Searching for serial refreshing in working memory

PSYCHONOMIC BULLETIN & REVIEW

Searching for serial refreshing in working memory: Using response times to track the content of the focus of attention over time

Journal:	*Psychonomic Bulletin & Review*
Manuscript ID	PBR-BR-15-267.R2
Manuscript Type:	Brief Report
Date Submitted by the Author:	n/a
Complete List of Authors:	Vergauwe, Evie; University of Missouri, Psychological Sciences Hardman, Kyle; University of Missouri, Psychological Sciences Rouder, Jeffrey; University of Missouri, Psychological Sciences Roemer, Emily; University of Rochester McAllaster, Sara; Chapman University Cowan, Nelson; University of Missouri, psychological sciences
Keywords:	working memory, focus of attention, refreshing, serial refreshing

Searching for serial refreshing in working memory:

Using response times to track the content of the focus of attention over time

Evie Vergauwe[1], Kyle O. Hardman[1], Jeffrey N. Rouder[1], Emily Roemer[2], Sara McAllaster[3], & Nelson Cowan[1]

[1]Department of Psychological Sciences, University of Missouri, Columbia, Missouri, USA

[2]University of Rochester, Rochester, New York, USA

[3]Chapman University, Orange, California, USA

Author Note

Funding for this project was provided by NIH Grant R01 HD-21338 to Nelson Cowan, NSF Grant BCS-1240359 to Jeffrey N. Rouder, and Swiss National Science Foundation Grant PA00P1_139604 to Evie Vergauwe.

Word count: 4’172 words in the main text including references

Abstract

One popular idea is that, to support maintenance of a set of elements over brief periods of time, the focus of attention rotates among the different elements thereby serially refreshing the content of Working Memory (WM). In the research reported here, probe letters were presented between to-be-remembered letters. Response times to these probes were used to infer the status of the different items in WM. If the focus of attention cycles from one item to the next, its content should be different at different points in time and this should be reflected in a change in the response time patterns over time. Across a set of four experiments, we demonstrate a striking pattern of invariance in the response time patterns over time, suggesting that either the content of the focus of attention did not change over time or that response times cannot be used to infer the content of the focus of attention. We discuss how this pattern constrains models of WM, attention, and human information processing.

Searching for serial refreshing in working memory:

Using response times to track the content of the focus of attention over time

People must often maintain a set of elements active in mind over brief periods of time. This information is purportedly stored in Working Memory (WM). One proposed mechanism to keep information active in WM is *refreshing*. In contrast to covert or overt verbal rehearsal, refreshing is assumed to be a domain-general mechanism that operates by bringing WM representations into the focus of attention (Barrouillet & Camos, 2012; Cowan, 1995; Higgins & Johnson, 2009). The act of refreshing presumably results in memory representations being reactivated which, in turn, protects the information from being forgotten.

Though considerable research has been devoted to the process of refreshing over recent years (e.g., Camos & Portrat, 2015; Loaiza, Duperreault, Rhodes, & McCabe, 2015; Souza, Rerko, & Oberauer, 2015; Vergauwe & Cowan, 2015), little is currently known about how refreshing operates to support the maintenance of a set of elements in WM. The present study tested the strong hypothesis that refreshing operates serially, with the focus of attention cycling from one item to the next (e.g., Barrouillet & Camos, 2012; Cowan, 2011; McCabe, 2008; Nee & Jonides, 2013; Vergauwe, Camos, & Barrouillet, 2014).

Evidence for a focus of attention in WM that is limited to one element at a time comes from studies showing that the element last processed has a privileged status of accessibility in WM compared to other to-be-remembered elements. For example, in an item-recognition task in which a list of items is followed by a probe to be judged present in or absent from the list, response times (RTs) to the last item are faster than to any other item of the list (e.g., Burrows & Okada, 1971; McElree & Dosher, 1989; Nee & Jonides, 2008; Oztekin, Dvachi, & McElree, 2010); the last item of the list is accessed at a faster rate (see McElree, 2006, for a review); and distinct brain regions are involved in judging the last item of the list (e.g., Nee & Jonides, 2008; Oztekin et al., 2010).

The RT benefit for the last item in WM may be leveraged to assess whether there is serial refreshing. The logic is as follows: When refreshing happens, the last-presented item is replaced in the focus of attention by another list item. This other item will consequently now have the focus-of-attention benefit if probed. The last-presented item, that was replaced, will presumably no longer have the focus-of-attention benefit. Refreshing thus should attenuate the last-presented-item benefit, and this attenuation can be assessed as an indirect index of serial refreshing.

Overview of the Study

To test theoretical assumptions about refreshing, we created the *probe-span task*. In four experiments, short series of red letters were presented for subsequent recall and black probe letters were presented between these memory items, with each probe to be judged present in or absent from the list presented so far, as quickly as possible (Figure 1). We manipulated the delay between each studied item and the subsequent probe. If the delay before the probe is very short, then we expect that refreshing has not yet occurred and the last item remains in the focus of attention. In this case, responses to the last-presented item should be speeded. If the delay is long, however, then according to common assumptions refreshing should have occurred; the most-recently-refreshed item is assumed to be in the focus of attention but its serial position should vary from trial to trial. In this case, there should be no advantage and responses to the last presented item should not be speeded. Of course, for the approach to work, the short and long delays need to be chosen carefully to test the common assumptions. These issues are addressed subsequently. In summary, based on the assumptions stemming from studies reviewed above, the serial refreshing hypothesis predicts a specific interaction pattern between serial position and the duration of the delay before the probe, in which the last presented item should be speeded only for short delays between the most recent memory item and following probe.

