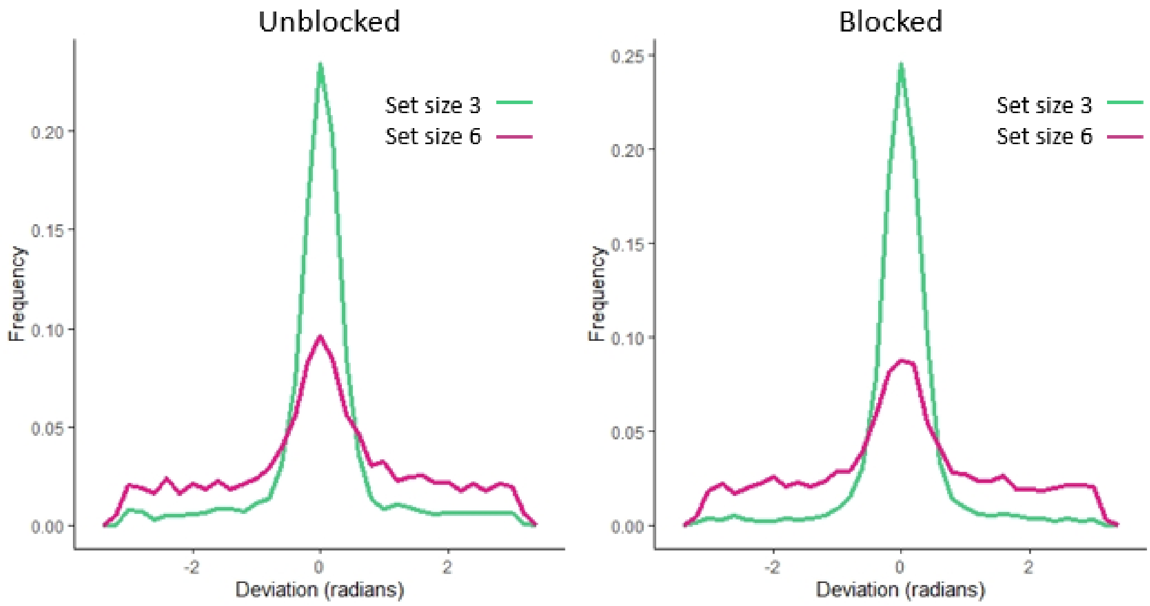


Figure 4.8. The frequency distribution for deviation for both set sizes 3 and 6 and each of the blocked and unblocked conditions.

To determine if there were attentional differences between the conditions, we compared the blocked and unblocked conditions on average number of fixations and average fixation length per trials (Figure 4.9). Examining Figure 4.9a, the average number of fixations appeared similar for the two conditions for set size 3, but for set size 6 there appeared to be fewer fixations on average in the blocked than unblocked condition. A 2 (condition) x 2 (set size) ANOVA of blocking condition and set size (grouped by participant) on the average number of fixations supported this by finding a significant interaction of condition and set size ($F(1,59) = 4.52$, $p = 0.04$). However there was no main effect of set size ($F(1,59) = 0.14$, $p = 0.71$) or of condition ($F(1,59) = 0.18.$, $p = 0.267$).

Average fixation duration appeared longer in the unblocked condition for set size 3, but was longer in the blocked condition for set size 6 (Figure 4.9b). For set size 6, it appeared that the unblocked condition had a larger number of average fixations with shorter durations; a result that was counter to our original predictions, as we expected more fixations with lower average duration in the blocked condition. This was following the results from Donkin et al. (2016) which suggested the blocked condition produced behavioural results more in line with a resources model which we believed to be more in line with a spreading of attention between items – more fixations with less average duration in the blocked condition. However, a 2 (condition) x 2 (set size) ANOVA of blocking condition and set size (grouped by participant) on the average fixation duration did not find the interaction to be significant ($F(1,59) = 0.39$, $p = 0.54$). Additionally, there was no main effect of condition ($F(1,59) = 0.87$, $p = 0.35$), but there was a main effect of set size on average duration ($F(1,59) = 522.72$, $p < 0.001$).

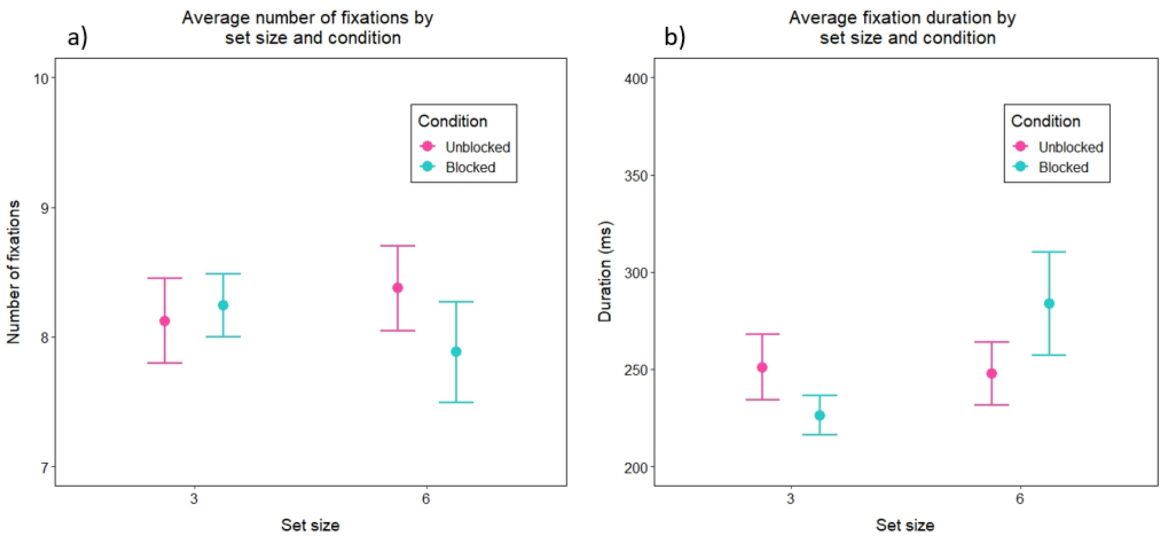


Figure 4.9. Average number of fixations and average fixation length for set sizes 3 and 6 and for the blocked and unblocked conditions. Error bars are standard error.

Strategy. In Chapter 3 we examined individual differences in eye-gaze behaviour and found that participants appear to have individual preferences as to whether they are more covert or overt with their attention when approaching a task. Beyond this, however, we were unable to examine more individual differences in task approach due to a lack of eye-gaze data. By using a gaze-contingent method as we have in the current experiment, we are better able to investigate differences in overt attention strategies. To begin our investigation, we wanted to determine if there were any consistent patterns in how participants explored the study arrays.

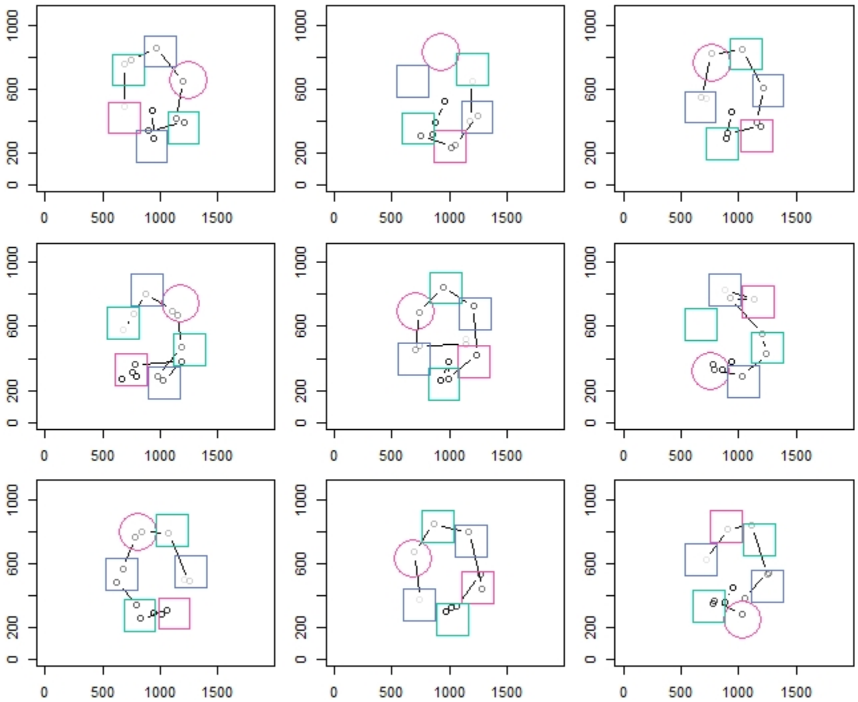


Figure 4.10. An example of a trace of a participant's eye movements across some set size 6 trials. It was common for participants to explore the scene in a clockwise or counter clockwise pattern, following the circular structure of presentation.

We performed a trace analysis of individual participant's fixation paths during each trial. As an illustrative example, we show one participants' pattern of fixations in a set of nine trials in Figure 4.10. In the figure, the target's location is indicated by a pink circle, and the path of fixations is indicated by the open circle points that start at full opacity and decrease in opacity with each subsequent fixation. Fixations are joined by lines to demonstrate the overall fixation path. Looking through a large set of such traces, we saw that, like in Figure 4.10, it seemed that participants commonly favoured exploring the display by looking in a clockwise or counter-clockwise fashion.

To quantify how common the clockwise strategy was we created an algorithm that numbered the locations of stimuli in each trial so that they would correspond to a particular systematic pattern of looking, and then coded participants looking patterns according to that numbering system. For the clockwise algorithm, the topmost item in each display was labelled at "Item 1" with the next stimulus in clockwise order (the next one down and to the right) was labelled as "Item 2", and the remaining items were labelled consecutively as they appear in clockwise order. In the case where two items were "topmost" items in the display, the leftmost of these was labelled as Item 1. Thus, a clockwise pattern would be defined as fixating on item 1, 2, 3 and so on up to the maximum number of items on which the participant fixated. A search pattern would also be defined as clockwise if, for example Item 3 was fixated on, followed by 4 and so on. Or if Item 6 was fixated on first, then 1, 2, 3 and so on. In other words, if the fixation order was ascending or ascending with one descent, then the trial would be classed as a clockwise search pattern[6]. Conversely, if the fixation pattern was descending or descending with one ascent then it was classed as anticlockwise. Trials not meeting either of these classifications were classed as "Neither".

[6] A stricter coding of a clockwise pattern would allow for only a descent between 6-1. However, we thought that participants may skip an item (for example the sequence: 3, 4, 5, 6, 2, 3) and that this too is probably consistent enough with a clockwise movement to be classified as such.

As shown in Figure 4.11, using clockwise or anticlockwise strategies was common, however, the amount of yellow in the figure makes it clear that they were not the only search patterns used. Interestingly, evidence by the horizontal bands of same colour in Figure 4.11, it is clear that there are participants with preferences for using clockwise or anticlockwise strategy. In Figure 4.12, we calculate the proportion of trials on which one of the three categories of strategy separately for each set size. The figure shows that the preference for different looking strategies depends on how many items need to be remembered.

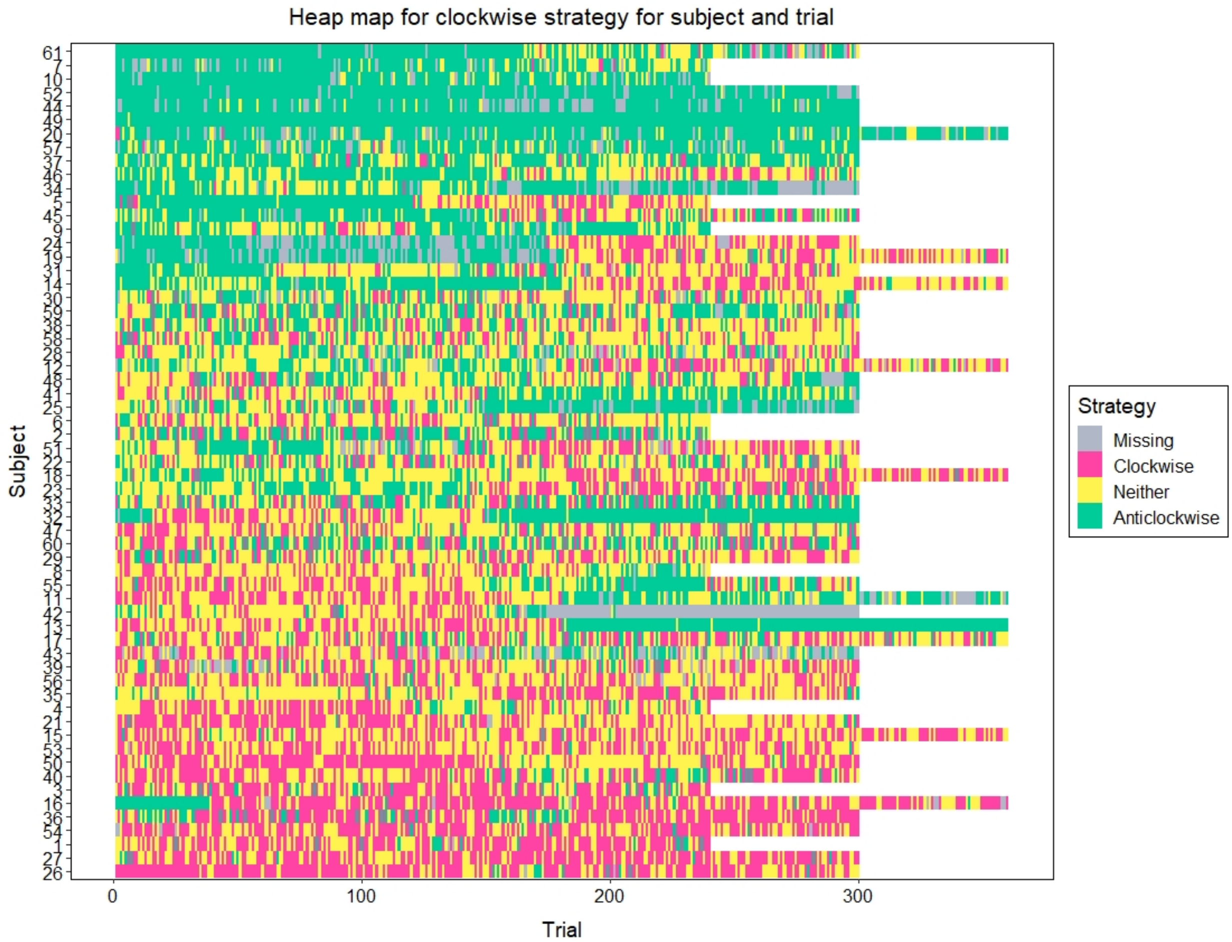


Figure 4.11. Heat map for clockwise strategy classification for each subject and trial. Horizontal stripes of the same colour suggest consistent strategy use over time. Subjects are sorted by individual strategy preference.

Interestingly, and quite unexpectedly, we see a large increase in the rate of anticlockwise looking strategies when set size was 6, while clockwise strategies were preferred in set size 3.

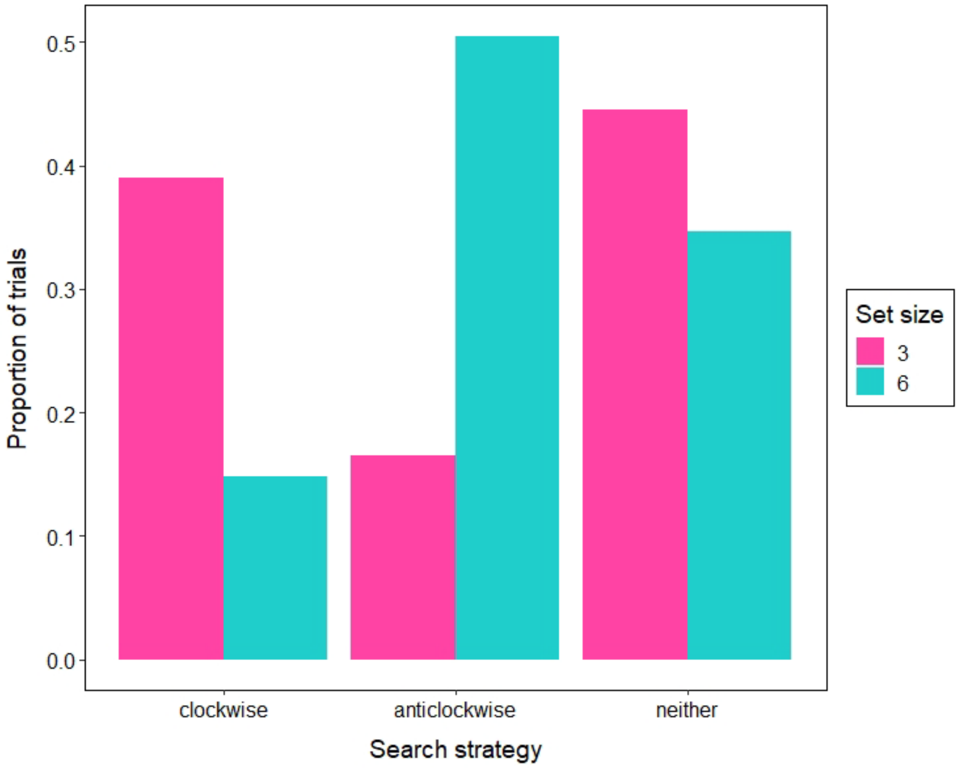


Figure 4.12. the proportion of trials on which clockwise, anticlockwise or neither study strategy type was used. Clockwise trials were more common for set size 3 and anticlockwise trials were more common for set size 6.

