In D. Reitter & F. E. Ritter (Eds.), Proceedings of the 14th International Conference on Cognitive Modeling (ICCM 2016). University Park, PA: Penn State.

The Representation of Visual Working Memory

Bella Z. Veksler, Rachel Boyd
(bellav717@gmail.com) (rachel.boyd.rb1@gmail.com)
Oak Ridge Institute for Science & Education at AFRL

Christopher W. Myers, Glenn Gunzelmann
(christopher.myers.29@us.af.mil) (glenn.gunzelmann@us.af.mil)
Air Force Research Laboratory, Wright-Patterson AFB, OH, USA

Hansjörg Neth (h.neth@uni-konstanz.de)
University of Konstanz, Konstanz, Germany

Wayne D. Gray (grayw@rpi.edu)
Rensselaer Polytechnic Institute, Troy, NY, USA

Abstract

Visual working memory (VWM) is a construct hypothesized to store a small amount of accurate perceptual information that can be brought to bear on a task. Much research concerns the construct's capacity and the precision of the information stored. Two prominent theories of VWM representations have emerged: slot-based and continuous-resource mechanisms. Prior modeling work suggests that a continuous resource that varies over trials with variable capacity and a potential to make localization errors best accounts for the empirical data. Questions remain regarding the variability in VWM capacity and precision. Using a novel eye-tracking paradigm, we demonstrate that VWM facilitates search and exhibits effects of fixation frequency and recency, particularly for prior targets. Whereas slot-based memory models cannot account for the human data, a novel continuous-resource model provides a better fit and identifies the relevant resource as item activation.

Keywords: visual working memory; visual search; ACT-R.

Introduction

Visual working memory (VWM) is a construct hypothesized to be a limited capacity system that maintains representations of visual information for temporary storage and manipulation for ongoing tasks (Luck & Vogel, 2013). This construct has garnered much attention and has been the focus of many studies and computational models. Even so, answers to fundamental questions, such as its capacity and representation precision, remain elusive (van den Berg & Ma, 2014).

Two theories of VWM representations dominate the literature: *slot* and *continuous resource* mechanisms. Slot theories generally posit a fixed capacity of 3 to 4 items with high to perfect precision (Luck & Vogel, 2013). A slot is a discrete memory container filled with an object representation with bound visual features (Luria & Vogel, 2011). Information stored within a slot can be accurately applied to a task regardless of its visual complexity, be it a single vertical line or a complex Chinese character.

By contrast, continuous resource theories of VWM posit a finite resource that can be spread across different areas of a scene or item. This resource is seen as a pool of mental processing power dedicated to VWM, which can be flexibly distributed across items in a display (Wilken & Ma, 2004). Fewer objects to be encoded lead to less distributed memory resources and allow for more precise object representations.

Recently, Donkin, Kary, Tahir, and Taylor (2016) have argued that a VWM system using a continuous resource may appear to support a slot interpretation when the number of items to remember varies from trial to trial. At times highly precise representations of a small number of objects appear to favor a slot-based model, but when set size is unpredictable participants are biased to focus on a small subset of items, leading to performance suggestive of a slot model. When set size was predictable (the same across multiple trials), resource models best characterized the data.

Van den Berg, Awh, and Ma (2014) varied precision, capacity, and the potential for spatial binding errors as three independent factors of VWM to test lingering questions. Using a 4x4x2 factorial design, all models were tested on 10 previously published empirical results from a change detection paradigm. The results indicated that a continuous model which varied both storage capacity and precision across trials, combined with the presence of the potential for spatial binding errors best accounts for the data. However, questions regarding the mechanisms behind the variance in precision and capacity remain unanswered.