Against our expectations, we observed that participants were the fastest to respond to the last-presented memory item at all probe delays. In fact, the duration of the delay before the probe did not affect the serial position function. This invariance of the serial position curves across time was replicated in three additional experiments that aimed at creating optimal conditions to detect serial refreshing by using *(1)* in Experiments 2-4, probe delays similar to the durations allowed for refreshing in studies providing evidence of it (Barrouillet & Camos, 2012); *(2)* in Experiments 3-4, a restricted set of phonologically similar letters as memoranda inasmuch as people strategically favor refreshing over speech-based rehearsal for such materials (Camos, Mora, & Oberauer, 2011); and *(3)* in Experiment 4, memoranda shortened from 1000 to 500 ms to ensure that refreshing could not occur during the latter part of the presentation time (cf. Oberauer & Lewandowsky, 2011).

Method

Participants

Subjects were undergraduate students at the University of Missouri-Columbia and were paid $15 for their participation or received course credit. All were native speakers of English and had normal or corrected-to-normal vision. In Experiments 1 through 4, respectively, there were 40 (24 female), 60 (36 female), 40 (21 female), and 40 (15 female) participants.

Materials and Procedure

The probe-span task was administered using E-prime software (Psychology Software Tools). Participants were asked to watch carefully and memorize series of four red letters presented sequentially on screen. In Experiments 1-2, all consonants, excluding Y, were used as stimuli. In Experiments 3-4, a pool of eight phonologically similar consonants was used as stimuli: B, C, D, G, P, T, V, and Z. In all experiments, the different consonants were used approximately equally often and no consonant was repeated within a series. These red letters were presented at the center of the screen in 48 point Courier New font (~2.29˚ of visual angle). Stimuli were presented on a standard CRT monitor and participants sat at a comfortable distance from the screen (~50 cm).

Each series began by a fixation cross, centrally displayed on screen during 750 ms. This fixation signal was replaced by the first red letter. Red letters were presented for either 1000 (Experiments 1-3) or 500 ms (Experiment 4). At the end of each series, an empty rectangle on screen prompted participants to recall the four red letters of that series in order of appearance by typing them on the keyboard. Participants were encouraged to fill in unknown letters with a guess. All entered letters appeared in the box, from left to right. Participants pressed enter to end the recall response and initiated the next series by pressing a button on the button box after recall.

After each red memory item, one black letter (probe) was presented in the center of the screen in 24 point Courier New font (~1.15˚ of visual angle). Participants were instructed to decide whether this black letter corresponded to one of the red letters they were to maintain on the current trial or not. This judgment was made by pressing the rightmost button of the button box when the black letter corresponded to one of the red letters in memory and pressing the leftmost button of when the black letter did not correspond to one of the red letters in memory.

The Delay Before Probe variable was manipulated within-subjects. Regardless of the Delay condition, the interval between two red letters was kept constant at 2000 ms. However, depending on the experimental condition defined by Delay Before Probe, the delay between the offset of the red letter and the onset of the black letter was different (100, 200 or 400 ms in Experiment 1; 200, 400, 600, or 800 ms in Experiments 2-3; and 100, 400 or 800 ms in Experiment 4). Black letters were always presented for 1000 ms. The remaining delay between the offset of the black letter and the onset of the next red letter differed as a function of Delay Before Probe (900, 800 or 600 ms, respectively, in Experiment 1; 800, 600, 400, or 200 ms, respectively, in Experiments 2-3; and 900, 600 or 200 ms, respectively, in Experiment 4).

Experiments 1 and 4 included 144 trials; Experiments 2-3 included 96 trials. For each trial and each participant, black letters were sampled randomly from a pool of potential probes in such a way that the likelihood of receiving a positive probe was 50% at each probe position. (Positive probes consisted of any of the letters presented in the series so far.) Thus, each trial could have 0 to 4 positive probes. For each probe position, the pool of possible probes consisted of all the letters presented in the series so far plus a random new letter for that series. Thus, across the entire experiment, and for each of the four probes, the black letter corresponded in half of the trials to one of the red memory items, and each red letter presented up to that point in the trial had equal chances of being used as target-present probe. Importantly, in each of these four pools, every different probe type was associated equally often with each of the possible levels of Delay Before Probe.

Before the experimental trials, participants received instructions that included a visualization of a trial. This was followed by five practice trials. Throughout the experiment, participants were asked to respond as fast as possible to the probes, without making errors, while maintaining the four red letters. They were not informed of the varying delays. Responses in the processing task were collected by button presses on a Serial Response box. Recall performance was scored by counting the number of letters that was correctly recalled with respect to serial order within each series (max = 4). Next, an average across all series was calculated per participant.

Performance-based exclusions

In each experiment, we applied exclusions as follows. First, we discarded the data of participants whose average recall score was less than 1 letter out of 4 (1, 2, 0, and 2 participants in Experiments 1 through 4, respectively). Next, to ascertain that participants paid sufficient attention to the probe task, we excluded the data of participants who performed below 55% correct (0, 1, 0, and 1 participants in Experiments 1 through 4, respectively). Finally, we verified participants' precise compliance with the instructions in the probe task. Because it is important that participants consider all of the red letters when judging the probe, we calculated the rate of correct responses to "not-last" probes (i.e., target-present probes that show any-but-the-last-presented red letter of a series) and excluded the data of participants who scored below 55% on these not-last probes (3, 3, 2, and 3 participants in Experiments 1 through 4, respectively). These exclusions resulted in a final sample of 36 (out of 40), 54 (out of 60), 38 (out of 40), and 34 (out of 40) participants in Experiments 1 through 4, respectively.