We were also interested whether other systematic aspects of looking behaviour, beyond clockwise or anticlockwise, could be identified. Similar to the clockwise algorithm above, we created search algorithms based on a left-to-right (L2R) style of searching (and the corresponding right-to-left search; R2L) as well as a top-to-bottom strategy (T2B and bottom-to-top B2T). For the L2R algorithm, the stimulus that was displayed at the leftmost position onscreen was labelled as ‘1’ with the next leftmost item labelled as ‘2’ and so on. Similarly, the T2B algorithm labelled the item presented highest above the centre of the screen was labelled as ‘1’, with the next highest item labelled as ‘2’ and so on. The R2L and B2T algorithms were defined as the opposite ordering for the L2R and T2B algorithms respectively. While these patterns would share some overlap with a clockwise algorithm, which has both a horizontal and vertical direction component, we wanted to see if we could account for other patterns of searching by comparing these three algorithms.

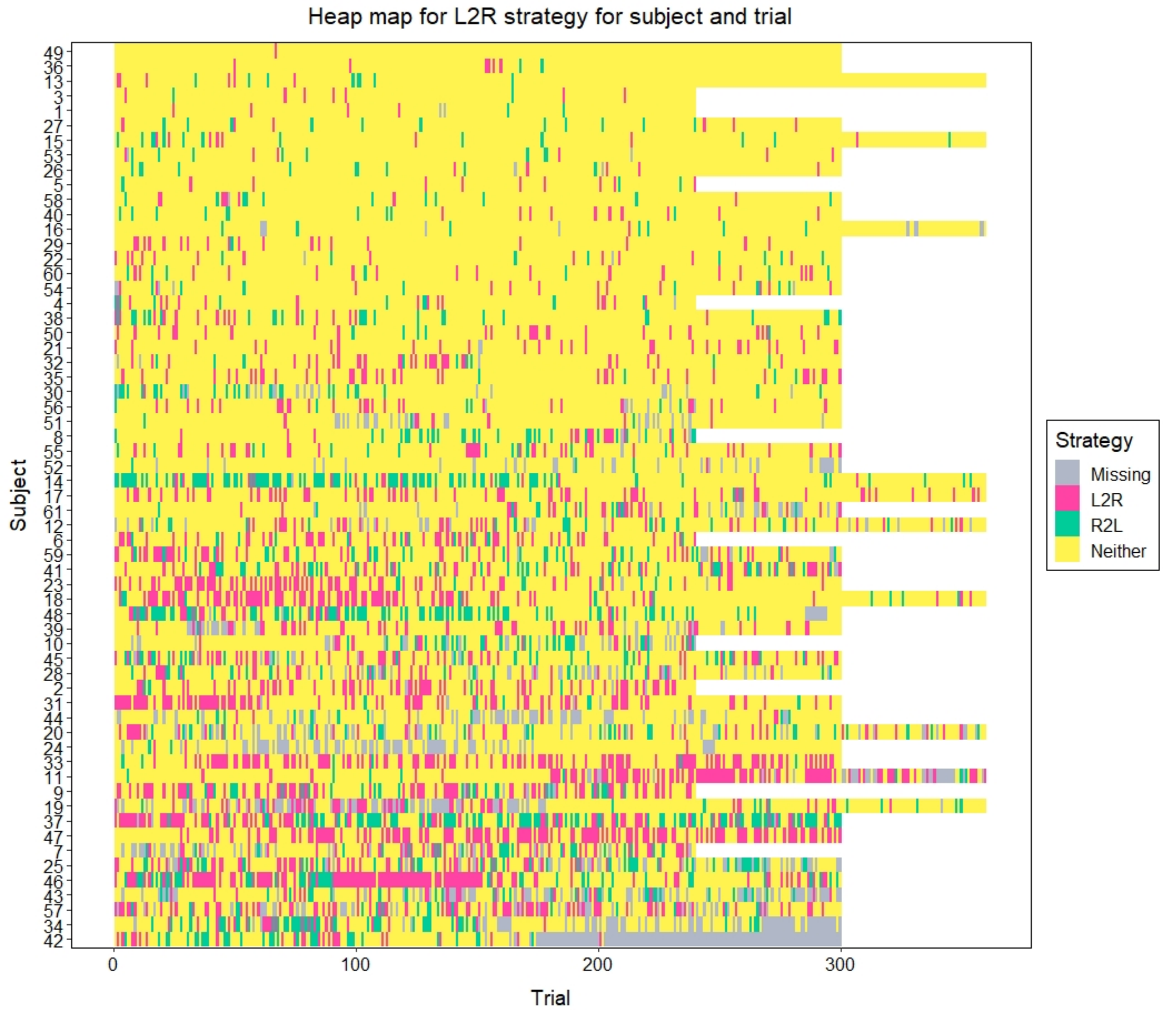


Figure 4.13. Left-to-right and right-to-left strategy use by participant and trial. Both strategies seem less common that the clockwise strategies however the L2R is more common than the R2L strategy. Subjects are sorted by individual strategy preference.

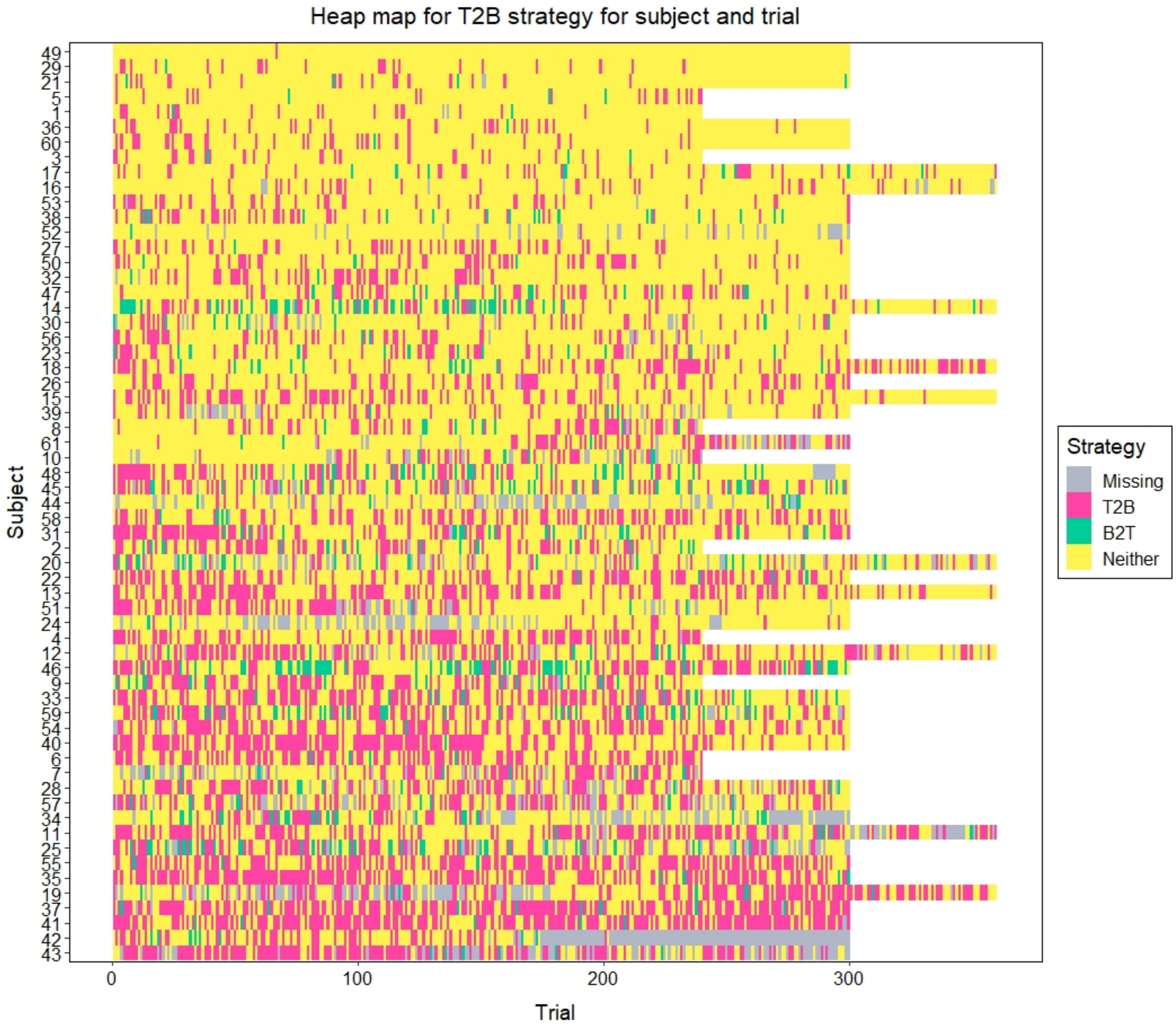


Figure 4.14. Top-to-bottom and bottom-to-top strategy use by participant and trial. Both strategies seem less common that the clockwise strategies however the L2R is more common than the R2L strategy. Subjects are sorted by individual strategy preference.

Displayed in Figure 4.13, we can see that the L2R and R2L strategies were uncommon with most trials classified as "Neither" by the algorithm, but with a slight preference for L2R strategies. Figure 4.14 shows that participants may have a preference towards T2B strategies, with very few trials classified as B2T.

The overall prevalence of each strategy type is summarised in Figure 4.15, separated by set size. For this analysis, we only counted a trial as being clockwise, anticlockwise, L2R etc., only if that strategy alone was associated with that trial. Some trials were associated with more than one strategy categorisation the most common being clockwise and T2B for set size 3 (11.8% of the set size 3 data) and anticlockwise and T2B for set size 6 (11.3%). In total, these combination trials made up approximately 24% of the data for both set sizes. It was also possible for trials to receive a category of "neither" by all algorithms (27.2% in set size 3, 16.8% in set size 6).

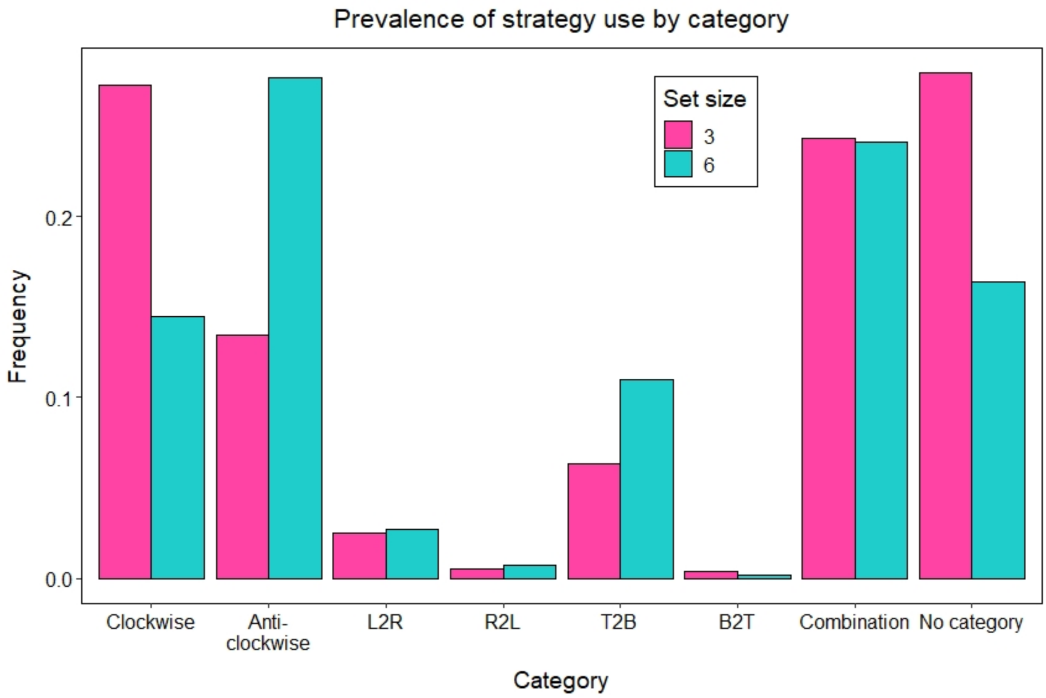


Figure 4.15. Prevalence of task strategies as classified by the clockwise, horizontal and vertical display exploration strategies. Trials in this analysis were assigned a strategy only if that trial was associated with that strategy alone. Clockwise and anticlockwise strategies appear to be the most prevalent.

Of the trials that received one categorisation, clockwise and anticlockwise strategies were the most favoured (again with the set size difference seen in Figure 4.12) with T2B as the next most common classification.

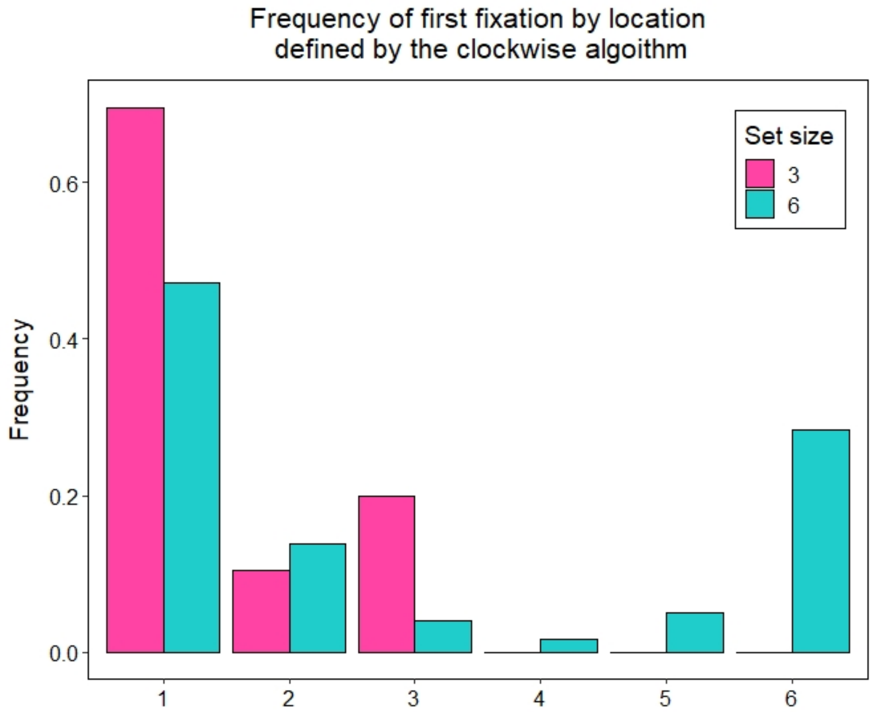


Figure 4.16. Frequency of first fixation location as defined by the clockwise algorithm. The top or top-rightmost item in the display is defined as "1" with the next stimulus down and to the right of the first classified as "2". In both set sizes there is a preference for first fixating the topmost stimulus in the display.

As well as overall pattern, we were interested to see if there was a consistent location at which participants would start their exploration of the display. Shown in Figure 4.16 is the frequency with which each item location was the first location fixated. The locations in Figure 4.16 are defined by the clockwise algorithm with the location "1" being the item displayed at the top or top-rightmost of the screen. “2” is defined as the stimulus immediately below and to the right of “1” and so on following a clockwise pattern.

For both set sizes there appears to be clear preference for first fixating on the item at the top of the display. This finding interacts with our analysis of serial position of fixation and duration, which found that the first item that is fixated typically receives the largest proportion of fixation time in the trial (Figure 4.7). This would suggest that items presented at the top of the display capture attention at the beginning of the trial. While a longer fixation on these items would be associated with an increase in performance (when this item is the target) - as we found in our analysis on total fixation time on the target (Figure 4.3) – fixating an item first means that it does not benefit from recency, which we found to enhance performance (Figure 4.5).

Revisiting and unique fixations. While examining the traces we noticed that on some trials, some participants would revisit stimuli on which they had already fixated. We were curious to see how often this would occur. We refer to a fixation as a revisit if the participant looked at an item they had already fixated at least once already during that trial. Examining the data shown in Figure 4.17, the overall tendency to revisit items is uncommon, especially for set size 6. In general, it is most common to make fixations on unseen items in this task.

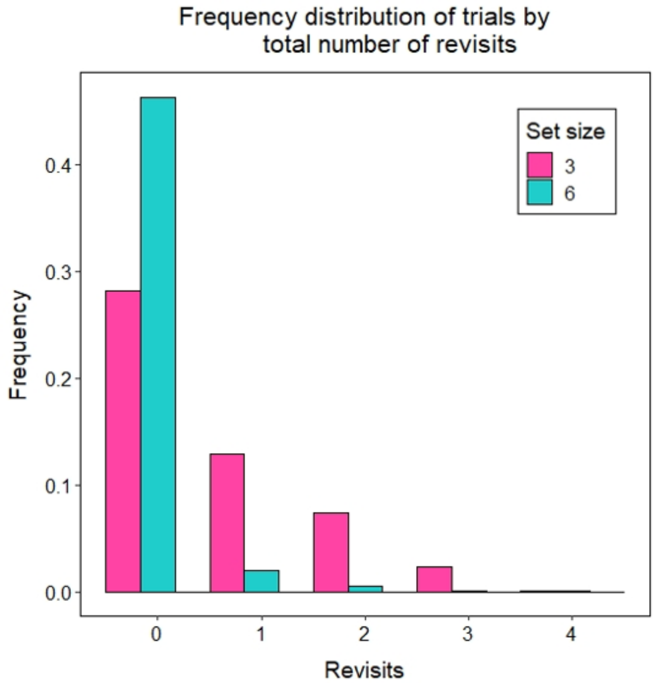


Figure 4.17. Frequency distribution of trials by the total number of revisits in that trial. Revisiting is generally more common in set size 3, as would be expected, however it is not relatively common.

