Cognitive Psychology 107 (2018) 1–21

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Does articulatory rehearsal help immediate serial recall?☆

Alessandra S. Souza*, Klaus Oberauer

Department of Psychology, University of Zurich, Switzerland

ARTICLE INFO

Keywords:
Articulatory rehearsal
Overt rehearsal
Serial recall
Working memory

ABSTRACT

Articulatory rehearsal is assumed to benefit verbal working memory. Yet, there is no experimental evidence supporting a causal link between rehearsal and serial-order memory, which is one of the hallmarks of working memory functioning. Across four experiments, we tested the hypothesis that rehearsal improves working memory by asking participants to rehearse overtly and by instructing different rehearsal schedules. In Experiments 1a, 1b, and 2, we compared an instructed cumulative-rehearsal condition against a free-rehearsal condition. The instruction increased the prevalence of cumulative rehearsal, but recall performance remained unchanged or decreased compared to the free-rehearsal baseline. Experiment 2 also tested the impact of a fixed rehearsal instruction; this condition yielded substantial performance costs compared to the baseline. Experiment 3 tested whether rehearsals (according to an experimenter-controlled protocol) are beneficial compared to a matched articulatory suppression condition that blocked rehearsals of the memoranda. Again, rehearsing the memoranda yielded no benefit compared to articulatory suppression. In sum, our results are incompatible with the notion that rehearsal is beneficial to working memory.

1. Introduction

Working memory (WM) for verbal materials is often tested through sequential presentation of short lists, with immediate forward serial recall of the list. Maintenance of these lists is often accompanied by the overt or covert repetition of the memoranda to oneself, a behavior known as *articulatory rehearsal*[1] (Baddeley, 1986). Rehearsal is the most common self-reported maintenance strategy in WM tasks, being reported in about one-third to one-half of the trials (e.g., Bailey, Dunlosky, & Kane, 2011; Dunlosky & Kane, 2007).

Researchers routinely assume that people rehearse because it helps them maintain information in WM. Yet, there is hardly any experimental evidence for a causal link between rehearsal and serial recall performance. The main aim of the present paper is to fill this gap by providing a first experimental investigation on the impact of different rehearsal schedules upon memory over the short-term. In the following, we will motivate our research questions by reviewing the role of rehearsal in WM models and the sparse extant evidence available linking rehearsal to WM recall.

1.1. Rehearsal in WM models

Articulatory rehearsal is usually assumed to be beneficial to WM, and several WM theories assign rehearsal a causal role in WM

☆ This research was supported by a grant from the Swiss National Science Foundation to K. Oberauer (project 149193).
* Corresponding author at: Department of Psychology, Cognitive Psychology Unit, University of Zürich, Binzmühlestrasse 14/22, 8050 Zurich, Switzerland.

E-mail address: a.souza@psychologie.uzh.ch (A.S. Souza).
[1] Hereafter we will use the more general (but also shorter) term "rehearsal" to refer to articulatory rehearsal.

https://doi.org/10.1016/j.cogpsych.2018.09.002
Accepted 28 September 2018
0010-0285/ © 2018 Elsevier Inc. All rights reserved.

maintenance. According to theories assuming time-based decay (Baddeley, 1986; Camos, Lagner, & Barrouillet, 2009; Cowan, 1999), rehearsal occurs within a phonological store which offsets trace decay by restoring representations to their initial level of activation. In the time-based resource sharing (TBRS) model, other types of reactivation are also possible via the use of an attention-based process known as refreshing. The effects of rehearsal and refreshing are assumed to be additive (Camos et al., 2009; Camos, Mora, & Barrouillet, 2013; Camos, Mora, & Oberauer, 2011; Hudjetz & Oberauer, 2007; Mora & Camos, 2013, 2015). It follows that when rehearsal is blocked, recall accuracy decreases because decay sets in.[2] Likewise, in the embedded processes model proposed by Cowan (2001), it is assumed that WM comprises a focus of attention that holds a limited number of chunks in a more semantic format, whereas other peripheral mechanisms can provide additional (and domain-specific) storage capacity. Among these additional mechanisms/processes, Cowan lists sensory memory and rehearsal (Cowan, 2011). Accordingly, in this model rehearsal is assumed to supplement the capacity of the focus of attention, but to be independent of it.

Recent computational modeling work has cast doubt on the presumed effectiveness of rehearsal for counteracting decay. Lewandowsky and Oberauer (2015) implemented rehearsal in a generic decay model of immediate serial recall and found that its beneficial effect is very limited. Two main problems were identified by this detailed analysis of rehearsal. First, to rehearse a list in correct order, it has to be retrieved in correct order. Any factor jeopardizing accurate list memory, such as decay, also introduces a non-negligible chance of erroneous retrieval during rehearsal. When list items are retrieved in the wrong order during rehearsal, then rehearsal damages the representation of serial order rather than protecting it. Second, rehearsal is not evenly spread among list items – typically, early list items are rehearsed more often, simply because they are available for rehearsal earlier. The uneven frequency of rehearsal introduces uneven strength among list items – in particular, early list items often become so strong that they interfere with retrieval of subsequent list items. This undercuts the beneficial effect of rehearsal: Rehearsal tilts the serial position towards better recall of list-initial items, but leads to little or no overall benefit.