Method of Analysis

We examined serial position curves for the RTs collected at probe position 2, 3, and 4 (following memory items 2, 3, and 4, respectively). Specifically, RTs to the target-present probes were analyzed as a function of the serial position of the matching memory item. For each experiment, a separate Bayesian analysis of variance (BANOVA; Rouder, Morey, Speckman, & Province,

2012) was run for each of the three probe positions (Probe 2, Probe 3, and Probe 4) with Delay Before Probe (100, 200, or 400 ms in Experiment 1; 200, 400, 600, or 800 ms in Experiments 2-3; and 100, 400, or 800 ms in Experiment 4) and Serial Position of the matching memory item as independent within-subjects variables. The BayesFactor package for the R statistical-analysis language with the default settings was used. For 5 out of the 12 ANOVAs (3 probe positions x 4 experiments), there was at least one participant with missing data in one or more cells because only correct RTs were analyzed. Per ANOVA, participants with missing data were omitted and we ran our analysis on the remaining participants. Table 1 presents the results of the analyses and reports for each analysis the number of participants included. Except for the ANOVA of the RTs collected at Probe 4 in Experiment 3, there was never more than one participant with missing data.

Using two variables, Delay Before Probe and Serial Position, models were specified for each combination of main effects and interactions and the BANOVA computes Bayes factors for each of these models. We used these Bayes factors to identify the best model (i.e., the model that yields the highest Bayes factor). As we will see, for the vast majority of our observations, the winning model included the two main effects of Serial Position and Delay Before Probe but not the interaction. We then assessed the strength of the evidence in the data against the interaction by computing a Bayes factor between the model that does not include the interaction (i.e., the Main effects-only model) and the alternative model in which the interaction term is included (i.e., the full model). The resulting Bayes factor quantifies the evidence in the data against an interaction between Delay Before Probe and Serial Position.

Results

General performance

As a validation of our task, as expected, participants correctly recalled several memory items at the end of the series and had high accuracy on the probes. For each experiment, recall performance and rate of correct responses to the probes of the participants in the final sets can be found in Table 2. The use of phonologically similar letters and the use of shorter presentation times resulted in slightly lower performances than in the other conditions but performance remained high. There were no recognition/recall trade-offs (Supplementary material 1).

Serial Position Curves

Serial position curves are shown for each probe delay in Figure 2. Similar to what is typically observed in the item-recognition task, we found RTs to the probe to be affected by the serial position of the matching memory item (e.g., Burrows & Okada, 1971); across the four experiments, and across the different probe positions, the curves show a clear benefit for the last presented memory item. Responses became somewhat faster after a longer delay but it is immediately clear that one can rule out the hypothesis that the shape of the serial position curves changes drastically over time. Participants were the fastest to respond to the last-presented memory item and this pattern was invariant across durations of the delay before the probe. This invariance of the serial position curves was confirmed by BANOVAs.

As can be seen in Table 1, across the three probe positions and the four experiments, the best model included the two main effects of Serial Position and Delay Before Probe but not the interaction. This model was favored over the model including the interaction with a Bayes factor ranging between 9.66 and 42.42 in Experiment 1, between 5.74 and 452.10 in Experiment 2, between 13.22 and 135.61 in Experiment 3, and between 4.37 and 15.84 in Experiment 4. Only for Probe 2 in Experiment 4 was the full model with an interaction best, and even then only weakly, preferred over the model including only the two main effects by a factor of 2.70. Moreover, as can be seen in the lower panel of Figure 2, responses were still fastest to the last-presented memory item (i.e., memory item 2) at the longest delay. Thus, a strong pattern of invariance emerges from the ensemble of our results; RTs to the last item remained the fastest over time.

Discussion

We tested the hypothesis that attentional refreshing operates serially to maintain a set of elements. Based on past research we assume that serial refreshing could not occur much by our shortest probe delay, but would come on line at longer delays. The result of this change would be a shift of the focus of attention away from dwelling on the last-presented list item, instead shifting between items one at a time to refresh them. Based on the assumption that participants are faster to respond to the item that is currently in the focus of attention, compared to any other item of the list, the serial refreshing hypothesis predicts that the item that receives the fastest RT should change over time. Our findings contrasted sharply with this prediction; RTs to the last item were the fastest, and this pattern remained invariant over time.

Note that the experiments created optimal conditions to detect the operation of refreshing, based on assumptions put forward in the literature. After observing the unexpected invariant pattern in Experiment 1, we used delays that were more similar to the time available for refreshing in studies providing evidence for refreshing (Experiments 2-4). To examine whether the invariance in Experiments 1-2 was due to people using articulatory rehearsal rather than refreshing, we aimed at minimizing the role of articulatory rehearsal using phonologically similar material (Experiments 3-4). Finally, after still not observing the expected change in the serial position curves over time in Experiment 3, we reasoned that perhaps part of presentation time might have been used for refreshing. To exclude this possibility, we used shorter presentation times (Experiment 4) but the serial position function did not change.

In what follows, we will discuss what we believe are three possible accounts for the observed invariance. To account for it, one must modify either the serial refreshing hypothesis or the hypothesis that speeded responses reflect the presence of an item in the focus of attention. Each of these accounts has potentially far-reaching implications.

By the first account, invariant serial position curves were observed because refreshing does not operate serially. Instead, the content of WM might be refreshed in parallel, with attention divided among the different items of the set at any point in time. To the best of our knowledge, there is currently no WM model that proposes parallel refreshing (but see Ratcliff, 1978, proposing parallel retrieval of items in WM, making parallel refreshing possible). It is worth noting that a few early studies have manipulated the delay between the memory list and the single probe in the item-recognition task and have found serial position curves to become somewhat flatter over time (e.g., Clifton & Birenbaum, 1970), suggesting the possibility of serial refreshing during a retention interval following list presentation, even if not during list presentation.