Revisiting is potentially interesting as a strategy because it may reflect that participants are aware that they had not encoded a stimulus properly. Whether participants have, or utilise, such meta-knowledge about the fidelity of their representations has implications for theories of VWM. To explore this idea, we examine the first revisit that happened in a trial. To probe this idea, we took the difference between the duration of the initial fixation on the to-be-revisited item and the average duration of the other items already fixated. If participants were revisiting items that were fixated for relatively less time, then we would expect the resultant duration differences in Figure 4.18 to be negative. Considering that the mass of the distribution is centred on zero, this indicates little difference in fixation duration between the item to be revisited and the other items seen before the revisit. This suggests that participants are likely revisiting for reasons other than selectively revisiting items that they had not fixated for as long.

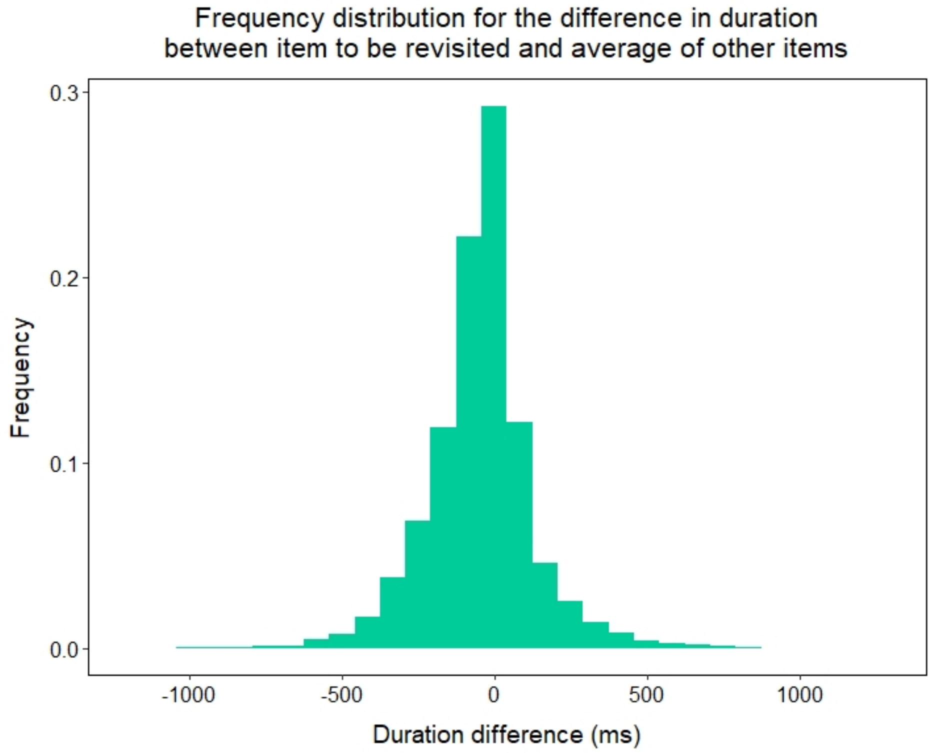


Figure 4.18. Frequency distribution for the difference in duration between the item to be revisited and the average of the other items preceding the revisit. As the distribution is centred close to zero, it seems unlikely that participants are choosing to revisit a stimulus based on lack of information on this item compared to others seen.

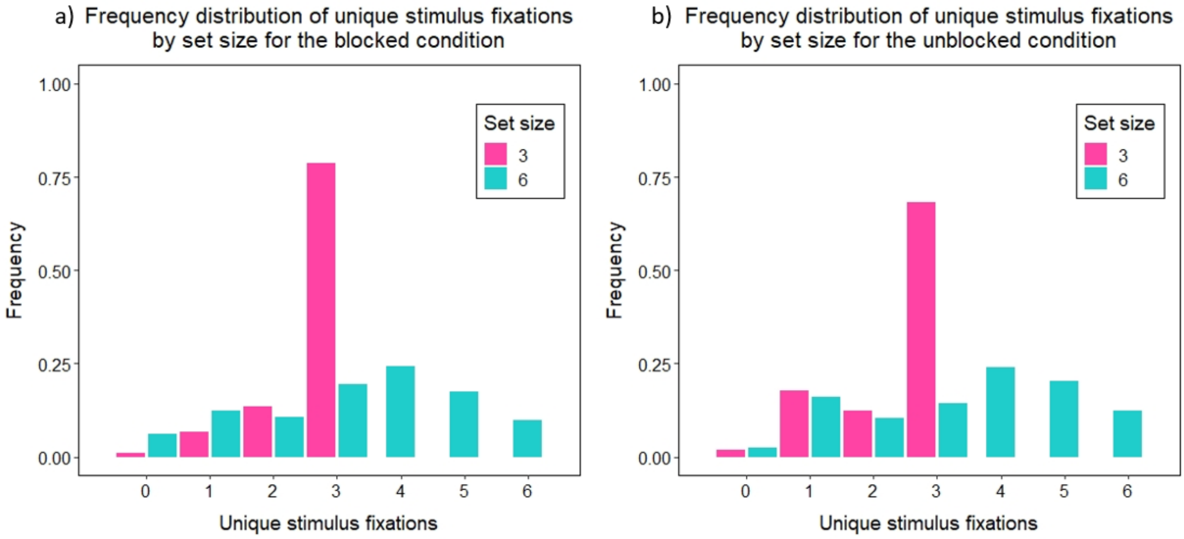


Figure 4.19. The number of unique fixations on stimuli for the a) blocked and b) unblocked conditions. Regardless of condition, the distribution of unique stimulus fixations is similar. In set size 3 it is most common to be able to fixate all the stimuli. For set size 6 is it most common to fixate 4 stimuli with fixating all 6 items being less likely.

We were also interested in how many of the to-be-remembered stimuli were fixated within a single trial. In particular, and especially for set size 6, we were interested to see on what proportion of trials participants were able to attend all the items. Figure 4.19 shows the distribution of unique stimulus fixations separately for the blocked (Figure 4.19a) and unblocked (Figure 4.19b) conditions. Comparing the figures, differences in the number of unique items fixated is minimal, suggesting again that there is unlikely to be any effect of whether set sizes are blocked or unblocked. In both Figure 4.19a and b, on most set size 3 trials participants were able to see all three stimuli. By contrast, seeing all stimuli happened less often in set size 6, with four unique items being the most common number of items to be fixated.

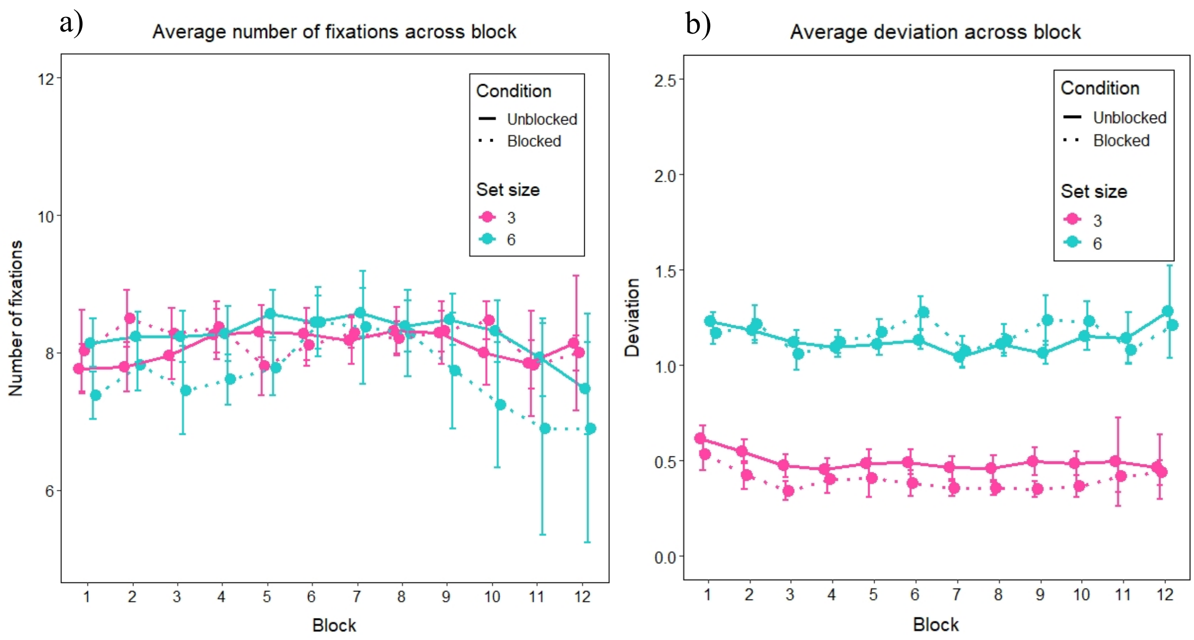


Figure 4.20. a) Average number of fixations over the course of the experiment for both set size and condition. Little difference is observed between the four groups as well as over time. b) Average absolute value of deviation across block for set size and condition. While there is an effect of set size, block and condition do not appear to impact average deviation. Each block contains 30 trials. Error bars show standard error.

Learning. We examined the number of stimulus fixations over time to assess if fixation strategy changed over the course of the experiment. Figure 4.20a shows the average number of unique fixations as a function of block number (using blocks of 30 trials). A three way ANOVA of set size, block and condition on the number of fixations revealed showed no significant main effects of block ($F(11,750) = 1.63$, $p = 0.09$), condition ($F(1,59) = 0.18$, $p = 0.68$) or set size ($F(1,59) = 0.02$, $p = 0.88$). No significant interaction was found between set size and block ($F(11,750) = 0.92$, $p = 0.52$) or between condition and block ($F(11,750) = 0.85$, $p = 0.59$). The three-way interaction of set size, block and condition was also not significant ($F(11,750) = 1.15$, $p = 0.32$). Overall, this suggests that there is no change in the total number of fixations per trial over the course of the experiment, between set sizes or due to condition. While the focus of this analysis was the number of fixations total, the same analysis was repeated for stimulus fixations alone and found similar findings – no significant effects.

We also examined the change in the response deviation over the course of the experiment (Figure 4.20b). A three-way ANOVA of set size, block and condition on the deviation revealed a significant main effect of block ($F(11,750) = 2.78$, $p = 0.0015$) and set size ($F(1,59) = 31.51$, $p < 0.001$), but not of condition ($F(1,59) = 0.30$, $p = 0.58$). No interactions were significant (all ps > 0.93). Examining Figure 4.20b, the effect of set size on deviation is clear, while the differences across the experiment are more subtle and not clearly indicative of an effect of learning (e.g., the uptick in deviance in block six is likely due to changing difficulty of the task). What is more apparent is that a potential effect of block is not impacted by set size or blocking/unblocking condition.

Discussion

The aims of this experiment were to see the eye-gaze behaviour map onto performance on the memory task, to determine if there was a blocking effect in this gaze-contingent design, and to examine the strategic deployment of overt attention used in the task.

We found that eye-gaze data related to memory performance in a number of ways. Not only did fixating on the target lead to a better memory performance (as would be trivially expected), so too did the amount of time spent looking at the target, as well as how recently the target was seen. We were also able to observe an effect of serial position of fixation, such that the first stimulus fixated in a trial is typically fixated longer compared to subsequent stimulus fixations. Our results, taken together, suggests a link between attention and encoding that could be investigated further with a gaze-contingent paradigm. We had hoped to investigate this relationship with the use of computational models of working memory, but the existing Bayesian methods we used to repeat the same analyses from Chapter 1 for this data failed to work, and we ran out of time to write our own code to perform equivalent analyses.

However, without modelling our results do provide us with some perspective on VWM models. In our analysis of total fixation time on the target, we saw a relatively continuous increase in accuracy as durations exceeded 500ms. On the other hand, for these shorter fixation durations, there was no apparent gradual increase in accuracy. In one sense, little change in performance with increasing fixation duration (up to 500 ms, at least) is consistent with the slots-like accounts in which accuracy is characterized as all-or-none. Similarly, in examining the number of item fixations we found that participants will most commonly make four unique item fixations in set size 6. This result is again arguably in line with a slots-like account, wherein a subset of items is encoded, rather than the whole array. That said, the theoretical implications of the results here for slot and resource models of VWM are not so straightforward, and we will return to this discussion in the General Discussion chapter (page 176).

One interesting difference between Experiments 1 and 3, is that the current experiment used a longer trial duration of 3000ms compared to the original 1000ms. While this gaze-contingent task was more demanding in that participants needed to actively control their fixations, it is still interesting that an increase in trial time did not make a larger impact. In Experiment 1 it was most common for three unique item fixations to be made. We do see a slight increase in unique fixations in the current experiment, but participants rarely looked at all items in the display, which, in this gaze-contingent design, also means that the stimulus is never seen.

Despite the connection we established between eye-gaze and memory, we did not see any effect of blocked and unblocked set size. The necessity of overt attention in this design should have, theoretically speaking, produced an effect of blocking, were one to exist. That is, in our gaze-contingent task, the reasons to spread attention more widely was larger than in the passive-viewing task. If, even under these conditions, we see no effect of blocked set size, it seems fair to conclude that this effect is not a key to strategy switching, as was implied by Donkin et al. (2016). At the very least, it appears that the effect of blocked presentation order does not generalise to tasks beyond the change-detection tasks used in their paper, since it did not hold for Experiment 1 (which was more similar in design as a standard VWM task) or Experiment 3 (in a gaze-contingent task). More concerning is the possibility that the original results in Donkin et al. may have little to do with a strategic deployment of attention. Clearly, further work is needed to understand the reason for the discrepancy between the two paradigms.

In regards to strategy we began to see some consistent features in participants' approaches to the task. Participants showed preferences for patterns of exploring the memory array in a clockwise or anti-clockwise fashion, which suggests the structure of the display could be quite important for understanding the strategic deployment of attention. This is something worth exploring in other data sets, which use different constraints on stimulus placement, as it is unclear with this data alone if this effect is due to participants searching out the nearest stimulus to examine next, or if they are following the structure of the display. One reason to think that strategy preference is not simply guided by stimulus placement is that looking behaviour appeared to depend on set size. In set size 3, clockwise movement patterns were preferred, while a counter-clockwise study strategy was more common in set size 6. The reason for this is not immediately apparent. It is possible that anti-clockwise movements feel easier to participants as a clockwise eye movement is more active in muscle recruitment (Sparks, 2002). That is, this finding may reflect participants choosing easier eye movements for larger set sizes.

The finding that sometimes participants choose to fixate on stimuli already seen was a potential area of interest. Since we found that total fixation time on the target increased performance, we might assume that revisiting could be employed tactically to improve encoding for items that were not encoded well initially. However, when we examined the difference in fixation duration between items that were chosen to be revisited and other items that were fixated on before the revisit occurred, we found there was no appreciable difference in average duration between the groups. This suggests that revisits occur for reasons we do not yet understand. Overall, it should be noted that revisiting is not particularly common in this task and thus could be due to fatigue or lapse in attention from participants.

Limitations. The reader may wonder whether the gaze-contingent nature of the current experiment makes it too different from standard VWM tasks, potentially compromising the implications of the results here for theories of VWM. The nature of the gaze-contingent task is that participants are forced to move their eyes, something that we have demonstrated they are not inclined (and may even learn not) to do. However, we would argue that if we see improvements in memory performance due to an eye-gaze measure in this task, this should relate back to the standard task in some way. That is, even in a gaze-contingent task, the same VWM system must be used as in the standard task – to assume otherwise would suggest two different systems for the tasks, which seems unlikely. We do, however, need to take care to consider whether the differences in the task create alternative explanations for any differences or similarities between our results and others in the literature.

While we believe this gaze-contingent task is informative of the standard experiment, it does still lack some information to give us a better understanding of a typical VWM experiment. The nature of this task dictates that participants can only begin to encode items when they fixate them. However, in the standard task our evidence so far shows that people prefer to use peripheral information. This seems to be despite the finding in this experiment that fixating on the target for longer durations leads to better memory performance. It appears that in the standard task, participants are performing a kind of trade-off. They are losing accuracy, for the sake of attempting to encode more items. However, with the present data, we are not able to make any strong claim about this. This is due to, first, not having eye-gaze data aligning with attention in the standard task means we don't know how participants are allocating attention in this task; and second, participants have no option to perform a trade-off in the current gaze-contingent task, as participants must fixate to encode items.

Our understanding of strategy in this task was also limited in part by the design. A dominant feature of most participants' strategies was utilising the circular structure of the items' presentation. While this feature was important to include in order to replicate Experiment 1, it gave participants an obvious way in which to explore the display. The effect of this is that our results seem unlikely to capture the full range of variation in strategy when participants are presented with a randomly (or otherwise structured) memory array.

Additionally, this task requires completely overt attention – it is not possible for participants to utilise their periphery in this task. However, as has been demonstrated by our previous experiments, there is obviously a role for covert attention in VWM which has not been captured in this experiment.

Conclusion and next steps. We believe we have demonstrated that this gaze-contingent paradigm is useful for examining attention in VWM tasks. This task allows for a link between eye-gaze data and attention that the standard task has not been able to provide. While computational modelling for this task was not successful, we believe this task has the benefit of allowing task strategy to be analysed.

