Journal of Experimental Psychology: Learning, Memory, and Cognition

Testing the Response Suppression Mechanism of Working Memory

Benjamin Kowialiewski and Klaus Oberauer

Online First Publication, June 3, 2024. https://dx.doi.org/10.1037/xlm0001359

Journal of Experimental Psychology:
Learning, Memory, and Cognition

ISSN: 0278-7393

https://doi.org/10.1037/xlm0001359

Testing the Response Suppression Mechanism of Working Memory

Benjamin Kowialiewski[1, 2] and Klaus Oberauer[1]
[1] Cognitive Psychology Unit, Department of Psychology, University of Zürich
[2] Psychology & Neuroscience of Cognition Research Unit (PsyNCog), University of Liège

Many working memory (WM) paradigms involve recalling multiple items from the same memory set. Participants rarely repeat items they have already recalled, avoiding repetition errors. To prevent these errors, WM models incorporate a response suppression mechanism that removes recalled items from the set of response options. Despite its importance for our understanding of WM, response suppression has received limited direct testing. To address this gap, we used computational models implementing two hypothetical mechanisms of response suppression to derive predictions and tested these predictions experimentally. Participants were asked to recall the same items multiple times during a single trial. If already recalled items are removed from the response set to prevent repetition errors, memory performance should be impaired when the same item is tested again. Contrary to this, we found that memory performance was unimpaired when the same item was tested a second time, and even displayed a recall advantage. Therefore, this study demonstrates the implausibility of response suppression to account for how people avoid repetition errors. We discuss alternative explanations.

Keywords: working memory, response suppression, recall, computational modeling

Many working memory (WM) paradigms involve recalling multiple items from the same memory set. The best-known version of such a task is the serial-recall task in which participants are asked to repeat back a list of items in the correct order. One commonly held assumption when modeling WM is that recalled items are discarded from WM to prevent repetitions, a mechanism called response suppression. Without this response suppression mechanism, current models of serial recall would not be able to realistically simulate human data because they would produce many repetition errors, which rarely occur. Despite the importance of this mechanism for models of WM, only a few empirical studies support its existence. Due to this empirical uncertainty, little is known regarding the way repetition errors should be prevented in models of WM. This study will test two different accounts of response suppression. We implement these two hypothetical mechanisms in a generic model of serial recall, derive predictions from the models, and test them experimentally.

When encountering a list such as "monkey, ball, and desk," the order of items is often new and arbitrary. Items' order must therefore be temporarily stored in some way. In many computational models, this is done via item–context binding (Burgess & Hitch, 1999, 2006; Farrell, 2012; Lewandowsky & Farrell, 2008; Oberauer & Lewandowsky, 2011; Oberauer et al., 2012; Schneegans & Bays, 2017). When encoding a memory list, it is assumed that people create new temporary associations between items and positional contexts. In the example above, these are associations between "monkey" and "Position 1," and between "ball" and "Position 2," and so on. An illustration of item–context association is presented in Figure 1. Items presented at adjacent serial positions share a proportion of contextual markers to represent the fact that adjacent positions are more similar to each other than distant ones. Retrieval is done by cueing an item with its position (i.e., trying to retrieve the item "monkey" using the cue "Position 1"), which is implemented by re-activating the original positional context. This cueing leads to the parallel re-activation of a set of retrieval candidates, as shown in the upper panel of Figure 1. Not only the target item is reactivated but also other items share a proportion of their positional markers with the target. In models of WM, items' activation does not completely determine which item will be selected. Instead, the selection process is stochastic. This is sometimes implemented by converting items' activation into recall probabilities using a selection rule (e.g., Luce's choice rule). Alternatively, selection for recall is simulated by selecting the candidate with the highest activation level after adding noise to it. In either case, the item receiving the highest activation from the positional retrieval cues is most likely to be selected, which in many cases is the correct (i.e., target) item. If another list item than the target item is selected, a transposition error occurs (i.e., recalling an item at the wrong serial position).

When the task requires multiple retrieval attempts in the same trial, such as serial recall, people sometimes repeat previously recalled items

Nash Unsworth served as action editor.

Benjamin Kowialiewski https://orcid.org/0000-0002-7743-761X

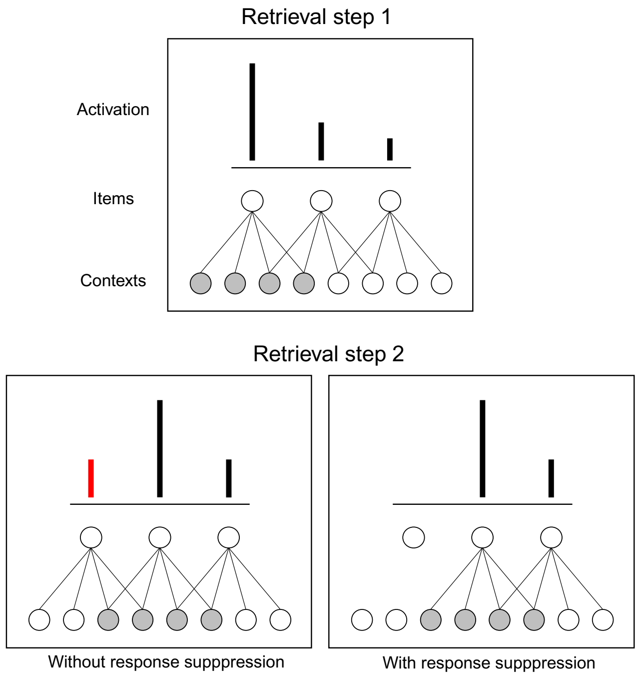
This research was supported by a grant from the Swiss National Science Foundation to Klaus Oberauer (Project 100014_192204). This is a registered report. All the data and codes have been made available on the Open Science Framework: https://osf.io/r2avt/ (Kowialiewski & Oberauer, 2023).

Benjamin Kowialiewski served as lead for conceptualization, formal analysis, software, writing–original draft, and writing–review and editing. Klaus Oberauer served as lead for resources and supervision, contributed equally to project administration, and served in a supporting role for writing–original draft and writing–review and editing. Benjamin Kowialiewski and Klaus Oberauer contributed equally to investigation, validation, visualization, and methodology.

Correspondence concerning this article should be addressed to Benjamin Kowialiewski, Cognitive Psychology Unit, Department of Psychology, University of Zürich, Binzmühlestrasse 14/22, 8050 Zurich, Switzerland. Email: bkowialiewski@uliege.be

Figure 1

Illustration of the Removal Mechanism of Response Suppression



Note. Encoding in the model is accomplished through item–context association, denoted by the straight lines connecting the contexts and the items. During retrieval, an item is cued by its position, such as the first position in this example. As contextual nodes are shared among positions, the cueing process triggers the simultaneous reactivation of multiple retrieval candidates. Following the successful retrieval of an item (in this case, the first item), the links with its associated context are subsequently removed (bottom right). This removal process prevents an item from being recalled twice. If response suppression does not occur (bottom left), there remains a probability for the already-recalled items to be erroneously recalled again, potentially resulting in a repetition error. See the online article for the color version of this figure.

(e.g., recalling ABCDEC), a type of error named repetition errors. Repetition errors are rare as they represent only about 5% of the total responses. To avoid the production of an unrealistically large number of repetition errors, models of WM include a response suppression mechanism. After being recalled, items are discarded from the competition. The phenomenon is illustrated in Figure 1, lower left panel (retrieval Step 2). As can be seen in the left panel, without this response suppression mechanism, repetition errors (i.e., recalling Item 1) would be as frequent as producing transposition errors (i.e., recalling Item 3), which is never observed. The inclusion of a response suppression mechanism solves this problem (lower right panel). Therefore, response suppression is a necessary mechanism for models of WM. Without it, models would produce too many repetition errors and would not realistically capture human data. Some work also suggests that response suppression could partially contribute to the recency effect observed when plotting recall performance across serial positions (Farrell & Lewandowsky, 2012).

Despite the importance of response suppression for models of WM, direct evidence supporting its existence is scarce. One line of evidence comes from the Ranschburg effect (Henson, 1998a; Maylor & Henson, 2000), which refers to the difficulty in recalling the second occurrence of an item. For instance, given the sequence to be remembered "ABCDBF," people fail to recall the second occurrence of the letter "B" more often than items presented in matched positions in sequences not including such repetitions (i.e., the letter "E" in "ABCDEF"). Response suppression is one plausible explanation for the Ranschburg effect. When people are asked not to recall the first occurrence of the repeated item, the Ranschburg effect disappears (Harris & Jahnke, 1972). Convergent outcomes have been found in partial recall procedures involving the recall of only one item or one part of the list (Jahnke, 1970). The fact that the Ranschburg effect appears only when the same item needs to be output twice is in line with the idea that people discard items after recalling them (Greene, 1991). Response suppression, however, is not the only plausible explanation for the Ranschburg effect. Recently, Roodenrys et al. (2022) reported strong interference effects occurring at recall. They manipulated items' phonological similarity in such a way that three items shared one phoneme with a target item that occurred later in the list (e.g., "noise cool bag push cash save"). As compared to matched sequences containing little to no phonemic overlap (e.g., "noise cool bag push teach save"), recall of the target item (i.e., "cash") was severely impaired. Further experiments showed that this interference effect was specifically because of recalling the items sharing the phoneme with the target item, as the effect disappeared when the target had to be recalled first. This result could be explained by a form of feature overwriting (Nairne, 1990) occurring at output, whereby producing a set of features overwrites the same set of features across all other WM representations. This mechanism could also explain the Ranschburg effect. Thus, the Ranschburg effect does not unambiguously speak in favor of response suppression.

Some studies provided evidence that a response suppression mechanism is required to model recall latencies realistically in immediate serial recall (Farrell & Lewandowsky, 2004). Farrell and Lewandowsky showed that response latencies for anticipation and postponement errors differ qualitatively: Whereas anticipation errors slow down with increasing displacement distance, the opposite pattern is observed for postponements, a pattern of results that can be captured only by a model including a primacy gradient of activation coupled with a response suppression mechanism. Recent investigations suggest however that the pattern of recall latencies observed by Farrell and Lewandowksy might be an artifact because of participants trying to segment the list into subgroups (Cowan & Elliott, 2023). Studies have indeed shown that grouping might occur spontaneously, even in the absence of a grouping structure imposed by the experimental setup (Farrell et al., 2011). This aspect, not controlled by Farrell and Lewandowsky, might be problematic, because recall latencies for items at the beginning of a group are always longer than in subsequent group positions. Hence, results from Farrell and Lewandowsky are subject to alternative explanations. Novel ways to test response suppressed are therefore warranted.

Response suppression can be implemented in two ways. The first implementation involves the removal of item–context associations. Such a mechanism is already implemented in models such as SOB-CS (Oberauer et al., 2012) and TBRS* (Oberauer & Lewandowsky, 2011). In these models, item–context binding is done via Hebbian learning. The removal of item–context associations is equivalent to Hebbian learning, except that a negative learning rate is used, a process called Hebbian anti-learning (Farrell & Lewandowsky, 2002). Therefore, the link between the item and its position is removed,[1] which is illustrated in Figure 1, right panel of the retrieval Step 2. As can be seen in the illustration, using the second position as a cue does not lead to partial retrieval of the first item, because there is no existing association anymore between the positional markers of Position 2 and the first item. The plausibility of this removal mechanism is supported by recent evidence showing that people can intentionally forget information in WM (Dames & Oberauer, 2022). For simplicity, we label this mechanism removal.

