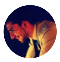
Encoding serial items depletes cognitive resource: An adversarial comparison of reconfiguration and resource-depletion accounts of working memory

Preprint · October 2025

DOI: 10.31234/osf.io/vj8pn_v1



Vencislav Popov

University of Zurich

Encoding serial items depletes cognitive resource: An adversarial comparison of reconfiguration and resource-depletion accounts of working memory

David J. Greeno[1], Candice C. Morey[1], and Vencislav Popov[2]

[1] School of Psychology, Cardiff University

[2] Department of Psychology, University of Zurich

All materials and data available to download at https://osf.io/d2zby

Word Count

7235

Correspondence:

David J. Greeno, School of Psychology, Cardiff University, United Kingdom

Email: greenodj@cardiff.ac.uk

University Switchboard: +44 2920 874000

When people encode items into working memory, the time taken to complete a secondary task slows as the memory load increases but becomes faster during retention intervals. The findings are best explained by either a reconfiguration process needed only during encoding, or by a resource-depletion model in which encoding of each item drains cognitive resource that recovers during retention. We tested diverging predictions of these accounts with a running span complex span task. List-length was unpredictable, and participants had to recall only the most recently presented N memory items. According to the reconfiguration account, after N items are presented, response times for the secondary task should stabilize. This should occur because the amount of updating required at each memory item beyond N should be equivalent. According to the resource-depletion account, each new encoding event beyond N should continue to deplete a limited resource. This depletion results in a continued slowing of response times for the secondary task. Across four experiments, secondary task response times increased with each new memory item, even after the required memory load was reached. The only exception was with memory loads of 2 where secondary task response times plateaued after 2 items. These findings place limits on reconfiguration as a comprehensive explanation for short-term memory but are mostly in line with a modified resource-depletion account.

Keywords: working memory; complex span; processing speed; resource-depletion; reconfiguration

Impact Statement

The complex span task involves a primary task (e.g. remembering digits) with a secondary task (e.g. maths or judgement tasks). This combination of storage and processing are considered to mimic real-world cognitive demands (e.g. remembering individual items from a shopping list while also working out the cumulative cost of those items). The theoretical underpinnings of performance in these tasks are still uncertain, but understanding is critical for predicting and explaining more complex, real-life behavior. The present series of experiments takes steps towards achieving this goal by directly comparing two accounts of complex span performance – one based on motor-response preparation, and one based on the depletion and recovery of an encoding resource. To promote intellectual rigor, we adopted a pre-registered adversarial collaboration with proponents of each account involved. Across 4 experiments we found evidence broadly in line with a modified version of the resource-based account. These findings help place limits on the range of explanations and further our progress towards understanding of the complex span task

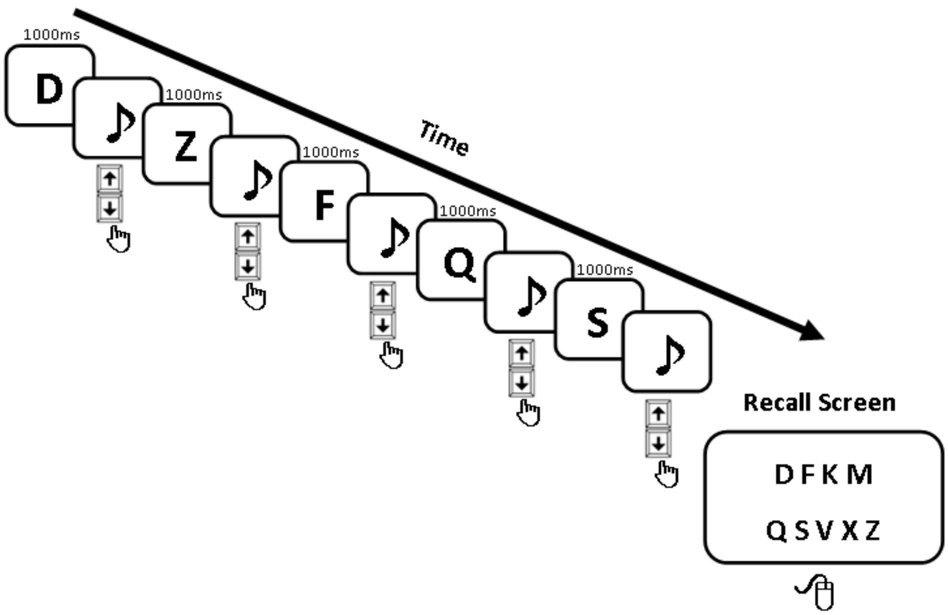
Introduction

Working memory is a system for concurrent maintenance and processing of information underpinning many cognitive functions such as reasoning and problem solving. The working memory system is often studied using complex span tasks. In these tasks, people have to remember some items, such as consonants, interleaved in a sequence with processing episodes like tone discrimination, followed by a recall test (Figure **1**). Results from complex span tasks have revealed bidirectional interference between maintenance of memoranda and processing (e.g. Vergauwe et al., 2022). Recall accuracy decreases when there are more processing episodes during the inter-item intervals (Barrouillet et al., 2004, 2011). Conversely, as presentation of the to-be-remembered list progresses, response times (RTs) to each of the interleaved processing episodes becomes slower (Joseph & Morey, 2022; Morey et al., 2018). Such results have been interpreted as support for domain-general resource models like the time-based resource-sharing model (TBRS), in which maintenance and processing both depend on the same attentional resource (Barrouillet et al., 2011), or more recently with the reconfiguration account of Joseph and Morey (2022). In this adversarial collaboration we recognize that these results can also be accounted for by the gradual resource recovery model of Popov and Reder (2020), and identify and test divergent predictions of these models.[1]

[1] The adversarial collaboration includes proponents of the reconfiguration hypothesis (C. C. Morey) and the resource-depletion-and-recovery hypothesis (V. Popov).

Figure 1.

Example of a complex span task trial whereby participants remember consonants for later recall (primary task) while also judging whether tones are high or low (secondary task).



The time-based-resource-sharing (TBRS) model (Barrouillet et al., 2004) was developed to account for complex span task data and makes the following assumptions. If left unattended, memory traces suffer from time-based decay. To prevent this decay during both encoding and retention intervals, attention needs to be repeatedly directed towards these traces to refresh them. However, since attention is limited, performing other operations will prevent or hinder the refreshing process, resulting in decayed memory traces and worse subsequent recall. According to TBRS, increasing the memory load (e.g. with each additional item added to the memory list) means more traces in memory need to be refreshed leaving less attentional resource for performing the secondary task. This causes poorer discrimination performance with increasing serial positions regardless of whether the secondary task is performed in between encoding of the memoranda or during a retention interval. This assumption was recently made explicit by Barrouillet and colleagues (2023), who stated that “the time sharing postulated by the TBRS model applies not only for storage and processing in complex span tasks, but also to any activities that involve attention and consequently compete for occupying the executive loop.” (p. 1541). While many TBRS predictions have been confirmed, it fails to account for some key aspects of the data. Since TBRS posits that memoranda are continuously refreshed to prevent decay, memory load should affect secondary task processing similarly regardless of whether processing occurs in between encoding events, or during a retention interval. However, this is not the case – there is no memory load effect during retention in a Brown-Peterson paradigm, except for the very first judgement (Joseph & Morey, 2022; Klapp et al., 1983; Oberauer, 2002).

Some of the strongest evidence against TBRS comes from Joseph and Morey (2022), who studied the effect of memory load on processing prior to encoding, during encoding and during retention. Participants had to perform a tone-discrimination task prior to the presentation of the consonants (“lead-in judgements”), in between presentations of the consonants (“load judgements”) or during the retention interval (“lead-out judgements”). The lead-in judgements were included to neutralize switching costs in response times, which had been observed in prior research after the first memory item (Morey et al., 2018). Their results showed three distinct patterns. Response times to the tone discrimination task were slow for the initial lead-in judgement, reflecting switching costs, but they quickly reached a baseline level prior to the presentation of the first memory item. After the first memory item, response times to the tone discrimination task increased substantially above baseline and continued to increase linearly with each additional memory item. Finally, during the retention interval, response times to each subsequent tone decreased gradually until they reached asymptote at nearly the lead-in baseline level.