By the second account, invariant serial position curves were observed because participants did not use refreshing in our experiments, even though we aimed to create optimal conditions to detect refreshing. At least, they might use no refreshing during the inter-probe intervals. Vergauwe and Cowan (2015) called on refreshing during probe performance to explain why a letter-probe task imposed less of a cognitive load than letter-processing tasks that did not require searching through the memoranda. However, it might be possible that in this procedure, refreshing during the probe task itself *takes the place of* refreshing between probes, in which case there would be no probe delay effect on refreshing. Here we did not require participants to carry out an unrelated processing task during inter-letter intervals as in most studies of refreshing (e.g., Barrouillet & Camos, 2012), and refreshing might take place in the unfilled periods between such processing episodes. Alternatively, people might not refresh the content of WM unless explicitly instructed to do so (e.g. Souza et al., 2015). In that case, an alternative explanation is needed to account for the much-replicated cognitive-load effects in WM by which recall performance depends on the attentional demands of the secondary task (e.g., Barrouillet & Camos, 2012). For example, an interference account does not require refreshing and would not expect an attenuation of the last-item benefit over time.

The first two accounts are conditional on accepting the assumption that the last-item benefit in RT reflects the item being in the focus of attention. If one accepts that fast responses to a particular item can be used to infer the existence of a one-item focus of attention, then one must accept that we have presented strong evidence against the prevailing view that attention refreshes one element at a time to support the maintenance of a set of elements in WM; the last-item benefit should have disappeared, even if one assumes very fast refreshing.

A third account of our data, however, could involve rejecting the assumption that fast responses to a particular item can be used to infer the content of the focus of attention. An account in terms of familiarity-based recognition rather than through focal attention seems unlikely, though, and is not consistent with our data (Supplementary material 2). Most evidence for a one-item focus of attention comes from studies showing a benefit in RT for the last presented or processed item (see Oberauer & Hein, 2012, for a recent review). If we reject the assumption that the last-item benefit results from the last item still being in the focus of attention, then most evidence for a one-item focus of attention in WM must be discarded. It is important to note here that similar last-item advantages have been observed when perceptual matching is prevented (e.g., McElree & Dosher, 1989; Nee & Jonides, 2008), excluding an account of the last-item benefit in terms of visual matching.

One might propose, though, that serial position effects in RT reflect different activation levels rather than items being in or out the focus of attention (e.g., because of retroactive interference of each item with prior items). Then, serial refreshing (without severe decay) might still occur, without changing the serial position curves over time in the current experiments. While the shape of the serial position curves beyond lag -1 in Figure 2 might be consistent with the idea of different activation levels, statistical analysis does not allow firm conclusions (Supplementary material 3).

To conclude, we have observed an unexpectedly stable serial position function that stands in contrast with what is expected based on the juxtaposition of two key assumptions about WM. The uncovered invariance puts important constraints on models of WM, attention, and human information processing; follow-up investigations should help disentangle alternative accounts.

References

Barrouillet, P., & Camos, V. (2012). As time goes by: Temporal constraints in working memory. *Current Directions in Psychological Science, 21*, 413-419.

Burrows, D., & Okada, R. (1971). Serial position effects in highspeed memory search. *Perception & Psychophysics, 10*, 305-308.

Camos, V., & Portrat, P. (2015). The impact of cognitive load on delayed recall. *Psychonomic Bulletin and Review, 22,* 1029-1034

Camos, V., Mora, G., & Oberauer, K. (2011). Adaptive choice between articulatory rehearsal and attentional refreshing in verbal working memory. *Memory & Cognition*, *39*, 231-244.

Clifton, C., & Birenbaum, S. (1970). Effects of serial position and delay of probe in a memory scan task. *Journal of Experimental Psychology, 86*, 69-76.

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.

Cowan, N. (2011). The focus of attention as observed in visual working memory tasks: Making sense of competing claims. *Neuropsychologia, 49*, 1401-1406.

Higgins, J. A., & Johnson, M. K. (2009). The consequence of refreshing for access to nonselected items in young and older adults. *Memory and Cognition, 37*, 164-174.

Loaiza, V. M., Duperreault, K. A., Rhodes, M. G., McCabe, D. P. (2015). Long-term semantic representations moderate the effect of attentional refreshing on episodic memory. *Psychonomic Bulletin and Review, 22,* 274-280.

McCabe, D. P. (2008). The role of covert retrieval in working memory span tasks: Evidence from delayed recall tests. *Journal of Memory and Language, 58*, 480-494.

McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The psychology of learning and motivation, Vol. 46*. San Diego: Academic Press.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General, 118*, 346-373.

Nee, D. E., & Jonides, J. (2008) Neural correlates of access to short-term memory. *PNAS, 15*, 14228-14233.

Nee, D. E., & Jonides, J. (2013). Trisecting representational states in short-term memory. *Frontiers in Human Neuroscience, 7*, 796.

Oberauer, K., & Hein, L., (2012). Attention to information in working memory. *Current Directions in Psychological Science, 21*, 164-169.

Oberauer, K., & Lewandowsky, S. (2011). Modeling working memory: a computational implementation of the Time-Based Resource-Sharing theory. *Psychonomic Bulletin & Review, 18*, 10-45.

Oztekin, I., Dvachi, L., & McElree, B. (2010). Are representations in working memory distinct from representations in long-term memory? *Psychological Science, 21*, 1123-1133.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59-108.

Rouder, J. N., Morey, R. D., Speckman, P. L. and Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56, 356-374.

Souza, A. S., Rerko, L., & Oberauer, K. (2015). Refreshing memory traces: Thinking of an item improves retrieval from visual working memory. *Annals of The New York Academy of Sciences*.

Vergauwe, E., Camos, V., & Barrouillet, P. (2014). The effect of storage on processing: How is information maintained in working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1072-1095.