Models that do not postulate decay may also ascribe a causal role to rehearsal by assuming that it increases the accessibility of list items in memory due to the creation of distributed traces of the rehearsed words at multiple time points. Such accounts have been common in the explanation of free recall data (Brodie & Murdock, 1977; Brown, Sala, Foster, & Vousden, 2007; Farrell, 2012; Tan & Ward, 2000). The effect of rehearsal in these models has been described as "repeating, re-ordering, and redistributing the study items" (Tan & Ward, 2000, p. 1606), which clearly is at odds with the goal of keeping track of their order of presentation as required in serial recall tasks. If anything, these theories should predict a cost of rehearsal to serial-order memory. The only way in which these models could account for a beneficial effect of rehearsal for serial order is if the rehearsal output itself preserves the order of the items in the list, as it is the case when participants are attempting forward cumulative rehearsals, that is, rehearsal cycling through the list in its order of presentation.

Interference models of working memory do not resort to rehearsal to explain performance in WM tasks. One such model is the serial-order in a box – complex span (SOB-CS) model (Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012). Lewandowsky and Oberauer (2015) also explored the role of rehearsal in SOB-CS, and found that rehearsal had essentially no effect on memory. Rehearsal, modelled as cycles of retrieving and re-encoding of items, added nothing to the memory representations. This is because in this model, representations are not getting weaker over time, so there is not much to gain from rehearsing items.

Of course, any computational model implements a number of assumptions about how information is encoded, maintained, and retrieved from WM that can be wrong. The final arbiter should always be, therefore, the empirical data. We will review next the empirical studies assessing the role of rehearsal in WM, and show that the evidence linking rehearsal to WM is lacking.

1.2. Linking rehearsal to WM performance

1.2.1. Blocking rehearsal

Several studies have observed that requiring concurrent articulation of irrelevant material during maintenance of a verbal list reduces recall (e.g., Baddeley & Lewis, 1984; Bhatarah, Ward, Smith, & Hayes, 2009; Camos et al., 2009). Concurrent articulation is assumed to prevent rehearsal, and is therefore often referred to as "articulatory suppression" (AS). The finding that AS impairs verbal serial recall has commonly been interpreted as evidence for the beneficial effect of rehearsal. This interpretation is not compelling, however, because the detrimental effect of AS can also be explained by interference: Articulating irrelevant material introduces representations of that material into WM, where it interferes with the memoranda (Gupta & MacWhinney, 1995; Nairne, 1990, 2002; Oberauer et al., 2012). Moreover, several studies have shown that increasing the length of the period in which participants perform AS (hence arguably preventing rehearsal, and allowing decay to set in) does not lead to more forgetting unless there is variation in the articulated materials (Humphreys et al., 2010; Lewandowsky, Duncan, & Brown, 2004; Lewandowsky, Geiger, & Oberauer, 2008; Lewandowsky, Geiger, Morrell, & Oberauer, 2010; McFarlane & Humphreys, 2012; Phaf & Wolters, 1993; Vallar & Baddeley, 1982) (reviewed by Lewandowsky & Oberauer, 2015). The latter pattern of findings is consistent with an interference explanation, because more variability in the articulated material introduces more interference, but it is not consistent with a decay explanation, which predicts a main effect of time irrespectively of the filler activities performed therein. Clearly, we cannot rely solely on the data from AS manipulations to establish whether rehearsal is beneficial to WM.

1.2.2. Overt rehearsal protocol

Corroborative evidence for a causal role of rehearsal for memory has come from studies in which rehearsal patterns were observed

[2] This prediction follows even if refreshing is not blocked because the effects of rehearsal and refreshing are thought to be additive.

directly by using overt rehearsal protocols (Rundus & Atkinson, 1970): Participants are asked to carry out any rehearsal they chose to engage in aloud during presentation and maintenance of the memory list. Tan and Ward (2008) were the first to report overt rehearsal patterns in a serial recall task. They presented lists of six words at varying rates (1, 2.5, or 5 s per word). In the fast-rate condition, which is the typical rate in WM tasks, participants mostly just read each given word aloud, as instructed (Tan and Ward classified this as a *fixed rehearsal* strategy). At slower presentation rates, in addition to reading the just presented word, participants often attempted to rehearse all words presented up to that point in forward order (a strategy called *cumulative rehearsal*) (see also Bhatarah et al., 2009). Slower presentation rates not only led to more cumulative rehearsal, they also yielded better recall performance. Tan and Ward reasoned that the increase in recall may be related to the increase in the opportunities to attempt cumulative rehearsal. To assess this possibility, they correlated overall recall accuracy with the degree of cumulative rehearsal participants used. The correlation between these two variables was positive and substantial. This observation is consistent with a beneficial effect of cumulative rehearsal on memory for serial order.

A positive correlation between recall and rehearsal, however, is not sufficient to establish that rehearsal is beneficial to WM. It is possible that rehearsal and recall are a function of a common cause. Both rehearsal and recall depend on the quality of memory: List items that have been forgotten cannot be rehearsed and cannot be recalled. Therefore, a condition that yields better memory should lead both to a high probability of rehearsing and of recalling items. By the same token, high capacity participants can be expected to be the ones who rehearse most often – not because rehearsal helps memory but because good memory is a prerequisite for rehearsing.