The second mechanism is an inhibition mechanism. Once recalled, an item is not removed from WM. Instead, it is discarded from the competition during the selection stage. In a model such as illustrated in Figure 1, this competition plays out in a competitive cueing layer, into which items' activation is fed. The competitive cueing layer selects the most likely candidate based on its activation level relative to other items. The inhibition mechanism can be modeled by adding twin units to the item units in the competitive cueing layer, as illustrated in Figure 2. The role of the twin units is to inhibit the activation of the items in the competitive cueing layer once recalled. At the beginning of a trial, all inhibitory units are deactivated (upper panel). Once an item is recalled, its corresponding inhibitory twin unit is triggered, thereby downregulating the activation of the item (lower panel).

The removal and inhibition mechanisms differ in one important aspect. The removal mechanism, by suppressing the item–context association, weakens the contents of WM. The inhibition mechanism, in contrast, keeps the contents of WM intact. This fundamental difference gives rise to diverging predictions. According to the removal account, people should have difficulties recalling items more than once, because the information should no longer be present in their WM after a first retrieval. According to the inhibition account, recalling items multiple times should still be possible if people can release their inhibition. One way the inhibition account could make people recall twice is by resetting all inhibitory units to zero after recalling all items, thus creating the opportunity to recall items a second time. To test this possibility, we will use a modified cued recall paradigm, in which participants' memory will be tested in a random order multiple times. In cued recall, items are cued from their position in random order. For instance, given the sequence "ABCDEF," participants could be asked to recall items in this order: "613524." Accordingly, participants would be first cued with "Position 6" and would recall the letter "F," followed by the cue "Position 1," and so forth. We will adapt this cued recall paradigm by including multiple bursts of retrieval attempts within the same trial.

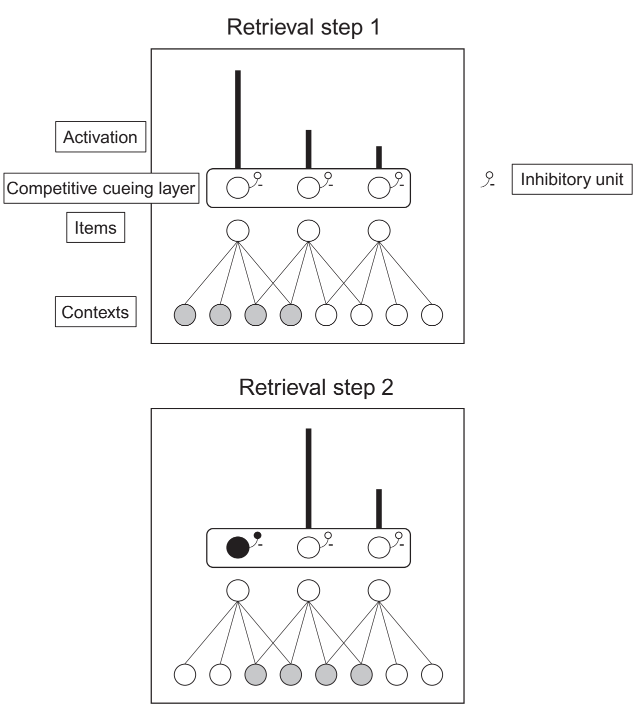
In Experiment 1, participants will be tested in a cued recall paradigm in which two bursts of retrieval attempts distinctively follow each other. For instance, participants could be asked to recall items in this order: "142635–263514." If items are removed from WM, recalling items twice should be problematic. If items are inhibited with a possibility to release this inhibition after the first burst of retrieval attempts, recalling items twice should be possible. In addition, we consider the possibility that participants have voluntary control over response suppression, and therefore stop suppressing items when they anticipate having to recall them twice. To test for this possibility, we included a control condition in which items will be tested only once (see methodological details of Experiment 1).

Experiment 2 used a variation of this paradigm, in which the multiple retrieval attempts were mixed in such a way that they could not be clearly identified anymore, for instance: "143242563561." The rationale of this second experiment is to test the inhibition account of response suppression. If people can release their inhibition after the first burst of retrieval attempts, confounding the two bursts of retrieval attempts should prevent people from recalling items twice. This is because, with one continuous series of retrieval attempts, there is no point at which response inhibition could reasonably be released to enable access to already recalled items again.

One potential limitation of the experiments described above is that memory performance is expected to decrease from the first burst of retrieval attempts to the second, regardless of the response suppression mechanism. Indeed, the mere fact of recalling items impairs memory performance for subsequent to-be-remembered information (Cowan et al., 2002), a phenomenon known as output interference. Therefore, decreased recall performance when items are recalled a second time is not likely to exclusively reflect the presence of response suppression. In addition, response suppression is not binary: The extent to which items are removed or inhibited might vary from one person to another, and this might be difficult to clearly disentangle from output interference. Experiment 2 implements a first solution to this problem, by comparing recall performance for repeated items with nonrepeated items in matched output positions (see methodological details of Experiment 2). A second solution is implemented using

Figure 2

Illustration of the Inhibition Mechanism of Response Suppression



Note. According to the inhibition account, items' activation is fed into a competitive cueing layer. Repetition errors are prevented thanks to the activation of inhibitory units, whose role is to inhibit the items' activation in the competitive cueing layer. In this mechanism, the core WM representation (i.e., the binding between item and position representations) remains intact after each retrieval attempt. WM = working memory.

[1] The process of removal is not an all-or-nothing phenomenon. It is controlled by a free parameter that varies continuously.

computational modeling. We collected data in a pilot experiment in which participants performed the cued recall paradigm as described above. In this pilot experiment, the cued recall paradigm involved only one burst of retrieval attempts. Based on the results of this first experiment, we fitted a standard serial recall model of WM including the most commonly established mechanisms in the literature (Cowan et al., 2002; Henson, 1998b; Hurlstone & Hitch, 2015; Lewandowsky, 1999; Page & Norris, 1998), in which response suppression was modeled using either the removal or the inhibition mechanism. The model also includes a mechanism for output interference. Predictions were then generated from these two mechanisms of response suppression, and from a model version without any response suppression (but still including output interference). By comparing experimental data to these predictions, we can determine whether the accuracy difference between the first and the second recall attempts is better explained by one of the models with response suppression, or by the model version with output interference alone. We compared the predictions of these three model versions to the data from Experiments 1 and 2 to determine which model fits the data best.

Preliminary Experiment

In this preliminary experiment, we asked participants to perform a cued recall paradigm. Each trial involved the encoding of a five-item list to be remembered. At retrieval, participants were given positions as cues and recalled the words associated with them. This experiment aimed at providing a set of data to fit a standard model of serial recall. This model was then used to generate new predictions for Experiments 1 and 2, based on the two response suppression mechanisms described in the introductory part.

Method

Participants

Thirty young adults aged between 18 and 35 years were recruited on the online platform Prolific (https://prolific.co/). Sample sizes were estimated based on a recent study of our own showing strong serial position effects in cued recall (Kowialiewski et al., 2023). As serial position effects are usually large, we were confident that this sample size would be sufficient to reach strong levels of evidence (Bayes factor [BF] > 10 for either the null or the alternative hypothesis) concerning the effects of interest. In case the BF did not reach a sufficient level of evidence, we planned to recruit thirty more participants; this turned out to be unnecessary. All participants were English native speakers, reported no history of neurological disorder or learning difficulty, and gave their written informed consent before starting the experiment. This experiment was carried out in accordance with the ethical guidelines of the Faculty of Arts and Social Sciences at the University of Zurich, as were Experiments 1 and 2 reported below.

Materials

The lists were constructed in such a way that we minimized the presence of semantic relationships among list items. As the presence of semantic relationships among items in a list has a large impact on memory performance, we reasoned that it should be particularly important to control for it when testing response suppression. The stimuli were drawn from a pool of 336 concrete words. The pool was composed of 42 different semantic categories (e.g., motorized vehicles, body parts, animals, beverages, etc.), with eight words in each semantic category. Each list was built by randomly sampling one item from five different categories that did not share semantic relationships with each other. For instance, we avoided sampling from the category "drinks" (e.g., whiskey) if a word from the category "container" (e.g., glass) was included in the list. For that purpose, the a priori semantic relationships between the categories were identified using a confusion matrix which identified the categories sharing one semantic relationship, a procedure we used in previous studies (Kowialiewski et al., 2022, 2023). We were concerned that related items would be particularly easy to memorize, which would potentially make them more difficult to remove. A total of 120 versions of the task were created. Participants were randomly assigned to one of these versions.

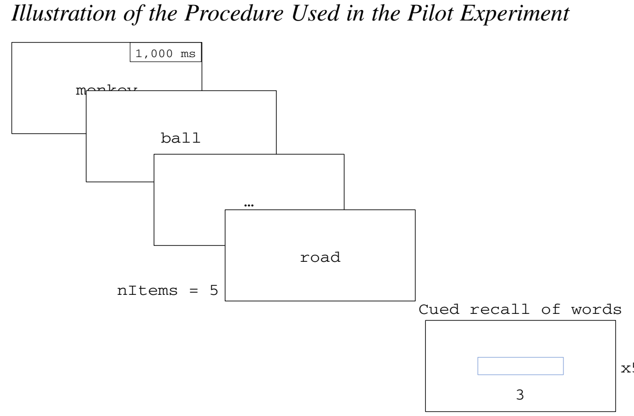
Procedure

Each trial started with the presentation of a five-item list to be remembered. Words were presented sequentially at the center of the screen in Courier font. Each word was presented for 1,000 ms, followed by the next word with no inter-stimulus interval. Directly after the presentation of the last item, the retrieval phase began. Participants were asked to perform cued recall. A prompt box appeared at the center of the screen, along with a number indicating the position of the to-be-recalled item. Participants were asked to type the word associated with each position. For instance, given the list to be remembered "cheek, scooter, freeze, son, and nova," participants could be asked to recall first the item in Position 3, thus recalling "freeze," followed by Position 5, recalling the word "nova," and so on until all items were tested. To validate each response, participants pressed the "Enter" key of their keyboard. If participants did not know a given item, they were invited to leave the prompt box empty and move on to the next item, resulting in an omission error. Participants performed four training trials. The task is illustrated in Figure 3.

Scoring Procedure

Recall performance was assessed using two scores. First, we used a strict serial recall criterion, in which an item was scored as correct if

Figure 3
Illustration of the Procedure Used in the Pilot Experiment



Note. Participants were presented with five items to remember. At retrieval, participants were cued with a position and were asked to type the word associated with it. All items were tested once. See the online article for the color version of this figure.

recalled at the correct serial position. Second, we computed repetition errors, which correspond to recalling an item twice in the list (e.g., recalling "dog, wall, and dog"). This second score is particularly informative for model fitting. Repetitions—more precisely, their absence—reflect the extent to which an item has been suppressed to remove it from the competition after being recalled. Therefore, the proportion of repetitions can be used to estimate response suppression parameters in the computational models. A high rate of repetition errors means that little suppression occurred, which should result in a low response suppression parameter. Conversely, if little to no repetition errors occur, the response suppression parameter should be very high.

Statistical Analysis

We conducted Bayesian analyses using the brms package implemented in R (Bürkner, 2017). We fitted a logistic regression model, with output position as a fixed effect, a random intercept, and a random effect of output position. To get the strength of evidence for a particular effect, we performed Bayesian model comparisons using a top-down testing procedure. We assessed each effect of interest by comparing the full model to the same model without the effect in question using the bayes_factor() function provided in the brms package.

All materials, scripts, and code have been made available on Open Science Framework: https://osf.io/r2avt/.