Vergauwe, E., & Cowan, N. (2015). Attending to items in working memory: Evidence that refreshing and memory search are closely related. *Psychonomic Bulletin and Review, 22,* 1001-1006.

Table 1

Evidence in the data against an interaction between Serial Position and Delay Before Probe in Experiments 1 through 4. Bayes factors are between the main effects-only model that does not include the interaction and the full model in which the interaction is included and describe the strength of the evidence for the absence of an interaction.

	Probe 2	Probe 3	Probe 4
Experiment 1	*36 participants*	*36 participants*	*36 participants*
	9.66 to 1	12.26 to 1	42.42 to 1
Experiment 2	*54 participants*	*53 participants*	*53 participants*
	5.74 to 1	9.26 to 1	452.10 to 1
Experiment 3	*38 participants*	*37 participants*	*29 participants*
	13.22 to 1	135.61 to 1	24.17 to 1
Experiment 4	*34 participants*	*34 participants*	*33 participants*
	1 to 2.70	4.37 to 1	15.84 to 1

Note: SP = Serial Position; Delay =Delay Before Probe

Table 2
Mean recall performance and mean probe task performance in Experiments 1-4. Standard deviations are reported in parentheses.

	Mean Items Recalled	Mean Probe Accuracy
Experiment 1	3.73 (.20)	.94 (.04)
Experiment 2	3.73 (.22)	.95 (.05)
Experiment 3	3.28 (.39)	.89 (.07)
Experiment 4	3.09 (.44)	.87 (.06)

Figure Captions

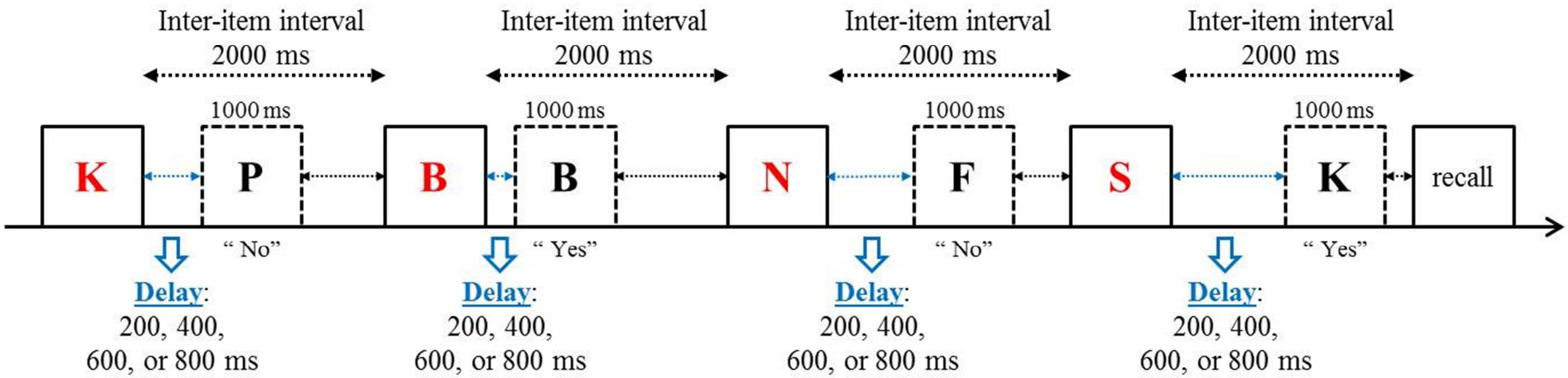
Figure 1

Illustration of a trial within the probe-span task. Series of four red letters were presented for subsequent recall and black probe letters were presented between the memory items, with each probe to be judged present in or absent from the list presented so far. At the end of the series, participants recall the four letters in order of appearance. The delay before the probe was manipulated. Here, delay durations are shown as used in Experiments 2 and 3 (between 200 and 800 ms).

Figure 2

Mean probe response RT in ms as a function of the serial position of the matching memory item (expressed as the lag between presentation and test; on the x axis) and probe position (Probe 2, Probe 3, or Probe 4 in the left, middle and right panels, respectively). The delay following the probe appears as the graph parameter. **Top row** (Experiment 1): regular letters were used as memoranda and were presented for 1000 ms; delay durations varied between 100 and 400 ms. **Second Row** (Experiment 2): regular letters were used as memoranda and were presented for 1000 ms; delay durations varied between 200 and 800 ms. **Third row** (Experiment 3): phonologically similar letters were used as memoranda and were presented for 1000 ms; delay durations varied between 200 and 800 ms. **Fourth row** (Experiment 4): phonologically similar letters were used as memoranda and were presented for 500 ms; delay durations varied between 100 and 800 ms. Error bars show standard errors of the mean.