A passive, tachistoscopic version of the change detection paradigm has been the dominant approach to establishing prominent theories of VWM (Alvarez & Cavanagh, 2004), and continues to be the paradigm most used in contemporary empirical research on VWM (Donkin et al., 2016). In this task, a participant is instructed to attend to and remember information within a stimulus display. The information is typically a set of unique objects that differ across features, such as shape and color. After some time the stimulus disappears and after a delay the object of the possible change is cued, or a new stimulus appears. If a change occurred, the participant must indicate the change in some manner, either by responding yes/no (c.f., Alvarez & Cavanagh, 2004), by identifying what has changed (c.f., D. E. Anderson, Vogel, & Awh, 2013), where the change occurred (c.f., Barton, Ester, & Awh, 2009), or some combination thereof. Researchers vary the number of items in a stimulus (i.e., set size) to evaluate VWM capacity and use change identification to evaluate VWM precision.

There are weaknesses in the passive change detection approach to understanding VWM (Rouder, Morey, Morey, & Cowan, 2011). Specifically, many, if not all, VWM studies rely on a passive approach to understanding VWM rather than an active one (Findlay & Gilchrist, 2003). Outside the experimental laboratory, visual search does not occur in a vacuum, but rather in the context of a task where targets contained in some visual array are distinguished from distractors. We argue that passive change detection with delayed responses (~2–3 s) does not tap into the functional importance of VWM — to facilitate the accurate completion of an active visual task through the temporary storage of readily available and accurate visual information.

In the current paper we provide an explanation for the variance in VWM precision and capacity. To do so, we introduce a new eye-tracking paradigm that moves away from the change detection tasks commonly used to investigate VWM. Our new paradigm of *repeated serial search* (Neth, Gray, & Myers, 2006) requires an individual to actively search for different (and sometimes repeating) targets within a stable visual display and thus represents a task that is more realistic and ecologically valid than a passive change detection paradigm. Importantly, it allows us to ask questions of VWM that inform how VWM drives search behavior and the potential differences in depth of encoding between targets and distractors since we have access to the full history of fixations.

Our empirical and modeling work leads to five important conclusions: (1) the variability in VWM capacity results from recency and frequency effects from selectively encoding visual information; (2) VWM precision variability results from the same recency and frequency effects; (3) memory facilitates search behavior; (4) targets have a stronger mnemonic trace than distractors; and (5) the relevant "resource" involved is memory activation. In the following sections we introduce our paradigm and present empirical results, followed by a model analysis of the empirical data.

Experiment

To determine the degree to which VWM facilitates visual search, we designed an experiment using a novel *repeated serial search* paradigm. In this paradigm, participants were required to search the same spatial configuration of 10 static items a total of 20 times. This paradigm taps into the VWM construct, motivating participants to retain a maximum amount of information in VWM to facilitate future searches.

Paradigm. On each trial, ten circular objects with a diameter of 60 pixels were distributed randomly over a centered white rectangular display area (measuring 1270-by-970 pixels). The objects were positioned at least 60 pixels away from any edge and the distance between the centers of any two objects was constrained to be at least 200 pixels. Each circle contained a hidden label (upper case letter, number, or monosyllabic four-letter word) that specified the target sought by the participant. On any given trial only one type of label was in the circles (letters, numbers, or words). The order of label types was randomized within each participant's task presentation.

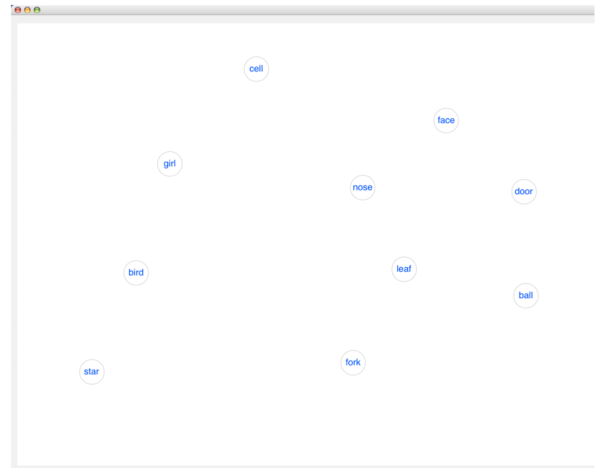

Figure 1. Example stimulus used in the experiment. Although all labels are visible here, they were hidden from participants' view until a cursor hovered within the circle.