Results

Recall performance as a function of output position was assessed using a Bayesian logistic regression model. When comparing the full model against the full model without the output position effect, we found decisive evidence ($BF_{10} = 4.287e+63$) supporting the output position effect. We found similar results when repetition errors were used as the dependent variable ($BF_{10} = 2.453e+28$). As can be seen in Figure 4, recall performance decreased across output positions, and repetition errors increased across output positions.

Discussion

We used a cued recall paradigm in which participants recalled items in random order by giving them positions as cues. We found clear output position effects on recall performance and repetition errors. In the next section, we fit a standard model of serial recall to these preliminary data.

Computational Modeling

Architecture

The computational architecture we used is a standard model of serial recall including established mechanisms in the literature. An advantage of these mechanisms is that they can be expressed in a simple mathematical form, which allows model fitting using maximum likelihood estimators. First, items are encoded via a binding mechanism that creates new item–context associations (Henson, 1998b; Oberauer et al., 2012). Thanks to these item–context associations, items can be retrieved by cueing them with their positional markers, as explained in the introduction and illustrated in Figure 1. Second, the strength of these item–context associations decreases with input position, following a primacy gradient (Page & Norris, 1998). Third, after each retrieval attempt, the WM content degrades by some constant proportion (Cowan et al., 2002; Oberauer, 2003), causing output interference. Finally, recalling an item results in its suppression (Lewandowsky, 1999). In the next paragraphs, we describe these mechanisms in a more detailed manner.

Encoding

Encoding in the model is done via a binding mechanism associating items to positional contexts. The target item is maximally encoded to its positional context. Positional contexts have overlapping representations, and this overlap decreases exponentially with rate *P* as positional distance increases. Due to positional overlap, items are partially associated with other positional markers. The association strength between the item *i* and the positional marker *j* is:

$$a_{i,j} = v_i P^{|i-j|}, \tag{1}$$

where v_i is the encoding strength for the input position *i*. The encoding strength decreases exponentially with each newly encoded item, generating a primacy gradient (Page & Norris, 1998):

$$v_i = \alpha\gamma^{i-1}. \tag{2}$$

In this formula, α is the peak of the encoding strength, fixed at 1.0 to model the fact that the first item is always maximally encoded. The γ term is a free parameter which controls the steepness of the primacy gradient. The lower the value of γ, the quicker the encoding strength decreases with each newly encoded item. To illustrate the model's behavior, Figure 5 displays the association strength of each item to all positions, using the best-fitting parameters of the first participant from our preliminary experiment.

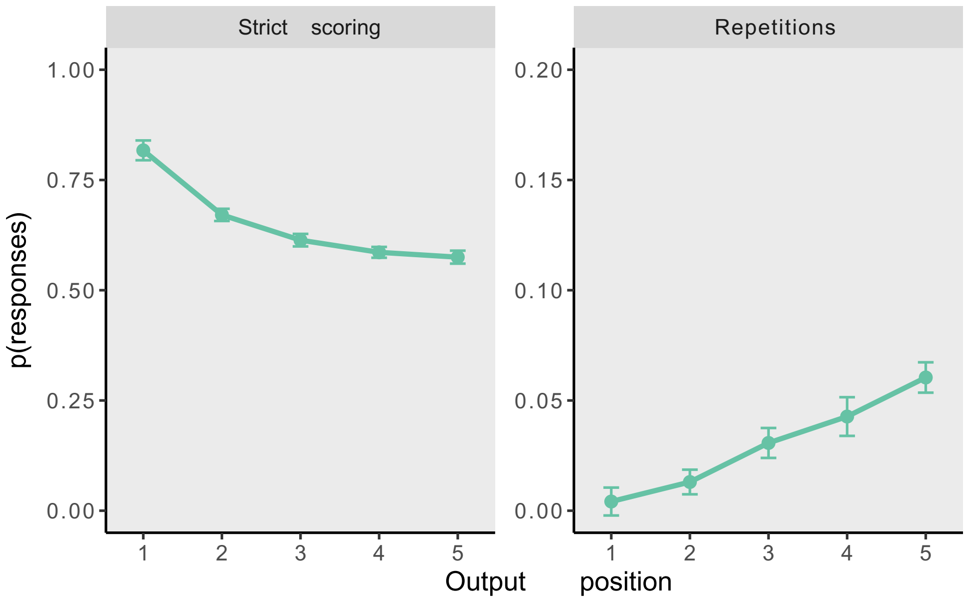
Retrieval

Retrieval is modeled as a competitive process between recall candidates. In cued recall, the recall candidates are the list items and all the nonlist items that could potentially be recalled. Activation of list item *i* for output position *k* is described as follows:

$$A_{i,k} = \beta + a_{i,j}(1.0 - \delta)^{k-1}, \tag{3}$$

where $a_{i,j}$ is the association strength between the cued position j and the item *i* associated with it, as computed in Equation 1. The second term of the formula scales the strength of the item–context association across output positions *k* via the free parameter δ, which models output interference. The term β is a free parameter that gives a constant boost of activation to all encoded items. It represents the general knowledge that an item was presented in the list. For instance, participants might have the general knowledge that the word "beach" was in the list, independently of where in the list this item was presented. The β parameter can be interpreted as sustained activation in the lexicon or semantic memory, in line with the idea that WM relies on activated long-term memory (Cowan, 1999; Oberauer, 2009). This base activation, by modulating the activation level of list items relative to nonlist items (see below), controls the proportion of extra-list intrusions produced by the model. In addition to the activation level associated with each list item, nonlist items ($A_{N+1,k}$) and

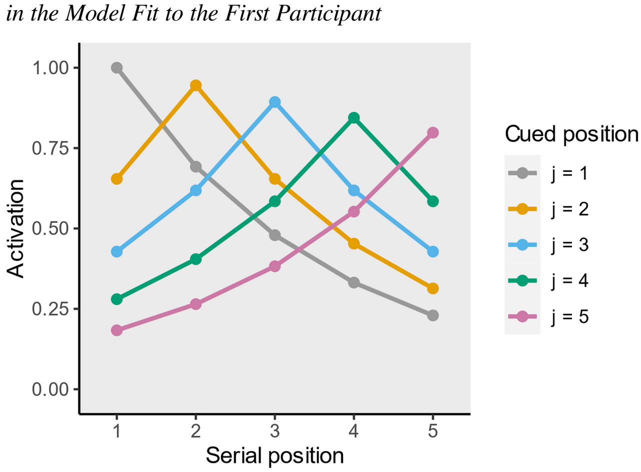
Figure 4
Results From the Pilot Experiment



Note. Left panel: Proportion of correct responses out of the total number of responses as a function of output position. Right panel: Proportion of repetition errors out of the total number of responses as a function of output position. Error bars indicate within-subject confidence intervals (Baguley, 2012). See the online article for the color version of this figure.

omissions ($A_{N+2,k}$) have their activation level, thus modeling extra-list intrusions and omission errors, respectively. Nonlist items have a base activation level fixed to 0.0, which, when combined with the exponential version of Luce's choice rule as described below, produces a nonnull probability of being recalled. Omission errors are often modeled by including an omission threshold. If an item's activation is above the omission threshold, the item can be recalled. Otherwise, the item is omitted. Here, we implement this assumption by treating omission as a retrieval candidate, which has its activation level θ.[2]

Figure 5
Illustration of the Association Strength for Each Item to All Positions in the Model Fit to the First Participant



Note. See the online article for the color version of this figure.

Recall probability of responding with the candidate i at the output position k is computed using the exponential version of Luce's choice rule:

$$p_i = \frac{\exp\left(\frac{A_{i,k}}{c}\right)}{\sum_{i=1}^{l=N+2} \exp\left(\frac{A_{i,k}}{c}\right)}, \tag{4}$$

where c is a noise parameter that controls how deterministic the selection process is. The lower the c value, the more likely highly activated candidates dominate the competition over other candidates.

Response Suppression

After an item has been recalled, its activation level is suppressed. This is done by multiplying its activation (i.e., when computing Equation 3) by a constant proportion, defined by the free parameter τ. Following a transposition error, it is not the activation of the target item which is scaled by τ, but the actually selected item. If no list item was recalled (i.e., extra-list intrusion or omission error), response suppression does not occur. The removal and inhibition mechanisms were implemented the same way, except that with inhibition, response suppression was released after a count of five retrieval attempts, which models the fact that people track how many items have been recalled so far and release their inhibition at once after the count.

[2] Mathematically speaking, this implementation is strictly equivalent to a recall threshold because an omission is selected as the response if and only if the omission threshold is higher than the activation of all other candidates. We decided to implement omissions this way for convenience.

In addition to this implementation of response suppression, we also considered the possibility that response suppression gradually wears off after each retrieval attempt. Duncan and Lewandowsky (2005) did not find any evidence that response suppression reduces over time, contrary to the initial implementation by Henson (1998a, 1998b). However, what Duncan and Lewandowsky did not consider is the possibility that the reduction of response suppression is event-based, as opposed to time-based. This additional implementation is reported in Appendix A. When we fit this model to our data, the parameter for gradual reduction of response suppression was estimated to a value at which response suppression is hardly reduced across output attempts. As a consequence, the model predictions are the same as for the model presented above. We therefore decided to keep the implementation of response suppression simple, not allowing it to wear off.

Note that when fitting the model as described below, response suppression was applied to items based on the observed responses produced by participants. For instance, if a person erroneously recalled the second item in Position 1 in a given trial, response suppression is applied to the second item in that trial, and the likelihood of all subsequent responses in that trial is evaluated on that basis. A list of all parameters is provided in Table 1.

Model Fitting

Model fitting was done for each participant using individual trials. For each recall attempt, we computed the probability p of recalling each of the recall candidates using Equation 4. The log-likelihood was then computed using the recall probability of the observed response o:

$$\log L = \log(p_o). \tag{5}$$

We used the deviance as loss function:

$$D = -2.0 \sum \log L, \tag{6}$$

where the sum operator applies to all trials and retrieval attempts.

Parameter estimation was done using the Nelder–Mead algorithm implemented in the Optim package (https://julianlsolvers.github.io/Optim.jl/stable/) of the Julia programming language (https://julialang.org/benchmarks/). To avoid that the algorithm would fall into local minima, each fitting attempt was repeated using 10 different starting points in the multidimensional parameter space. These starting points were randomly selected by sampling values from a uniform distribution. Only the set of parameters minimizing the deviance was kept.

Table 1
List of Fixed and Free Parameters of the Model

Symbol	Role	Value
P	Positional overlap	[0.0–1.0]
α	Maximum encoding strength	1.0
γ	Steepness of the primacy gradient	[0.0–1.0]
β	Base activation level	[0.0–10.0]
δ	Output interference	[0.0–1.0]
ω	Activation of the nonlist items	0.0
θ	Activation of the omission threshold	[0.0–10.0]
c	Noise parameter used during the selection rule	[0.0–1.0]
τ	Response suppression	[0.0–1.0]

Note. Fixed parameters are indicated by a single value. Free parameters are indicated by a range.