A

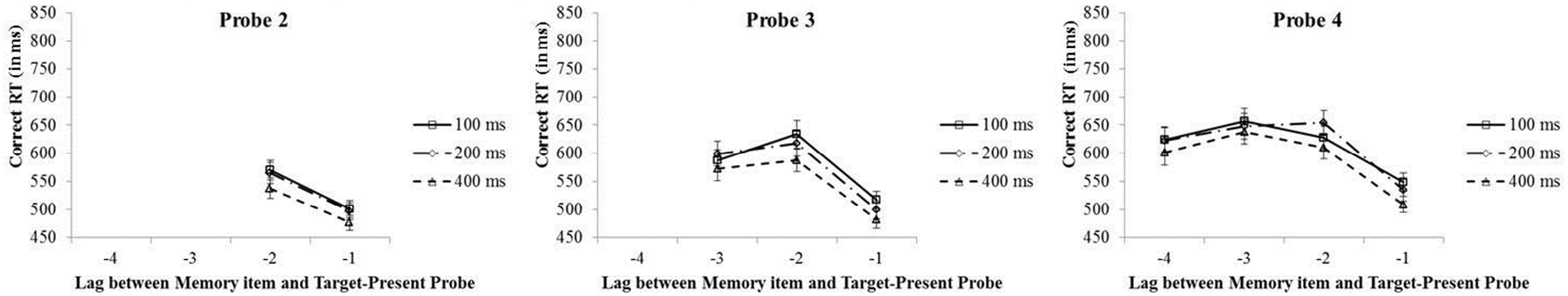


B

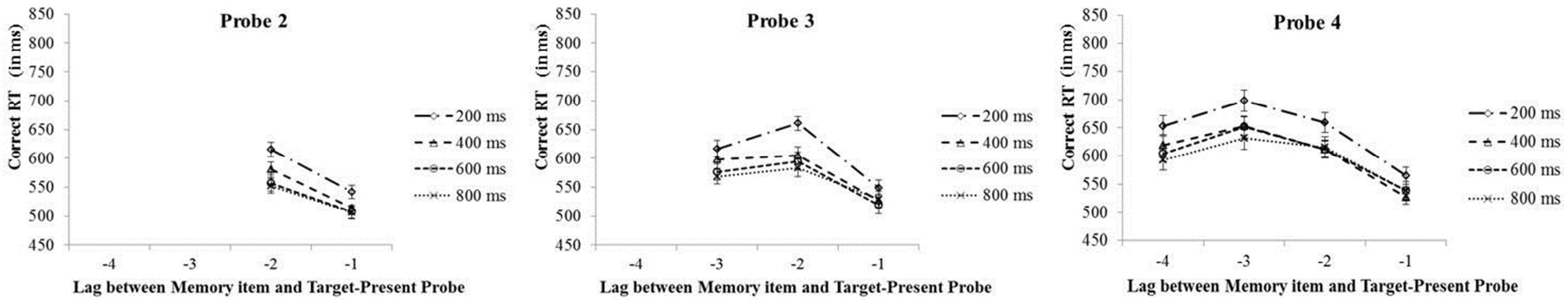
	Memory material	Delay before probe	Presentation duration memory items
Experiment 1	*Regular letters*	*100, 200, 400 ms*	*1000 ms/item*
Experiment 2	*Regular letters*	*200, 400, 600, 800 ms*	*1000 ms/item*
Experiment 3	*Phonologically similar letters*	*200, 400, 600, 800 ms*	*1000 ms/item*
Experiment 4	*Phonologically similar letters*	*100, 400, 800 ms*	*500 ms/item*

Figure 1. A) Illustration of a trial within the probe-span task. Series of four red letters were presented for subsequent recall and black probe letters were presented between the letters to be remembered, with each probe to be judged present in or absent from the list presented so far. At the end of the series, participants recall the four letters in order of appearance. The delay before the probe was manipulated. Here, delay durations are shown as used in Experiments 2 and 3 (between 200 and 800 ms). B) Table reporting experimental factors that could change from one experiment to another: memory material, delay before probe, and presentation duration of the memory items.

Experiment 1 (regular letters presented during 1000 ms; delays from 100 to 400 ms)



Experiment 2 (regular letters presented during 1000 ms; delays from 200 to 800 ms)



Experiment 3 (phonologically similar letters presented during 1000 ms; delays from 200 to 800 ms)

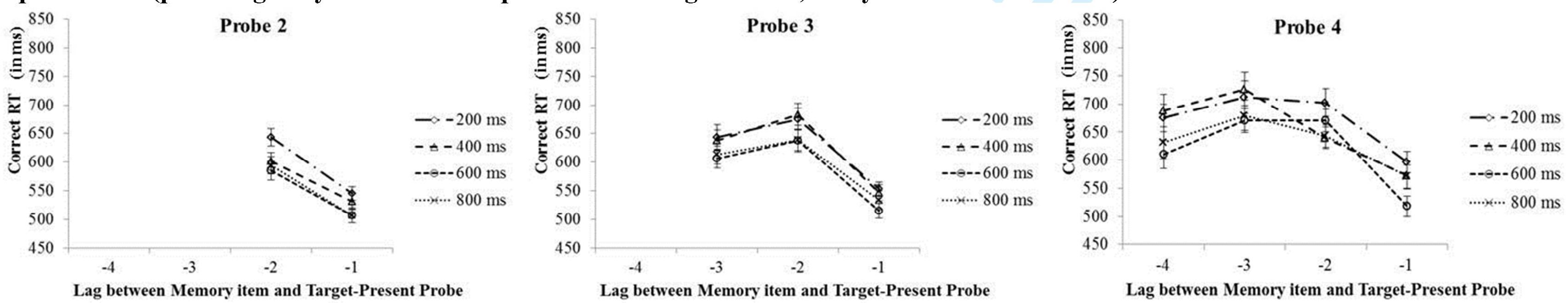


Figure 2 continues below

Experiment 4 (phonologically similar letters presented during 500 ms; delays from 100 to 800 ms)