Each trial was composed of 20 searches through the display. At the beginning of each search, the experimental software announced the current target label to the participant (e.g., "cell" in Figure 1). Participants could hover with the mouse cursor over each circle to uncover its hidden label. Once the cursor was moved off the circle, the corresponding label was hidden again. Participants were instructed to click on the circle corresponding to the target label. If the clicked circle indeed contained the correct target label, a new target was announced; however, if a clicked circle contained a different label the software recorded an error and the current target was announced again to provide a reminder to the searcher. Consequently, searchers typically uncover nontargets (distractors) in the process of searching for targets and these distractors may turn into targets in subsequent searches.

There were three within-participants information presentation types that manipulated the number of intervening targets between identical targets. While they are of theoretical interest, we collapse across these presentation types for the current analyses to save space and mitigate complexity.

Participants. A total of 13 Rensselaer Polytechnic Institute undergraduates (3 females) volunteered for course credit. Their mean age was 18.92 years ($SD = 1.04$).

Procedure. Participants signed informed consent forms, viewed a slideshow of the instructions, and were calibrated to an LC Technologies eye tracker prior to beginning the study. Every participant completed 60 trials in total. Each trial consisted of a series of 20 searches. Every search commenced when a computer generated voice announced the next target to be found.

Results

A timeline of the sequence of fixations during every search within a trial was created for each participant. This was made possible through the collection of visual point of regard and mouse click data while participants performed the visual search task. Given this sequence of fixations, we can determine the frequency of fixations across labels and how long ago — in terms of duration and the number of intermediate items — each label was last viewed to investigate recency and frequency effects in finding a target. We can also determine differences that are due to the functional role of labels (i.e., whether labels were previously seen and encoded as targets or as distractors).

Recency effects. For this analysis we restricted the data to the first two times an item was a target of a search. A 2 (label-type)-by-10 (recency) ANOVA was performed to evaluate the effect of label encoding and recency of last fixation. There was an interaction between whether an item was a target before and how recently it was last fixated, $F(9, 108) = 3.76, p < .001, \eta^2=0.24$. There was also a significant main effect of recency, $F(9, 108) = 11.94, p < .001, \eta^2=0.50$ (see Figure 2 top). This effect was greater for labels that had not been previous targets, $F(1, 12) = 73.42, p < .001, \eta^2=0.86$. In general, labels that were prior targets were less impacted by recency of fixation.

One explanation for the inverted U-shape of the items that were only distractors prior to the current search is that as participants are searching the display, they are only encoding whether or not the current item is the target, rather than the identity of the item. This depth of encoding may result in an inhibition of return effect for more recently fixated items (2–5 fixations ago) leading to longer search times than when the distractor was seen longer ago.

Frequency effects. A 2 (label-type)-by-7 (frequency) AVOVA was performed to evaluate the effect of label encoding frequency. There was insufficient data in frequency bins 1 and 2 (i.e., in cases where a second search for a target was preceded by one or no fixations on the item prior to the search), leaving bins 2–8 for analysis (see Figure 2 bottom). Nonetheless, these bins reflect the general trend in the data. There was a significant interaction between fixation frequency and label-type on the number of fixations to find the target, $F(6, 72) = 5.92, p < .001, \eta^2=0.33$, where searches required fewer fixations when a label had been a target before despite being seen less than 5 times, $F(1, 12) = 38.37, p < .001, \eta^2=0.76$, (Figure 2 bottom). Further, there was a main effect of frequency on number of fixations to find the target, $F(6, 72) = 3.25, p < .01, \eta^2=0.21$. In particular, items that were not prior targets show a benefit of having seen the item more times, whereas items that were previously targets seem to be encoded sufficiently enough that it takes roughly the same number of fixations to find the target regardless of the number of previous fixations.

Recency and frequency effects. In order to provide a more robust description of the human data, we examined the proportion of all searches in which a target was last seen R (Recency) fixations ago or was seen F (Frequency) times prior to the search and was found within N fixations. Figure 3 illustrates the respective distributions generated by analyzing the human data in this way. In particular, in the recency graph, the peak of each distribution shifts to the right (more fixations to find target) as R increases. It should be noted that the human data exhibits a bell-shaped curve across all recency values, with the proportion of searches in which target is found in a higher number of fixations falling off gradually.