On the face of it, these results are inconsistent with TBRS. There is only one reason in TBRS why the cost of refreshing a memory list on concurrent judgement should disappear after the start of the retention period. Barrouillet and colleagues (2023) argued that participants might be spontaneously prioritizing the processing task, interrupting the refreshing of memory items whenever a processing stimulus appears during retention. While plausible, this explanation is problematic. If participants prioritize processing judgements during retention, there should be a clear cognitive load cost on memory performance instead. However, most studies find no such cost in the Brown-Peterson paradigm, and the few that do show a substantially smaller effect compared to the Complex Span Task (for a recent review, see Langerock et al., 2025). The prioritization explanation cannot account for the absence of both effects, unless used to claim that in every study, participants spontaneously prioritize the task that shows no load effect.

As an alternative, Joseph and Morey (2022) proposed that the observed slowing of the processing task during encoding is due to the *reconfiguration* of to-be-remembered information in preparation for a later response. Instead of using distinct resources meant only for temporary maintenance, remembering may be explained by the co-opting of mental resources associated with other more general processes (Jones & Macken, 2018). Effects observed in short-term memory tasks result from participants' ability to use internal processes for manipulating and preparing information in anticipation for later recall. The processes used for preparing speech can also be applied to preparing and recalling a list of randomly ordered letters, even though this task is less natural (MacDonald, 2016). During encoding, participants convert memoranda into a response-ready speech format. Each new item prompts them to update the output sequence from the beginning of the list, which incrementally increases the time needed for concurrent processing as the list grows. After the final item is added, no further reconfiguration is required, so response preparation is complete and can wait until execution, resulting in no expected impairment in processing during the retention interval.

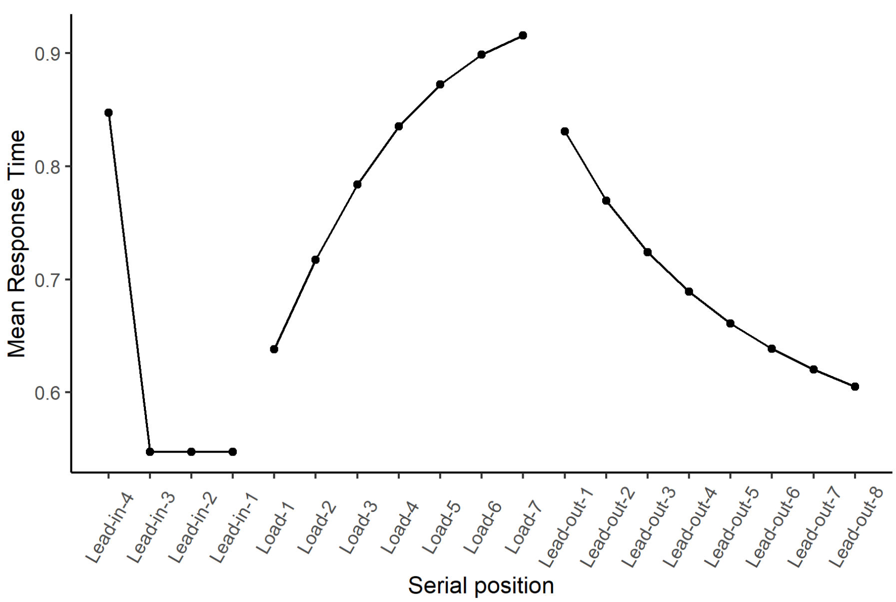
While Joseph and Morey (2022) argued that their results could not be easily accounted for by resource-based models like TBRS, resource-based models vary in their assumptions. In particular, the Source of Activation Confusion (SAC) model implements a gradual resource-depletion-and-recovery assumption (Popov & Reder, 2020; Popov et al., 2019; Reder et al., 2007). In this model, the encoding of information into working and long-term memory depletes a limited resource that *recovers gradually over time.* In contrast to the TBRS model, the resource is not related to the maintenance of memoranda. Instead, each time a new item is encoded into memory, its encoding depletes a fixed proportion of the existing resource pool with the strength of the memory representation proportional to the amount of resource depleted. Crucially, the depleted resource recovers gradually over time. This resource-depletion-and-recovery assumption can successfully model word-frequency effects and primacy effects (Popov & Reder, 2020), sequential directed forgetting effects (Popov et al., 2019), free time effects (Mizrak & Oberauer, 2021), semantic similarity effects (Kowialiewski et al., 2021), and EEG findings (Lohnas et al., 2020; Ma et al., 2024).

The resource-depletion-and-recovery assumption provides a simple explanation for the gradual slowing of tone judgement response times during encoding and speeding during retention found by Joseph and Morey (2022). During encoding, each item depletes a fixed proportion of the remaining resource pool, resulting in less resource dedicated to each subsequent memory or processing item. Since the amount of resource dedicated to an item determines the strength of its representation, which in turn determines recall and discrimination performance, the model predicts both a decrease in recall with serial position (a primacy effect for the memoranda) and a slow-down for tone judgements with serial position. Crucially, because the resource recovers gradually over time, the model, in contrast to TBRS, also predicts the gradual speeding of tone judgements during the retention interval. These predictions hold if we assume that distractor items deplete less resource than memory items, and that the recovery rate is faster than the depletion rate from distractors during the retention interval.

To show that the resource-depletion-and-recovery hypothesis is consistent with the data, we simulated a simplified version of the SAC model (see the accompanying model code and Popov & Reder, 2020 for a full explanation of the model). Response times were simulated via a drift diffusion process, with the drift rate for each processing item determined by the amount of available resource at the time of encoding it. We did not try to fit the data directly; rather, we selected parameters that reproduced the qualitative pattern observed in Joseph and Morey (2022). The only ad hoc assumption added to the model is that the initial lead-in judgement reflects a switching cost, consistent with the interpretation of Joseph and Morey. As can be seen from Figure 2, the model captures well the gradual slowing of the processing judgements with each serial position during encoding followed by a speeding of processing judgements during the retention interval. In summary, both the reconfiguration account and the resource-depletion-and-recovery model of Popov and Reder (2020) provide plausible accounts of Joseph and Morey's findings.

Figure 2.

Simulation of Joseph and Morey's (2022) Experiment 3 results with the SAC model. Note the correspondence with the right panel of Figure 2 of Joseph and Morey.



The current experiments examine a critical point where the predictions made by the two hypotheses differ. What would happen if new items were encoded, but the length of the output plan remained constant? We created a novel task combining the unpredictable series length of a running span task (Pollack et al., 1959) with interspersed judgments of the complex span task. In this new task, participants do not know when the series of memoranda and judgments will finish but know that they must always recall the most recent N items presented. Once the number of presented items exceeds N, the task requires continuous updating of the response set. With each additional item, participants must update their response set by removing the earliest item in it and adding the new item at the end.

The reconfiguration hypothesis predicts that if the number of items presented are less than N, tone judgement response times should slow down, just like in the traditional complex span task. This is because there are an increasing number of items that must be transformed into a response-ready format. However, once the number of presented items exceeds N the response set no longer increases in number and the reconfiguration time should no longer increase with serial position. Thus, tone judgement response times should gradually increase up to serial position N, possibly up to N+1 to reflect the cost of updating, after which they should remain at the same level until the retention interval, when updating is no longer required.

In contrast, the resource-depletion-and-recovery hypothesis posits that each encoded item depletes a proportion of hypothetical available resource. Popov and Reder (2020) envision this resource as limited fuel for memory formation. In contrast to the reconfiguration hypothesis, the resource-depletion-and-recovery hypothesis predicts that the slowing in tone judgement response times should continue as long as new items are presented for encoding, because each new item requires a portion of the encoding resource.

The present experiments tested these predictions by incorporating a running span version of the complex span task whereby the lists varied in length (up to a total of 9 items) meaning that the list lengths were always unpredictable to the participants. Participants then recalled the last N items presented which was either the last 2 or 4 (Experiment 1a/1b) and 3 or 4 (Experiment

2a/2b) items. Additionally, to ensure generality of findings across domains an auditory (1a/2a) or visual processing task (1b/2b) was interleaved with the presented items. Neither account explicitly predicts that the nature of the processing task will affect the outcome. Confirming whether encoding a verbal sequence for response impacts auditory and visual processing similarly will help to better describe any shared resource that needs to be coordinated to support multi-tasking.