1.2.3. Experimental manipulation of rehearsal

To help distinguishing between a spurious correlation and causality, it is necessary not only to observe rehearsal but also to manipulate its occurrence. A handful of studies have used instructions to manipulate rehearsal. For instance, Palmer and Ornstein (1971) instructed participants to rehearse items according to one of three schedules: (a) forward, cumulative rehearsals of all items; (b) repetition of only the last two presented items; or (c) fixed rehearsals of the last presented item. The cumulative instruction improved accuracy in a probed recall test, particularly for items from the beginning of the list, and this benefit survived a 15 s distraction period. In a similar vein, a recent study by Nishiyama and Ukita (2013) showed a rehearsal benefit for the maintenance of non-words for a free recall test. In this study, rehearsing items cumulatively, or even fixed rehearsals, improved performance compared to a no-rehearsal control condition. In the rehearsal condition the rehearsal time was added to the retention interval, whereas in the control condition no such time was added. Therefore, the rehearsal benefit cannot have arisen merely from protecting items against decay, suggesting instead that the sheer repetition of the memory items improved item memory beyond its state immediately after encoding (see also Hellyer, 1962).

Whereas the preceding studies provide evidence that rehearsal improves memory for individual items, it is unclear whether rehearsal can benefit memory for serial order. One study testing this was carried out by Estes (1991). Participants were presented with lists of five consonants, followed by a random list of 6 digits (distractors). The experimental conditions differed regarding the events following the end of the distractor list. In the *immediate-test* condition, participants were immediately prompted to recall the consonants. In the *filled* condition, they engaged in a period of articulatory suppression (repeatedly saying "la") before recalling. In the *rehearsal* condition, participants were instructed to rehearse the list aloud once before recalling. The addition of the overt rehearsal period prior to serial recall yielded worse performance than when participants immediately recalled the list. Performance in the rehearsal condition was only slightly better than when participants performed AS during the same period.

Turley-Ames and Whitfield (2003) tested participants in a complex span task before and after presenting a cumulative rehearsal instruction. The rehearsal instruction emphasized that participants should work quickly through the distractor task and use the remaining time to repeat the to-be-remembered words aloud as many times as possible (in a cumulative fashion) every time a new word was presented. The rehearsal group showed a greater improvement in performance between the first and second phases than a control group. This effect, however, arose because the rehearsal group spent more time encoding new items than the control group. In a subsequent experiment, Turley-Ames and Whitfield controlled for the time devoted to encoding the memoranda in all conditions, and the beneficial effect of the rehearsal instruction disappeared.

In sum, the experimental evidence so far, although supporting a beneficial effect of rehearsal on free recall, is ambiguous when it comes to serial recall.

1.3. The present study

The goal of the present series of experiments was to investigate the causal effect of rehearsal on immediate serial recall.

1.3.1. How to examine the causal role of rehearsal

Our review has shown three possible routes to examine the role of rehearsal in WM. The first one is to block rehearsal with AS. We have argued that this is not a suitable approach because AS does not only prevent people from rehearsing, it also adds interference to WM.

The second possibility is to vary the opportunities for rehearsal by changing the rate of presentation of the words. This is the approach taken by Tan and Ward (2008), who found that with slower presentation rate participants rehearsed more, and recall improved. Although induced experimentally, this relation between rehearsal and recall is still correlational, and therefore open to multiple interpretations: As presentation rate is reduced, participants have more time not only to rehearse but also to engage in other strategies to improve memory. For instance, they could attempt to elaborate on the presented words. Strategy-report studies have shown that elaboration is the second most common strategy, just behind rehearsal, and self-reported elaboration has been associated with better recall accuracy compared to self-reported rehearsal (Dunlosky & Kane, 2007). To explore whether elaboration – rather than articulatory rehearsal – could explain the beneficial effect of slower presentation rates on serial recall of words, in all experiments we varied the concreteness of the list words as an attempt to influence the difficulty of elaboration.

The third and least ambiguous route to study the causal effect of rehearsal is to manipulate rehearsal directly via instructions. This is the main route we chose to pursue here. Given prior reports that cumulative rehearsal is correlated with better memory (Nishiyama & Ukita, 2013; Palmer & Ornstein, 1971; Tan & Ward, 2008), we chose this rehearsal strategy as the main target of our manipulation. To assess whether this rehearsal instruction improves memory, we have to select a performance baseline. Ideally, in this baseline people would not be rehearsing at all, such that we could compare a no-rehearsal to a cumulative-rehearsal condition. However, instructing people not to rehearse has two drawbacks. First, there is no way to ensure that they heeded the instruction: If they cease to rehearse overtly, it is still possible that they rehearse covertly. Second, instructing people not to rehearse could force them to suppress their memory representations. Rehearsal is a frequent and spontaneous behavior, and one explanation for its occurrence is that a representation of a verbal list in WM functions like a representation of an action – in this case, a speech action. In ideomotor theory, representing the action outcome in WM is assumed to cause the action (Lewandowsky & Oberauer, 2015). If this is the case, then suppressing one's tendency to rehearse the contents of WM requires suppressing those contents themselves. As a consequence, instructing people not to rehearse could damage their representations of the memoranda. This would lead to worse performance in the "no-rehearsal" baseline, not because rehearsal helps memory, but because a no-rehearsal instruction would effectively be a "forget" instruction.