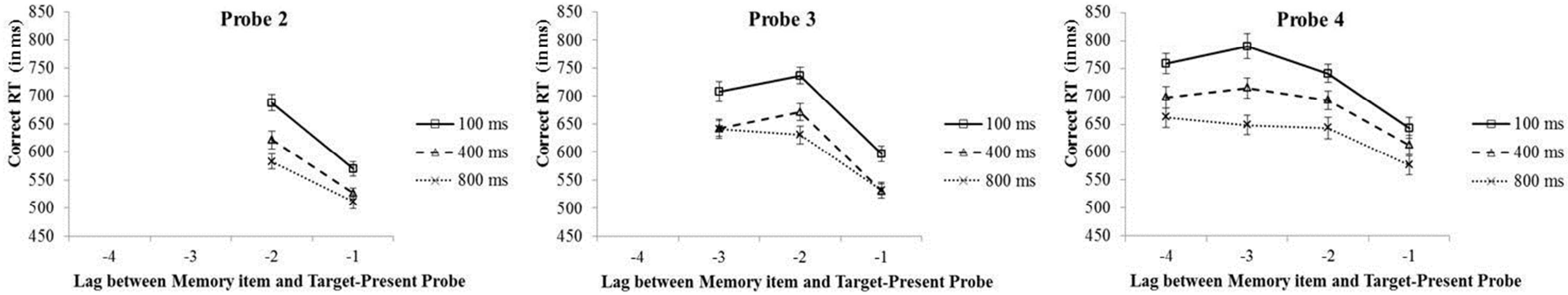
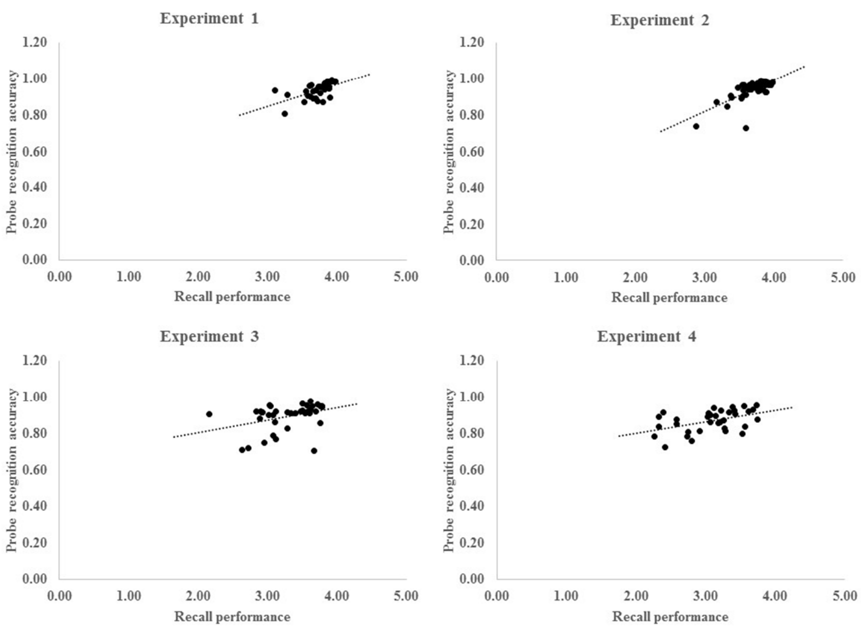


Figure 2. Mean probe response RT in ms as a function of the serial position of the matching memory item (expressed as the lag between presentation and test; on the x axis) and probe position (Probe 2, Probe 3, or Probe 4 in the left, middle and right panels, respectively). The delay following the probe appears as the graph parameter. **Top row** (Experiment 1): regular letters were used as memoranda and were presented for 1000 ms; delay durations varied between 100 and 400 ms. **Second row** (Experiment 2): regular letters were used as memoranda and were presented for 1000 ms; delay durations varied between 200 and 800 ms. **Third row** (Experiment 3): phonologically similar letters were used as memoranda and were presented for 1000 ms; delay durations varied between 200 and 800 ms. **Fourth row** (Experiment 4): phonologically similar letters were used as memoranda and were presented for 500 ms; delay durations varied between 100 and 800 ms. Error bars show standard errors of the mean.

Supplementary material 1 (recognition/recall trade-offs)

We examined the relationship between recall performance and probe recognition performance across our participants by plotting recognition performance as a function of recall performance. As can be seen in the Figure below, in none of the four experiment did we observe a recall/probe recognition trade-off. In all four experiments, we observed that people who had better recall scores also had better recognition scores. This seems to go against the idea that some people might give priority to the recall task, to the detriment of probe recognition, while others would give priority to the probe recognition task, to the detriment of recall performance. Instead, the observed pattern is consistent with the idea that the sources underlying recall performance are the same as the sources underlying recognition performance such that a better memory representation results in both better probe recognition and better recall performance.