In the frequency graph, when the item has never before been fixated (F==0), the proportion of all searches in which the target is found stays at roughly 10% across all N. Items which have been seen more times (F==9) have a slightly

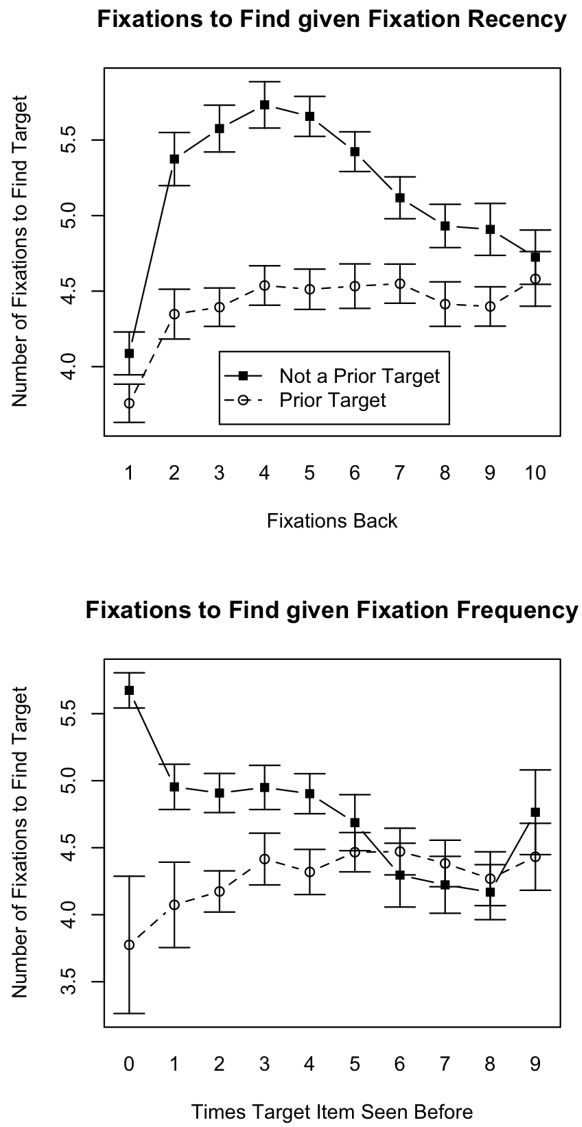

Figure 2. Mean number of fixations needed to find a target as a function of recency and frequency of seeing the target before.

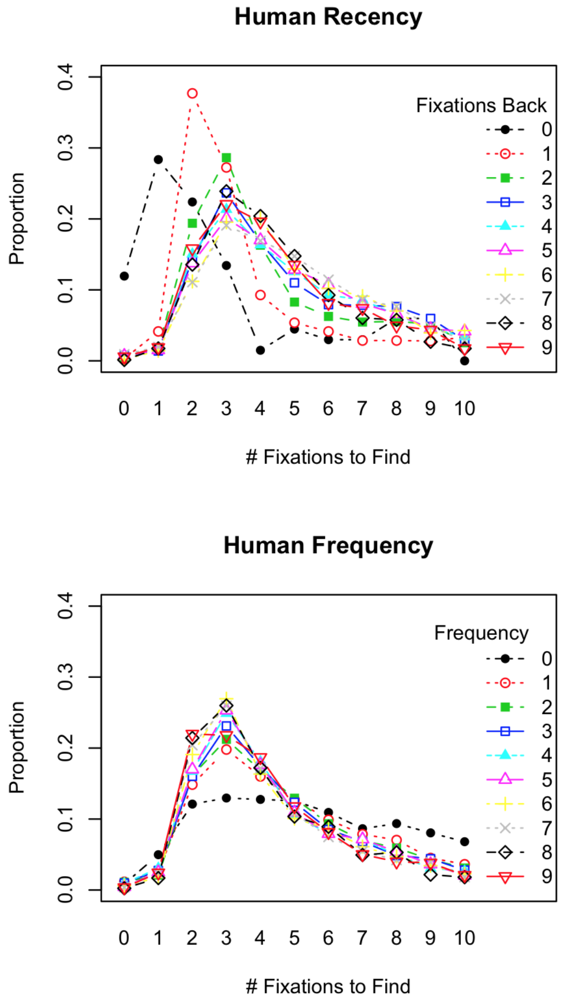

Figure 3. Proportions of recency and frequency effects in the human data.

more pronounced peak at N=3 as compared to items which were fixated fewer times.