To avoid these problems, we decided to compare performance under our cumulative-rehearsal instruction to the performance obtained when people are free to rehearse as they wished in Experiments 1 and 2. We see two main advantages in this choice. First, prior research has shown that the variability in the use of cumulative rehearsal in this free-rehearsal baseline correlates with recall (Tan & Ward, 2008). This suggests to us that at least some people rehearse less than they could rehearse. If the correlation reflects a causal effect of rehearsal on recall, then making these people rehearse more should improve their performance. Second, this approach does not require people to suppress any of their natural rehearsal tendencies. Hence we avoid the risk of conflating rehearsal benefits with suppression costs.

The above benefits of using the free-rehearsal condition as a baseline notwithstanding, it could be criticized in the following way: Free rehearsal probably constitutes a mixture of several strategies, and the instructed rehearsal condition may shift the strategy towards more uniformly cumulative rehearsal. One possible objection is that the freely chosen mixture of strategies is, for some reason, more beneficial than a pure strategy. According to this view, rehearsal according to a person's spontaneously chosen schedule would be beneficial if compared to a condition that blocks rehearsal. To test this conjecture, we had each participant in Experiment 3 rehearse according to the freely chosen schedule of another, yoked participant. We compared this yoked-rehearsal condition to an AS condition matched to the rehearsal condition in all regards but the type of information that is articulated, namely the memoranda vs. irrelevant information.

To foreshadow our results, we have not obtained evidence in any of the three experiments that rehearsal benefits WM performance.

2. Experiment 1

Previous research converges on cumulative rehearsal being the most beneficial rehearsal strategy (Guttentag, Ornstein, & Siemens, 1987; Palmer & Ornstein, 1971; Tan & Ward, 2008). Theoretical considerations also point to cumulative rehearsal as the rehearsal schedule best suited for helping memory for serial order. First, it preserves serial order information, whereas most other rehearsal schedules – with the exception of fixed rehearsal – rehearse items in an order different from the one that needs to be remembered, thereby entailing the risk of introducing erroneous order information. Second, to protect early list items from decay, rehearsal has to revisit them throughout list presentation – otherwise early list items would decay more than later list items. For instance, a fixed rehearsal schedule stops rehearsing the first item once the second item is presented – from that point in time, the first item would suffer unmitigated decay. This would lead to a negative primacy effect, contrary to what is observed in serial recall (Lewandowsky & Oberauer, 2015). It is probably for these reasons that most decay-rehearsal models of serial recall feature cumulative rehearsal as the rehearsal schedule in their computational implementations (Burgess & Hitch, 1999; Daily, Lovett, & Reder, 2001; Page & Norris, 1998). It follows that, if rehearsal is beneficial to memory for serial order, increasing the amount and length of cumulative rehearsal should improve immediate serial recall. This finding would provide evidence for a causal role of rehearsal on WM performance.

Accordingly, the first goal of the present study was to investigate whether increasing cumulative rehearsals yields a corresponding increase in serial recall performance. Experiment 1 involved three conditions: The *Fast* and the *Slow* conditions were replications of the overt-rehearsal experiment of Tan and Ward (2008) with their fast and their slow presentation rate, respectively. In these conditions participants were asked to try to memorize the words in any way they liked, but if they felt they want to articulate the words, they should do so overtly. The *Slow* condition served as the baseline for the new *Slow-C* (*instructed cumulative rehearsal*) condition, in which we presented words at the slow presentation rate and asked participants to always engage in overt cumulative rehearsal to the best of their abilities throughout the memory list.

A second goal of Experiment 1 was to explore why recall is better with a slower presentation rate. Specifically, we tested the possibility that people use the longer inter-word intervals for elaboration. Elaboration refers to the enrichment of to-be-remembered stimuli by relating them to each other, to semantically related contents of long-term memory, or to visual images. Elaboration is a frequently reported strategy in complex-span tasks, and people reporting using it outperform individuals reporting other strategies, including articulatory rehearsal (Dunlosky & Kane, 2007; Kaakinen & Hyönä, 2007). A beneficial role of elaboration could explain why serial recall is often better at slower presentation rates (e.g., Tan & Ward, 2008), because the fast presentation rate arguably leaves insufficient time for an effective elaboration strategy. To this end, we varied the word sets used to construct the memory lists. We used concrete and highly imageable words in one half of trials, and abstract words with low imageability in the other half. We reasoned that concrete, highly imageable words may facilitate elaboration more than abstract words with low imageability. Concrete words are usually better recalled than abstract words, yielding the so-called concreteness effect (Bourassa & Besner, 1994; Romani, Mcalpine, & Martin, 2008). Hence if the presentation rate effect is due to the use of elaboration, then a larger concreteness effect would be observed in the slow rate compared to the fast rate conditions, at least for the conditions in which participants are free to rehearse as they wish.