Model Assessment: Preliminary Experiment

The first necessary step is to test if response suppression is necessary at all. To do this, we compared two models. The first model was the model including response suppression as a free parameter. The second model was the same model with the response suppression parameter fixed to 1.0 implying no response suppression. As both the inhibition and removal models could not be distinguished at this stage (see below), we arbitrarily chose the removal model for model comparison. The fit of the model i was obtained using the Akaike information criterion (AIC):

$$\mathrm{AIC}_i = 2K + \sum D. \tag{7}$$

The sum runs over the deviances across all participants and $2K$ reflects a penalty term for the number of free parameters. As we fit the model separately to each participant, the number of free parameters that were estimated, K, corresponds to the number of participants times the number of free parameters of the model. Models with and without response suppression have seven and six free parameters, respectively. To obtain the relative likelihood between both models, we transformed the AICs into Akaike weights (Wagenmakers & Farrell, 2004). The difference in AIC between each model relative to the best model is first computed:

$$\Delta_i(\mathrm{AIC}) = \mathrm{AIC}_i - \min(\mathrm{AIC}). \tag{8}$$

The relative likelihood for each model is then normalized to obtain the Akaike weights $w_i(AIC)$:

$$\frac{w_i(\mathrm{AIC}) = \exp\left\{\frac{-1}{2}[\Delta_i(\mathrm{AIC})]\right\}}{\sum_{k=1}^{K}\exp\left\{\frac{-1}{2}[\Delta_i(\mathrm{AIC})]\right\}}. \tag{9}$$

Both models can then be compared using their Akaike weights:

$$\frac{w_1(\mathrm{AIC})}{w_2(\mathrm{AIC})}. \tag{10}$$

The model comparison showed that the data were more likely under the model with than the model without response suppression. Akaike weights' ratio between the two models led to a value of $4.163\mathrm{e}+97$. Therefore, the model including response suppression better accounted for participants' data.

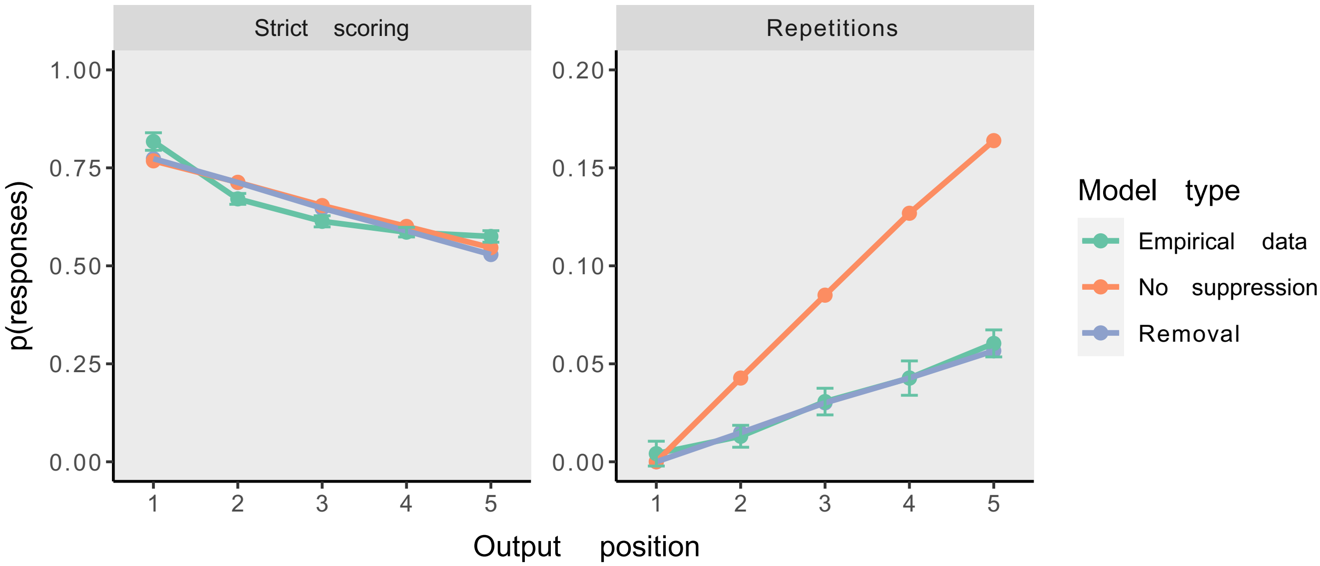
Output position curves produced by both models are plotted in Figure 6. As can be seen, the model not including response suppression produces too many repetition errors compared to what is observed. Despite this, both models make very similar predictions in terms of recall performance. Not including response suppression does not impair memory performance in this model, because repetition errors are simply a specific case of transposition errors: If response suppression prevents the model from producing a repetition error, it produces another transposition error instead, leading to similar memory performance as the model without response suppression, even if producing a different distribution of errors.

Discussion

Using a standard model of serial recall we compared model versions with and without a response suppression mechanism. The

Figure 6

Results From the Standard Model of Serial Recall



Note. Left panel: Proportion of correct responses out of the total number of responses as a function of output position. Right panel: Proportion of repetition errors out of the total number of responses as a function of output position. See the online article for the color version of this figure.

model comparison clearly favored the need for response suppression to accommodate the low rate of repetition errors.

It could be argued that this need for response suppression is specific to the standard architecture we used to fit the empirical data. The architecture we used is a model that represents serial order through position coding, that is, through binding each item to a content-independent position representation. A model not relying on position coding, the context retrieval and updating (CRU) model, has been the object of recent development (Logan, 2021; Logan & Cox, 2021; Osth & Hurlstone, 2021). The CRU does not include a response suppression mechanism. Therefore, we tested whether CRU could solve the problem of repetition errors without the need for a response suppression mechanism. We report these additional simulations in Appendix B. To summarize these simulation results, we observed that CRU produces an excess of repetition errors. We therefore conclude that preventing repetition errors is a general problem for current models of WM, regardless of the architecture considered. In the next sections, we describe two experiments that will allow us to test the removal and inhibition mechanisms.

Experiment 1

The purpose of Experiment 1 is to test the removal and inhibition mechanisms. We adapted the cued recall paradigm by including a second burst of retrieval attempts. After recalling all items once, participants' memory for the items was tested a second time. The two bursts of retrieval attempts were split across two clearly distinct retrieval phases. The order in which items were cued in the two phases was different. For instance, given the list "ABCDEF," participants could have been cued a first time in the order "25143," followed by "51324." From this manipulation, the removal and inhibition mechanisms make opposing predictions. While the former predicts that the second burst of retrieval attempts should lead to a dramatic drop in performance, the latter predicts that participants should not have difficulties recalling items a second time.

In addition to this scenario, we imagined another plausible outcome. Participants, being aware that they would have to recall items a second time, could stop removing once-recalled items from memory to maximize their performance in the second testing burst. This scenario is only valid under the assumption that removal exists. To test this, we included a control condition in which items were recalled only once. The two conditions were blocked and counterbalanced across subjects. If participants stop removing items from their WM when required to recall items twice, we should observe increased repetition errors in the first burst of retrieval attempt in this condition, as compared to the control condition in which they know that each item will be tested only once. As a consequence, a model not including any response suppression mechanism should now be favored when tested during the first burst of retrieval attempts.

Method

Power Analysis for Model Comparison

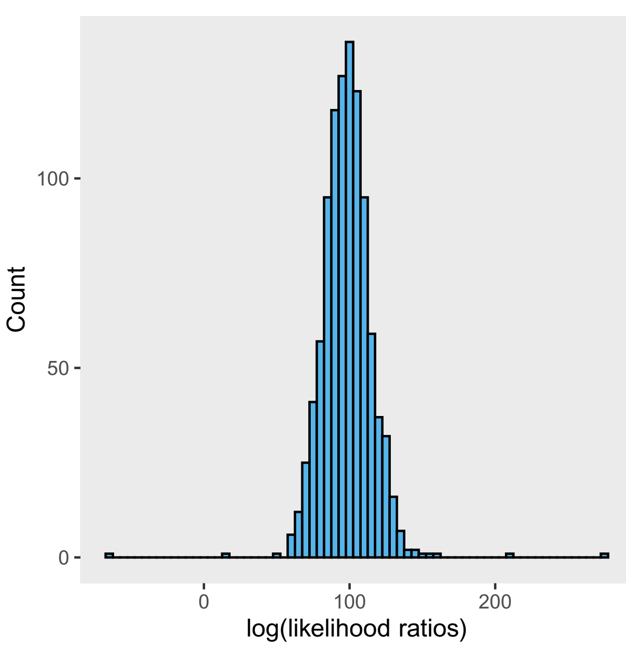
The preliminary experiment included 64 experimental trials and lasted approximately 1 hr. This preliminary experiment had only one experimental condition. Experiment 1 included two experimental conditions: One condition including one burst of retrieval attempts, and one condition including two bursts of retrieval attempts. We were concerned that doubling the total number of trials would make the experiment uncomfortable for participants. Therefore, to keep the experiment at a reasonable length, we planned to reduce the number of trials to 30 per condition, resulting in a total of 60 trials. This reduction of trials comes however with a cost in terms of statistical power, as intra-individual variability being a critical factor in within-subject designs (Smith & Little, 2018). To test to what extent this reduction of trials would have been detrimental to our modeling approach, we used a bootstrapping procedure. For each participant, we randomly selected with replacement[3] a subset of 30 trials. We then fitted both models based on this subset of trials. This fitting procedure was repeated 1,000 times. For each fit, we performed a model comparison between the model including the removal, and the model including no response suppression mechanism following the same procedure as explained previously. This resulted in a distribution of relative likelihoods between the model including the removal mechanism and the model not including it. If 30 trials were sufficient to replicate what we observed so far, we expected that the distribution of relative likelihoods between the two models should have had a systematic bias towards the model including response suppression, indicated by a value >1.0. Figure 7 displays the bootstrapping distribution of relative likelihoods between the two models on a log scale. As values on the x-axis are log-scaled, a value of 0 indicates that the two models cannot be distinguished, and a value above 0 indicates that the removal model had a higher likelihood than the model without response suppression. As can be seen, the removal model was almost always favored ($p = .998$). Therefore, reducing the number of trials to 30 per experimental condition was sufficient to replicate what we observed so far.

Participants

We recruited forty young adults aged between 18 and 35 years on the online platform Prolific. Despite the previous section suggesting that the study would have already been sufficiently powered with $N = 30$, we nonetheless increased the sample size to 40 participants

Figure 7

Bootstrapping Distribution of Relative Likelihoods Between the Model With and Without Response Suppression



Note. Values of the relative likelihood were log-transformed for better visualization. See the online article for the color version of this figure.

to compensate for the reduction of trials. In case a BF for a theoretically important effect is between 1/5 and 5, we planned to continue increasing N until we reach BF $<$ 1/5 or BF $>$ 5, or until we reach a maximum of $N = 80$.

Material

The lists were generated the same way as the preliminary experiment, with the exception that only 30 lists were generated. Those 30 lists were used in the condition including two bursts of retrieval attempts. The lists were then re-used in the condition including one burst of retrieval attempt, except that their presentation order was reversed from [1, 2, 3, 4, 5] to [5, 4, 3, 2, 1] to neutralize potential learning effects while keeping the structure of the whole sequence equivalent. This way of generating the lists means that each word was presented twice across the whole experiment: once in each experimental condition.

The recall orders were generated using the following procedure. We first generated two bursts of retrieval attempts using the same constraints as those used in the preliminary experiment, except that (a) the two bursts of retrieval attempts were always different from each other (e.g., recall orders such as "15423–15423" were not allowed), and (b) the same position was never probed twice in a row, including when transitioning between the first and second bursts (e.g., recall orders such as "42531–13245" were not allowed). For each participant, we generated 30 of those recall orders, which were directly used in the condition including two bursts of retrieval attempts. Recall order in the condition including only one burst of retrieval attempt was identical to the condition including two bursts, except that participants were never tested twice. This way of building the recall directions ensures that lists are matched in terms of retrieval difficulty.

Procedure

The procedure was identical to the one used in the preliminary experiment, with the exception that participants performed the two experimental conditions (i.e., involving one or two bursts of retrieval attempts) in separate blocks. Participants were informed before each block about the recall condition.

In the condition involving two bursts of retrieval attempts, participants were not warned that the second burst started. Participants were presented with the first burst, followed directly by the second burst without any interruption.