Experiments 1a and 1b

Experiment 1 incorporated a running span version of the complex span task with list lengths up to 9 items. Participants were unaware how long each list would be but required to recall the last N items (either 2 or 4). Experiment 1a used an auditory processing task interleaved between memory items and Experiment 1b used a visual processing task. This novel task enabled us to tease apart the reconfiguration and resource-depletion hypotheses with each able to make distinct predictions about the response times for the processing task beyond the critical Nth item in the list. To summarize, the reconfiguration account predicts an asymptote in judgement response times at the value reached around the second to third item in the Recall Last 2 condition and around the fourth to fifth item in the Recall Last 4 condition (see Figure 3). The resource-depletion account predicts that judgement response times will increase gradually with each list position, regardless of how many items need to be recalled (see Figure 4). These predictions were pre-registered (see https://osf.io/d2zby/registrations).

Figure 3.

Predicted interleaved judgement response times as a function of serial position (x axis), overall list length (differentiated by color), and recall condition (per panel) according to the reconfiguration account.

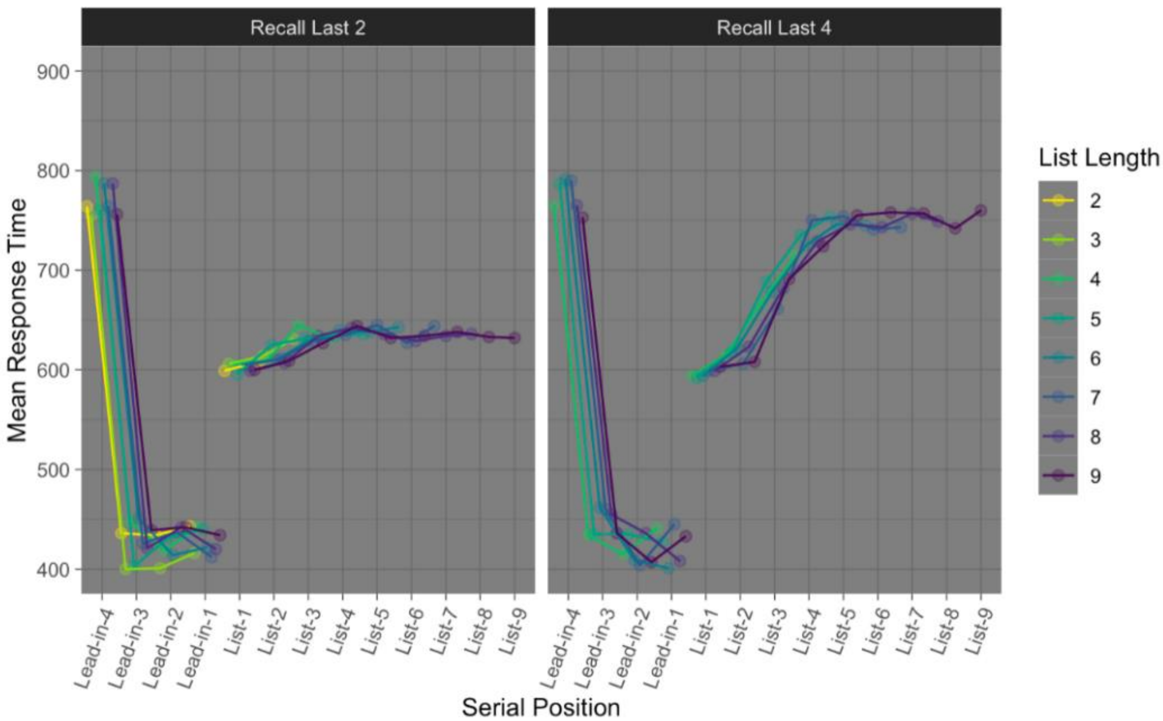
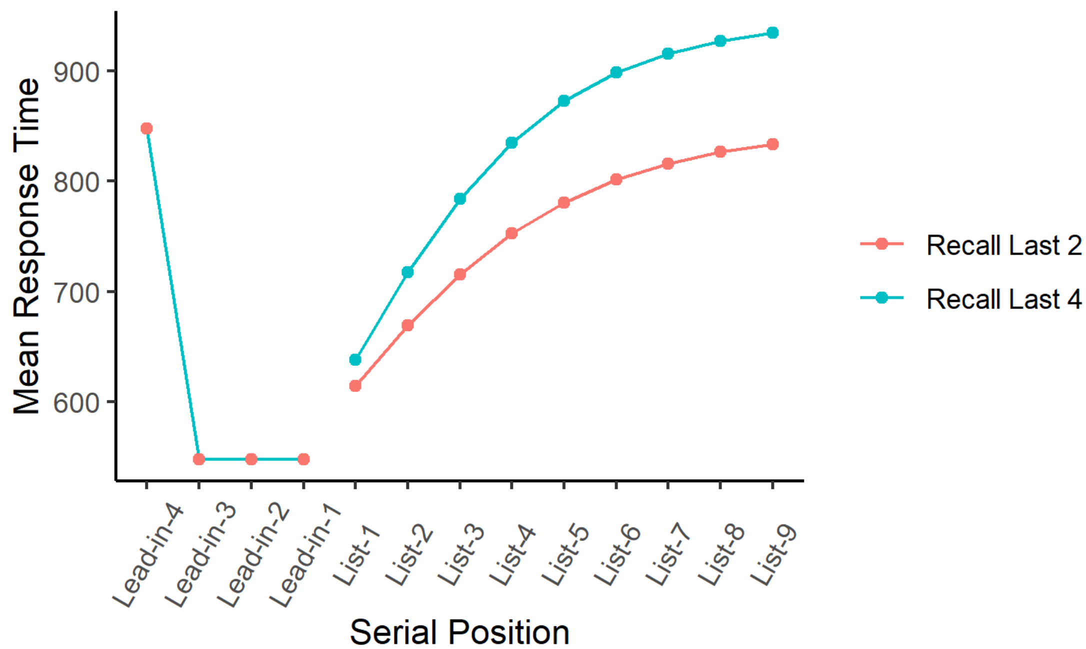


Figure 4.

Predicted interleaved judgement response times as a function of serial position (x axis), overall list length, and recall condition according to the resource recovery and depletion account.



Note. Based on the work by (Bunting et al., 2006), the model assumes that people put less effort into the "Recall Last 2" relative to the "Recall Last 4" condition. In the model this is achieved by lowering the learning rate (in this case, from 0.3 to 0.25), which also reduces the amount of resource depleted by each memory item. As a result, more resources are available for processing the distractors in the "Recall Last 2" condition, causing faster overall response times, and a smaller slope over serial position. List Length predictions overlap, thus only one line is shown, for the maximum list length condition. Predictions were simulated from the Popov & Reder (2020) model.

Method

Participants

The final sample comprised 60 participants for Experiment 1a (*M* = 21.6 years, *SD* = 3.57) and 60 participants for Experiment 1b (*M* = 20.82 years, *SD* = 3.11). The final sample only included participants that had passed our pre-registered exclusion criteria which stipulated a minimum of 85% correct on the tone judgements and a minimum of 75% correct recall in the 2-item condition. Thirty-three additional participants from Experiment 1a data collection and 21 from Experiment 1b data collection failed to meet the inclusion criteria and so were not included in our final samples of 60.

Participants were recruited from the Cardiff School of Psychology participation panel and awarded course credits for their participation or were recruited via the *Prolific* Academic Online Participant Panel (https://www.prolific.co) and received monetary compensation. To maintain consistency between the sample demographics drawn from each panel, only those based in the UK and aged between 18-30 were permitted to take part via Prolific. Those recruited via Prolific were also required to have a minimum approval rate given to them by investigators from previous studies of 90%. Previous research has confirmed that comparable data quality can be expected from participants recruited from Prolific and university samples (Uittenhove et al., 2023).

All participants were required to have normal/corrected vision and hearing, and all stages of the experiments were conducted in accordance with the Cardiff School of Psychology ethics procedures (approval number EC.18.09.185339GR).

Sample Size

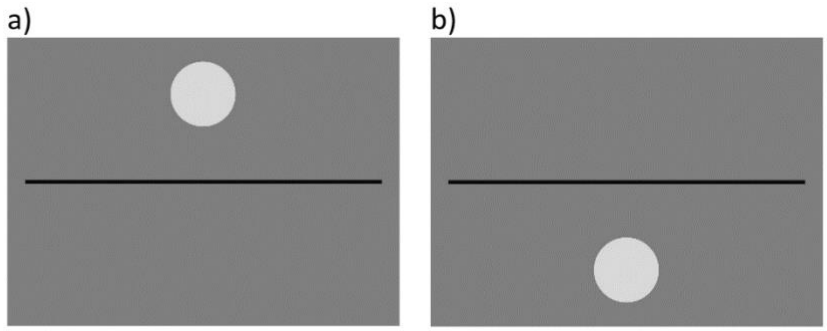
With samples of around 40, Joseph and Morey (2022) observed robust slowing of tone responses during list presentation. We decided on a larger sample size of 60 because of our use of longer lists of up to 9 items. To keep the experiment duration under one hour, we needed to decrease the number of observations per list length. However, we applied Bayesian analysis techniques so that if our results had been inconclusive with N=60, we would have continued to increase the sample in batches of 20 until reaching a decisive outcome. This sampling strategy was preregistered.