We conducted two versions of this study: In Experiment 1a immediate serial recall was performed orally as in Tan and Ward (2008), whereas in Experiment 1b immediate serial recall required typing the words using the keyboard. Everything else was the same in both experiments. We have no specific prediction regarding how output mode interacts with rehearsal. This factor is usually not systematically investigated; it mostly represents one of the researcher's degree of freedom in setting up an experimental design. We wished to simplify data analysis by using typed recall (for which scoring can be done quickly and automatically). This was the main reason for us to make the comparison between the two recall modes.

2.1. Method

2.1.1. Participants

Forty-eight students of the University of Zurich took part in a single one-hour session for course credit or a reimbursement of 15 Swiss francs. Half of the participants took part in Experiment 1a and the other half in Experiment 1b (each with $n = 24$). For all experiments reported in this article, participants signed an informed consent form prior to the start of the experiment, and were debriefed at the end. The experimental protocol was approved by the Institutional Review Board.

2.1.2. Materials and procedure

We compiled two sets of German words from the "Semantischer Atlas" data base (Schwibbe, n.d.). One set consisted of words with high ratings of concreteness and imageability (henceforth concrete words) whereas the other consisted of words with low ratings on both dimensions (henceforth abstract words). The word sets were equated with regard to their mean word length (mean = 7.8 characters) and frequency (mean log frequency among 4.5 million words = 4.9). Half of the lists in each condition were sampled from each set. Lists with concrete and abstract words alternated in random order within each block.

Before the start of the experiment, participants were informed about the three conditions: *Fast*, *Slow*, and *Slow-C*. For the fast and the slow condition, they were instructed that it was up to them whether and how they rehearsed the words, but they should carry out any rehearsal aloud. In contrast, in the instructed-rehearsal condition (Slow-C), they were told to always rehearse in cumulative fashion as best as they could, and cumulative rehearsal was illustrated with an example.

The three conditions were run in three blocks, administered in counterbalanced order across participants. At the beginning of each block participants were informed by a brief on-screen message of the upcoming condition. This was followed by four practice trials and 20 test trials in each block.

In the fast condition, words were presented centrally on the screen, in black, at a rate of one per second (0.9 s display, 0.1 s blank-screen gap), whereas in the two slow conditions, one word was presented every 5 s (0.9 s display, 4.1 s blank); these times match those of the fast and the slow presentation-rate conditions of Tan and Ward (2008). After the last word's blank period, the forward serial recall test started, which was marked by the onset of a question mark in the middle of the screen. In Experiment 1a, participants were instructed to recall all studied words in forward serial order by speaking them into a microphone. When they could not remember a word in a given serial position, they were instructed to say *"Weiss nicht"* ("don't know" in German) or to guess. When participants were finished recalling the words, they were instructed to press the spacebar. Recall was recorded for offline accuracy check. In Experiment 1b, participants were told to type the words using the keyboard. They were instructed to type at least the first three letters of each word, finishing each recalled word with the return key. They could skip recall of some words by simply pressing the enter key.

2.1.3. Data analysis

A research assistant coded the utterances occurring during each time gap following a memory item, creating a data file in which each spoken word was identified by its serial position in the list, and extra-list words by additional numbers; these data were then analyzed with a custom-written Matlab script. In a first step we counted the total number of words rehearsed in each trial across all time gaps. In the second step, we classified the sequence of words spoken in each gap in terms of the rehearsal strategies used by participants. We used the four categories used by Tan and Ward (2008): A *Fixed* strategy was coded when participants simply read aloud the words as they were presented; or when they repeated the currently presented word multiple times, but no other word was rehearsed in that time gap. A *Cumulative* strategy was coded when somewhere during the gap participants rehearsed all list items presented so far in their correct serial positions. For example, at gap 4, rehearsing the word sequence 1-2-3-4 somewhere during the gap was deemed cumulative (e.g., 1-1-2-3-4, 4-4-1-2-3-4). A *Partial Cumulative* strategy was coded when somewhere during the gap participants started rehearsing from the beginning of the list and proceeded in forward order, but did not manage to rehearse all of the items presented so far (e.g., 4-4-1; 4-4-1-2; 4-3-1-2). Any other type of rehearsal strategy was classified as *Other*. Most of the patterns that fell into the latter category comprised rehearsal of pairs or triplets of items (e.g., 4-3-4; 2-3-3-4). Whereas Tan and Ward

(2008) instructed participants to read each word aloud (thereby enforcing a fixed rehearsal strategy as the default), we did not. Therefore, we added a fifth strategy called *Silence* which was registered when no word was spoken during a gap.

Finally, following Tan and Ward (2008), we computed the maximum length of the cumulative-rehearsal sequence in a given trial. We determined the length of cumulative rehearsal in each gap of a given trial; non-cumulative rehearsals and fixed rehearsal were assigned a length of 1. Then we maintained the maximum length in any gap for that trial.