For our next experiment we want to run a second gaze-contingent experiment, with the hope of addressing some of the limitations of Experiment 3. Firstly, we noted in this task that participants are required to fixate on items to encode them and this may not be preferred by participants. As a result, the behaviour we see here might not be indicative of general VWM strategy. Therefore, in the next experiment we plan to add peripheral information back into the display. We planned to do this by making an additional peripheral item available when a stimulus is fixated. We also plan to make items appear at random locations around the display. This will remove the inherent structure we have in Experiment 3 and gives us a chance to see if such clear strategies for eye movements persist.

Chapter 5: Investigating the role of peripheral information in VWM using a gaze-contingent paradigm

In the previous chapter, we saw that a number of eye-gaze variables, such as how recently a fixation was made, or total dwell time on the target, had an impact on task performance. Such results showed a much clearer link between fixation on the target and performance than in either Experiment 1 or Experiment 2. We were also able to see a preference for a clockwise search of the display, though with notable individual differences.

Following Experiment 3, there were two main features of the task that required further examination. Firstly, by using the gaze-contingent paradigm we intentionally restricted participants to view just one item at a time. While this made the connection between eye-gaze, attention, and memory much clearer, this is obviously a much more restrictive task than the standard task. As we have seen, people tend to use covert, peripheral attention in the standard task. The gaze-contingent task gives us a nice way of exploring the quality of memory information stored using covert attention, while allowing for some control over what can be encoded. The experiment we use in this chapter is a compromise between the standard task, during which participants apparently mostly use covert attention, and the gaze-contingent task, in which they can only use overt attention.

Secondly, since Experiment 3 was based on Experiment 1, we presented the stimuli equidistant from the centre of the screen and in a circle formation. While this allowed for good comparison between these experiments, such arrays limit our ability to study the attentional deployment strategies people use more generally. That is, it may be that participants use clockwise or counter-clockwise exploration of items simply because the stimuli were presented in a circle. We wanted to see how people would explore the scene if there was no systematic structure to the study array across trials. While also being arguably more natural, it is also in line with other VWM tasks, which sometimes utilise a less systematic presentation of study items.

The current task

Experiment 4 was a second gaze-contingent task that differed from the third experiment with respect to the above two key ways. With regard to the first point of the preceding section, we added peripheral items back into the design by revealing a second item whenever one was fixated. As in the previous experiment, when an item's location was fixated upon by the participant, the stimulus at that location was revealed. However, in this experiment the stimuli were paired such that whenever one stimulus was looked at, its pair would also be revealed. By having this additional item, we enabled participants to utilize their peripheral vision if they wished, thus allowing us to see how covert attention would impact task performance compared to overt attention.

The second point from the last section was addressed by presenting items in a less systematic structure compared to Experiment 3. Having stimuli scattered (seemingly) randomly across the display creates ambiguity as to how participants should explore the items. We hoped to see some differences in individual strategy emerge as a result.

As before, participants were able to explore the test array as they wished. It would be possible, therefore, that participants could vary in how much they used peripheral information. Some participants might only look directly at one item in each pair. This would lead them to uncover all the items but only see half directly, with the other half present only in the periphery. By contrast, it is possible that some participants would still endeavour to fixate individual items. It is also possible that by revealing a stimulus in the periphery, a participant's attention would be captured by this item and they would look at it directly after looking at the initial item. There is a consistent precedent of attention capture of sudden onset or salient stimuli (Le Pelley et al., 2015; Christ & Abrams, 2006; Posner, 1980), so it is possibly that the sudden onset of the peripheral stimulus would capture attention. However, it is also possible participants could engage attentional control to suppress this impulse, as has been seen elsewhere (Woodman & Luck, 2007; Dickinson & Zelinsky, 2005). As seen in other VWM studies that reference attentional control, we expect a reasonable amount of individual variation (Vogel et al., 2005; Cusack et al., 2009).

Method

Participants. 27 first year psychology students from UNSW agreed to participate for course credit. Participants were recruited using the student-specific experiment sign-up system, SONA. Approval for this study was obtained from the UNSW Human Research Ethics Approval Panel – C (HREAP-C).

Apparatus. A Tobii Pro Spectrum 600 mounted on a 23.8-inch widescreen monitor (1,920 x 1,080 resolution, refresh rate 60 Hz) was used for tracking eye movements and displaying the experiment. Participants were seated 40cm away from the screen. A chinrest was used to help participants keep their heads still.

Stimuli. The stimuli and the patches used to obscure stimuli were identical to those used in Experiment 3 with one variation. Unlike previous experiments, the stimuli were not presented on an invisible circle but instead were allowed to appear anywhere on screen (within a margin of 160 pixels from the edge of the screen and no closer than 160 pixels to another stimulus). The same definitions that constituted what defined a fixation in Experiment 3 were the same for the current experiment.

Design. The full experiment consisted of 300 trials – 30 blocks of 10 trials. While 21 participants were able to complete the full experiment, 6 participants experienced a shorter version of the task due to issues with participants running late and limited testing space. This shorter version consisted on 240 trials total – 24 blocks of 10 trials. Each trial contained 6 stimuli; there was no variation in set size. As in Experiment 3, items were initially obscured by a pink patch and the presentation duration of the study items was 3000ms. The main manipulation of this experiment was the pairing of stimuli. As previously, when a participant fixated on the location of an item it was revealed to the participant. However, in this experiment a second stimulus was also revealed in addition to the fixated stimulus. The creation of the three pairs of stimuli was random (i.e., we did not attempt to control for the distance between the pair of the items, or their features in any way), and the pairing was consistent for the entire trail. That is, these stimuli were paired such that for the duration of the trial, if one item in the pair was fixated on, they would both be revealed.

Procedure. Before beginning the main task, participants were given both verbal and written instructions on how to do the task as well as some practice trials to adjust to the experiment. On each trial, participants were shown a fixation cross (500ms), a brief interval of a black screen (400ms) and then the stimulus screen was presented for 3000ms. Each trial contained six gem stimuli. As in the previous gaze-contingent experiment, the location of each item was indicated by a pink patch which obscured the stimulus from view. In order to view the stimulus, participants needed to fixate on a pink patch to reveal the item underneath. Whenever an item was fixated on and revealed, its paired item would also be revealed simultaneously. Participants were free to "uncover" as many or as few items as they wished and were able to revisit items if they wished. The only restriction was the 3000s duration of the presentation interval. Following this interval, a mask was presented, followed by a blank screen and finally the test screen. On this screen the ring from one of the stimuli was presented at the location it appeared during study. Participants then needed to place the bead on the ring as it had appeared during that trial. Participants received feedback on their performance. ("OUTSTANDING!" for deviations less than 10°, "Very good!" between 10° and 20°, "Good" between 20° and 35°, "OK" between 35° and 45° and deviations greater than 45° were labelled with "Poor"). The trial sequence is depicted in Figure 5.1. After each block, participants were given a break in which they rested for a minimum of 20 second before continuing.

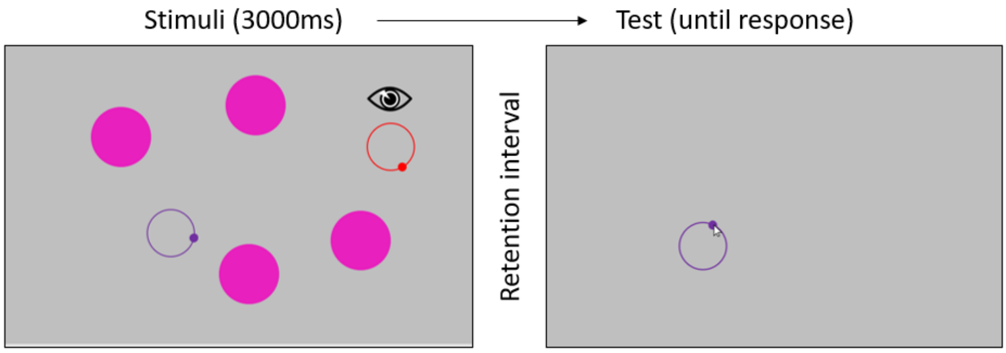


Figure 5.1. The trial sequence in Experiment 4. Similar to Experiment3, study objects are obscured from view by filled pink circles. If a circle is fixated on, it reveals a study stimulus to the participant. Each stimulus has a pair such that when one in the pair is fixated on, the pair is also shown. The memory task of placing the ring on the bead is as before.

Results

Across all participants, 7860 trials were recorded. Similar to Experiment 3, we removed trials that recorded more than 30 fixations during the presentation window (786 trials, 10%). We also removed trials if they had an average fixation duration less than 100ms (451 trials, 5.7%). No trials were removed for having no eye-gaze data recorded. This resulted in a total of 1237 trials being removed (15.7% of the data).

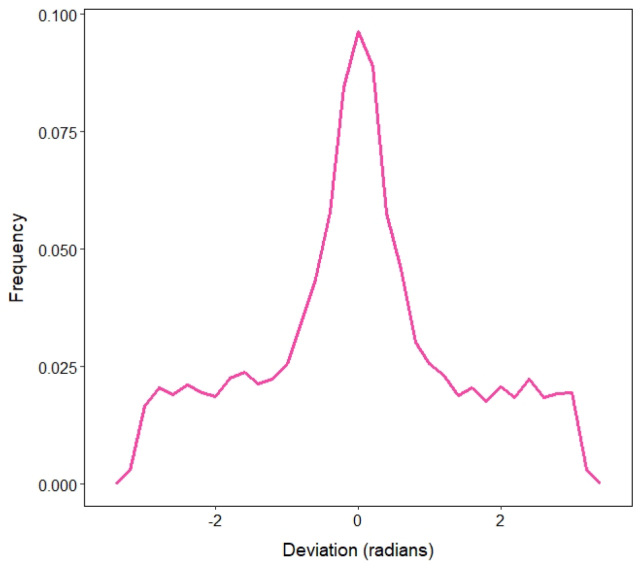


Figure 5.2. The frequency distribution of deviation. Note: this experiment only contains set size 6. The general distribution of responses is comparable to set size 6 from previous experiments (e.g., see Chapter 4).

Accuracy. Examining the deviation of responses, (Figure 5.2) it appears that the response distribution for accuracy in this experiment was similar to previous tasks. In particular, the results for this experiment were very similar to the results for set size six in Experiment 3. This suggests that, in terms of difficulty this task is comparable to previous VWM tasks.

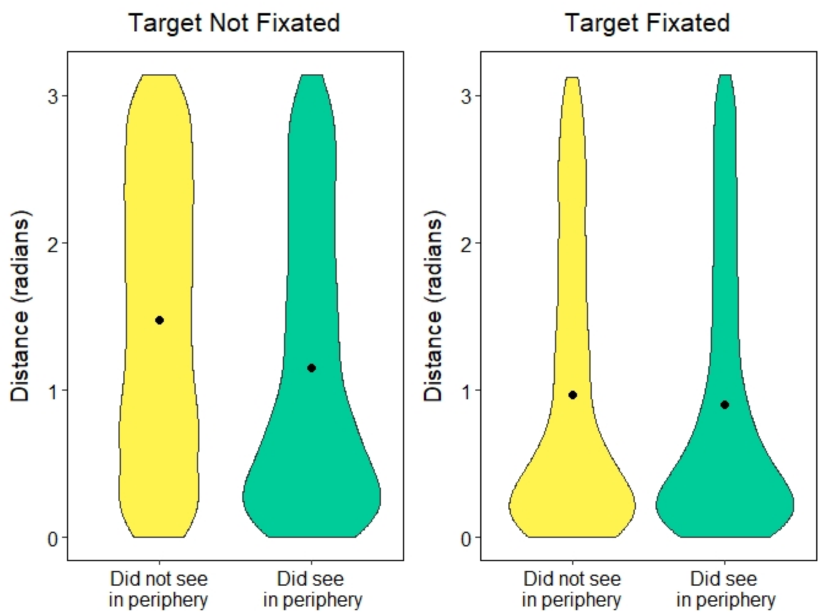


Figure 5.3. The distribution of response duration by whether the target was fixated or not and by whether the target was seen in the periphery or not. Fixating the target is associated with an improvement in accuracy. Even when the target is not fixated, seeing the target in the periphery alone is associated with an improvement to performance.

Eye-gaze data. As with previous experiments, task performance was measured as the absolute value of the deviation (distance, in radians) from the participant's response to the correct answer. Figure 5.3 shows the distribution of response accuracy according to whether the target was fixated during the experiment or not, and by whether the target was shown in the periphery or not. The target was counted as being fixated if the participant looked directly at that stimulus during the trial. A target was counted as "seen in the periphery" if the target's pair was directly fixated, thus also revealing the target elsewhere in the display. Note it was possible to see the target both in the periphery and to also fixate it directly, just as it was possible never to see the target at all.

As displayed in Figure 5.3, accuracy in this task was improved if the target was fixated directly. If the target was fixated, whether it was also seen in the periphery did not appear to improve performance to any large extent. However, if the target was not fixated but was seen in the periphery, recall of the target improved significantly compared to not seeing the target at all.

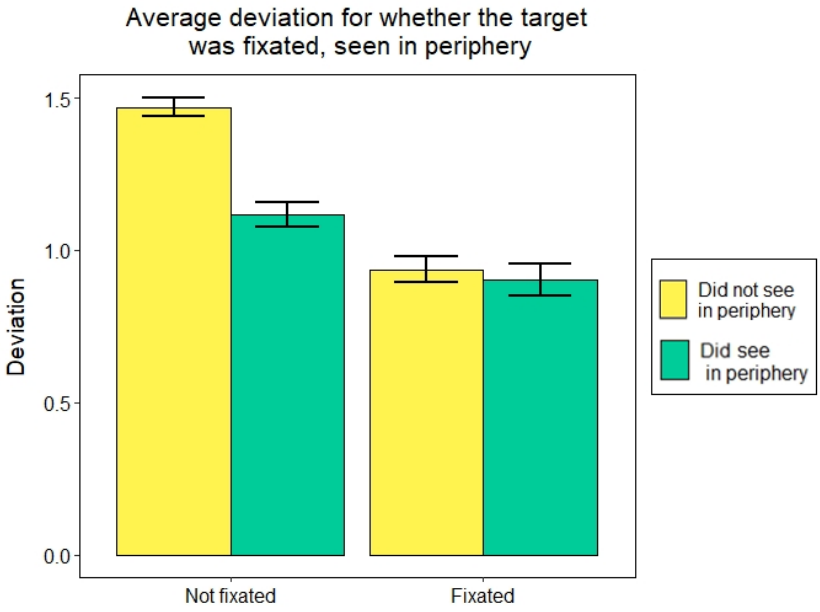


Figure 5.4. Average deviation for whether the target was fixated or not and for whether it was visible in the periphery or not. Seeing an item in the periphery is associated with a reduced in deviation even when it is never fixated in that trial. Error bars are standard error.

To see whether this pattern holds across participants, we first calculate the average deviation for each individual. A 2 (target fixated or not) x 2 (target seen in periphery or not) ANOVA revealed significant main effects of target fixation ($F(1,26) = 85.13$, $p < 0.001$) and viewing the target in periphery ($F(1,26) = 32.04$, $p < 0.001$). From Figure 5.4 we can see that viewing the target, whether it is seen directly or indirectly in the periphery reduces response deviation. There was also a significant interaction such that when the target is not fixated, the effect of seeing it in the periphery is significantly greater ($F(1,26) = 33.07$, $p < 0.001$).

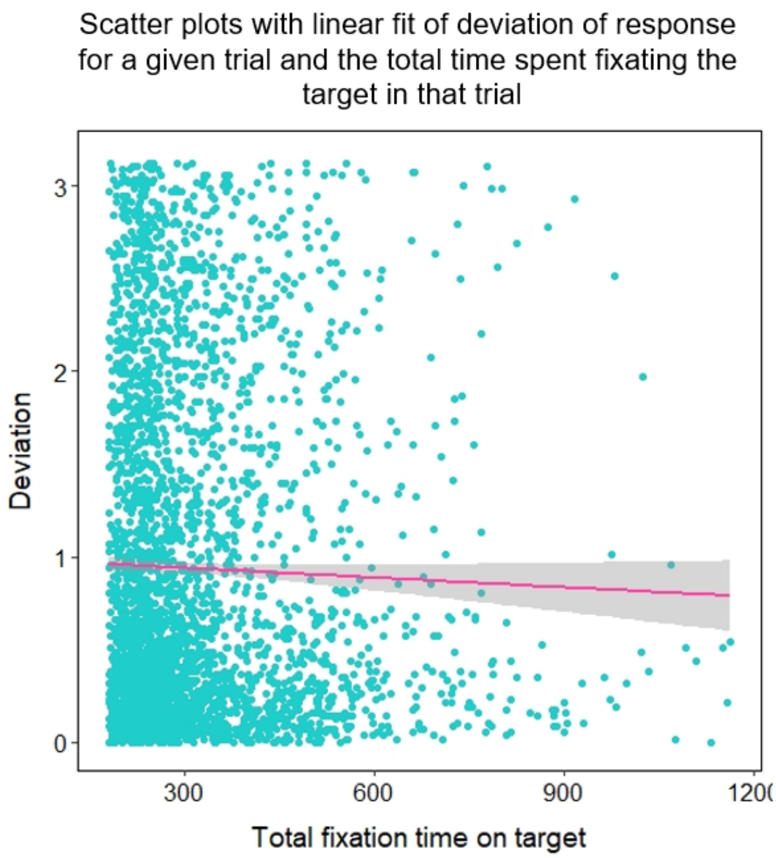


Figure 5.5. Deviation by the total fixation duration on the target in a trial. Longer total durations are associated with more accurate responses however this relationship is much less clear when compared to Experiment 3. The pink line is a linear fit with the shaded region showing error.