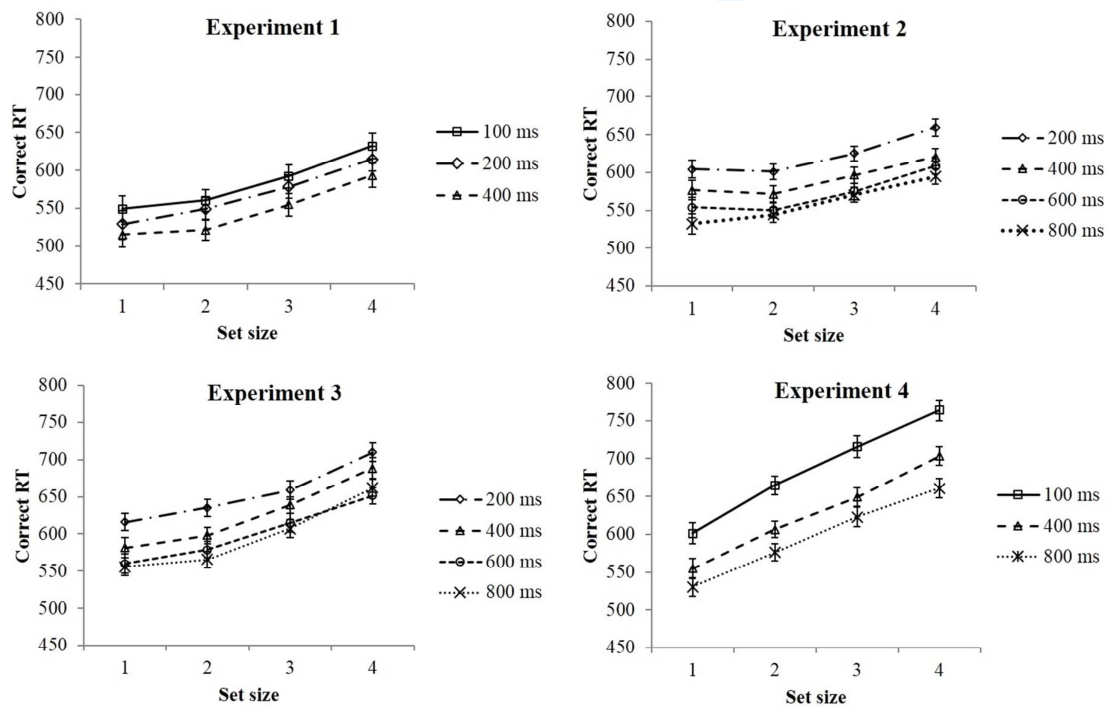


Footnote 1: We thank an anonymous reviewer for this suggestion.

Supplementary material 2 (familiarity-based recognition in probe task)

One could argue that, to perform the probe task, participants were relying on some kind of familiarity-based recognition rather than search-related processes in the focus of attention. This seems unlikely. It is, for example, difficult to see how familiarity would help recognition in Experiments 3 and 4, in which we used a restricted set of 8 phonologically similar letters as memory material. By the time one is judging the fourth probe, four different memory items have been seen and potentially 3 other letters (as probes) that are very similar to the items that have to be remembered. It seems unlikely that pure familiarity-based recognition would be able to distinguish between a probe letter representing a memory item and a probe letter that is very similar but is not representing a memory item; participants would be suffering from severe proactive interference from all the letters they have been seeing. Familiarity would be worthless in the face of the high degree of proactive interference. Recognition performance was, however, in all experiments, very high, indicating that people used working memory rather than familiarity to respond to the probes.

Furthermore, the use of familiarity-based recognition is not consistent with our data. It has been proposed that set size effects can be used to differentiate between probe recognition based on familiarity arising from activated representations in LTM and probe recognition based on an active search in the central region of WM. In particular, it has been proposed that the presence of set-size effects (increasing RTs as a function of the number of items in WM) can be used as a behavioral marker of the memory list being represented in the central, capacity-limited component of WM, as opposed to LTM (e.g., Burrows & Okada, 1975; Souza, Rerko, & Oberauer, 2014; see Vergauwe & Cowan, 2015, for a similar argumentation). We can examine this in our task in which the number or items in WM increases across probe positions (one item in WM when judging the first probe, two items in WM when judging the second probe, and so on). Using probe position as a proxy for the number of items in WM, we observed that RT increased as a linear function of set size, see Figure below. These functions suggest that participants engaged in an active search through the central component of WM in all four experiments, rather than making a familiarity-based decision.



Supplementary material 3 (Serial Position Curves beyond lag -1)

When it comes to the shape of the serial position curve, and specifically when it comes to the last-item benefit, there seem to be at least two possibilities. According to the first possibility, the fast response to the last item reflects that item being in the focus of attention, as opposed to the other items not being in the focus of attention (the focus-of-attention hypothesis). According to the second possibility, the shape of the serial position curve is reflecting the different activation levels of the different items in working memory, with the last item having the highest activation level (the activation-level hypothesis). One could argue that the serial position curve beyond lag -1 (i.e., beyond the last-presented item) might help to distinguish between the two possibilities. While the focus-of-attention hypothesis has no straightforward way of accounting why the portion of the serial position curve beyond lag -1 might not be flat, the activation-level hypothesis can quite easily account for a non-flat shape because different items might have different activation levels. It can thus be argued that observing a non-flat serial position curve beyond lag -1 might be more consistent with the activation-level hypothesis than with the focus-of-attention hypothesis.[1]

We examined the serial position curves beyond lag 1 for probe positions 3 and 4 in each of the four experiments. As can be seen in Figure 2, most curves do not seem to be flat but the pattern is not so consistent across experiments. This was confirmed in the statistical analysis. We used the same BANOVAs as reported in the main text, with Delay Before Probe and Serial Position as within-subject factors, excluding the last-presented memory item. As can be seen in the Table below, in three out of the four experiments, there is good evidence that the serial position curve is not flat beyond lag -1. In the last experiment, however, there is some evidence for the serial position curve being flat beyond lag -1. Thus, while, across experiments, there is more evidence for the serial position curves not being flat beyond lag -1, the current analysis does not allow us to draw any firm conclusions as to whether serial position curves are reflecting activation levels or the focus of attention. Moreover, it should be noted that, when supplemented with some additional mechanism responsible for some primacy gradient, the focus-of-attention hypothesis could also account for non-flat portions of serial position curves. Evidence for non-flat curves was found when participants saw each memory item for 1000 ms (i.e., in the first three experiments) while there was more evidence for flat curves when participants saw each memory item for only 500 ms (i.e., Experiment 4). It is thus possible that some sort of strategic encoding strategy is used for early list items, requiring more than 500 ms and resulting in somewhat faster responses to these items. One possibility might be that a second code is formed, different from whatever kind of code is used for the later items. It is worth noting that the other kind of code, in any case, exerted only a small effect compared to the larger, last-item benefit.

	Probe 3	Probe 4
Experiment 1	33.84 to 1 Against flat curve	3.17 to 1 Against flat curve
Experiment 2	7.26 to 1 Against flat curve	1265.32 to 1 Against flat curve
Experiment 3	17.63 to 1 Against flat curve	2.31 to 1 Against flat curve
Experiment 4	1.51 to 1 For flat curve	3.30 to 1 For flat curve

Footnote 1: We thank an anonymous reviewer for this suggestion.