We analyzed the data of all our experiments with Bayesian mixed effect models (LME) using the *lmBF* function from the BayesFactor package (Morey & Rouder, 2014) implemented in R (R core team, 2014). This function computes the strength of the evidence for the specified model (M_1) against a Null or reduced model (M_0). The ratio of the likelihood of these two models is the Bayes factor (BF_{10}). The BF is the multiplicative factor by which our ratio of prior beliefs in the two models should be updated in light of the data. BFs below 3 are usually regarded as "weak evidence"; BFs between 3 and 10 are regarded as providing "substantial evidence"; BFs between 10 and 100, as providing "strong evidence"; and above 100, "decisive evidence" in favor of one model over the other (Kass & Raftery, 1995).

Each analysis considered a set of linear models, starting from a full model with all of our predictors and interactions thereof. We obtained the BF_{10} for each model by comparing it to the Null model, which includes only a random intercept. Across several steps, we systematically assessed the evidence for a given fixed predictor by dropping this term from the full model. By computing the ratio of the BF_{10} for the model including the effect of interest with the BF_{10} of the model excluding it, we obtained the BF for the comparison of these two models, which reflects the strength of the evidence supporting the effect of interest in the data. When this ratio was below 1, we removed the term from the full model – meaning that this term was excluded from the full model that we subsequently used to assess the evidence for all remaining predictors. When the ratio was above 1, the term was retained in the full model. We repeated these steps until we tested the evidence for all predictors. We always started by assessing the evidence for the interaction terms and moved up to the main effects. All models included a random intercept for participant, and random slopes over participants for the effects of the variables manipulated within-subjects (Barr, Levy, Scheepers, & Tily, 2013).

The materials (including word lists and instructions), data, and the analyses scripts for all the experiments reported here are available in the Open Science Framework at: https://osf.io/5q7nd/.

2.2. Results

2.2.1. Rehearsal

Fig. 1 presents the proportion of trials in which each of the four rehearsal strategies identified by Tan and Ward (2008) was observed, alongside the proportion of trials in which participants remained silent. As there were no discernible differences between rehearsals of concrete and abstract words, we collapsed the data over word type.[3] The results are in line with those of Tan and Ward (2008): At the fast presentation rate, participants did not rehearse consistently. On about half the occasions, they did not say any words aloud, and in the remaining cases they mostly just read aloud the word just presented (i.e., engaged in the "Fixed" strategy). Towards the end of the list, they occasionally rehearsed several words, though usually not in a cumulative fashion, so that most of these rehearsal events were classified as "Other". At the slow presentation rate, cumulative rehearsal predominated. Towards the end of the list, there was a shift from full to partial cumulative rehearsal. In the Slow-C condition, participants rehearsed on a larger proportion of trials than in the Slow condition, and engaged in cumulative rehearsal more often.

To quantify differences between the two slow conditions we counted the total number of words rehearsed in each trial, summing across all six time-gaps following the six words. In addition, we determined the length of the longest sequence of cumulative rehearsal for each trial. For Experiment 1a, the average number of words rehearsed was 21.9 [95% within-subjects CI: 19.1, 24.8] in the Slow condition, and 31.3 [28.9, 33.7] in the Slow-C condition, a difference strongly supported statistically ($BF_{10} = 905.6$). The same pattern was observed in Experiment 1b: The average number of words rehearsed was 19.2 [15.6, 22.8] and 30.0 [27.5, 32.6] for the Slow and Slow-C condition, respectively ($BF_{10} = 516.9$). For comparison, in the Fast condition people rehearsed on average 5.0 words in both experiments [Exp. 1a, 95% CI: 2.7, 7.9; Exp. 1b: 2.0, 8.0].

When instructed to rehearse cumulatively, participants produced longer sequences of cumulative rehearsal. The average maximum cumulative rehearsal length for the Slow condition was 3.5 [95% CI: 3.0, 3.9] and 2.7 [2.1, 3.2] in Experiments 1a and 1b, respectively. In contrast, in the Slow-C condition this value increased to 4.4 [4.1, 4.7] and 3.9 [3.6, 4.1], respectively. There was strong statistical support for the increase in cumulative rehearsal length in both experiments, Exp. 1a: $BF_{10} = 18.5$; Exp. 1b: $BF_{10} = 135.6$. In the Fast condition, the average rehearsal length was just 0.90 [0.5, 1.2] and 0.59 [0.15, 1.02] in Exps. 1a and 1b. We conclude that our cumulative rehearsal instruction increased people's rehearsal activity overall, and their disposition to rehearse cumulatively in particular.

Fig. 2 shows the frequency with which words in each serial position were rehearsed, on average, across an entire trial in Experiments 1a and 1b. The same pattern was observed for both experiments. At the slower presentation rate, it was mostly the earlier list words that received more rehearsals, as would be expected from a cumulative rehearsal strategy. The instruction to use cumulative rehearsal accentuated that effect, but it did not lead participants to neglect rehearsal of list-final words more than in the other conditions.

[3] Rehearsals split by word-list type can be found in the OSF.