We repeated a number of analyses run in Experiment 3 in order to compare that experiment to the current one. Firstly, we examined accuracy as a function of total fixation duration on the target[7]. As seen in Figure 5.5 the response deviation decreases on average as total fixation time on the target increases ($r(7072) = -0.17$, $p < 0.001$), ($r(25) = -0.50$, $p = 0.007$). We also examined whether effect of exposure was also present for viewing the target in the periphery (without fixating) it. However, there was no significant correlation between the duration the target was visible in the periphery and average deviation of response ($r(25) = -0.35$, $p = 0.06$) $r(7072) = -0.07$, $p < 0.001$).

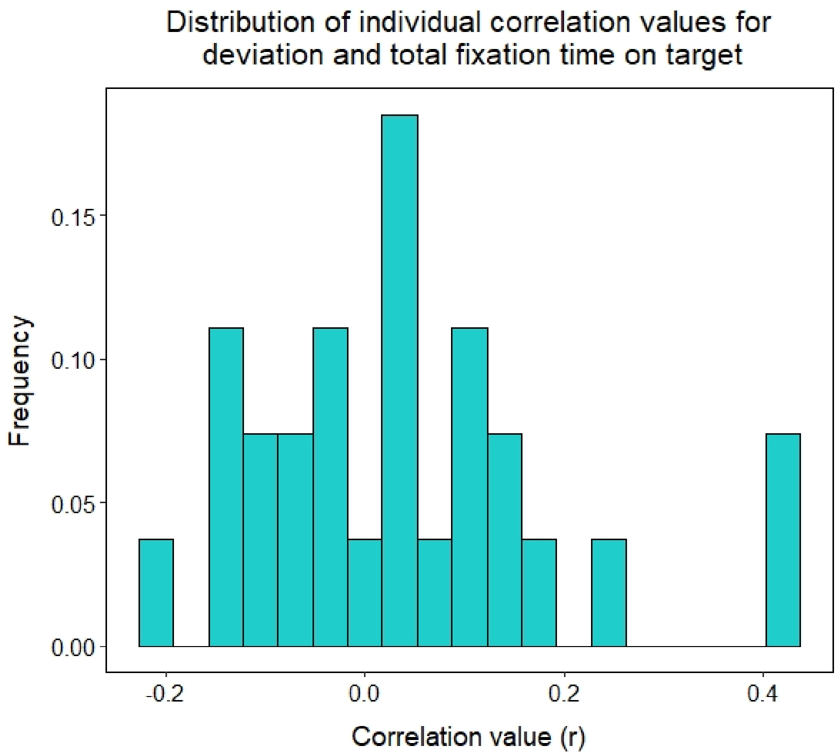


Figure 5.6. Distribution of individual correlation values for deviation and total fixation time on the target.

Next, we examined the distribution of individual correlation values for the absolute value of deviation and total fixation time on the target. Displayed in Figure 5.6, the

[7] As with our analysis of Experiment 3, the analysis of Experiment 4 was limited by our models failing to run (see footnote on page 128). Ideally, we would have liked to run this analysis with more appropriate methods instead of correlation, however due to time constraints this was not possible. As a result, caution in interpreting the results is advised

distribution of these individual correlation values appears to centre around 0. This suggests that longer fixation durations on the target is less clearly associated with an increase in accuracy for individuals in this experiment compared to Experiment 3.

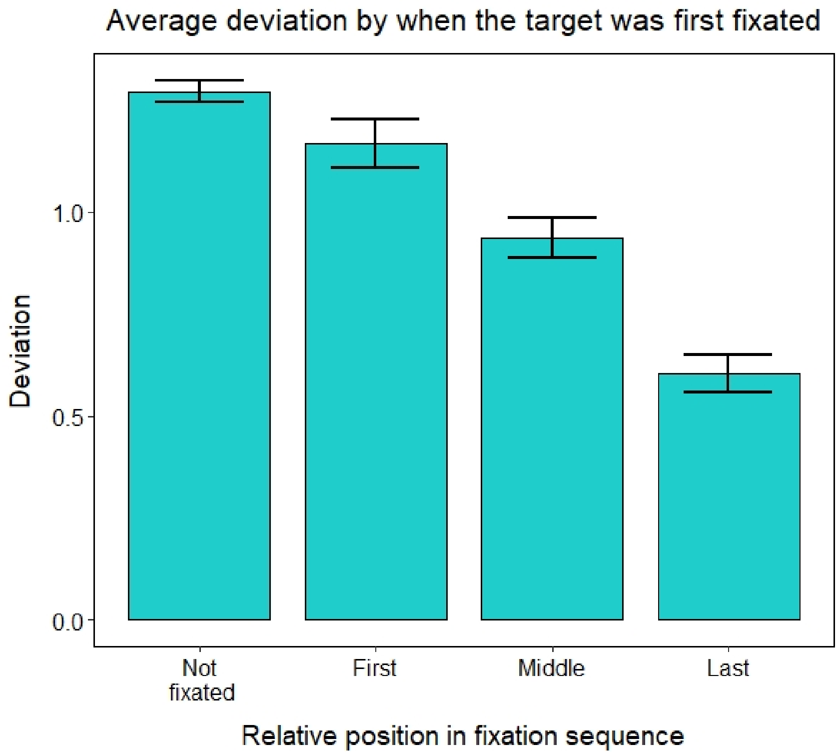
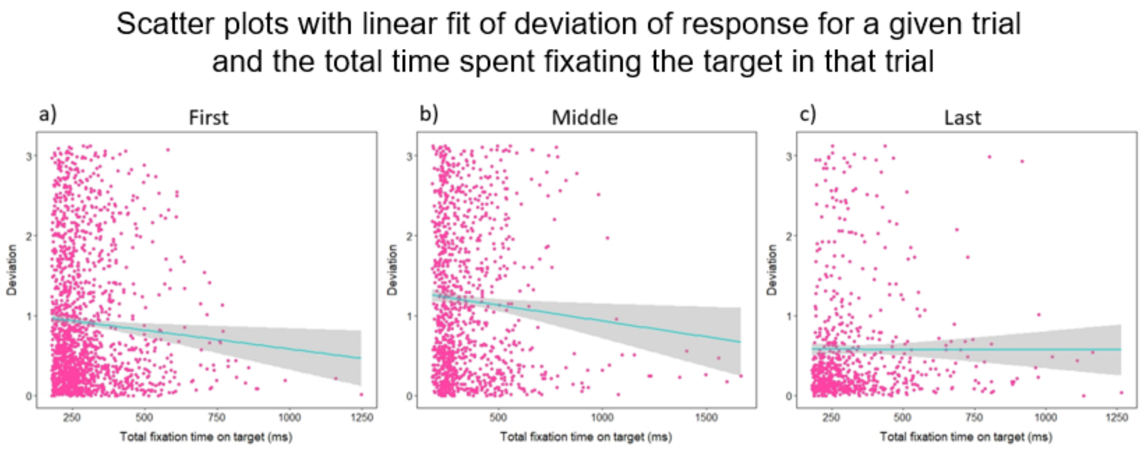


Figure 5.7. Average deviation of response as a function of when a target was seen. Trials were grouped by whether the target was the first item seen, the last item seen or seen somewhere in the middle of the fixation sequence. If there was more than one target fixation made, only the first instance was counted. When the target is seen more recently it is associated with lower average deviation. Error bars show standard error.

Continuing our replication of our analysis in Experiment 3, we looked at the effect of recency. Figure 5.7 shows the (absolute value of) response deviation by when the target was seen in the trial. As with previous iterations of this analysis, we only counted the first instance of when the target was seen[8]. As Figure 5.7 shows, average deviation was smallest when the target was the last item fixated in the trial. Similar to Experiment 3, there appears to be an effect of recency, such that items fixated closer to test were recalled with more accuracy.



c)

Figure 5.8. Deviation and total fixation time on the target for a) when the target was the first item fixated b) when the target was seen somewhere in the middle of the sequence of trial fixations and c) when it was the last item seen. The blue line is a linear fit with the shaded region showing error.

Again, to disentangle the effect of recency and total fixation time on the target, we examined the correlation of deviation and total target fixation time for each fixation sequence category shown in Figure 5.7. Figure 5.8 shows the correlation for when the target is the first item fixated in the trial (Figure 5.8a), when the target is viewed in the middle of the fixation sequence (Figure 5.8b) and when the target is the last item fixated in a trial (Figure 5.8c). For each of the groups, there was a negative correlation between deviation and total fixation time

[8] Again, we re-ran this analysis but with preferencing the most recent instance of fixating the target and results were very similar.

on the target. This correlation was overall small, but largest for when the target was seen in the middle of the fixation sequence (r = -0.09), then when the target was fixated first (r = -0.07) and least for when the target was viewed most recently (r = -0.04).

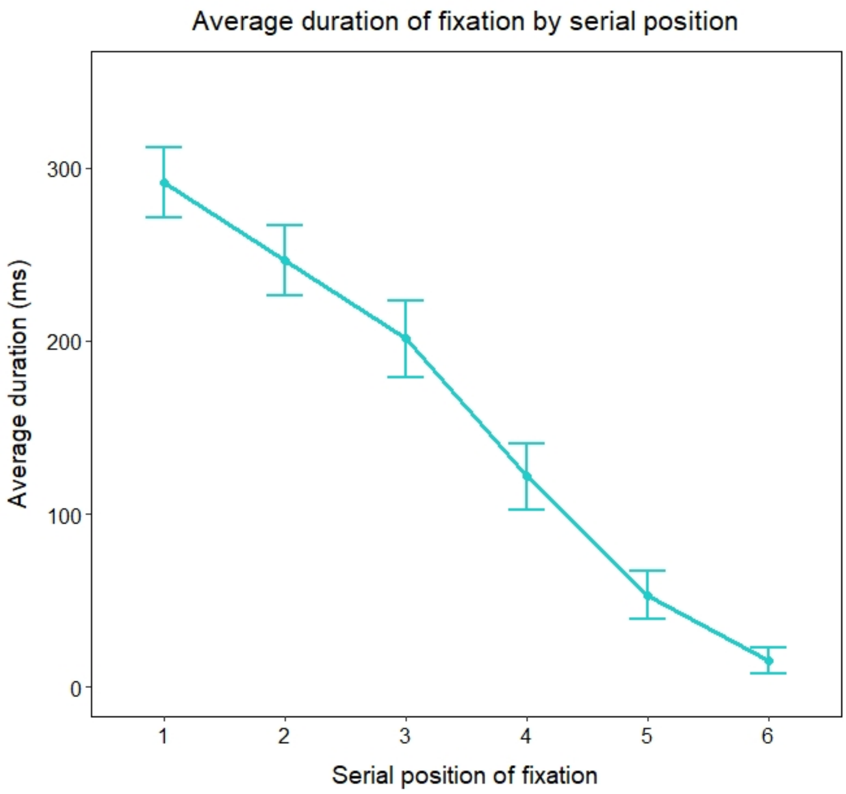


Figure 5.9. Average duration as a function of serial position of fixation. The first fixation is on average longer in duration than subsequent fixations. Error bars show standard error.

Serial position. As with previous experiments, we also examined the connection between fixation duration and the serial position of fixation. Similar to all prior experiments, we see that the first fixation on a stimulus is the longest with each subsequent fixation being shorter. A one-way ANOVA of serial position on fixation duration revealed a significant effect ($F(2,54) = 180.68$, $p < 0.001$) with average duration of fixation reducing with serial position as seen in Figure 5.9.

Strategy. Similar to Experiment 3, we examined the fixation paths that individual participants created for each trial of the experiment with the purpose to examining strategy. Figure 5.10 shows examples of some of these traces from individuals who exhibit different approaches to their attention deployment strategy. The target's location is denoted with the pink circle and the target's pair's location is depicted with a pink square. The remaining pairs are also colour coded; the green items form one pair; the purple for the other. Fixations are shown by the points and the line between them shows the path. The fixation points having opacity that varies with time; the first fixation in the trial as the highest opacity and the last trial has the least.

a) b)

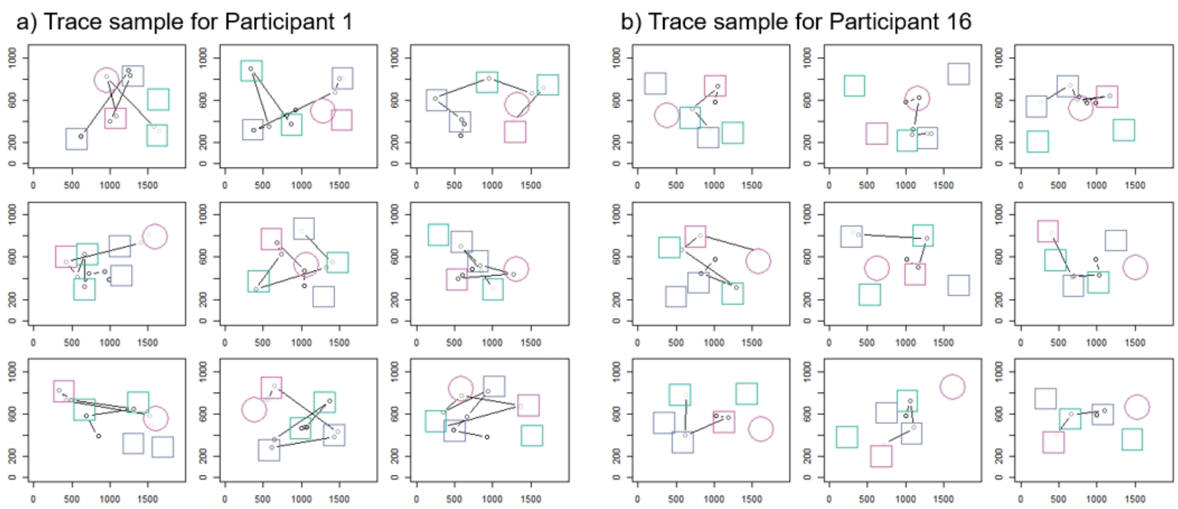


Figure 5.10. Selection of eye-gaze traces from two different subjects. a) data from participant 1 who consistently attempts to visit all items. b) data from participant 16 who typically checks only one item in each pair, thus utilising peripheral vision in the task.

Figure 5.10a shows data from a participant who consistently fixated on four to five of the six items presented in the task. From this small set of examples, it appears that they preferred to fixate on individual items, rather than utilizing information gathered in the periphery. Indeed, they seem to first fixate both items in a pair, before proceeding to the next pair. This can be contrasted with Figure 5.10b, which shows data from a participant who typically fixated on 3 items in the trial. Notably, the participant visits only one item from each colour, suggesting that they relied on their covert attention to encode the pairs of items. This behaviour seems consistent with utilizing peripheral information and thus reducing the number of fixations and more consistent with a covert style.

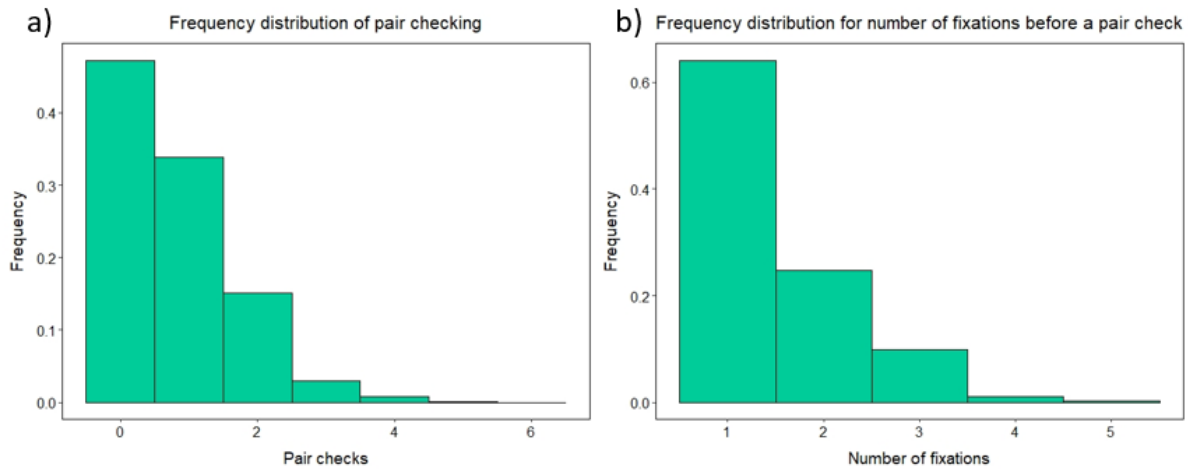


Figure 5.11. Frequency distributions of pair checking behaviour. a) the frequency with which participants make a given number of pair checks. Most frequently no pair checks are made. b) when a pair check is made, the frequency with which the check is made after a given number of fixations. Typically, the first pair check occurs on the first pair presented.

Pair checking. We sought to explore whether this difference between participants was systematic, or merely anecdotal, by quantifying aspects of these strategies. One way in which we did this was by measuring the number of times in a trial a participant would, after revealing the fixated item, immediately look to that item's pair. We named this behaviour "pair checking". Figure 5.11a depicts frequency of pair checking behaviour across all participants. It is most common for no pair checks to be made during a trial, however over half of all trials (53.7%) contain at least one pair check.