Behavioral Section

This section reports behavioral analyses. The range of theoretical questions that can be tested using solely the behavioral data is limited. The modeling approach fills this gap.

We first assessed recall performance (using the strict scoring criterion) as a function of output position (1 through 10) and test occasion (i.e., first test vs. second test), separately the condition including two retrieval attempts using a Bayesian logistic regression model. We included output position and test occasion as fixed and random

[3] Bootstrapping with replacement increases the variance compared to sampling without replacement. We chose this option because it is more conservative.

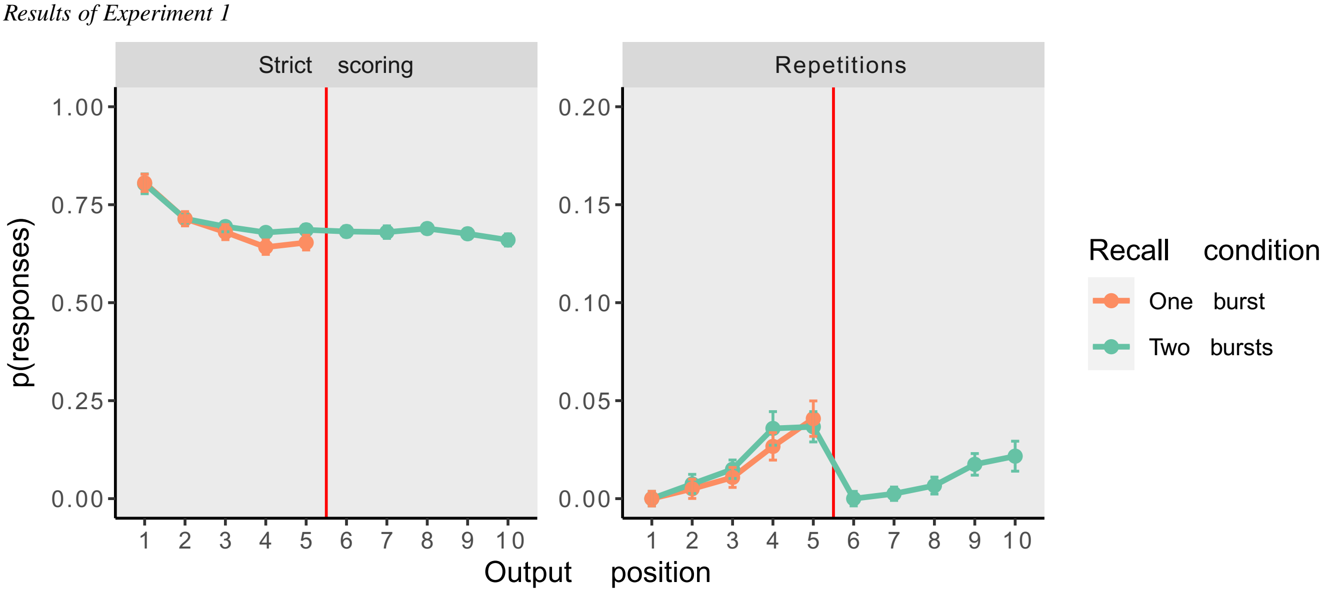
effects, as well as random intercepts. We expect a main effect of output position, as demonstrated many times. In addition, if people remove items after recalling them, we should expect a main effect of test occasion, with items tested a second time being associated with poorer recall performance, after taking output position as regressor. This analysis showed that there were two best models: The model including both the main effects and the interaction, and the model including only the main effect of the test occasion and the interaction. Comparison between these models led to ambiguous evidence ($BF_{10} = 1.003$). We selected the model without the main effect of output interference for further comparisons to keep the analyses simple. Additional comparison showed decisive evidence supporting the interaction term ($BF_{10} = 1.354e + 10$). Further exploration of this interaction using additional Bayesian logistic regression models showed that there was decisive evidence in favor of an effect of output position across Positions 1 through 5 ($BF_{10} = 6.033e + 10$) and moderate evidence against it across Positions 6 through 10 ($BF_{01} = 7.9$). Finally, we found strong evidence supporting the effect of test occasion ($BF_{10} = 21.407$). This main effect of test occasion comes from the fact that recall performance in the second bursts of retrieval attempt was higher compared to what should be expected under a linearly decreasing effect of output position. These results are illustrated in Figure 8, where we plot recall performance (left panel) and repetition errors (right panel) as a function of output position and recall condition. As can be seen, recall performance did not substantially decrease in the second relative to the first test.

Next, we test the possibility that people stop removing once they are asked to recall items twice, in order to avoid the steep drop in performance that removal would cause for the second recall burst (see Figure 9, modeling part). According to this hypothesis, there should be fewer repetition errors across output Positions 1 through 5 in the condition including one burst of retrieval attempts, as compared to the condition including two bursts of retrieval attempts. We tested this using a Bayesian logistic regression model with output position (1–5), recall condition (recalled once vs. recalled twice), and their interaction as fixed effects, with the proportion of repetition errors as the dependent variable. The random effect of output position and recall type was also included in the analysis, as well as a random intercept. The best model included the main effect of output position only. As compared to this model, there was ambiguous evidence against the effect of the recall condition ($BF_{01} = 1.8$), and moderate evidence against the interaction term ($BF_{01} = 9.848$). These results are also reflected in Figure 8, right panel.

Model Assessment: Experiment 1

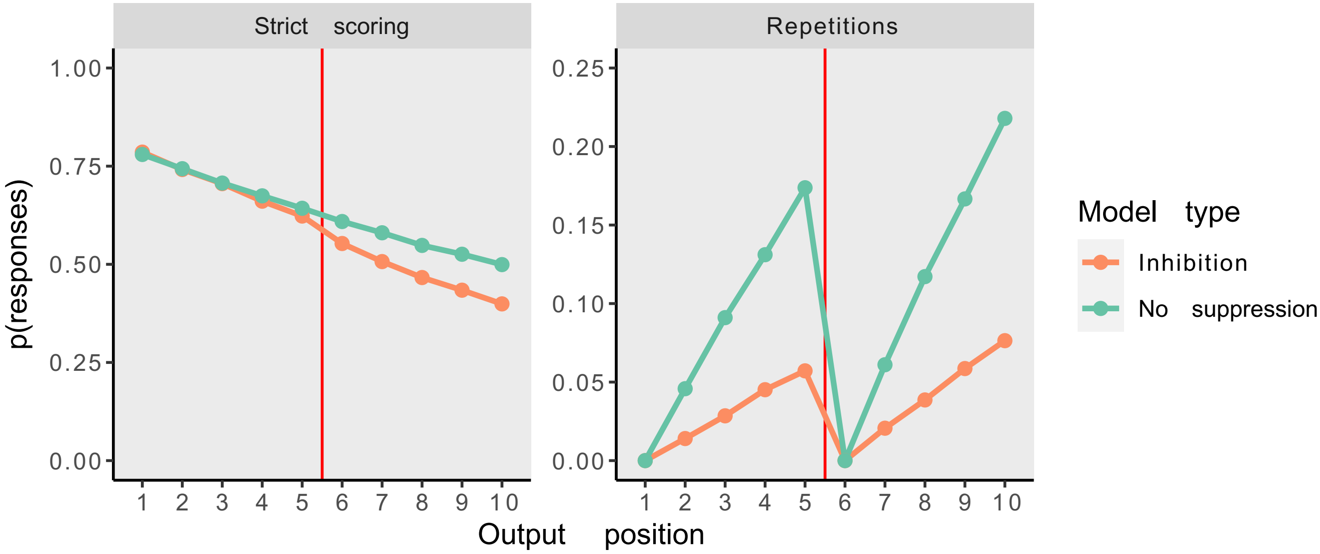
The three models as described above were fitted to the data of Experiment 1. Each participant was fitted individually to both recall conditions using the same parameter values, except for the response suppression parameter which we varied depending on the model. We identified four models to compare. The first model is the NoRS (i.e., no response suppression) model, for which response suppression was fixed to 1.0. The second model is the removal model, for which one response suppression parameter was fitted across both recall conditions. We labeled this model Removal1 (i.e., removal with one response suppression parameter). The third model is a variation of the removal model, in which we used different response suppression parameters for the conditions involving one versus two bursts of retrieval attempts. We labeled this model Removal2 (i.e., removal with two response suppression parameters). The inclusion of this second model informs us whether participants adopted a different way to remove items depending on the recall requirements (i.e., recalling items once or twice). The last model is the inhibition model, in which we used only one response suppression value across both recall conditions. All models were compared to each other after computing the Akaike weight as described previously. All models along with

Figure 8
Results of Experiment 1



Note. Left panel: Proportion of correct responses out of the total number of responses as a function of output position. Right panel: Proportion of repetition errors out of the total number of responses as a function of output position. The red vertical line indicates the separation between the first and second burst of retrieval attempts. Error bars indicate within-subject confidence intervals (Baguley, 2012). See the online article for the color version of this figure.

Figure 9
Predictions From the Inhibition and NoRS Model, Experiment 1



Note. The data were fitted using the first five output positions. The red vertical line separates the two bursts of retrieval attempts. NoRS = model without response suppression. See the online article for the color version of this figure.

their number of free parameters, AIC values, and Akaike weight values are listed in Table 2.

We fitted the models to the first bursts of retrieval attempts of both conditions and used the best-fitting parameters estimated in that way to generate predictions for the second burst of retrievals in the condition that had a second burst. We computed the AIC and Akaike weights of each model when tested against the entirety of the data set, including the second burst of retrieval attempts. Using this approach, we found that the best-fitting model was the inhibition model (see Table 2). The model comparison showed that the data were more likely under the inhibition model than under the model without response suppression. The Aikaike weights' ratio between the inhibition and NoRS model approached infinity. Predictions from both models are displayed in Figure 9. The inhibition model fits best for the same reasons as in the pilot experiment: Without a way to exclude the items that have already been recalled, the NoRS model necessarily produces an excess of repetition errors.

Next, we compared the inhibition model to both removal models. Results indicate that the data were more likely under the inhibition model than the model including one removal parameter (i.e., Removal1). Likewise, the data were more likely under the inhibition model than the model including two removal parameters (i.e., Removal2). Both comparisons led to Akaike weight ratios approaching infinity. Predictions from both removal models are displayed in Figure 10. As can be seen, both models predict a drop in performance during the second burst of retrieval attempts, which is not observed in the empirical data.

Table 2
List of Models, Along With Their Number of Free Parameters and Associated AIC and Akaike Weight Values

Models	Number of parameters	AIC_i	W_i(AIC)
NoRS	6	41,230	→0
Removal1	6 + 1	50,509	→0
Removal2	6 + 2	49,806	→0
Inhibition	6 + 1	39,806	→1

Note. AIC = Akaike information criterion; NoRS = model without response suppression; Removal1 = model with one response suppression parameter for all conditions; Removal2 = model with one response suppression parameter for each recall condition; Inhibition = model with global inhibition as response suppression.