Materials

To-be-remembered items in the complex-span task comprised nine consonants (D, F, K, M, Q, S, V, X, Z). These consonants were selected by Joseph and Morey (2022) for their variability in place of articulation in the vocal tract and to avoid alphabetically consecutive consonants within to-be-remembered lists. For the processing element of the complex span task the auditory processing task consisted of tones that were either the note B (high tone; 308 Hz) or G (low tone; 245 Hz). A visual processing task was used in Experiment 1b consisting of a grey circle placed either above (analogous to a high tone) or below (analogous to a low tone) a black line (see Figure 5.).

Figure 5.

Visual judgement task. Participants indicate via key press whether the grey circle is above (a) or below (b) the black line.



Design and Procedure

Both variables, recall condition and list length, were manipulated within-participants. After consenting to take part, participants undertook a short training session in which they learned to identify the low and high tones (1a) or whether the circle was above or below the line (1b). Participants were then guided through instructions of the primary recall task, namely that they must recall either the most recent 2 or 4 letters in the order that they had appeared.

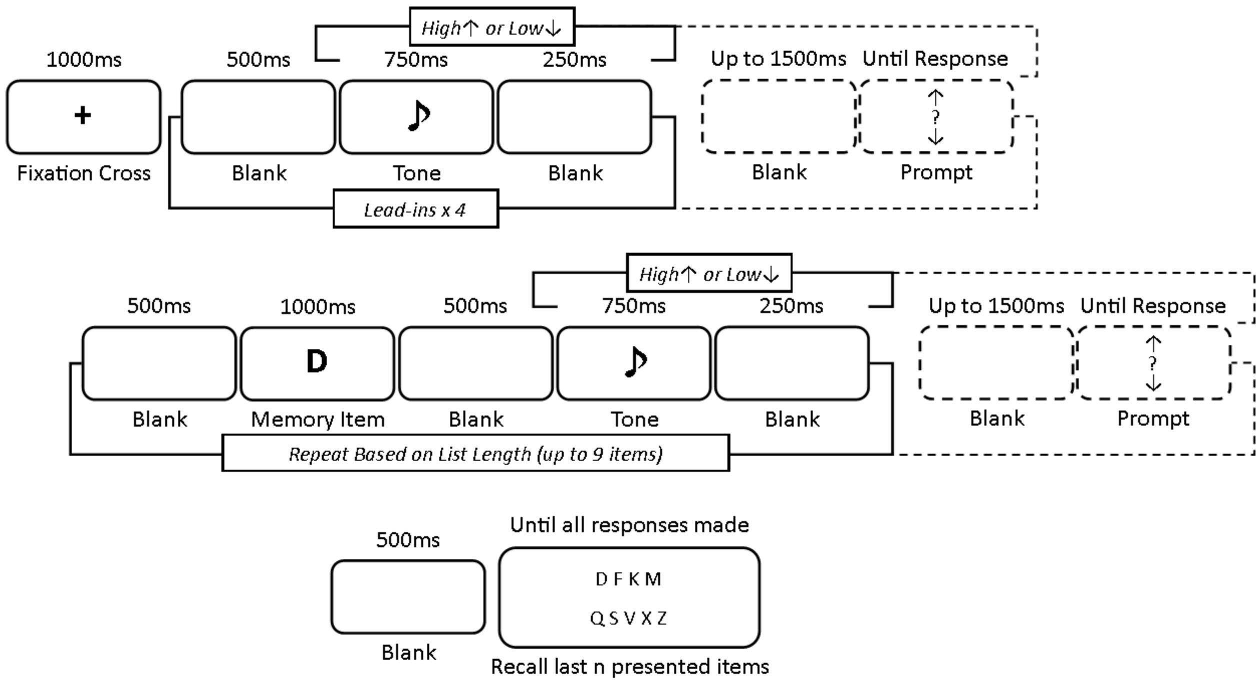
For each trial, a fixation cross appeared for 1 second followed by four consecutive lead-in processing trials. A tone (1a), or image (1b) was presented for 750ms followed by a 250ms inter stimulus interval (blank screen). In a slight departure from the methodology used by Joseph and Morey (2022), upon presentation of the tone/image participants could immediately indicate their response to the tone/circle (i.e. was it high/above or low/below). The response was collected via an up or down arrow keypress. If the response was made within 750ms a visual cue (up or down arrow) informed the participant of their response and that it had been registered. Assuming a response given, the next item was presented 1000ms after initial presentation of the previous processing task stimuli. This kept timings between stimuli consistent. However, on any occasions where no response was made within 1000ms, the program continued to await a response for a further 1500ms (blank screen) with a visual prompt to respond (up and down arrow in the center of the screen) appearing 2500ms after initial presentation of processing task stimuli. The experiment only proceeded once a response was given. This slight change in methodology was incorporated to capture the actual time taken for participants to perform the processing task judgement and indicate a response (i.e. within 1000ms of stimuli presentation), rather than the time taken to indicate a decision in response to a later prompt (i.e. >1000ms from stimuli presentation) as had been the case in Joseph and Morey.

Following four lead-in judgements, participants then completed a complex-span sequence (to-be-remembered digit followed by processing trial) up to a maximum list length of 9. Participants were required to either recall the final two items or the final four items of the to-be-remembered list. This to-be-remembered list-length variable was blocked with the presentation order of those blocks counterbalanced. Within a block, list lengths were randomly mixed and then drawn without replacement. Once all list lengths were used, they were replaced with this process repeating until the end of the block. This ensured no more than two consecutively presented lists of the same length. Within each recall block, each list length was presented an equal number of times. For two-item recall, list-lengths ranged from 2-9 items with five trials at each list length. For four-item recall, list-lengths ranged from 4-9 items with five trials at each list length. This gave a total of 70 experimental trials (40 two-item recall and 30 four-item recall).

In the complex-span element, to-be-remembered consonants were presented for 1 second with a 500-ms inter-stimulus interval. Each consonant was followed by the processing task (same presentation parameters as lead-in trials), and this continued up to the length of the list. Following presentation of all consonants and processing tasks, all nine consonants were re-presented on screen in alphabetical order with participants required to click on the final N letters from the list in the same serial order they saw them presented. See Figure 6 for a procedural schematic including timings. Prior to beginning each experimental block, participants first completed practice trials comprising 1 instance of each list length. The experimental programs for all 4 experiments are available to download at https://osf.io/d2zby.

Figure 6.

Trial schematic of the running span version of the complex span task.