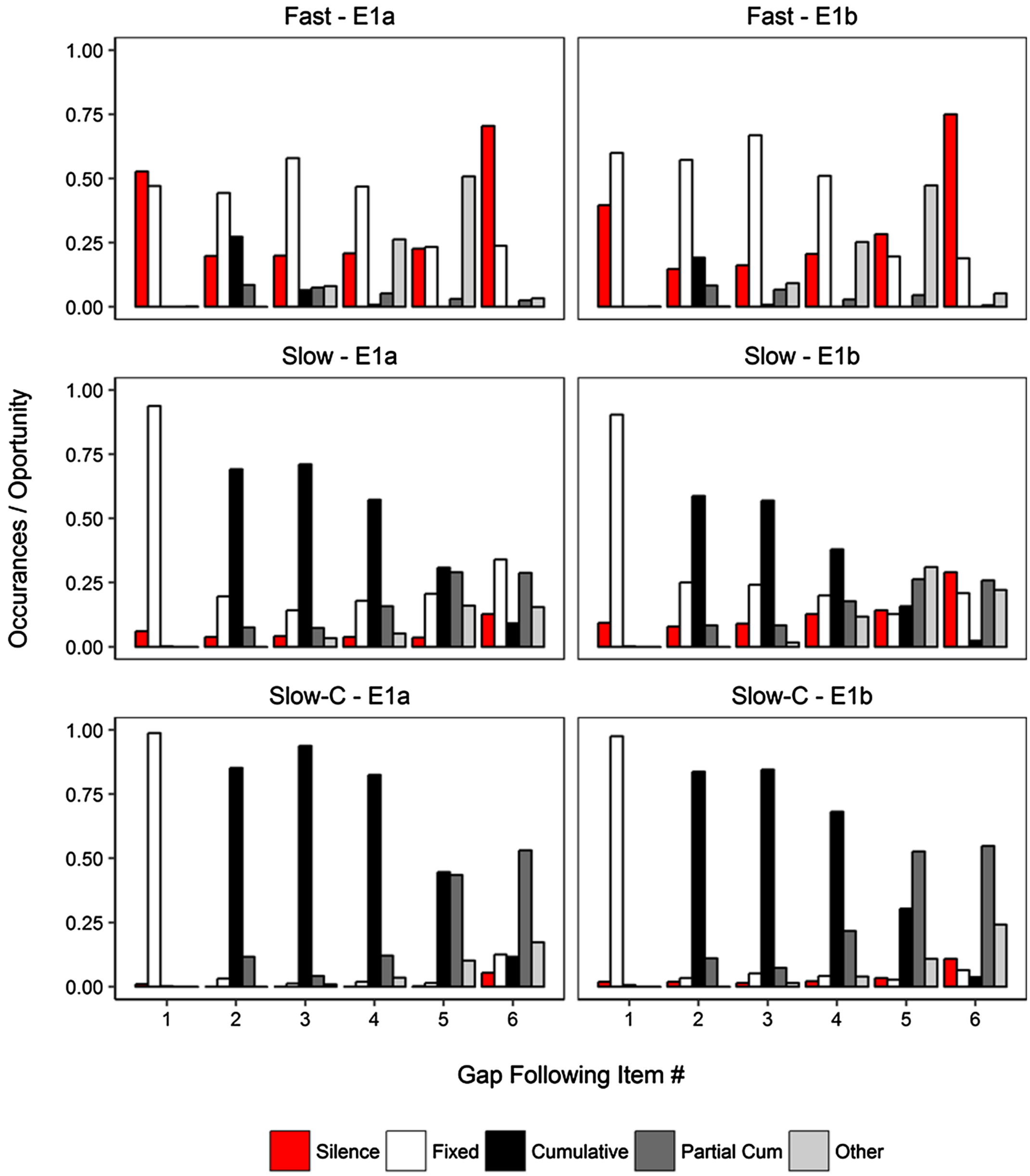

Fig. 1. Relative frequency of each type of rehearsal strategy plotted separately for the three conditions of Experiments 1a and 1b.

2.2.2. Immediate recall

Fig. 3 shows the serial-position curves for the three rehearsal conditions separately for the two word sets (i.e., concrete and abstract words). Panel a shows the results for Experiment 1a and panel b shows the results for Experiment 1b.

We analyzed the data of the two experiments separately. For each experiment, we conducted two sets of analyses on the proportion correct data. One analysis was focused on comparing the Fast vs. Slow presentation rate conditions. In these conditions, participants were free to rehearse as they wished. The relation between presentation rate, rehearsal, and recall has served as one pillar to claim that rehearsal is beneficial to WM. In Experiment 1 we included the manipulation of concreteness to assess whether the presentation rate effect could also be explained by elaboration. The second set of analyses targeted the effect of cumulative rehearsals by directly comparing the Slow to the Slow-C condition. This contrast is germane for our main aim of testing for the effect of cumulative rehearsal in WM.

We analyzed the data of the targeted conditions with Bayesian regression models having Condition (either Fast vs. Slow; or Slow vs. Slow-C), Concreteness (concrete vs. abstract), Serial Position, and interactions thereof as fixed (categorical) predictors. Table 1 presents the evidence for each predictor for Experiments 1a and 1b.