Figure 5.11b shows the frequency distribution for the number of fixations made before the first pair check in a trial. Most commonly, the pair check occurs after fixating one item. That is to say, the first pair check is usually made on the first pair revealed.

Figure 5.12. Pair checking strategy use by participant and trial. Participants are sorted by average number of pair checks per trial.

In viewing traces such as those in Figure 5.10, we saw an indication of frequency of pair checking varying with individual participants. Indeed, the strategy demonstrated by participant 16 shows relatively low pair checking as they appear to be purposefully looking at only one item in each pair. By contrast, participant 1 shows a high amount of pair checking. Examining pair checking over the course of the experiment for each individual, we found that use of this strategy was fairly consistent for individual participants (Figure 5.12).

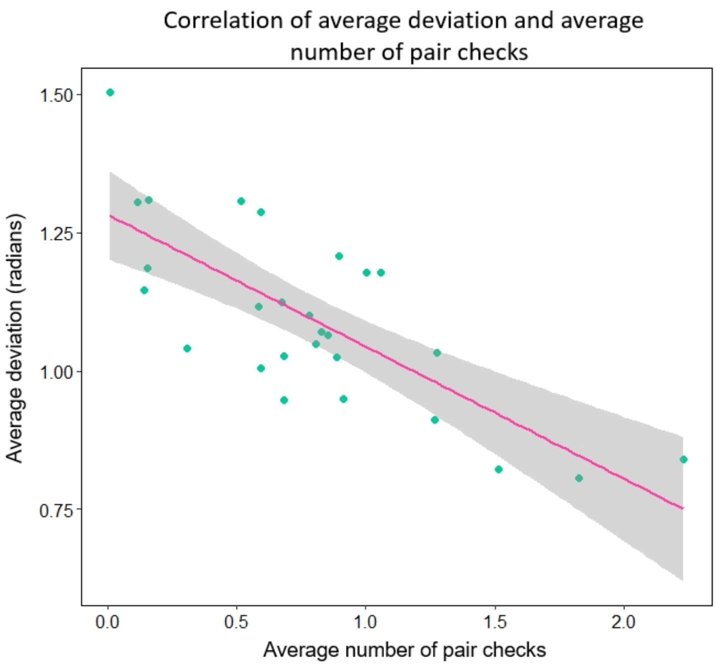


Figure 5.13. Average deviation and the average ratio of pair checks to number of stimulus fixations for each participant. Participants who have pair checks make up a higher proportion of trial fixations tend to be more accurate.

We were interested in seeing if pair checking was associated with task performance. Displayed in Figure 5.13, we found that an increase in pair checking as a proportion of all the fixations in a trial was associated with lower average deviations, i.e. better task performance. A test of Pearson's correlation revealed that there was a significant correlation ($t(25) = -5.75$,

p <0.001) such that participants who checked pairs more often, on average, also produced more accurate responses, on average.

Figure 5.16. Left-to-right and right-to-left strategy use by participant and trial. Participants are sorted by most common strategy use.

Other strategies. In addition to pair-checking behaviour, we were interested to see if there were other systematic ways in which participants explored the display. Using the algorithm described in Chapter 4 (page 136), we examined whether participants searched in horizontal or vertical strategies as well as if they explored items according to the distance between them (that is, choosing the shortest or longest path through the display). As before, we created an algorithm that remapped the location of all items in the display to those corresponding to the sequence of fixations expected according to each given strategy type. For a left-to-right exploration strategy, for example, we coded the strategy as starting at the leftmost item, then proceeding to the next leftmost item and so on. This sequence was compared to the fixation data for that trial and thus each trial was categorised whether it fit the strategy or not.

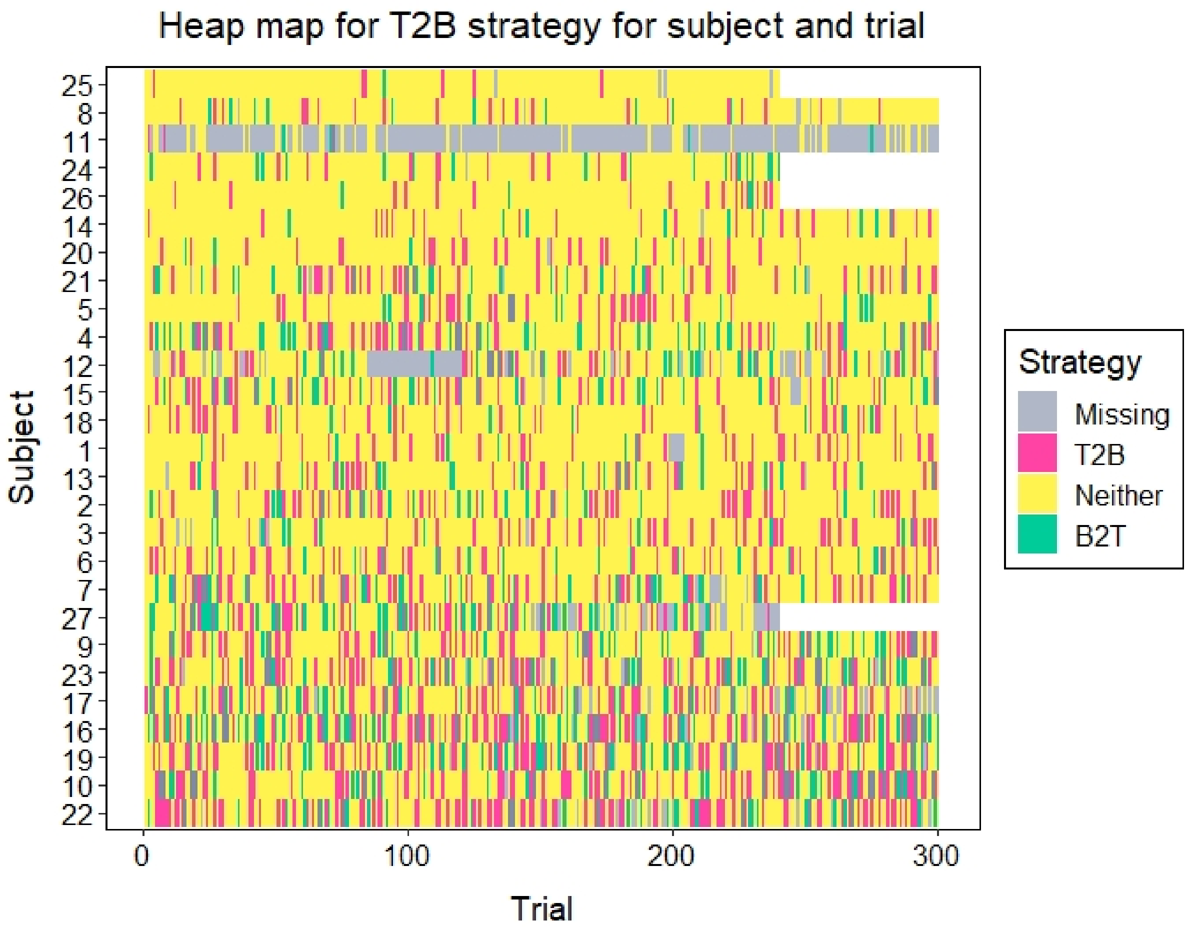


Figure 5.17. Top-to-bottom and bottom-to-top strategy use by participant and trial. Participants are sorted by most common strategy use.

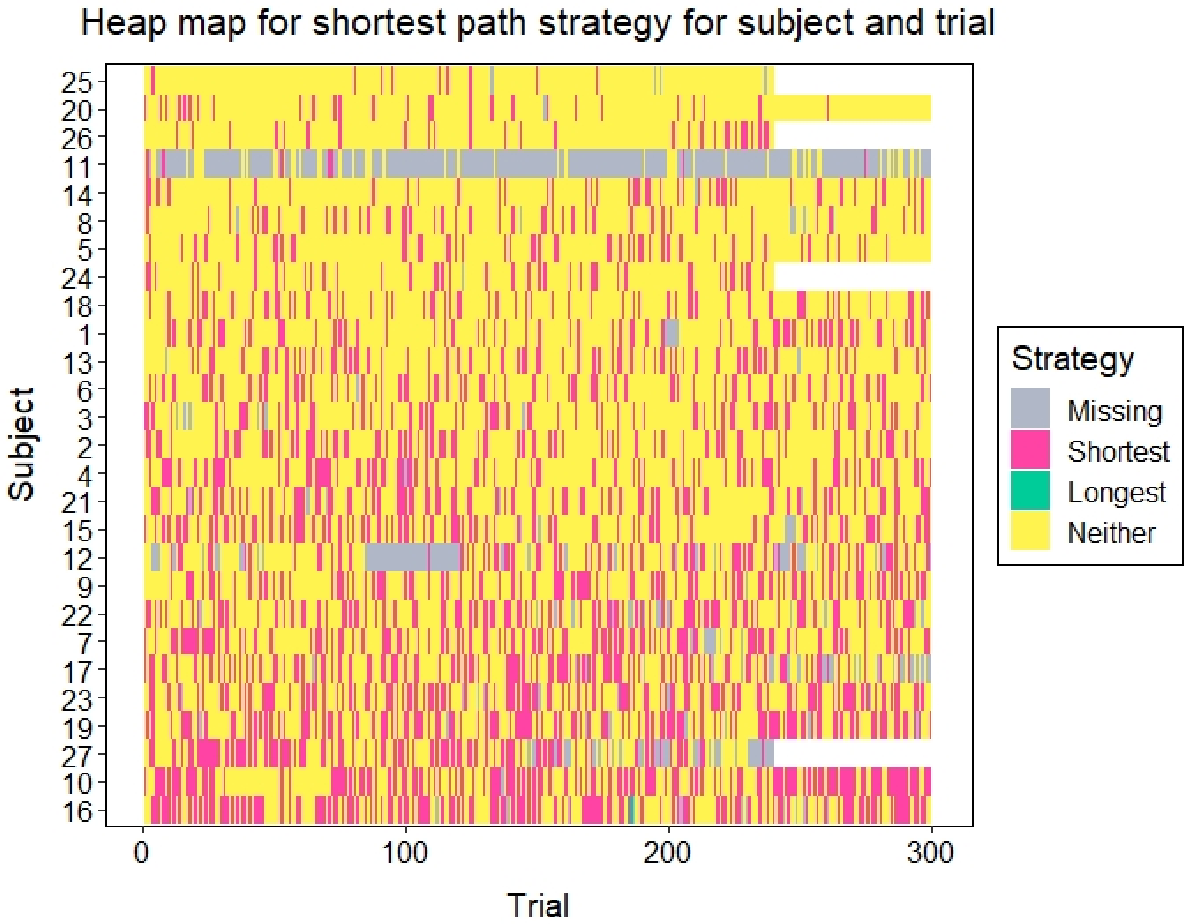


Figure 5.18. Shortest and longest path strategy use by participant and trial. Participants are sorted by most common strategy use.

Figure 5.16 depicts the prevalence of horizontal exploration strategies, while Figure 5.17 shows the prevalence of vertical exploration strategies. Figure 5.18 shows how often participants explored the display by the distance between items. Overall, it looks as though the most common strategy was to take the shortest path (26% of the data) between subsequent items.

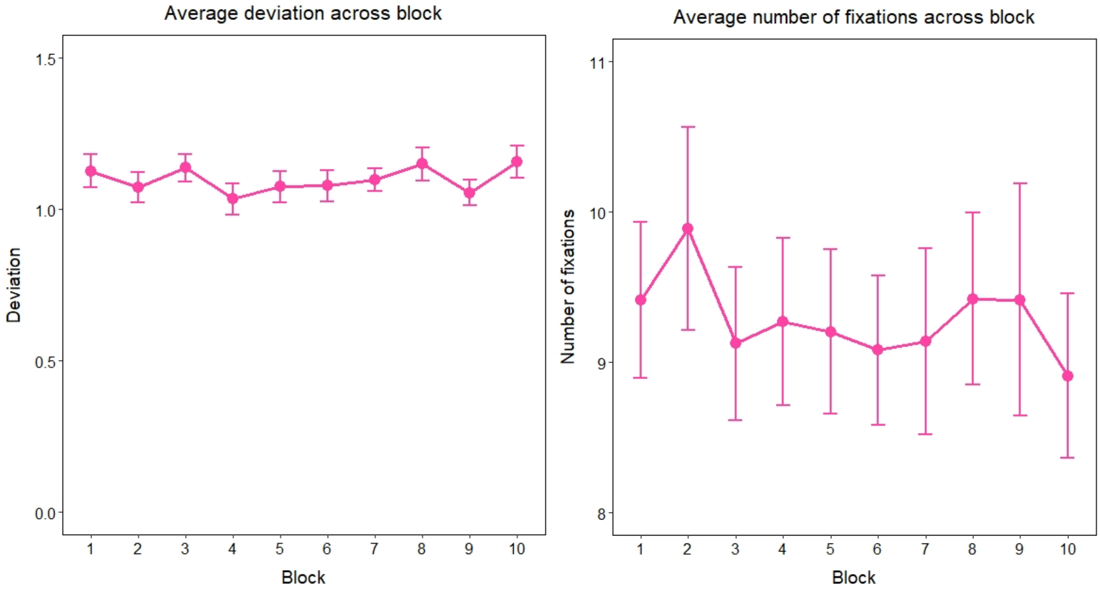


Figure 5.19. Average number of unique fixations across the experiment. Number of fixations is averaged by block (groups of 30 trials). a) Average across participants. b) individual participant averages. Error bars show standard error.

Learning. As with Experiment 3, we examined the average number of unique fixations (Figure 5.16b) and average deviation over time (Figure 5.16a). A one way ANOVA of block on the number of fixations revealed that there was no significant change in the number of fixations over time ($F(3,80) = 1.47$, $p = 0.22$). A one-way ANOVA of block on the average deviation across the experiment revealed that average deviation did not vary across the course of the experiment ($F(9,198) = 1.13$, $p = 0.34$).

Discussion

The aims of this experiment were to examine the effect of additional peripheral information on memory recall in this task; and to examine the strategies used by participants. In addition, we were hoping to repeat some of the analyses completed in Experiment 3, to see whether certain results were consistent across experiments.

Our results again showed effects of target-fixation duration and recency on performance. However, both of these effects were smaller compared to those in Experiment 3. For example, when comparing the relationship between deviation and target-fixation duration for fixations longer than 600ms in Experiment 3, we rarely saw highly inaccurate responses (see Figure 4.3). In the current experiment however, the fixations longer than 600ms are sometimes associated with high deviation responses (Figure 5.5). The distinction between high and low precision responses in Experiment 3 may be interpreted as more in line with a slot-like accounts of VWM that rely on the idea of all-or-none representations. However, this distinction is less clear in the current experiment, where we see a more gradual change in precision, even for shorter fixation durations, which seems more consistent with a resources account. One possibility is that the additional information in the periphery diverted attention away from the target item, covertly, which seems to have had a detrimental effect on the effect of overt attention.

We predicted that memory performance would decrease when participants could rely on only peripheral information. We found that seeing an item in the periphery led to more accurate performance than the chance-level behaviour that occurs when the target is never presented. However, the improvement in performance observed when the participant had also fixated on the target was relatively minor. This result seems to suggest why participants, on average, sometimes choose not to fixate on any items in the standard task: the extra benefit of fixating on items is perceived as less valuable than being able to see more items at once. As we speculated in the discussion of Experiment 3 (Chapter 4, page 150), it appears as though many participants are engaging in a trade-off between accuracy and efficiency. This seems to suggest that when items are presented peripherally, there is a tendency for participants to use a more diffuse attention to attempt to encode items in the display rather than focusing or fixating on a subset – a finding that seems more in line with a resources-like characterization of VWM.

When looking at the individual data, however, it is clear that there are a variety of different approaches to deploying attention across the display. In our results section we explored a few possible strategies, some identified in Experiment 3 and some novel. We found that individuals varied in terms of whether they used the strategies of checking pairs and revisiting items. The pair-checking behaviour was defined as, after fixating on an item, the next fixation was on the paired item that was just revealed in the periphery. Participants differed in their application of this approach, but we found that pair checking was associated with improved accuracy in the task. This supports the idea that, overall, more direct fixations, whether they are on previously seen stimuli or not, is related to better accuracy. This also ties in nicely to the correlation between response deviation and total fixation time on the target: the more time spent looking at the target, generally, the better the response accuracy. Together, these results suggest that fixations (even multiple fixations) will be associated with better task performance. However, while fixation leads to more accurate responses, the benefit it provides over seen an item in the periphery is not large, even though it is significant. We speculate that it is for this reason we see a divergence in participant behaviour. Some participants favour maximising accuracy at the cost of more eye movements and effort. Other participants however will trade off slightly lower accuracy for less effort expended, thus being more efficient.

Despite a potential advantage to accuracy, our results did not seem to suggest that pair checking was associated with awareness of memory precision for different items. When pair checking occurred, it most often occurred on the first item pair (Figure 5.11b). This means that after fixating on the first item, participants would check the paired item that was revealed in the periphery rather than seeking out unseen items. Given the small difference in accuracy between fixating an item and only seeing an item in the periphery, one might expect that a participant who was aware of the quality of their item memories to make three fixations to reveal all unique items, then make a pair check. This way, a participant would at least see all the items, even if they don't have time to fixate them all. This does not seem to be the case with participants seemingly preferring to pair check after the first item fixation. This might seem to imply that participants may not be aware of the precision of their memory representations, something that may be considered inconsistent with some theories of VWM (e.g., Van den Berg et al., 2012). However, it could be the case that participants who pair check are aware of the benefit to memory given by fixating an item directly and pair check for this reason. This explanation does seem to complement our finding that pair checking behaviour is associated with a benefit to accuracy on the memory task.