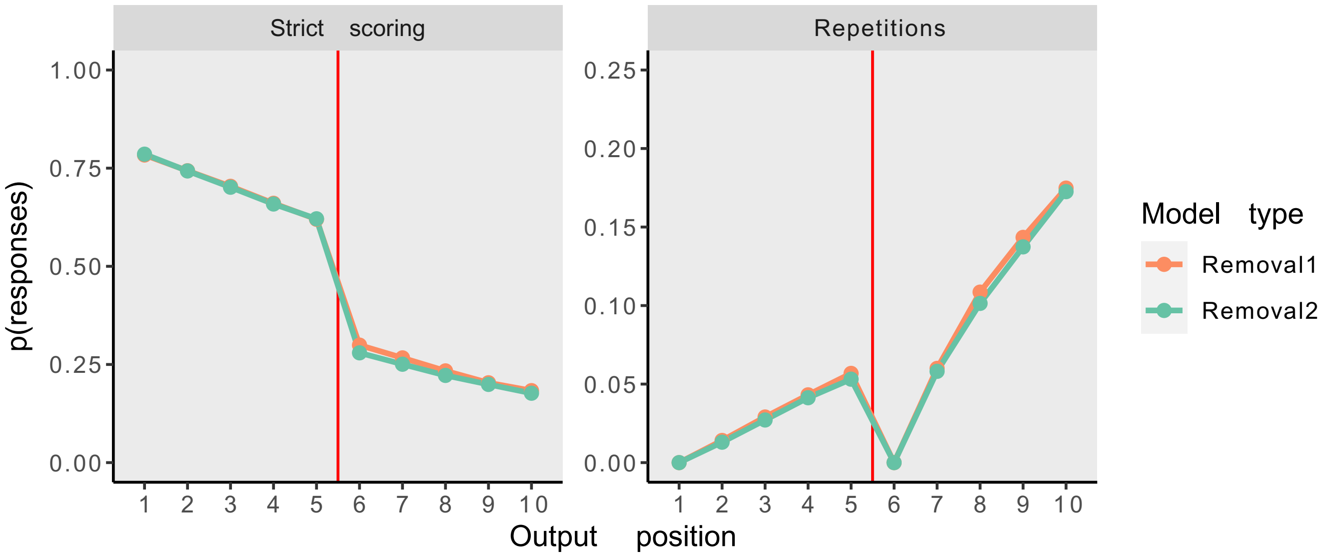
Discussion

Results from this first experiment show no decline in recall performance during the second burst of retrieval attempts. This pattern of results was predicted only by the inhibition model and a model including no response suppression. The model comparison clearly favored the inhibition model, as it is the only model preventing an excess of repetition errors. It is worth noting that, in the empirical data, output interference was virtually absent during the second burst of retrieval attempts, as shown in Figure 8. In contrast to the data, all models predict that output interference should still affect items during the second burst of retrieval attempts. Experiment 2 tests the inhibition account more directly, by using an experimental setup for which inhibition has less flexibility about what to predict than for Experiment 1.

Experiment 2

Experiment 2 was similar to Experiment 1, except that the two bursts of retrieval attempts were mixed with each other instead of following one after the other. This mixing should prevent people from releasing their response inhibition for all items after the first burst of retrieval attempts, as there is never a point in the output sequence at which all items have been tested once, and none have been tested twice. With the mixing of first and second retrieval bursts, we assume that

Figure 10
Predictions From the Removal1 and Removal2 Models, Experiment 1



Note. The data were fitted on the first five output positions. The red vertical line separates the two bursts of retrieval attempts. Removal1 = model with one response suppression parameter for all conditions; Removal2 = model with one response suppression parameter for each recall condition. See the online article for the color version of this figure.

inhibition is never released. Consequently, models including the removal and inhibition mechanisms make identical predictions. Both predict that people will have difficulties recalling an item a second time.

The design of Experiment 2 allows us to clearly disentangle the effect of response suppression from output interference at the behavioral level, by matching the output position of items recalled the first time with the output position of items recalled a second time. For instance, by comparing output sequences beginning with ABCDC and ABCBD, we can compare the accuracy in output Positions 4 and 5. In each of these positions, one sequence tests an item for the first time, whereas the other sequence tests an item for the second time. Response suppression entails that the item tested for a second time is less likely to be recalled correctly.

If we find equivalent recall for items tested first and items tested second, that would be incompatible with the predictions of any known implementation of response suppression in models of WM. If this this combination of null effects turns out to be observed, it simply means that there is something fundamentally wrong with the way we prevented repetition errors in our models so far.

Method

Power Analysis

To keep the experiment at a reasonable length, we planned to reduce the total number of trials from 60 to 50. To test if this would be sufficient to observe an effect of response suppression, we first fitted a Bayesian logistic regression model using the WM model's predictions reported in the next section as a data set. The WM model's parameters were those obtained when fitting the model to the preliminary data. We then used the predict() function available in the brms package to generate a new set of data using the fitted logistic regression model. We considered a total of 40 participants, with 50 experimental trials per participant. Finally, we fitted a new set of logistic regression models to these simulated data. The model comparison showed decisive evidence supporting the main effect of test occasion (first vs. second test) on memory performance ($BF_{10} = 2.053e + 102$). We were therefore confident that this experiment was sufficiently powered to observe the effect of test occasion that is predicted by both mechanisms of response suppression.

Participants

We recruited forty young adults aged between 18 and 35 years on the online platform Prolific. In case a BF for a theoretically important effect is between 1/5 and 5, we planned to continue increasing N until we reach BF < 1/5 or BF > 5, or until we reach a maximum of $N = 80$.

Material

The lists were generated the same way as the preliminary experiment. We included a total of 50 lists to keep the experiment at a reasonable length, as described above. Recall order was defined by random sampling without replacement from the vector [1, 2, 3, 4, 5, 1, 2, 3, 4, 5], with the further constraint that a given item could not be cued with its positions two consecutive times. This way of constructing recall order allowed us to compare memory performance for items recalled once and twice but sharing the same output position.

Procedure

The procedure was identical to the one described in Experiment 1. Participants were told during the instructions that they would have to recall items twice. There was only one experimental condition, in which participants recalled each item twice in a random order, following the constraints as explained in the Material section.

Behavioral Section

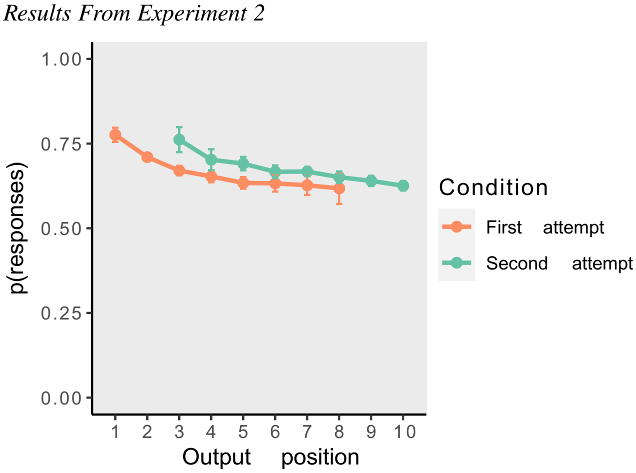
We assessed recall performance as a function of output position (1 through 10) and test occasion (first, second) using a Bayesian logistic mixed-effects model. If a form of response suppression exists, we should find the main effect of the test occasion, with the second test resulting in poorer recall compared to the first one. If response suppression does not exist, we should find evidence against the main effect of the test occasion. This analysis showed that the best model was the model including the interaction term only. As compared to this model, we found moderate evidence against the effect of test occasion ($BF_{01} = 3.385$). These results, displayed in Figure 11, show that there was no decline whatsoever for items recalled a second time. If anything, these items were better recalled as compared to the items recalled for the first time, matched for output position.

Model Assessment: Experiment 2

We performed a model comparison using the same logic as in Experiment 1. We identified two models: one model without response suppression, and one model with response suppression, labeled NoRS and RS, respectively. We compared the two models through their Akaike weights as described previously. The two models with their number of free parameters, as well as AIC and Akaike weight values are listed in Table 3.

We fitted each model to the data from the first recall attempt of each item and used the best-fitting parameters to compute predictions for the second recall attempt of each item. Akaike weight values were then computed by testing the model against the entire data set. Using this approach, we found that the data were more likely under the NoRS model than under the removal and inhibition models. The ratio between their Akaike weights approached infinity. Predictions from these models are displayed in Figure 12. As can be seen, the NoRS model was favored because it does not predict a substantial drop in performance for items tested a second time, contrary to the removal and inhibition models.

Figure 11
Results From Experiment 2



Note. The proportion of correct responses out of the total number of responses as a function of the output position. Error bars indicate within-subject confidence intervals (Baguley, 2012). See the online article for the color version of this figure.

Table 3
List of Models, Along With Their Number of Free Parameters and Associated AIC and Akaike Weight Values

Models	Number of parameters	AIC_i	W_i(AIC)
NoRS	6	37,178	→1
RS	6 + 1	52,212	→0

Note. AIC = Akaike information criterion; NoRS = model without response suppression; RS = any model including a response suppression mechanism (i.e., inhibition or removal).

Discussion

Results from Experiment 2 showed no decline in recall performance for items tested a second time. This pattern of results contradicts the prediction from response suppression mechanisms. This is the first experiment in which a model not including a response suppression mechanism was clearly favored as compared to models including a response suppression mechanism.

General Discussion

In WM tasks, participants rarely repeat items they have already recalled. Models of serial recall typically use response suppression as a way to prevent an excess of repetition errors (Henson, 1998b; Lewandowsky, 1999). Through response suppression, items are removed from memory (Lewandowsky, 1999; Oberauer & Lewandowsky, 2011; Oberauer et al., 2012) or inhibited (Farrell & Lewandowsky, 2004; Hurlstone & Hitch, 2015) once recalled. This leads to the prediction that the second test of an item should yield impaired memory performance. Contrary to this hypothesis, we observed that people can recall the same item twice without loss. Therefore, this study rejects response suppression as a plausible mechanism.

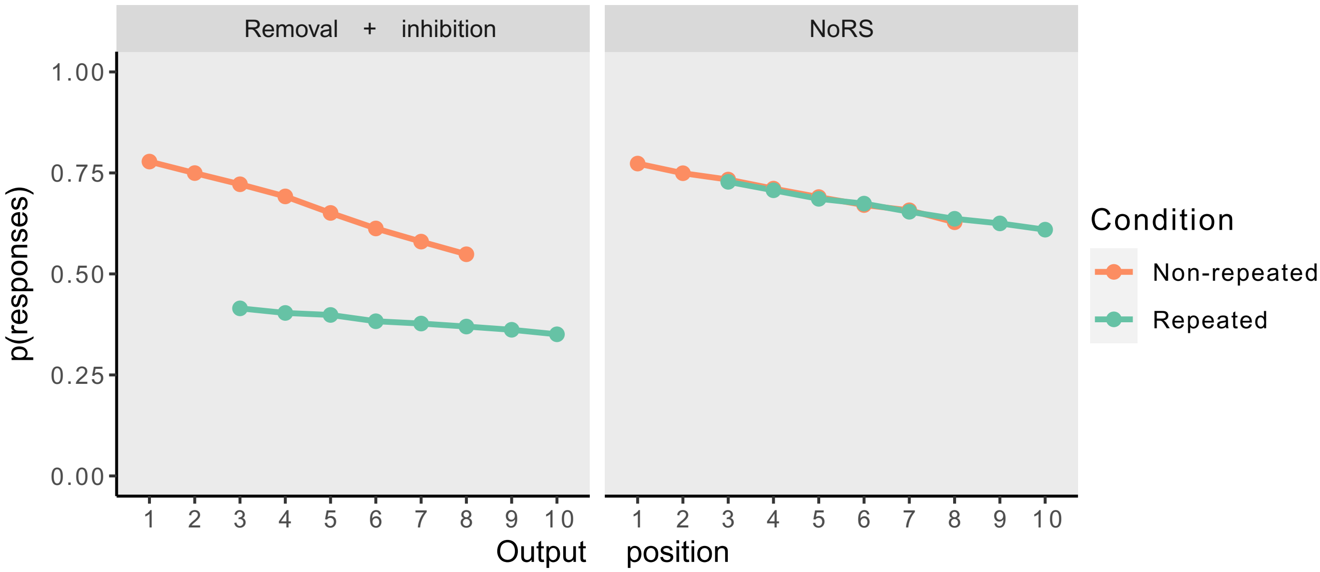
Should Response Suppression Be Completely Abandoned in Serial-Recall Models?