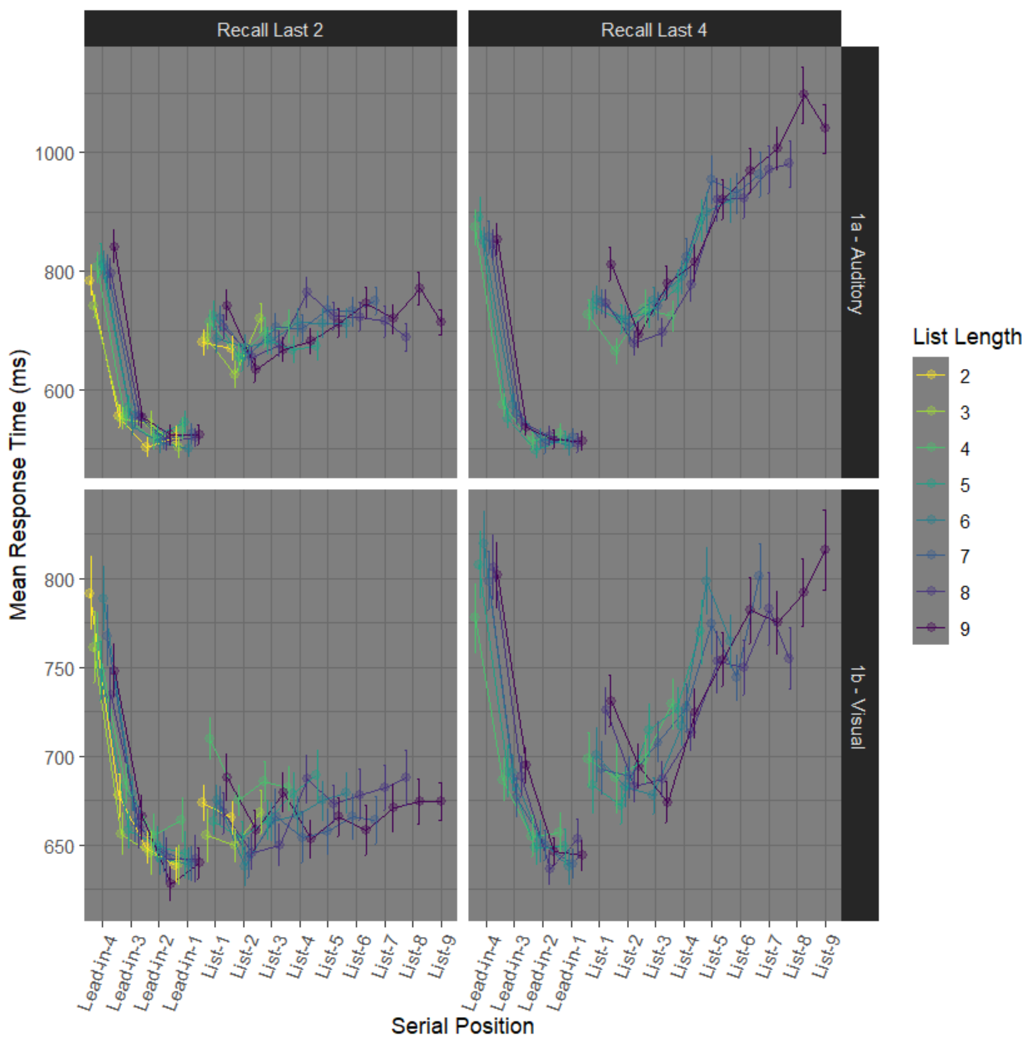
Note. Task elements in dashed lines only occurred when a processing task response took longer than 1000ms with those elements ending immediately upon a response. Task parameters were identical across all 4 experiments except for using a visual, rather than an auditory processing task (represented by tone symbol in figure), in experiments 1b and 2b.

Results

All analyses were conducted in line with the pre-registration with any exceptions noted. The primary analyses across all experiments focused on the judgement response times, particularly whether (or when) response times reached asymptote by serial position, and whether tone judgement response times differed depending on the number of items to be recalled or the overall list length. Only response times from correct judgements were included in the analyses. Before proceeding with the analysis, we excluded responses faster than 200ms (considered to indicate that participants were anticipating a response), responses more than 3 SDs from each participant's mean and any responses over 5000ms. This meant that 3.06% of the otherwise valid responses were excluded from the 1a dataset and 2.72% from 1b. Figure 7. Shows the response times (in milliseconds) for the processing tasks, plotted by serial position and experiment.

Figure 7.

Response times (in milliseconds) for the secondary processing task for the 2-item and 4-item recall conditions in Experiment 1a and 1b.



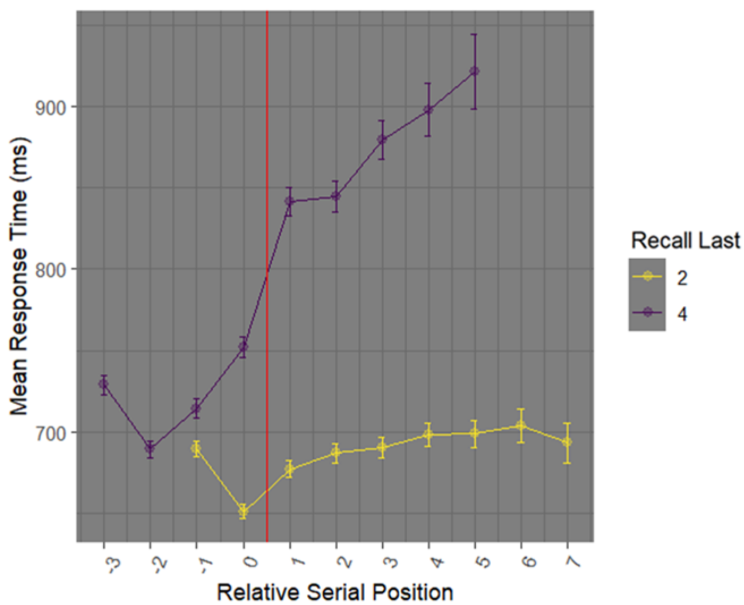
We used Bayesian multi-level regression as implemented in the *brms* R package to analyze response times after position N+1. Because descriptive patterns appeared comparable regardless of the stimulus modality of the processing task, we collapsed across experiment for increased power. Response times were log transformed and then z transformed with continuous predictors also z transformed and categorical predictors set to sum contrasts. Three models were tested. Model 1 included list length and recall amount (i.e. predicting no increase after position N+1). Model 2 included list length, recall amount and *relative* serial position. The serial position was coded relative to the recall amount – for example, item in position 4 was coded as having relative position 2 in the “recall 2 condition”, but as having relative position 0 in the “recall 4 condition”. In this way, Model 2 tested if there is an increase in processing times after position N+1 for both 2-item and 4-item recall. Model 3 included list length, recall amount, serial position and the interaction between relative serial position and recall amount. To note, in a slight deviation from the initially pre-registered analysis we ran all three models with and without random slopes. Including random slopes and only removing if there is strong evidence against their inclusion can help reduce the risk of drawing false positive conclusions (see Oberauer, 2022). All output below is from the models including random slopes but the models without also yielded similar results (see analysis files on our OSF, https://osf.io/d2zby).

Model 3 was the best model with *BF = 35.47* in favor compared to model 2 and model 2 having a *BF* = 155.87 in favor over model 1. Because of the strength of evidence in favor of the interaction between serial position and list length, models were then run separately for 2-item recall and 4-item recall. For 4-item recall the model incorporating serial position was favored over the model without, *BF* = 181.6. This is strong evidence in favor of a continued increase in

RTs after position N+1. However, for 2-item recall the model not incorporating serial position was favored over the model including it, *BF* = 4.24, providing moderate evidence against there being an increase in response times after N+1. Figure 8. shows the response times collapsed across experiments and list length and plots the data as a function of serial position relative to the critical position N.

Figure 8.

Response times (in milliseconds) collapsed across experiments for the secondary processing task plotted by serial position relative to the critical item N.



Discussion

Results from Experiments 1a and 1b did not provide conclusive support for either hypothesis. While the reconfiguration hypothesis correctly predicted the asymptote after item N+1 for 2-item recall it did not predict the continued increase in response times after item N+1 for 4-item recall. Similarly, while the resource-depletion hypothesis correctly predicted the continued increase in response times at item N+1 for 4-item recall it did not predict the asymptote at item N+1 for 2-item recall. However, despite neither account predicting the full set of data the present findings are not yet reason to dismiss either hypothesis.

Firstly, when predicting likely performance under the reconfiguration hypothesis the difficulty in preparing and continually updating a 4-item motor-plan may have been underestimated. When accounting for the pattern of response times in Joseph and Morey (2022) reconfiguration of to-be-remembered information was considered additive in nature (i.e. item 1, item 1 + item 2, item 1 + item 2 + item 3 etc.) with this process continuing until the final item in the list. However, in the present experiment, once N is reached in the list, the reconfiguration process instead requires updating (i.e. removing item 1 and beginning a new reconfiguration process at item 2 which then comprises the most recently presented item at the end). This continual updating process is necessary until the final item in the list. These could be considered somewhat different processes with the effort required for updating, particularly for 4 items, far exceeding the effort required for the additive reconfiguration thought to occur in Joseph and Morey. As such, the point at which an updating process for longer lists would reach an asymptote may exceed item n + 1 meaning there were not enough items after the critical N in which to see an asymptote.