Finally, the strategies we examined in this task were much less common than the clockwise/counter-clockwise strategies in Experiment 3. This suggests that with a random distribution of items in the display, there is perhaps less coherence of participant strategies.

Conclusions. The current task successfully recreated the effects we saw in the previous experiment indicating that the fixation duration on the target and recency of fixation both affect memory performance. As expected, we saw that, compared to items fixated at study, the items only studied in the periphery were remembered less well. However, some participants favoured a strategy in which they preferred to view items in the periphery over fixating them. This suggests some participants are willing to trade accuracy for efficiency in eye movements. Likewise, there are also participants who favoured direct fixation over only seeing items in the periphery.

General discussion

Our initial motivation for this thesis was to find a connection between eye-gaze, attention and memory in VWM tasks in order to inform a process model of encoding in VWM. We had hoped to use eye-gaze data to inform our understanding of encoding so that we could describe the memory process in a novel way, and thus help to resolve the mimicry issues that make it hard to discriminate between certain theoretical accounts. Additionally, we hoped to investigate individual strategies within VWM tasks and how changes in task design might alter participants' approach to encoding stimuli. Ultimately, our initial goal of developing a process model of encoding in VWM was thwarted by the observation that participants often use covert attention in standard VWM tasks, obfuscating the link between eye-gaze and performance in VWM. However, despite the setbacks, each of our experiments do offer some insight into the nature of VWM as well as participants approaches to these tasks.

Summary of Results

Experiment 1 consisted of an orientation-recall task with eye tracking and blocked and unblocked set sizes. Based on Donkin et al. (2016), we believed that whether or not trials were grouped by set size would affect participants' deployment of attention. This appeared not to be the case, with participant behaviour being roughly equivalent in both conditions. It is possible that the lack of effect was due to a kind of floor effect with respect to the eye-gaze data. In this first experiment, we identified that the majority of fixations occur at the centre of the display rather than on any particular item. The impact of this pattern of fixations was demonstrated in our analysis of memory performance, where we observed an unexpectedly small improvement in recall accuracy due to fixating on the target. Regression analyses of the impact of time fixating the target and number of fixations on accuracy showed no effect. Overall, it appeared that the connection between eye-gaze and performance was weak. However, we were able to reveal some interesting patterns in gaze-fixation behaviour. The first stimulus fixation made in a trial was typically the longest. In contrast to this, however, recall was best when the target was fixated most recently during test – in other words, a recency effect. We also found a delay of around 300ms before the first item was fixated and that participants most commonly fixated on only three unique items in a trial, regardless of set size. Finally, we also saw that participants seemed to significantly reduce the number of fixations they made across the experiment.

In Experiment 2, we followed the task design of Smith et al. (2016). The behaviour in this task had been found to be explained well by an attention-weighted sample-size model (Smith et al., 2016), and we wanted to replicate these results and investigate whether eye-gaze behaviour could provide insight into the distribution of attention in such a task. The results of this experiment were, again, restricted by a lack of useful eye-gaze data. However, we were able to make some connections between the fixation behaviour in this experiment and that seen in Experiment 1. Again, we found that the first item fixated is generally looked at for the longest duration, and that when items were presented sequentially, the same was true of the first item presented. This finding was consistent with the predictions of the attention-weighted sample size model, which suggested the first item receives most of the attention in the display. However, unlike the predictions of the model, this increase in attention did not translate clearly to task accuracy. Contrary to the primacy effect found by Smith et al. (2016), accuracy was similar regardless of how recently the item was presented. This result seems to be related to the fact that the stimuli were presented for only short durations, and so the first item presented was often the last fixated, suggesting that there was insufficient time for participants to make additional fixations. To support this, we also found that participants, when they did fixate on items in the display, did not fixate all the items and that there was a delay between the start of the trial and the first item fixation. Due to this, there was a confounding of primacy and recency in this experiment. Finally, we found an effect of learning such that the number of fixations made by participants decreased across the course of the experiment.

Following the disconnect between eye-gaze and task performance found in the first two experiments, we decided to use a gaze-contingent design in Experiment 3. This experiment copied the design of Experiment 1, but made it such that participants needed to fixate on an item's location in order for a stimulus to be revealed. Once a stimulus was no longer fixated, it was hidden from view again. Using this design, we found a much clearer connection between eye-gaze and task performance. We found that total fixation time on the target was associated with what appeared to be a continuous improvement in task performance. Again, we saw an effect of recency of fixation, such that when the target was seen just before test, it was most accurately identified. Like Experiments 1 and 2, we also saw that the first item fixated was typically the item fixated the longest. This again suggests a less than straightforward relationship between fixation, primacy, recency, and task performance. Like Experiment 1, we did not see a difference in behaviour between the blocked and the unblocked conditions. This suggests that counter to the original finding by Donkin et al. (2016), participants' approach to encoding stimuli does not varying depending on whether the number of items to be remembered was known. Further, we did not see the reduction in fixations over time we had seen in our previous two experiments, but we did again see a delay between presentation and fixation of the first stimulus. We also saw that participants favoured exploring the display in either a clockwise (for set size 3) or an anticlockwise (for set size 6) pattern. However, participants seemed to differ in terms of the degree to which such strategies were employed.

In our final experiment, we wanted to create a design that was an intermediary between the standard task and our gaze-contingent design used in Experiment 3. Experiment 4 used similar gaze-contingent rules as Experiment 3 with the exception that when one item was fixated, a second item was revealed along with the fixated item. With this experiment we saw some similar effects as in Experiment 3: an increase in accuracy due to direct fixation on the target, an increase in accuracy with more time fixated on the target, as well as a beneficial effect of recency of fixation on task performance. Additionally, we found that having the target presented in the periphery, even if it was not fixated, was associated with an increase in performance. Like all of our previous experiments, we saw that the first item fixated was typically the longest fixation in the trial. There was no evidence of a reduction in number of fixations over the course of the experiment. Regarding strategy, we saw that participants sometimes chose to fixate on the item that appeared peripherally with an item they had just fixated on (which we termed 'pair checking'). This behaviour was not common, but when it did occur, the first pair check was made typically after the first fixation. This may be due to attention capture or because such strategies are easy to employ earlier in a trial, and suggest that pair checking is not typically due to a participant eventually going back to fixate items they had previously only seen in the periphery. When we analysed this on a person-by-person basis, it seemed that there was considerable variability in how much individuals would pair check. Examining other strategies, we did not find as much consistency in strategy with this task as with Experiment 3. However, the most popular strategy appeared to be to move one's eyes to the stimulus closest to your existing fixation.

Eye tracking and VWM tasks

Taken together, the similarities and differences in these experiments offer a range of insights into VWM, but perhaps the simplest conclusion is that eye tracking is of limited utility in terms of "standard" VWM tasks. From both Experiments 1 and 2 it is clear that participants frequently elect to fixate on the centre of the display rather than any particular item. This preference was more pronounced in Experiment 2, though even in that experiment, with just six participants, we saw a wide spectrum of possible behaviours – from making multiple fixations on every trial, to never leaving the central fixation. When analysing the data in these experiments, any connection between eye-gaze variables and memory performance was weak, at best. Experiment 1 found that fixation significantly improved accuracy, but beyond this, there was no indication that looking at the target longer, or looking at more or fewer items in a trial was associated with an increase in task accuracy, the probability of an item being in memory, or memory precision. In Experiment 2, not even fixating on a target was clearly associated with improved memory performance.

For these reasons, using passive eye tracking in these more standard tasks seems ill advised. With some task modifications, however, eye tracking in visual working memory tasks may indeed be useful. One key difference between Experiment 1 and 2 is trial duration. It seems likely that the longer durations in Experiment 1 encouraged more item fixations from participants. When considering the consistent delay between the start of the trial and the first stimulus fixation, having a stimulus presentation window greater than 300ms appears to be required in order for participants to make one stimulus fixation. Additionally, the spatial separation between the stimuli could be an important factor. It could be that the more spatially separated stimuli in Experiment 2 resulted in participants abandoning fixating individual stimuli due to it being too difficult (which was directly counter to our initial intention of trying to illicit more fixations by increased item separation). Alternatively, it is possible that since both experiments presented items in predictable locations (either on a circle in Experiment 1, or in the centre of each screen quadrant in Experiment 2) encouraged fewer direct fixations, as participants were more easily able to tune their peripheral vision to these stimuli. Both Experiment 1 and 2 show a reduction in fixations over time, though we do not see an increase in accuracy. This is to say that people are not learning how to do the task better, but they do seem to be learning to do it more efficiently (i.e., moving their eyes less). Another possibility is that participants are becoming increasingly disengaged over time, leading them to look at fewer items. However, the lack of a decrease in number of fixations in Experiments 3 and 4 suggests the former may be a better explanation. Regardless, this decrease in eye movement across the task helps create a disconnect between eye-gaze data and task performance, potentially undermining the utility of eye-gaze in any standard VWM task.

Given these findings, we recommend the use of gaze-contingent tasks to continue to investigate the connection between eye movements, attention, and VWM. It could be argued that a sequential presentation design with stimuli presented at the centre of the screen would also elicit a connection between eye-gaze and memory performance, but would also have the benefits of a controlled presentation order and stimulus exposure durations. However, this design does not allow participants a freedom of choice in how they approach the task. This freedom allowed by a gaze-contingent design gives insight into participants' strategies for encoding simultaneously presented stimuli. For this reason, we believe gaze-contingent designs offer unique insights into VWM.

While the gaze-contingent design we used in Experiments 3 and 4 is a departure from the standard task, the findings in these experiments seem informative. For example, while Experiment 3 is a straight gaze-contingent design, Experiment 4 reintroduces peripheral information into the task thus moving the experiment design closer to a standard VWM experiment. By examining encoding in these intermediary experiments, we believe it is possible to reveal the way in which participants utilise their peripheral vision in standard, non-gaze-contingent VWM tasks. Indeed, in Experiment 4 we saw a significant improvement in performance when the target was seen in the periphery but not fixated, compared to when it was not seen at all. Future experiments could continue to address this by adding additional peripheral stimuli to our Experiment 4 design. For example, fixating an item could reveal multiple peripheral stimuli instead of just one. Comparing the results of this experiment with our Experiment 4 would be interested to see how attention is divided between the two peripheral items - both may receive equal attention, or one may be attended over the other. Furthermore, asking whether increasing covert attention demands decreases the utility of overt attention seems an interesting and relevant question for understanding VWM. Such results would give insight into how participants allocate memory to items in their periphery in standard experiments.

Comparing slots and resources accounts of VWM

Despite our lack of explicit modelling, our results do shed some light on the current theories of VWM. Superficially, some of our results suggest a slot-like interpretation of VWM. For example, we find that people most commonly only fixate on three to four unique items on any given trial in Experiments 1, 3, and 4. This finding seems consistent with the purported capacity limit in VWM– a phenomena first introduced by slots accounts of memory (Luck & Vogel, 1997), and only recently proposed to be a necessary adjunct to resources account (Van den Berg et al., 2014). However, the general findings in our experiments are not particularly supportive of the slots-based account.

Most inconsistent with the idea that VWM should be considered an all-or-none, slots-like capacity is that there are a number of apparently gradual effects of attention on performance. For example, we see across most experiments that the total target fixation time seems to improve task accuracy relatively continuously. This gradual nature seems consistent with a resource-like VWM system, as it suggests an increase with precision as exposure to the target increases. By contrast, a slots model says that responses either come from memory, and are therefore of a fixed precision, or the response is a guess. However, there seems to be no indication of discrete changes in accuracy with fixation duration. Additionally, while fixation improves accuracy, so too does seeing an item in the periphery. Indeed, the accuracy of items only seen in the periphery are not as accurate, overall, as those fixated. However, the difference between the two types of entry into VWM does not appear to be discrete. There seems no way for a slots-like theory of VWM to explain the continuous nature of the difference between fixated and peripheral stimuli.

Overall, it would seem that a resources model provides a better account of our data than an all-or-none (or even some-or-none) slots model. For example, the variable-precision model (Van den Berg et al., 2012) can account for the gradual memory effects seen in our experiments. In their model, items in the display are encoded with a random precision assigned to each item (Van den Berg et al., 2012), which is consistent with our observation that even when an item has been long fixated by the participant, we sometimes see low precision responses. However, the variable-precision model does not provide a particularly satisfying explanation of the nature of such mistakes, since low precision responses in this model 'just happen'. If we were to incorporate fixation duration into the variable-precision model, assuming that longer fixations yield more precise representations, then such long-fixation, low-precision errors would no longer occur. That is, the variable-precision model, in its current form, is lacking a reason for why such errors happen. We believe this failure is indicative of another feature of the memory process that is not accounted for by the variable-precision model – namely, interference.

Interference and the focus of attention

The interference model developed by Oberauer and Lin (2017) provides an account of VWM which posits that memory performance is limited by confusion or overlap of stimulus representations in memory. The model is specified for items with continuous similarity spaces (such as the stimuli in our Experiments 1, 3 and 4) and contains three components: a long-term memory that maintains relevant memory representations, a region of direct access that holds information for a small number of stimuli at a time, and a focus of attention that selects single features within the region of direct access. In this model, items are stored in memory with noisy representations of their locations and features, which are linked by a binding parameter. This model has been shown to be able to replicate the set size effect demonstrated by slots and resources models, but was also able to make unique predictions regarding performance and item similarity. Previous research has found decreased performance when non-target items are less similar in colour (the feature to be recalled) to the target (Bona & Silvanto, 2014; Huang & Sekuler, 2010; Magnussen & Greenlee, 1999), but this result was never accounted for by existing slots and resource models. However, the interference model was able to explain this effect, as well as demonstrating the commonly found effect that non-target features are more likely to be recalled when they were presented in a location that was spatially similar to the target (Oberauer & Lin, 2017).

The strength of the interference model is that the source of errors in this account is not solely driven by a lack of attention. While holding an item in the focus of attention protects its representation in memory to some extent, all memories are vulnerable to interference from representations of other items (Oberauer & Lin, 2017). As a result, while a target may be fixated for a long duration and thus given a larger proportion of attentional resources, it is still able to be recalled incorrectly due to proximity and similarity with other items in the display. The interference model therefore provides an explanation for our finding that a target can be recalled incorrectly even if it has been fixated for a relatively long duration. Therefore, compared to the account presented by the variable-precision model, the interference model appears to be able to be integrated with our findings on attention.

Oberauer and Lin (2017) also provide a number of model variants that offer explanations of other phenomena, including the effect of recency. In their decay-refreshing interference model, when the focus of attention highlights a feature in memory, the strength of the bindings to that feature are increased. Thus, whole representations of items are better maintained if they are revisited, or, rehearsed. However, if an item is not rehearsed, the binding parameter decays, decreasing the ability of the location cue at test to be associated with the correct feature to recall. As such, the last item to be selected with the focus of attention will always have the most strength in the binding parameter and is thus the most likely to be recalled correctly, thus yielding the desired recency effect.

This finding is mirrored in the consistent recency findings within our data. Aside from Experiment 2, all of our experiments displayed a clear benefit to task performance when the target is seen more recently within the trial. Critically, this effect appears independent of the total time spent fixating on the target or the number of revisits an item has. This is in line with the decay-refreshing interference model, which says that the most recently focused item should have the strongest representation, not the item that has been accessed the greatest number of times or received the most focus overall.

The concept of a focus of attention was also examined by McElree (2001), who found that recall was strong for items that were focused on by participants, but this attention could typically only be applied to one item at a time. McElree (2001) used a speed/accuracy trade-off variant of an *n*-back task. On each trial, participants were presented with a sequence of 6-

15 letters presented at the centre of the screen, followed by a mask, before a test letter was presented for an interval ranging between 43 and 3000 ms. The participant was then required to respond if the target letter appeared in the *n*-back (*n* = 1, 2, 3) position in this trial or not. Related to our own results, McElree (2001) found that larger *n* and more time between the presentation array and recall had a negative effect on task performance. More than this, when fitting models to the accuracy and processing time data (target presentation time + response time), he found that the data was not well accounted for by a model that searched memory serially in order of recency. Instead, the best explanation came from a mixture model that combined this backwards search with a focal attention component that enabled a fast match on trials when participants were able to maintain the *n*-back item in their focus of attention. The implication of McElree's model is that the performance boost for the most recently presented item is not simply due to recency of exposure, since recency is also favoured in the backwards search model. Rather, the improved performance of the mixture model suggests the need to posit a focus of attention process.

Unlike the *n*-back task, in which participants are specifically trying to maintain a memory of an item *n*-back from the test item, in our experiments participants did not have any information to help guide them as to what they should remember. This is consistent with the decay-refreshing interference model, which says that recency and focus of attention align in the absence of any preference for which items are refreshed (Oberauer & Lin, 2017).