Subsequent model runs were compared on the basis of these distributions. We wanted to be able to capture both the magnitude of the proportions in both recency and frequency, as well as the general shape of the distributions as proportions gradually tapered off for the higher N. Note that this collapses across whether or not the item was a prior target.

Experiment Discussion

The results from the study indicated that the number of fixations to find a target is affected by (1) whether that label had been a prior search target, (2) the recency of a label's previous fixation, and (3) the frequency of a label's previous fixations. Each of these effects contributes to the variability in VWM capacity and precision. A label more recently encoded will lead to the appearance of a larger VWM capacity and higher VWM precision. Similarly, a label more frequently encoded will lead to the appearance of a a larger capacity with greater precision. In passive change detection, the probe is chosen at random and may sometimes select a target that has neither been recently or frequently encoded. This could naturally lead to the perception of capacity and precision variability of VWM. By looking at the selective attention process during the search, we can more concretely point to the mechanisms leading to this variability.

Model-based Analysis

Given the debate in the literature between slot based and continuous resource models of VWM, we chose to run a factorial combination of models and search strategies. The three classes of models were: No memory, Slot-Based Memory, and Continuous Resource Memory. The search strategies were either Nearest First or Random. For each memory-strategy combination, scan-paths were generated for each of the 20 searches within a trial. In all models, the assumption is that once an object was visited, it was removed from the set of possible next visits until the next target was announced.

No Memory Model

This model served as a theoretical baseline for the other models and searched the display for every search within a trial without any memory for previous targets or distractors. In the *random search* version, the model searched the display in a random fashion. In the *nearest first* version, the model allocated attention to the closest object to the one currently being fixated. No parameters were varied in this model.

Slot-Based Memory Models

This class of models had a slot-based memory and the number of slots available ranged from 0 to 10. Slots were instantiated as a queue (FIFO) based on the human fixation history prior to the current search (see Figure 4). Uncovering a label reinstantiates slot 0 and pushes the labels contained in slot i into slot $i + 1$. This corresponds closely to the R denotation in the Recency human data analysis. At the beginning of every search, the model queried its slot-based memory to determine whether the target was already present in one of the slots. If it was, the model immediately directed its attention to the location of the target. If it was not in one of the slots, the model searched the display in either a Random or Nearest First manner. Only the number of slots parameter was varied in this model type.

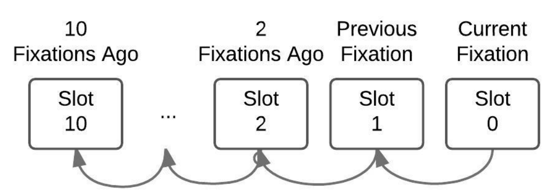

Figure 4. Slots are instantiated corresponding to the timeline of fixations in the human data.

Continuous Resource Memory Models

This class of models relied on human eye fixation history of the trial prior to the current search, taking into consideration both the time stamps of when the item was fixated and how many previous fixations were made to the item. The ACT-R memory equation (J. R. Anderson, 2007) was applied to each of the items. It specifies the activation of a given item *i* in memory as

$$A_i = ln \sum_{j=1}^{n} t_{ij}^{-d} + \beta_i + \epsilon_i \quad (1)$$

where j is a fixation on the item and t_{ij} is a time stamp of how long ago the item was seen on fixation j, $-d$ is a decay value, β is a base level constant offset, and ϵ is logistically distributed transient noise with a mean of 0 and standard deviation of σ.

Activation of the target item was recalculated at the beginning of each search and the model checked whether the activation of the target was above threshold, T, and if so, moved attention directly to the known location of the item. If $A_{target} < T$, then the model selected and encoded another item based on either a *random search* or a *nearest first* strategy. If the target item was still not found, activation was recalculated for the target at each additional movement of attention.