The results of this study strongly suggest that response suppression is not the appropriate mechanism to prevent the presence of repetition errors in models of serial recall. However, its exclusion leaves us without a way to prevent repetition errors. We identify two potential solutions to address this problem. The first solution is to find an alternative mechanism to response suppression which would allow already recalled items to be discarded from the set of response options, while remaining in memory. In the next section, we discuss ideas for how that could work.

In the absence of such an alternative, a second solution is to discard already recalled items from the competition in an ad-hoc fashion. This can be done by reducing the item's activation to zero or some other low value at retrieval, just as the way it is currently done with response suppression. By doing so, researchers should keep in mind that this should be a temporary solution for a future, yet unknown mechanism. Such a temporary solution would come in handy when working with tasks in which items must be retrieved only once. However, as soon as one implements a design involving multiple testing occasions, this temporary solution would not work anymore, because it would lead to impaired recall performance for the second test occasion. In such a design, the recommendation would be to not include any form of

Figure 12

Predictions From the Removal, Inhibition, and NoRS Models, Experiment 2



Note. The data were fitted from the first retrieval attempt of each item. NoRS = model without response suppression. See the online article for the color version of this figure.

downgrading of recalled items. This would however come at the cost of predicting an excess of repetition errors. Finding a new mechanism with these aforementioned constraints is therefore needed, otherwise WM models would be limited to explain people's behavior in a restricted set of experimental paradigms.

Implications for the Ranschburg Effect

People have difficulties reporting the second occurrence of an item, a phenomenon termed the Ranschburg effect (Harris & Jahnke, 1972). Depending on the assumptions included in models of serial recall, response suppression can predict a Ranschburg effect, because the suppression of an item hinders its subsequent recall (Henson, 1998a). In the absence of evidence supporting response suppression, it is difficult to conceive that the Ranschburg effect is to be explained by response suppression. In addition to this, some evidence shows that people can recall multiple occurrences of the same item in a list without compromising overall performance (Cowan & Hardman, 2021), a result which contrasts with the Ranschburg effect itself. One way to better understand these contradictions would be to identify the boundary conditions in which a Ranschburg effect can or cannot be observed. All in all, the present results, along with other empirical evidence, suggest that response inhibition is hardly a viable option to incorporate in models of serial recall.

Alternatives to Response Suppression

When starting this project, we considered an item-specific inhibition mechanism, rather than one relying on a global release of inhibition. In this mechanism, an item is inhibited once recalled. When tested a second time, the item's inhibition is released. This hypothetical mechanism would allow people to recall items twice in every situation. This idea, although appealing, has one major flaw. As inhibition hinders people's ability to retrieve an item, how can they know which item they should release inhibition for? Since the item is inhibited, and therefore inaccessible to memory, we find it difficult to imagine that people know which item they should release inhibition for. Unless a good solution for this contradiction is provided, we find such a mechanism too implausible to be considered.

If items are neither inhibited nor removed from memory after being recalled, how can we explain the fact that people rarely repeat an item they already recalled? One mechanism we envision involves participants remembering recall events. Following the initial recall, this memory record could be used during subsequent recall to check whether a retrieved item is among those already recalled before, and if it is, reject it before overtly recalling it. In scenarios requiring a second retrieval attempt, this record could be used to identify whether the list position currently tested is among those that have already been tested before. If so, that would be a reason not to reject an item even if that item has already been recalled before.

The assumption that people can hold a memory of a recall event does not currently exist in serial-recall models and warrants future investigations. One important question to address is how people could maintain such recall events without overloading WM. One plausible explanation involves using episodic long-term memory to store this record. Many studies have shown the role of episodic memory during some WM tasks (e.g., Rose et al., 2014). If people use episodic long-term memory to maintain some of the memoranda during WM tasks, it is also possible that they could use it as a way to store and retrieve past recall events. However, a recent series of experiments provides evidence against the involvement of episodic long-term memory in immediate serial recall tasks (Oberauer & Bartsch, 2023). An alternative way to keep a record of which items have already been recalled could be to associate a "recalled" representation with the position marker of items already recalled. Every time an already recalled item is accidentally retrieved again, the "recalled" tag would likely be retrieved with it, providing the necessary information to a decision process to either reject that item (in a standard serial-recall task) or accept it (in a situation where a second recall demand is expected).

Conclusion

We tested response suppression in list-recall experiments involving multiple retrieval attempts. In contrast to the predictions following from response suppression, participants can recall items more than once without loss. Therefore, an alternative to response suppression is required. One possible alternative is that people store memories of their recall events, and decide whether to accept or reject a retrieved item based on these memories.

References

Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for ANOVA. *Behavior Research Methods*, *44*(1), 158–175. https://doi.org/10.3758/s13428-011-0123-7

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*(3), 551–581. https://doi.org/10.1037/0033-295X.106.3.551

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, *55*(4), 627–652. https://doi.org/10.1016/j.jml.2006.08.005

Bürkner, P.-C. (2017). Brms: An R Package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge University Press.

Cowan, N., & Elliott, E. M. (2023). Deconfounding serial recall: Response timing and the overarching role of grouping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(2), 249–268. https://doi.org/10.1037/xlm0001157

Cowan, N., & Hardman, K. O. (2021). Immediate recall of serial numbers with or without multiple item repetitions. *Memory*, *29*(6), 744–761. https://doi.org/10.1080/09658211.2021.1942920

Cowan, N., Saults, J. S., Elliott, E. M., & Moreno, M. V. (2002). Deconfounding serial recall. *Journal of Memory and Language*, *46*(1), 153–177. https://doi.org/10.1006/jmla.2001.2805

Dames, H., & Oberauer, K. (2022). Directed forgetting in working memory. *Journal of Experimental Psychology: General*, *151*(12), 2990–3008. https://doi.org/10.1037/xge0001256

Duncan, M., & Lewandowsky, S. (2005). The time course of response suppression: No evidence for a gradual release from inhibition. *Memory*, *13*(3–4), 236–246. https://doi.org/10.1080/09658210344000233

Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review*, *119*(2), 223–271. https://doi.org/10.1037/a0027371

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review*, *9*(1), 59–79. https://doi.org/10.3758/BF03196257

Farrell, S., & Lewandowsky, S. (2004). Modelling transposition latencies: Constraints for theories of serial order memory. *Journal of Memory and Language*, *51*(1), 115–135. https://doi.org/10.1016/j.jml.2004.03.007

Farrell, S., & Lewandowsky, S. (2012). Response suppression contributes to recency in serial recall. *Memory & Cognition*, *40*(7), 1070–1080. https://doi.org/10.3758/s13421-012-0212-6

Farrell, S., Wise, V., & Lelièvre, A. (2011). Relations between timing, position, and grouping in short-term memory. *Memory & Cognition*, *39*(4), 573–587. https://doi.org/10.3758/s13421-010-0053-0

Greene, R. L. (1991). The Ranschburg effect: The role of guessing strategies. *Memory & Cognition*, *19*(3), 313–317. https://doi.org/10.3758/BF03211155

Harris, C. J., & Jahnke, J. C. (1972). Effect of partial recall on the Ranschburg phenomenon. *Journal of Experimental Psychology*, *93*(1), 118–122. https://doi.org/10.1037/h0032491

Henson, R. N. A. (1998a). Item repetition in short-term memory: Ranschburg repeated. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(5), 1162–1181. https://doi.org/10.1037/0278-7393.24.5.1162

Henson, R. N. A. (1998b). Short-term memory for serial order: The start-end model. *Cognitive Psychology*, *36*(2), 73–137. https://doi.org/10.1006/cogp.1998.0685

Hurlstone, M. J., & Hitch, G. J. (2015). How is the serial order of a spatial sequence represented? Insights from transposition latencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 295–324. https://doi.org/10.1037/a0038223

Jahnke, J. C. (1970). Probed recall of strings that contain repeated elements. *Journal of Verbal Learning and Verbal Behavior*, *9*(4), 450–455. https://doi.org/10.1016/S0022-5371(70)80087-1

Kowialiewski, B., Krasnoff, J., Mizrak, E., & Oberauer, K. (2022). The semantic relatedness effect in serial recall: Deconfounding encoding and recall order. *Journal of Memory and Language*, *127*, Article 104377. https://doi.org/10.1016/j.jml.2022.104377

Kowialiewski, B., Krasnoff, J., Mizrak, E., & Oberauer, K. (2023). Verbal working memory encodes phonological and semantic information differently. *Cognition*, *233*, Article 105364. https://doi.org/10.1016/j.cognition.2022.105364

Kowialiewski, B., & Oberauer, K. (2023). *Response suppression* [Data set]. https://osf.io/r2avt

Lewandowsky, S. (1999). Redintegration and response suppression in serial recall: A dynamic network model. *International Journal of Psychology*, *34*(5–6), 434–446. https://doi.org/10.1080/002075999399792

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 49, pp. 1–48). Elsevier. https://doi.org/10.1016/S0079-7421(08)00001-7

Logan, G. D. (2021). Serial order in perception, memory, and action. *Psychological Review*, *128*(1), 1–44. https://doi.org/10.1037/rev0000253

Logan, G. D., & Cox, G. E. (2021). Serial memory: Putting chains and position codes in context. *Psychological Review*, *128*(6), 1197–1205. https://doi.org/10.1037/rev0000327

Maylor, E. A., & Henson, R. N. A. (2000). Aging and the Ranschburg effect: No evidence of reduced response suppression in old age. *Psychology and Aging*, *15*(4), 657–670. https://doi.org/10.1037/0882-7974.15.4.657

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251–269. https://doi.org/10.3758/BF03213879

Oberauer, K. (2003). Understanding serial position curves in short-term recognition and recall. *Journal of Memory and Language*, *49*(4), 469–483. https://doi.org/10.1016/S0749-596X(03)00080-9

Oberauer, K. (2009). Design for a working memory. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 51, pp. 45–100). Elsevier. https://doi.org/10.1016/S0079-7421(09)51002-X

Oberauer, K., & Bartsch, L. M. (2023). When does episodic memory contribute to performance in tests of working memory? *Journal of Cognition*, *6*(1), Article 44. https://doi.org/10.5334/joc.311

Oberauer, K., & Lewandowsky, S. (2011). Modeling working memory: A computational implementation of the Time-Based Resource-Sharing theory. *Psychonomic Bulletin & Review*, *18*(1), 10–45. https://doi.org/10.3758/s13423-010-0020-6

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, *19*(5), 779–819. https://doi.org/10.3758/s13423-012-0272-4

Osth, A. F., & Hurlstone, M. J. (2021). *Do item-dependent context representations underlie serial order in cognition? Commentary on Logan (2021)* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/g3qp2
Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*(4), 761–781. https://doi.org/10.1037/0033-295X.105.4.761-781
Roodenrys, S., Miller, L. M., & Josifovski, N. (2022). Phonemic interference in short-term memory contributes to forgetting but is not due to overwriting. *Journal of Memory and Language*, *122*, Article 104301. https://doi.org/10.1016/j.jml.2021.104301
Rose, N. S., Buchsbaum, B. R., & Craik, F. I. M. (2014). Short-term retention of a single word relies on retrieval from long-term memory when both rehearsal and refreshing are disrupted. *Memory & Cognition*, *42*(5), 689–700. https://doi.org/10.3758/s13421-014-0398-x
Schneegans, S., & Bays, P. M. (2017). Neural architecture for feature binding in visual working memory. *The Journal of Neuroscience*, *37*(14), 3913–3925. https://doi.org/10.1523/JNEUROSCI.3493-16.2017
Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design. *Psychonomic Bulletin & Review*, *25*(6), 2083–2101. https://doi.org/10.3758/s13423-018-1451-8
Wagenmakers, E.-J., & Farrell, S. (2004). AIC Model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196. https://doi.org/10.3758/BF03206482

Appendix A

Recovery of the Response Suppression Parameter

We implemented an additional mechanism in which response suppression was modulated in the inhibition model by another free parameter, *R*, which progressively reduces the response suppression value *r* for the item *i* after each retrieval attempt *k* following the initial response suppression of that item:

$$\Delta r_i = \tau R^k. \tag{A1}$$

The *R* parameter ranges between 0.0 and 1.0. When $R = .0$, response suppression is removed entirely with the next retrieval attempt. When $R = 1.0$, response suppression is never reduced. We fitted the model including this new mechanism to our pilot experiment. The distribution of parameter values across individual participants indicates that most participants had very large values of *R*, as can be seen in Figure A1.