Secondly, within the resource-depletion hypothesis there is an encoding cost associated with each item (known as the learning rate). It was assumed that while the learning rate would be slightly lower for 2-items compared to 4-items it would still be high enough to prevent full recovery of resource after each presented encoding event and prior to completing the processing task. However, as part of estimating an appropriate learning rate for 2 and 4-items it was not considered that 2-item recall may be completed without any item to serial position binding. Usually, it would be considered necessary to bind to-be-remembered items to their correct serial position to maintain correct ordering of that information. However, for 2-item recall the last item can be held in the focus of attention with participants knowing it is the last item. All that is then required is that the most active other item should be selected first, followed by the item in the focus of attention. For the 2-item recall condition it could be optimally completed via this passive process leaving sufficient resource for the processing task irrespective of the list length. However, as soon as the number of to-be-remembered items increases, as binding to serial position once again becomes necessary, then this passive process quickly becomes suboptimal. Experiment 2a and 2b again sought to differentiate the two competing hypotheses while adjusting for the new considerations suggested by Experiment 1's outcome.

Experiments 2a and 2b

Given the findings of Experiment 1 and the two possible explanations for the data not fitting either hypothesis, Experiments 2a and 2b utilized a 3-item recall condition along with a 4-item recall condition. For the reconfiguration hypothesis this should reduce the effort of updating, relative to 4-item recall, and bring the task closer to the effort required for the continuous updating of 2-items. As such, due to the effort required to update 4-item lists, an asymptote was no longer predicted after serial position 4 in the Recall 4 condition. However, due to a relative decrease in effort, it was predicted that response times would not increase after serial position 3+1 in the Recall 3 condition. For the resource-depletion hypothesis, replacing 2-item recall with 3-item recall means that binding to serial position becomes a requirement for all the featured list-lengths. Optimal performance of the task can no longer be achieved via a passive process. As such, the learning rate will be higher for 3-items, relative to 2-items, and again a continued increase in response times after serial position N+1 was expected.

Participants

The final sample comprised 60 participants for Experiment 2a (*M* = 25.57 years, *SD* = 3.34) and 60 participants for Experiment 2b (*M* = 26.62 years, *SD* = 2.88). The final sample only included participants that had passed our pre-registered exclusion criteria which stipulated a minimum of 85% correct on the tone judgements and a minimum of 66.69% correct recall in the 3-item condition. This second exclusion figure was calculated from the average recall performance achieved in the 4-item recall condition in Experiment 1. Recall accuracy for 3 items should be higher or, at the very least, equivalent to recall of 4 items.

Twenty participants from Experiment 2a data collection and 16 from Experiment 2b data collection failed to meet the inclusion criteria and so were not included in our final samples of 60. Additionally, one participant was also excluded from the final sample in Experiment 2b due to them incorrectly following the task instructions (not responding immediately but rather waiting for the prompt to respond after 2500ms elapsing).

Participants were all recruited via *Prolific* and received monetary compensation. They were recruited from the same demographic described in Experiment 1 and had not participated in Experiment 1.

Materials

The consonants and processing tasks were identical to those used in Experiment 1.

Design and Procedure

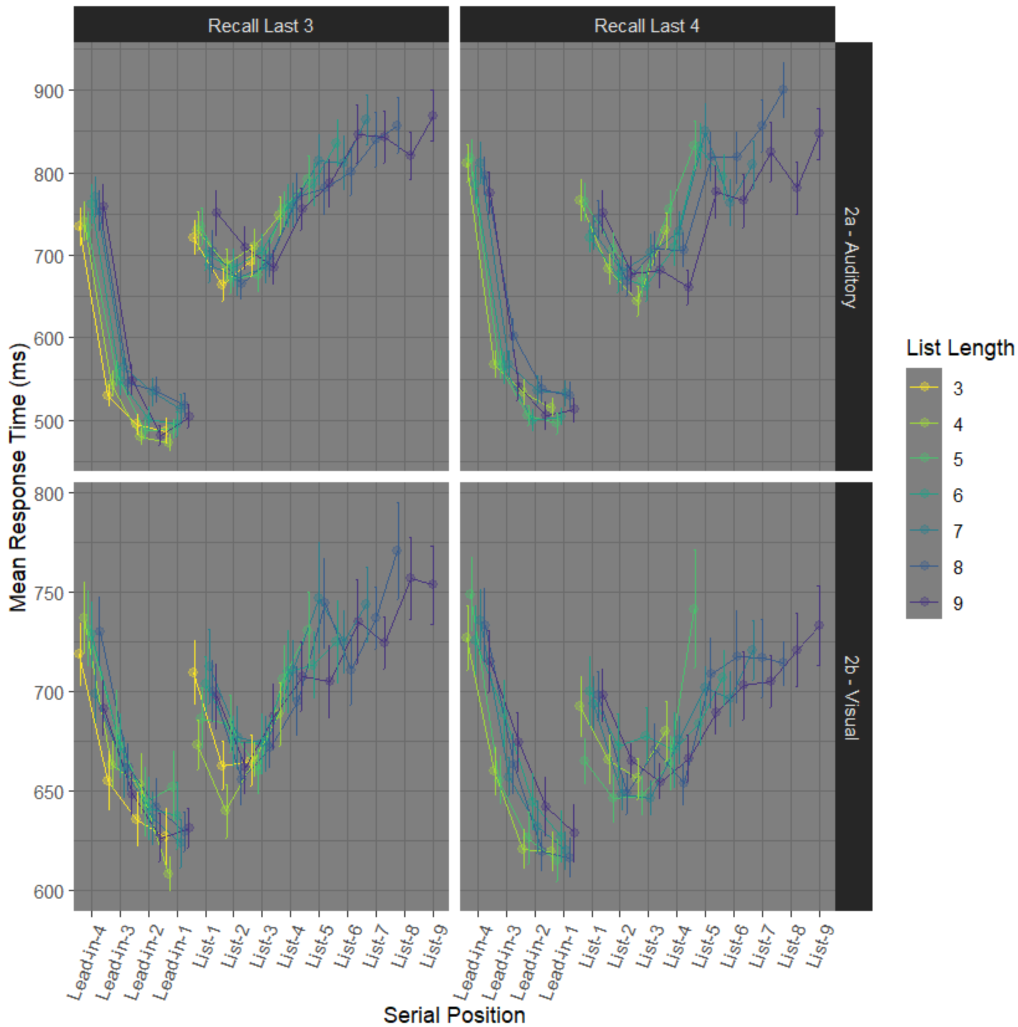
This was similar to Experiment 1 except the 2-item recall condition was replaced with 3-item recall. The 4-item recall condition remained. All other presentation parameters were identical with the auditory judgement task used in Experiment 2a and visual judgement task in 2b.

Results

Data preparation and response exclusions were identical to Experiment 1 resulting in 2.71% of otherwise valid responses excluded form Experiment 2a and 2.12% from 2b. Figure 9. shows the response times for the processing tasks, plotted by serial position and experiment.

Figure 9.

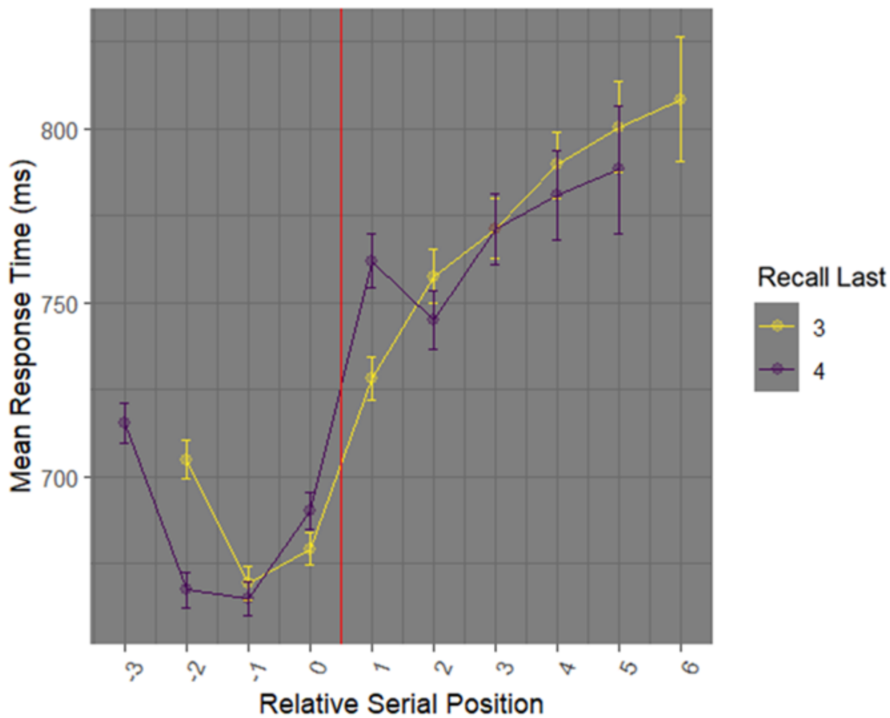
Response times (in milliseconds) for the secondary processing task for the 3-item and 4-item recall conditions in Experiment 2a and 2b.