Four parameters were varied in the context of ACT-R's memory equation (d, β, T, and σ) to find the best fit to the human recency and frequency data using MindModeling.org (Harris, 2008). In particular, we varied the parameters as follows: d: [0,1], β: [0,10], T: [0,20], and σ: [0,5]. This created a total of $27,951$ combinations of parameters for each search strategy.

Model Evaluation

Each of the above models was run through all trials (and searches) obtained from human data. The ACT-R memory equation uses recency and frequency information as sources of activation for a given chunk in memory. Thus, we examined the human data as a function of both the recency and the frequency of previous fixations to current targets. In this case, recency refers to how many fixations ago the item was last fixated, R with respect to the current fixation. For each parameter set, summary statistics were calculated to determine the percentage of all trials on which the target was seen R fixations ago or was previously seen F times and found in N fixations. This resulted in 10 distributions for recency and another 10 for frequency, each with 11 data points (one for each N of fixations to find the target, see Figure 3 for human data). Then Root Mean Squared Error (RMSE) and R^2 scores were calculated for each target recency curve and for each target frequency curve.

A composite goodness-of-fit measure was created to combine the R^2 and RMSE measures to capture both the shape and the magnitude of the differences between human data and model predictions. Because best fits according to R^2 are values closer to 1, and best fits according to RMSE are values closer to 0, we re-scaled the R^2 measure (1-R^2) and computed an average of all curves for each parameter setting.

Table 1
Best fits for all model types.

Memory	Strategy	Composite Score*
Continuous resource	Nearest first	0.07
Continuous resource	Random	0.09
Slot (2)	Random	0.35
Slot (2)	Nearest first	0.36
None	Nearest first	0.37
None	Random	0.37

Note: *Lower composite scores indicate better model fits.

The best fitting slot-based model was one which contained 2 slots ('remembered' the last two items previously fixated; see Table 1). The best fitting continuous resource model resulted from the following parameter settings: d=1, β=1, T=10, and σ=4.0 (see Figure 5).

The no memory model established a baseline with which the other memory models could be compared. As can be seen in Table 1 and Figure 5, the continuous resource memory model did a much better job of capturing human performance. In particular, whereas a slot-based memory with 2 slots was the best fitting in this particular class of models, it failed to capture the shape of both the recency and frequency distributions. The continuous resource model, on the other hand, exhibited the bell-shape curve with gradual drop-off seen in the human data for both recency and frequency. Furthermore, a *nearest-first* search strategy was marginally better at capturing the effects than a *random search* model, suggestive of the type of strategy participants may have used as they conducted their search of the display.

We further evaluated the flexibility of all the model types to determine how convincing the fits actually are (and whether they could have been achieved merely by searching such a large space). Model Flexibility Analysis (MFA) was used to calculate the proportion of all empirical outcomes that each model could have potentially fit (Veksler, Myers, & Gluck, 2015). Although the *slot* model only has one parameter (number of slots), it's actually more flexible than the *continuous resource* model which has 4 parameters. MFA revealed flexibility for the *slot* model to be $\phi = .14$ and for the *continuous resource* model to be $\phi = .014$. Thus the *continuous resource* model makes more precise predictions and is less flexible.

Discussion & Conclusions

In the current work, we explored why variability in VWM capacity may at times exhibit variable precision and capacity. The new paradigm of *repeated serial search* allowed us to more readily observe the specific shifts of visual attention that occur during natural search. Human data suggests that the variability in VWM precision and capacity is closely tied to selective attention as search progresses.