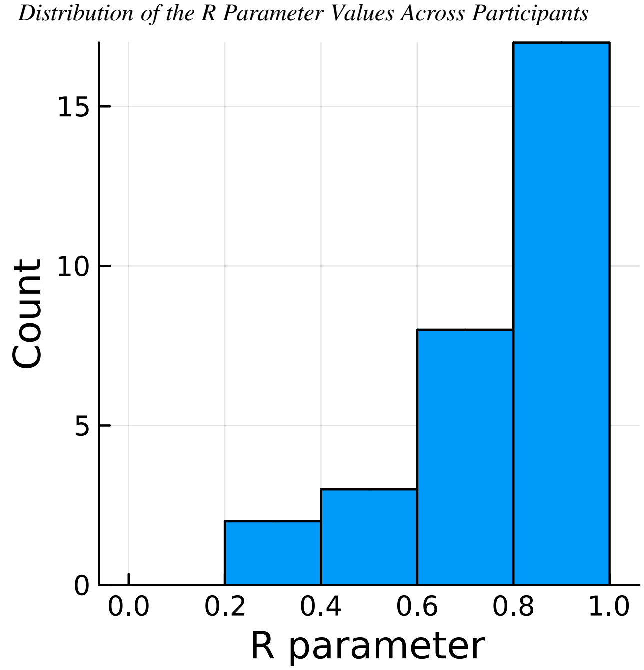
Using this newly fitted model, we generated predictions for the two experiments we planned to conduct. Generating predictions for Experiment 1 gave the following results (Figure A2).

As can be seen, introducing the possibility that response suppression wears off did not change the overall qualitative pattern of predictions, with a dramatic drop of performance in the second burst of retrieval attempt, just as in the model including the constant response suppression mechanism (i.e., the "removal model"). The same pattern was observed when generating predictions for Experiment 2 (Figure A3).

These new simulations indicate that a gradual recovery of response suppression is unlikely to change the conclusions from our planned experiments.

Figure A1

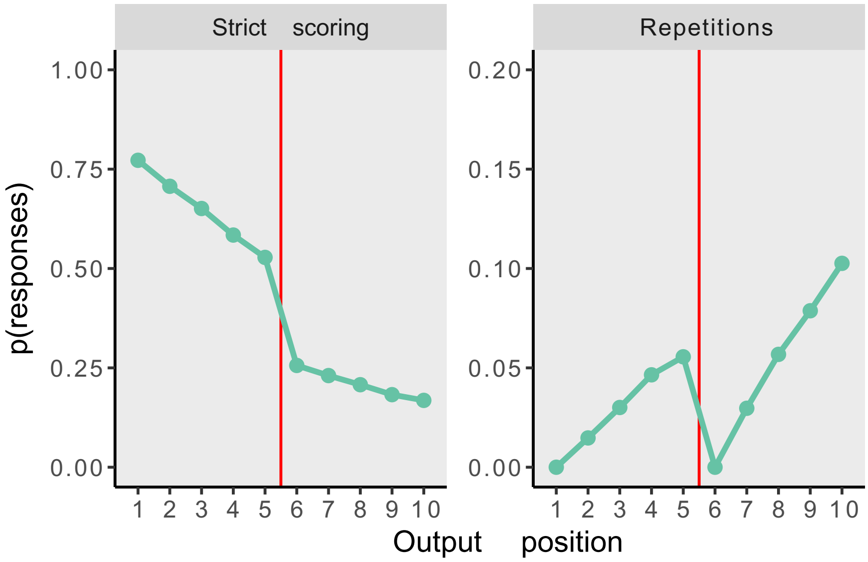
Distribution of the R Parameter Values Across Participants



Note. The model was fit on the data from the pilot experiment. See the online article for the color version of this figure.

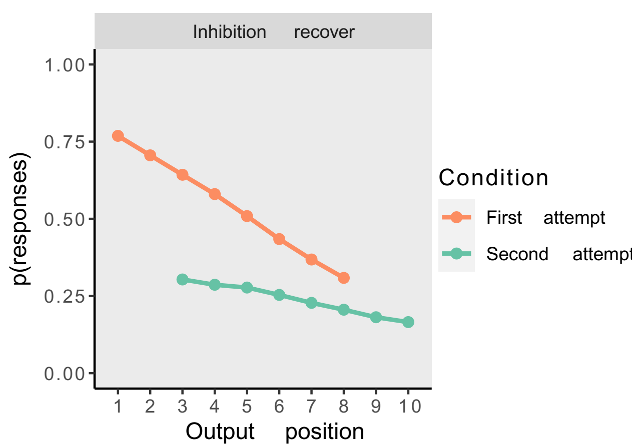
(*Appendices continue*)

Figure A2
Model Results



Note. The model was fit on the data from the pilot experiment. Predictions were then generated based on the Experimental design planned for Experiment 1. See the online article for the color version of this figure.

Figure A3
Model Results



Note. The model was fit on the data from the pilot experiment. Predictions were then generated based on the Experimental design planned for Experiment 2. See the online article for the color version of this figure.

(*Appendices continue*)

Appendix B

CRU Model

In these simulations, we tested whether the CRU model could prevent an excess of repetition errors without a response suppression mechanism. The interested reader can access the mathematical details of this architecture in Logan (2021).

After closer examination of CRU's properties, we concluded that CRU requires a form of response suppression. During the comparison process between the current context C_c and the stored context C_s, CRU includes the previously recalled contexts as part of the retrieval candidates. Without a way to discard the previously recalled contexts, CRU will necessarily produce an excess of repetition errors compared to what is observed in the empirical data. Consider a list to be remembered as "ABC" in an immediate serial recall paradigm. After recalling "A," the model attempts to retrieve the second item. Given that "A" has been correctly recalled, and $\beta = .6$, this gives us the following dot product for each value i in v (note that a similar figure can be found in Logan, 2021, Figure 5; Figure B1):

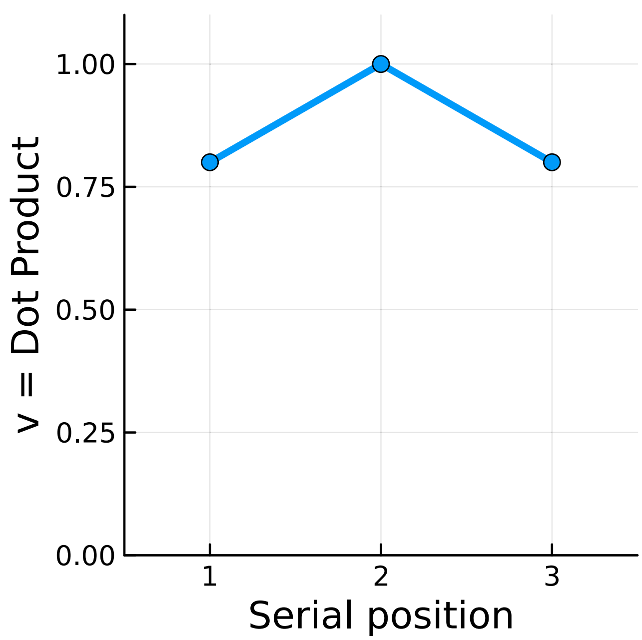
Therefore, when trying to retrieve "B," the model is equally likely to erroneously retrieve either "A" or "C." This property is precisely the same problem that we illustrated in Figure 1 of the present article.

If we simulate a seven-item list to be remembered in an immediate serial recall task, using $\beta = .6$, it can be seen in the figure below that without response suppression (left panel), CRU produces a substantial amount of repetition errors. Implementing response suppression (right panel) prevents repetitions (Figure B2).

These simulations were run using the same choice rule as the one used in the present article, with a noise parameter $c = 0.1$. Response suppression was implemented by scaling the dot product of the already-recalled items by 0.0. For simplicity, we did not allow omission and extra-list intrusions, as these simulations were simply run for illustrative purposes and not actually fitting the model. We conclude that response suppression is a necessary mechanism, regardless of the computational architecture considered.

Figure B1

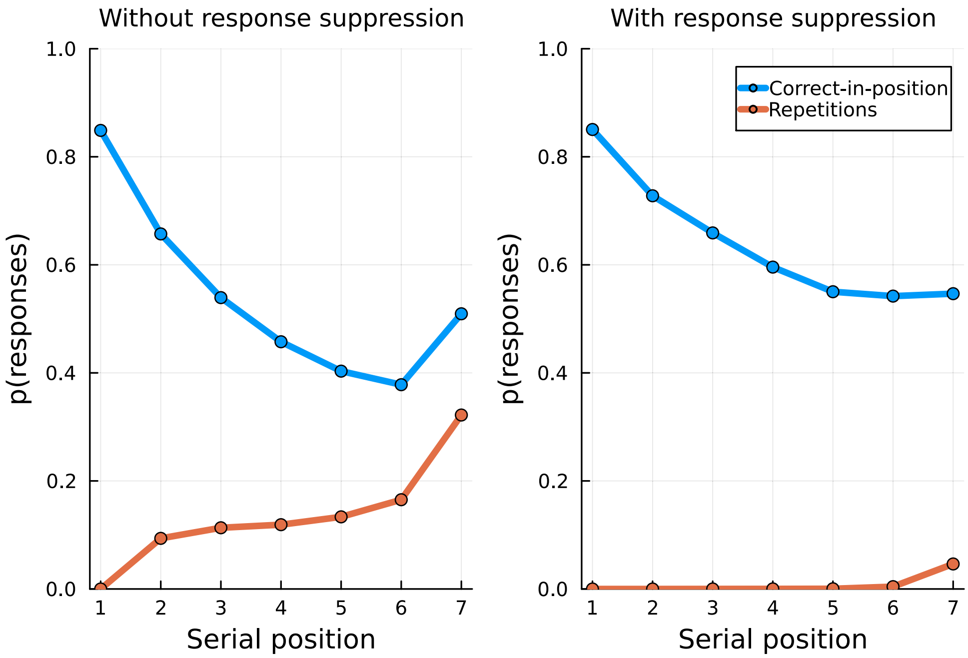
Dot Product Between Vectors in the CRU Model When Retrieving the Second Item



Note. These values were obtained by simulating a three-item list to be remembered. CRU = context retrieval and updating. See the online article for the color version of this figure.

(*Appendices continue*)

Figure B2
Results From the CRU Model



Note. Datapoints from these figures were generated using 100,000 simulated trials. CRU = context retrieval and updating. See the online article for the color version of this figure.

Received May 31, 2023
Revision received March 6, 2024
Accepted March 7, 2024 ■