The analysis process was identical to Experiment 1 with three models tested. Model 1 included list length and recall amount. Model 2 included list length, recall amount and relative serial position and model 3 included list length, recall amount, relative serial position and the interaction between relative serial position and recall amount. Model 2 was the best model with *BF = 1432569.44* in favor compared to model 1 with model 2 having a *BF* = 15.52 in favor over model 3. The strong evidence against the interaction included in model 3 indicates that the increase in response times after N+1 is identical irrespective of whether there are 3 or 4 memory items to recall. Figure 10. shows the response times collapsed across experiments and list length and plots the data as a function of serial position relative to the critical position N.

Figure 10.

Response times (in milliseconds) collapsed across experiments for the secondary processing task plotted by serial position relative to the critical item N.



Note. Relative serial position 0 is the 3rd item in the 3-item recall condition or the 4th item in the 4-item recall condition.

Discussion

The results from Experiment 2 were more clearcut in separating the competing hypotheses than those of Experiment 1. Contrary to the reconfiguration hypothesis, response times for the secondary processing task continued to rise after item N+1 in both the 3 and 4 item recall conditions. This continued rise in response times after item N+1 was in line with the predictions made by the resource-depletion account.

General Discussion

Across the two sets of experiments, when participants were required to recall the last 3 or 4 items from lists of varying length, we observed a continued slowing of response times for an interleaved auditory or visual processing task after memory item N+1. However, for recall of 2 items we saw an asymptote in response times after item N+1.

The findings lend support for the suggestion that working memory operates via depletion and recovery of a fixed pool of resource, adding further credence to a now rapidly establishing evidence base (e.g. Kowialiewski et al., 2021; Lohnas et al., 2020; Mizrak & Oberauer, 2021; Popov et al., 2019; Popov & Reder, 2020; Ma et al., 2024). As the experimental task proceeded, each encoding event depleted a fixed proportion of available resource leaving insufficient time for full resource recovery prior to the next encoding event. With incrementally less resource available for each subsequent encoding event, the processing task response times continually increased.

Reconciling competing theoretical accounts

While the resource-depletion-and-recovery account accommodates performance with a memory load of 3 or more items, the present data from our "recall 2" condition provoked a new assumption within the resource model. For a passive task requiring memory of just 2 items whereby no position binding is necessary, the learning rate should be set very low. Conceptually, this implies that not all encoding is equivalent; rather, there are circumstances under which adequate encoding can take place without taxing the resource, namely when contextual binding is not required. The specific value assumed for the learning rate depends on which version of the model is used. The full model (Popov & Reder, 2020) simulates the entire process of item activation, binding to positional cues and retrieval via reactivation of positional cues. In that model, two separate parameters govern a) the strength of activation of item nodes and b) the strength of binding, and both operations deplete the same resource pool. To accommodate the "recall 2" condition, the model would set the binding strength parameter to 0 and rely only the relative item activation to perform the task, as described previously. While the full model provides the most complete account of the mechanisms underlying memory tasks, in practice we often use a simplified model that is more mathematically tractable while producing identical predictions for many purposes (for a more detailed discussion, see Ma et al., 2024). In the simplified model (which was used for the simulations in this paper), the binding and retrieval process is not modeled explicitly. Instead, we use a single composite learning rate to derive predicted item activation values during recall. This simplified version can accommodate the "recall 2" condition by setting the learning to any value that results in equilibrium between depletion and recovery (which will depend on the recovery rate). Although this simplification yields adequate approximations, there may be other task scenarios like our "recall 2" condition where separately considering item activation and binding demands enhances the explanation of how the task is performed. The more complex assumption of a distinction between item activation and contextual binding in the full encoding resource model could be essential for conveying the boundaries of depletion of the resource and accurately predicting memory limits.

For tasks requiring memory of 3 or more items then a consistent, but higher value, is required. Essentially, the current data suggests that the learning rate is dependent on whether a task requires binding. Once established that binding is required then the learning rate is consistent irrespective of the amount to-be-remembered. However, while providing a plausible explanation for the present data, whether these assumptions hold across different working memory tasks is something which will need to be tested further.

For the reconfiguration hypothesis the results are more difficult to accommodate, and likely point to some fundamental limits in how far perceptual and motor processes can explain short-term memory phenomena on their own. Given that the length of the required response remains consistent after item N+1, there is no reason to suppose that reconfiguration of memoranda into a motor plan would take incrementally longer with presentation of each new item. It could perhaps be suggested that in the present series of tasks participants were not actively updating the list via a cumulative reconfiguration process with preparation of the entire motor response instead occurring after presentation of all items (e.g. Palladino & Jarrold, 2008). For verbalizable information, within the reconfiguration hypothesis, it is assumed speech motor planning functions are co-opted to prepare a serially ordered sequence of letters for later recall. Response preparation only occurring at the end of the list within the present experimental tasks is entirely possible.

However, this means that the increasing response times during the encoding period still need explaining, because if participants adopt a passive approach to apprehending the incoming memory items, why should responses on the corresponding choice task slow at all? From a strong theoretical standpoint whereby it is suggested all short-term memory tasks can be explained via perceptual and motor functions (e.g. Jones & Macken, 2018), preparation of the letter responses should have been possible without presuming any specialized short-term memory resource, regardless of whether response preparation occured during or after list presentation. Within this view, a parsimonious explanation for the continued increase in choice response times during list presentation becomes very difficult to sustain.

An alternative is to instead consider that while perceptual and motor processes must contribute towards short-term memory task performance, it is perhaps not sufficient to only draw reference to those processes. Mental effort is considered inherently aversive (e.g. David et al., 2024) and Morey et al. (2025) suggest that children become more likely to rehearse when the task is perceived as more challenging. A possibility is that participants within a given task setting are attempting to successfully complete that task while trying to keep mental effort to a minimum. Relying on perceptual and motor processes is one such way in which mental effort can be reduced and it should be assumed that this low-effort method will be adopted when it will work. For some short-term memory tasks, (e.g., for the 2-item recall condition reported here and possibly for short lists in Joseph & Morey, 2021), relying on perceptual processes and motor preparation can be a successful strategy. However, at the point where the mental effort required for successful task completion increases (e.g. in the case of 3 and 4-item recall in the present experiments) then the reliance on perceptual and motor processes may not be sufficient, and mental effort must increase. The difference in response times between the 2-item recall and the 3- and 4-item recall at item N+1 observed in the present experiments may indicate that as task demands increase, and more mental effort becomes necessary, then perceptual and motor processes alone are no longer sufficient for successful task completion.

More broadly speaking the present data also has implications for other theories of working memory. For the TBRS model (Barrouillet et al., 2004) an asymptote in response times after item N+1 would also be predicted as the amount of content to be attended to and refreshed remains consistent from that point. Along with the previous findings of Joseph & Morey (2022) of rapid speeding of response times after completion of encoding in the retention interval, it seems necessary to reconsider some of the assumptions that form the TBRS as, at present, they are not compatible with the data. Similarly, initial consolidation (e.g., Ricker et al., 2018), or strategic decisions to elaborate (e.g., Bartsch et al., 2018) could plausibly explain an increase in response times as memory load increases but there is no reason to suppose that those processes would continue to take longer despite the overall memory load that those processes apply to remaining the same.

Task-switching costs

Surprisingly, including lead-in tones here did not completely eliminate the switch cost at the first serial position compared to subsequent positions, as lead-in tones seemed to do in Joseph and Morey (2022). Joseph and Morey found that using lead-in tones eliminated the initial delay in response times at the first serial position. In this experiment, lead-ins were included to help participants adjust to task switches and prevent slow responses at position 1 from affecting the measurement of response time increases across serial positions. When preparing the current pre-registration, we ran versions of the experiment that collected responses 1000ms after presentation of the processing task stimuli had begun (identical parameters to Joseph and Morey). In those data (see supplementary figure on our OSF page, https://osf.io/d2zby) the response times at serial position 1 were not obviously slower than at position 2. However, in the present series of experiments, when collecting response times immediately upon presentation of the processing task stimuli this nonlinear pattern re-appeared. Neither hypothesis predicted this, and such non-linear patterns of response times, if taken at face value, would be incompatible with many models of working memory. However, rather than over-interpreting these patterns and assuming that memory load for some reason exerts a truly non-linear effect here, it seems likely that the reduction in time before response collection in these experiments meant that the lead-in tasks were no longer fully neutralizing the substantial slowing of responses that can be caused by a task switch (e.g., Koch et al., 2018; Vandierendonck et al., 2010). Future research could verify this further by varying the time from memoranda offset to collection of the processing task response time, or increasing the number of lead-in judgments.