The results of our experiments appear largely consistent with the account just described, with the exception of Experiment 2, where we did not find clear primacy or recency effects. Furthermore, the study by Smith et al. (2016), on which Experiment 2 was based, found an effect of primacy rather than recency. While this seemingly provides evidence against the interference model, we believe that in both Experiment 2 and the Smith et al. (2016) study, recency and primacy are confounded by the relatively short stimulus presentation time. As a result, it is hard to compare the results of these experiments with our other experiments. Across our experiments we found a consistent delay between the start of the trial and the first fixation on an item of around 300ms. In the 200ms simultaneous presentation condition in Experiment 2, we found that this meant participants were very unlikely to fixate on any items. This meant that by the time participants would make their first fixation on an item, the trial would be over. This creates the aforementioned confound, as the first item seen is also the last item seen in these short duration trials. Therefore, primacy and recency attention effects overlap in this design. Given that the presentation durations in the experiments by Smith et al. (2016) were at longest equal to our shortest presentation times in Experiment 2, this confound is likely also present in their experiment. Ignoring the semantic question of whether these are better considered recency or primacy effects, it does seem that the account of the attention-weighted sample-size model is consistent with the notion of a focus of attention, applied to one item in the display, with remaining attentional resources divided between the other items (i.e. those seen in the periphery).

Strategy in VWM

In our exploration of strategy, we examined to covert and overt strategies, how fixation behaviour changed across the experiment, as well as attempting to categorise patterns participants used to explore the display. We found that participants tended to prefer efficiency – showing, for example, a reduction in average fixations across the experiment, which highlighted a potential reason for more covert strategies. However, while individuals did appear to vary in the degree to which they used covert and overt attention, there was not often a clear delineation between participants of these strategy types. Our attempts to categorise different eye-movement patterns (or strategies) – e.g., into clockwise or top-to-bottom – also showed a good amount of individual variation. However, in this case, strategies were much more consistent, depending the nature of the display. When stimuli were presented in a circular formation, participants most commonly explore the display in a clockwise or anticlockwise pattern.

As discussed in the findings of Experiment 3, there was a strong preference for a clockwise search pattern. Subsequently we performed the same analysis of strategy on the data from Experiment 1, which shared the same design features as Experiment 3 (ring and bead items, presented in a circle around the centre of the screen and equidistant from each other), but did not use the gaze-contingent component.

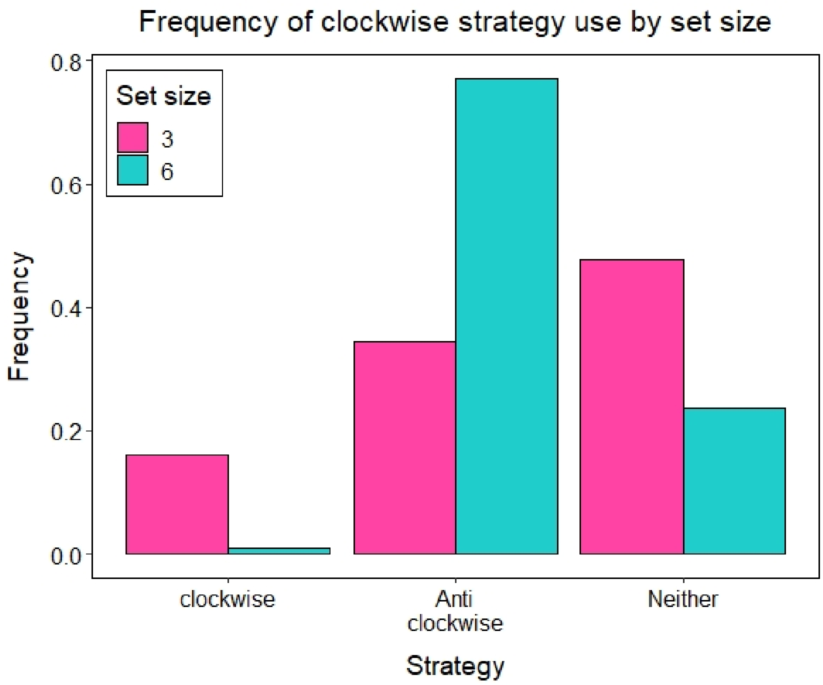


Figure 6.1. Frequency of clockwise strategy use by set size in Experiment 1. Similar to Experiment 3, an anti-clockwise strategy is preferred in set size 6 and a clockwise strategy is more likely to be used in set size 3.

The results, shown in Figure 6.1, revealed that similar to Experiment 3, an anti-clockwise strategy was more common for set size 6 than for set size 3. It was also more likely to see a clockwise strategy used in set size 3. However, overall, in Experiment 1 it was more likely to see an anticlockwise strategy used compared to Experiment 3. These results suggest that there are some consistent patterns in strategy that emerge between experiments and these strategies seem dependent on how items are presented on screen.

When the stimuli were presented in a less clear structure (c.f., Experiment 4), we were not able to find an overwhelmingly common strategy, though the shortest path between subsequent stimuli was most popular. Overall, it seems that the pattern of exploration of a display is dependent on how the items are arranged onscreen. This suggests that strategy, as suggested by Bengson and Luck (2016), could vary significantly with task design.

Our analysis of strategy in this thesis is only a starting point of exploration in this area. Currently, there is a relative paucity of dedicated research on this topic in the VWM literature. However, recently, an article by Gonthier (2020) reviewed a variety of articles relating to visuospatial memory processes, with the aim of categorising the types of strategy used. Holistic processing was one such category, which is associated with keeping gaze fixed on the centre of the screen. Gonthier remarks that keeping one's gaze on the centre of the display may be seen as an attempt by participants to encode all items at once, as a single picture. Indeed, in a study that included self-report it appeared that participants could explain that they kept their gaze centred for this purpose (Pearson & Sahraie, 2003). Under this interpretation, it could be that the participants in our passive eye tracking studies are attempting a similar strategy. With a gaze-contingent paradigm, this strategy is not possible, and instead participants must necessarily adopt a more overt strategy. Within our categorisation of exploration patterns, however, it is difficult to determine how to categorise these overt strategies with the current eye-gaze data alone. For example, we saw that participants in Experiment 1 and 3 favoured using an anticlockwise strategy for set size 6, but it is not clear as to why. It could be that participants are still attempting to create a holistic picture of the display or it could be that they are trying to remember items with respect to their relative positions – what Gonthier terms a relational strategy. Different again, participants could also be, for example, using a strategy that enlists long term memory by imagining the items as positions on a clockface.

Future experiments may benefit from creating a design that could distinguish such strategies from each other – something that is challenging when trying to avoid relying on measures of self-report. However, the exploration of strategy self-report is interesting in itself. Specifically, it would be interesting to know how consistent participants self-reports of strategy are with their eye movements. For example, if a participant reports using a holistic strategy in a standard VWM task, do they tend to move their eyes less than a participant who reports a different strategy? When exploring overt strategies, examining eye-gaze behaviour and self-report with variety of different display types (in a gaze-contingent task) would also show the degree to which participants change their reported strategy or behaviour with a change in task presentation. Using a study array with, for example, three items grouped together spatially while one item is separated from this group might encourage participants to use a chunking strategy (Gonthier, 2020; Miller, 1956). This might be seen in prioritising fixations on the grouped items first, showing a priority for information density. Throughout the task the spatial distance between all items could vary such that grouped items would be closer together or further apart. Such a design could be used to investigate at what point participants show preference for grouped items, and at what point they use an alternate strategy to approach the task.

Conclusions

This project began by attempting to integrate encoding information into a computational model of VWM. We sought to use eye-gaze as our proxy for attention to investigate how participants go about learning memory arrays and which account of VWM is more supported with these patterns of eye-gaze. However, with the covert strategy use in our first two experiments, we discovered more questions about the role of attention and strategy which were difficult to address with standard VWM tasks. With the use of gaze contingent designs, we were able to better understand the connection between attention and memory. Our results suggest that more than a slots or resources based account, an interference model (Oberauer & Lin, 2017) of VWM provides the best theoretical explanation of the results seen in our experiments. When regarding eye-gaze and memory what seems to be most important is what is held in attention closest to test and for an array without any cued stimulus, this appears to be the item attended most recently.

In regards to understanding strategy in VWM, our results offer the suggestion that design of the experiment is important in how participants approach the task. Because of this, strategy between different working memory tasks is unlikely to be uniform. In fact, this could be a reason that while some studies have shown support for a slots account (e.g. Zhang and Luck, 2008) other studies have found their results more consistent with a resources account of VWM (e.g. Wilken & Ma, 2004). In summary, participants appear quite flexible in how they could approach a task but take cues from the design to direct their strategy and this should be taken into account in future experiments. However, more work needs to be done to understand both the scope of strategies used as well as the variables that encourage strategy change in participants.

In terms of both the role of attention and strategy in visual working memory, it seems clear that the standard VWM tasks are not suited to answering many of these questions. However, these questions are vital if we are to create a complete understanding of VWM. Given this, gaze-contingent tasks appear to be a useful resource in beginning to answer many of these questions.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological science, 15*(2), 106-111.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*(5890), 851-854.

Beesley, T., Pearson, D., & Le Pelley, M. (2015). Implicit learning of gaze-contingent events. *Psychonomic Bulletin & Review, 22*(3), 800-807.

Bengson, J. J., & Luck, S. J. (2016). Effects of strategy on visual working memory capacity. *Psychonomic Bulletin & Review, 23*(1), 265-270.

Bona, S., & Silvanto, J. (2014). Accuracy and confidence of visual short-term memory do not go hand-in-hand: behavioral and neural dissociations. *PloS one, 9*(3), e90808.

Boot, W. R., Becic, E., & Kramer, A. F. (2009). Stable individual differences in search strategy?: The effect of task demands and motivational factors on scanning strategy in visual search. *Journal of vision, 9*(3), 7-7.

Boot, W. R., Kramer, A. F., Becic, E., Wiegmann, D. A., & Kubose, T. (2006). Detecting transient changes in dynamic displays: The more you look, the less you see. *Human Factors, 48*(4), 759-773.

Christ, S. E., & Abrams, R. A. (2006). Just like new: Newly segregated old objects capture attention. *Perception & Psychophysics, 68*(2), 301-309.

Connor, C. E., Egeth, H. E., & Yantis, S. (2004). Visual attention: bottom-up versus top-down. *Current biology, 14*(19), R850-R852.

Cowan, N., & Morey, C. C. (2006). Visual working memory depends on attentional filtering. *Trends in cognitive sciences, 10*(4), 139-141.

Cusack, R., Lehmann, M., Veldsman, M., & Mitchell, D. J. (2009). Encoding strategy and not visual working memory capacity correlates with intelligence. *Psychonomic Bulletin & Review, 16*(4), 641-647.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision research, 36*(12), 1827-1837.

Dickinson, C. A., & Zelinsky, G. J. (2005). Marking rejected distractors: A gaze-contingent technique for measuring memory during search. *Psychonomic Bulletin & Review, 12*(6), 1120-1126.

Donkin, C., Kary, A., Tahir, F., & Taylor, R. (2016). Resources masquerading as slots: Flexible allocation of visual working memory. *Cognitive psychology, 85*, 30-42.

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual review of psychology, 48*(1), 269-297.

Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review, 12*(6), 1127-1133.

Findlay, J. M., & Gilchrist, I. D. (2001). Visual attention: The active vision perspective. In *Vision and attention* (pp. 83-103): Springer.

Fougnie, D. (2008). The relationship between attention and working memory. *New research on short-term memory, 1*, 45.

Frick, R. W. (1988). Issues of representation and limited capacity in the visuospatial sketchpad. *British Journal of Psychology, 79*(3), 289-308.

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends in cognitive sciences, 16*(2), 129-135.

Geisler, W. S., Perry, J. S., & Najemnik, J. (2006). Visual search: The role of peripheral information measured using gaze-contingent displays. *Journal of vision, 6*(9), 1-1.

Gonthier, C. (2020). Charting the Diversity of Strategic Processes in Visuospatial Short-Term Memory. *Perspectives on Psychological Science*, 1745691620950697.

Gorgoraptis, N., Catalao, R. F., Bays, P. M., & Husain, M. (2011). Dynamic updating of working memory resources for visual objects. *Journal of Neuroscience, 31*(23), 8502-8511.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics, 57*(6), 787-795.

Huang, J., & Sekuler, R. (2010). Distortions in recall from visual memory: Two classes of attractors at work. *Journal of vision, 10*(2), 24-24.

Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception & Psychophysics, 64*(6), 882-895.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, 105*(19), 6829-6833.

Lange, E. B., & Engbert, R. (2013). Differentiating between verbal and spatial encoding using eye-movement recordings. *Quarterly journal of experimental psychology, 66*(9), 1840-1857.

Lange, E. B., Starzynski, C., & Engbert, R. (2012). Capture of the gaze does not capture the mind. *Attention, Perception, & Psychophysics, 74*(6), 1168-1182.

Lawrence, B. M., Myerson, J., & Abrams, R. A. (2004). Interference with spatial working memory: An eye movement is more than a shift of attention. *Psychonomic Bulletin & Review, 11*(3), 488-494. doi:10.3758/BF03196600

Lawrence, B. M., Myerson, J., Oonk, H. M., & Abrams, R. A. (2001). The effects of eye and limb movements on working memory. *Memory, 9*(4-6), 433-444. doi:10.1080/09658210143000047

Le Pelley, M. E., Pearson, D., Griffiths, O., & Beesley, T. (2015). When goals conflict with values: Counterproductive attentional and oculomotor capture by reward-related stimuli. *Journal of Experimental Psychology: General, 144*(1), 158.

Liversedge, S. P., Rayner, K., White, S. J., Vergilino-Perez, D., Findlay, J. M., & Kentridge, R. W. (2004). Eye movements when reading disappearing text: is there a gap effect in reading? *Vision research, 44*(10), 1013-1024.

Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science, 240*(4853), 740-749.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279.

Magnussen, S., & Greenlee, M. W. (1999). The psychophysics of perceptual memory. *Psychological research, 62*(2), 81-92.

Marr, D. (1982). Vision: a computational investigation into the human representation and processing of visual information. W. H. *WH San Francisco: Freeman and Company*.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics, 17*(6), 578-586.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(3), 817.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81.

Nosofsky, R. M., & Donkin, C. (2016). Response-time evidence for mixed memory states in a sequential-presentation change-detection task. *Cognitive psychology, 84*, 31-62.

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D., Conway, A., Cowan, N., . . . Hurlstone, M. J. (2018). Benchmarks for models of short-term and working memory. *Psychological bulletin, 144*(9), 885.

Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review, 124*(1), 21.

Over, E., Hooge, I., Vlaskamp, B., & Erkelens, C. (2007). Coarse-to-fine eye movement strategy in visual search. *Vision research, 47*(17), 2272-2280.

Palmer, J. (1988). Attentional limits on the perception and memory of visual information. *Journal of experimental psychology: human perception and performance, 16*(2), 332.

Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision research, 34*(13), 1703-1721.

Pearson, D., & Sahraie, A. (2003). Oculomotor control and the maintenance of spatially and temporally distributed events in visuo-spatial working memory. *The Quarterly Journal of Experimental Psychology Section A, 56*(7), 1089-1111.

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology, 32*(1), 3-25.

Postle, B. R., Idzikowski, C., Sala, S. D., Logie, R. H., & Baddeley, A. D. (2006). The selective disruption of spatial working memory by eye movements. *The quarterly journal of experimental psychology, 59*(1), 100-120. doi:http://dx.doi.org/10.1080/17470210500151410

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The quarterly journal of experimental psychology, 62*(8), 1457-1506.

Rayner, K., Smith, T. J., Malcolm, G. L., & Henderson, J. M. (2009). Eye movements and visual encoding during scene perception. *Psychological science, 20*(1), 6-10.

Ricker, T. J., & Hardman, K. O. (2017). The nature of short-term consolidation in visual working memory. *Journal of Experimental Psychology: General, 146*(11), 1551-1573. doi:http://dx.doi.org/10.1037/xge0000346

Rouder, J., Thiele, J., & Cowan, N. (2014). *Evidence for guessing in working-memory judgments.* Paper presented at the 55th annual meeting of the Psychonomic Society.

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences, 105*(16), 5975-5979.

Sewell, D. K., Lilburn, S. D., & Smith, P. L. (2014). An information capacity limitation of visual short-term memory. *Journal of experimental psychology: human perception and performance, 40*(6), 2214.

Smith, Lilburn, S. D., Corbett, E. A., Sewell, D. K., & Kyllingsbæk, S. (2016). The attention-weighted sample-size model of visual short-term memory: Attention capture predicts resource allocation and memory load. *Cognitive psychology, 89*, 71-105.

Smith, & Ratcliff. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review, 116*(2), 283.

Sparks, D. L. (2002). The brainstem control of saccadic eye movements. *Nature reviews neuroscience, 3*(12), 952-964.

Thornton, T. L., & Gilden, D. L. (2007). Parallel and serial processes in visual search. *Psychological Review, 114*(1), 71.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology, 12*(1), 97-136.

Van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review, 121*(1), 124.

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.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2005). Pushing around the locus of selection: Evidence for the flexible-selection hypothesis. *Journal of cognitive neuroscience, 17*(12), 1907-1922.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of vision, 4*(12), 11-11.

Woodman, G. F., & Luck, S. J. (2007). Do the contents of visual working memory automatically influence attentional selection during visual search? *Journal of experimental psychology: human perception and performance, 33*(2), 363.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: evidence from visual search. *Journal of experimental psychology: human perception and performance, 10*(5), 601.

Zang, X., Jia, L., Müller, H. J., & Shi, Z. (2015). Invariant spatial context is learned but not retrieved in gaze-contingent tunnel-view search. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(3), 807.

Zelinsky, G. J., & Loschky, L. C. (2005). Eye movements serialize memory for objects in scenes. *Perception & Psychophysics, 67*(4), 676-690.

Zhang, & Luck. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233-235. doi:10.1038/nature06860

Zhang, & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological science, 22*(11), 1434-1441.