Selective attention directly affects the ease with which subsequent targets can be found, with both recency and fre-

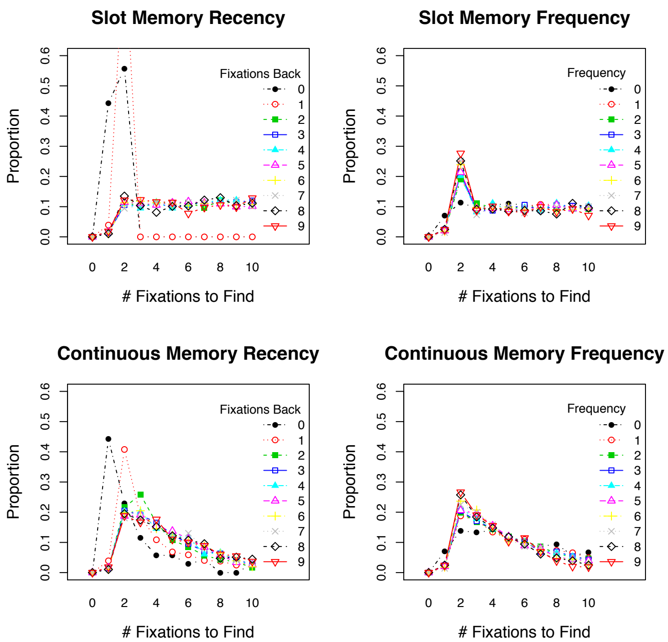

Figure 5. Model fits for best fitting slot and continuous resource models, nearest first search strategy.

quency playing a role. Items that were previously fixated more recently resulted in faster search times and boosted the likelihood of recalling the location of the target. Likewise, items which had previously been fixated more often were easier to find. Importantly, there was a stronger mnemonic trace for items which were previous targets as these items were found faster than those which were only fixated as distractors during previous searches.

We compared two models of VWM: a *slot*-based and a *continuous resource*-based model. In the case of the slot-based model, the recency of an item's encoding is taken into consideration to facilitate subsequent searches. However, this was not sufficient to account for the human data as it failed to capture the shapes of the distributions in both recency and frequency domains. A continuous resource model, on the other hand, directly incorporated both effects of selective attention. The continuous resource was instantiated as the item's activation, computed by taking into account both the frequency and recency of previous item fixations.

One limitation of the current approach is that none of the models explicitly account for the stronger mnemonic trace for prior targets. The continuous resource model could potentially account for this difference by including an item's fixation duration in its computation of activation - target items typically have longer fixations and more opportunity for rehearsal. While such models are beyond the scope of the current work, they are an interesting avenue for future research. Another possible concern is that humans may use an 'adaptive avoidance' strategy in which items known to *not* be the target are actively not gazed at. Future work will need to address the degree to which this type of strategy may drive behavior in visual search.

In conclusion, the repeated serial search paradigm elucidates the variability seen in VWM capacity and precision by taking into account selective attention considerations. Future work could apply the same continuous resource model to other data sets to explore the robustness of the model in accounting for various VWM results, as well as incorporating potentially hybrid models which combine slots and continuous resources.

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), 106–11.

Anderson, D. E., Vogel, E. K., & Awh, E. (2013). Selection and storage of perceptual groups is constrained by a discrete resource in working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 824–835.

Anderson, J. R. (2007). *How can the human mind exist in the physical universe?* (F. E. Ritter, Ed.). Oxford University Press.

Barton, B., Ester, E. F., & Awh, E. (2009). Discrete resource allocation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1359–67.

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

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. Oxford Univ. Press.

Harris, J. (2008). Mindmodeling@home: a large-scale computational cognitive modeling infrastructure. In *Proceedings of the sixth annual conference on systems engineering research 2008* (pp. 246–252).

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400.

Luria, R., & Vogel, E. K. (2011). Shape and color conjunction stimuli are represented as bound objects in visual working memory. *Neuropsychologia*, *49*(6), 1632–9.

Neth, H., Gray, W. D., & Myers, C. W. (2006). Memory models of visual search - searching in-the-head vs. in-the-world. *Journal of Vision*, *5*(8), 8–9.

Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin & Review*, *18*(2), 324–30.

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

van den Berg, R., & Ma, W. J. (2014). "plateau"-related summary statistics are uninformative for comparing working memory models. *Attention, Perception, & Psychophysics*.

Veksler, V. D., Myers, C. W., & Gluck, K. A. (2015). Model flexibility analysis. *Psychological Review*, *122*(4), 755–769.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 1120–35.