Conclusion

The patterns of response times reported here lend support for a process of resource recovery and depletion that affects both memory encoding and simple choice reaction tasks. By considering the overall number of encoding events in a task (both the memory items and interleaved secondary task processing) the account can successfully accommodate a continued slowing of response times to a secondary processing task as lists progress, and for the gradual speed up of responses during a subsequent retention interval. Importantly, these results also answer a long-standing question about the resource at the center of the SAC model. Popov and Reder (2020) conceptualized the resource specifically as a fuel for memory formation operations, but at the time it was unclear whether the same resource would be shared with other executive functions. The present study makes it clear that both auditory and spatial simple choice reaction tasks tap into this same resource, which suggests a broader functionality.

Constraints on Generality

All our participants were based within the United Kingdom, recruited from the Cardiff University School of Psychology participant pool or via Prolific. Participants were within a narrow age band (Undergraduate Psychology students or those aged 18-30 on Prolific) with expertise in the English language. As such, we cannot be certain that our results and their implications would apply to children, the elderly, or across cultures. However, our aim was to compare two hypotheses using a difficult and dull task, which necessarily means recruiting a narrow range of participants in peak cognitive health. We have no reason to suppose that the results reported in our article do not reflect general cognitive processes but as views about memory converge, it will become more important to confirm this using a broader range of tasks and populations.

References

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*(1), 83.

Barrouillet, P., Camos, V., Minamoto, T., Nishiyama, S., Chooi, W. T., Morita, A., Logie, R. H., & Saito, S. (2023). Time sharing in working memory processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(10), 1539–1556. https://doi.org/10.1037/xlm0001261

Barrouillet, P., Portrat, S., & Camos, V. (2011). On the law relating processing to storage in working memory. *Psychological Review*, *118*(2), 175.

Bartsch, L. M., Singmann, H., & Oberauer, K. (2018). The effects of refreshing and elaboration on working memory performance, and their contributions to long-term memory formation. *Memory & Cognition*, *46*(5), 796–808. https://doi.org/10.3758/s13421-018-0805-9

Bunting, M., Cowan, N., & Scott Saults, J. (2006). How does running memory span work? *Quarterly Journal of Experimental Psychology*, *59*(10), 1691–1700. https://doi.org/10.1080/17470210600848402

David, L., Vassena, E., & Bijleveld, E. (2024). The unpleasantness of thinking: A meta-analytic review of the association between mental effort and negative affect. *Psychological Bulletin*. https://doi.org/10.1037/bul0000443

Jones, D. M., & Macken, B. (2018). In the Beginning Was the Deed: Verbal Short-Term Memory as Object-Oriented Action. *Current Directions in Psychological Science*, *27*(5), 351–356. https://doi.org/10.1177/0963721418765796

Joseph, T. N., & Morey, C. C. (2022). Impact of memory load on processing diminishes rapidly during retention in a complex span paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition* , *48*(10), 1400–1419. https://doi.org/10.1037/xlm0001061

Klapp, S. T., Marshburn, E. A., & Lester, P. T. (1983). Short-term memory does not involve the" working memory" of information processing: The demise of a common assumption. *Journal of Experimental Psychology: General*, *112*(2), 240.

Koch, I., Poljac, E., Müller, H., & Kiesel, A. (2018). Cognitive structure, flexibility, and plasticity in human multitasking—An integrative review of dual-task and task-switching research. *Psychological Bulletin*, *144*(6), 557–583. https://doi.org/10.1037/bul0000144

Kowialiewski, B., Lemaire, B., & Portrat, S. (2021). How does semantic knowledge impact working memory maintenance? Computational and behavioral investigations. *Journal of Memory and Language*, *117*, 104208.

Langerock, N., Oberauer, K., Throm, E., & Vergauwe, E. (2025). The cognitive load effect in working memory: Refreshing the empirical landscape, removing outdated explanations. *Journal of Memory and Language*, *140*, 104558. https://doi.org/10.1016/j.jml.2024.104558

Lohnas, L. J., Davachi, L., & Kahana, M. J. (2020). Neural fatigue influences memory encoding in the human hippocampus. *Neuropsychologia*, *143*, 107471. https://doi.org/10.1016/j.neuropsychologia.2020.107471

Ma, S., Popov, V., & Zhang, Q. (2024). A neural index reflecting the amount of cognitive resources available during memory encoding: A model-based approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/xlm0001364

MacDonald, M. C. (2016). Speak, Act, Remember: The Language-Production Basis of Serial Order and Maintenance in Verbal Memory. *Current Directions in Psychological Science*, *25*(1), 47–53. https://doi.org/10.1177/0963721415620776

Mizrak, E., & Oberauer, K. (2021). What is time good for in working memory? In *Psychological Science*. https://doi.org/10.31234/osf.io/ahqwj

Morey, C. C., Hadley, L. V., Buttelmann, F., Könen, T., Meaney, J.-A., Auyeung, B., Karbach, J., & Chevalier, N. (2018). The effects of verbal and spatial memory load on children's processing speed. *Annals of the New York Academy of Sciences*, *1424*(1), 161–174. https://doi.org/10.1111/nyas.13653

Morey, C. C., AuBuchon, A. M., Attwood, M., Castelain, T., Cowan, N., Crepaldi, D., Fjerdingstad, E., Fredriksen, E., Jarrold, C., Koch, C., Lelonkiewicz, J. R., Lupyan, G., Mendenhall, W., Moreau, D., Schonberg, C., Tamnes, C. K., Vlach, H., & Elliott, E. M. (2025). Is Verbal Rehearsal Strategic? An Investigation into Overt Rehearsal of Nameable Pictures in 5- to 10-Year-Old Children. *Journal of Cognition and Development*, *26*(1), 50–69. https://doi.org/10.1080/15248372.2024.2389123

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *In*, 411–421.

Palladino, P., & Jarrold, C. (2008). Short Article: Do Updating Tasks Involve Updating? Evidence from Comparisons with Immediate Serial Recall. *Quarterly Journal of Experimental Psychology*, *61*(3), 392–399. https://doi.org/10.1080/17470210701664989

Pollack, I., Johnson, L. B., & Knaff, P. R. (1959). Running memory span. *Journal of Experimental Psychology*, *57*(3), 137–146. https://doi.org/10.1037/h0046137

Popov, V., Marevic, I., Rummel, J., & Reder, L. M. (2019). Forgetting Is a Feature, Not a Bug: Intentionally Forgetting Some Things Helps Us Remember Others by Freeing Up Working Memory Resources. *Psychological Science*, *30*(9), 1303–1317. https://doi.org/10.1177/0956797619859531

Popov, V., & Reder, L. M. (2020). Frequency effects on memory: A resource-limited theory. *Psychological Review*, *127*(1), 1–46. https://doi.org/10.1037/rev0000161

Ricker, T. J., Nieuwenstein, M. R., Bayliss, D. M., & Barrouillet, P. (2018). Working memory consolidation: Insights from studies on attention and working memory. *Annals of the New York Academy of Sciences*, *1424*(1), 8–18. https://doi.org/10.1111/nyas.13633

Uittenhove, K., Jeanneret, S., & Vergauwe, E. (2023). From Lab-Testing to Web-Testing in Cognitive Research: Who You Test is More Important than how You Test. *Journal of Cognition*, *6*(1), 13. https://doi.org/10.5334/joc.259

Vandierendonck, A., Liefooghe, B., & Verbruggen, F. (2010). Task switching: Interplay of reconfiguration and interference control. *Psychological Bulletin*, *136*(4), 601–626. https://doi.org/10.1037/a0019791

Vergauwe, E., Von Bastian, C. C., Kostova, R., & Morey, C. C. (2022). Storage and processing in working memory: A single, domain-general resource explains multitasking. *Journal of Experimental Psychology: General*, *151*(2), 285–301. https://doi.org/10.1037/xge0000895