We will describe first the comparison of the Fast vs. Slow conditions. For both experiments, there was overwhelming evidence for

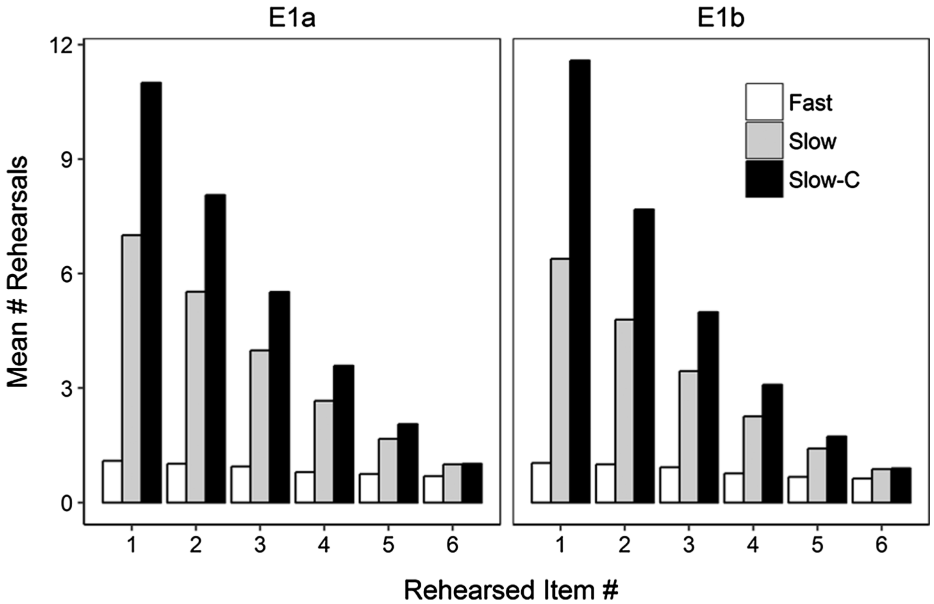

Fig. 2. Average frequency of rehearsing words in each serial position during an entire trial, separately for the three conditions of Experiments 1a and 1b.

the main effects of condition (reflecting better memory in the slow than fast condition), concreteness (better memory for concrete than abstract words), and serial position (reflecting the typical patterns of primacy and recency). In Experiment 1b, but not Experiment 1a, the serial-position curve was somewhat flattened with slower presentation rate. There was evidence for an interaction between condition and concreteness in both experiments: Recall of concrete words benefited more from the slow presentation rate than recall of abstract words. This is the pattern predicted from the hypothesis that participants use the longer inter-word gaps to elaborate on the presented words, a strategy that is facilitated by the concrete words. This finding lends support for an alternative explanation of the presentation rate effect, thereby weakening the case for rehearsal as causing this effect.

To test the focal hypothesis that cumulative rehearsal improves serial-order memory, we compared the two conditions with slow presentation rate, which differ only in whether people were instructed to rehearse in a cumulative fashion or were free to rehearse as they liked (see Table 1). There was weak evidence *against* a main effect of condition (which reflects the manipulation of rehearsal instruction) in either experiment (the Null was supported by a factor of 2–3).

To more closely test for the possibility that cumulative rehearsal increases overall serial recall, we compared the mean proportion of correctly recalled words between the Slow and the Slow-C conditions in each experiment using one-sided Bayesian t-tests. A one-sided test can be used to specifically test the hypothesis that increasing cumulative rehearsal *improves* performance (whereas a two-sided test provides evidence for differences in any direction). There was substantial evidence *against* a benefit of cumulative rehearsal in both Experiments (Exp. 1a, $BF_{10} = 0.14$; Exp. 1b, $BF_{10} = 0.10$). Together, these results show that the instruction to perform cumulative rehearsal did not increase the overall number of words that participants recalled in correct serial order.

The instruction to rehearse cumulatively was, however, not completely inconsequential. In Experiment 1b, there was very strong evidence for an interaction between condition (Slow vs. Slow-C) and serial position, indicating that the cumulative rehearsal instruction induced larger primacy and smaller recency compared to the condition in which participants rehearsed as they wished. This interaction was not observed in Experiment 1a, for which the evidence against an interaction between condition and serial position was strong.

2.3. Discussion

Experiment 1 used an experimental manipulation of rehearsal to determine the causal effect of cumulative rehearsal on serial-recall performance. The rehearsal instruction substantially increased participants' inclination to rehearse in a cumulative fashion. Although this instruction increased the length of cumulative rehearsals by about one item, this did not lead to an increase in the number of words participants could recall. The one-sided Bayesian t-tests provided substantial evidence *against* a beneficial effect of rehearsal in both experiments.

Furthermore, when recall was typed, increasing the length of cumulative rehearsals tilted the serial position curves towards more primacy and less recency. This tilting of the serial-position curve is to be expected if rehearsal increases an item's chance of being recalled, because cumulative rehearsal implies that early list items receive more rehearsal than later list items (see Fig. 2). The tilting of the serial position curves was only observed for typed recall; when recall was spoken, cumulative rehearsal had no effect at all. We have no ready explanation for the subtle differences yielded by the two recall modalities. Notwithstanding these minor differences, both experiments converge on the same conclusion regarding our main question: Increasing cumulative rehearsal does not benefit immediate forward serial recall.

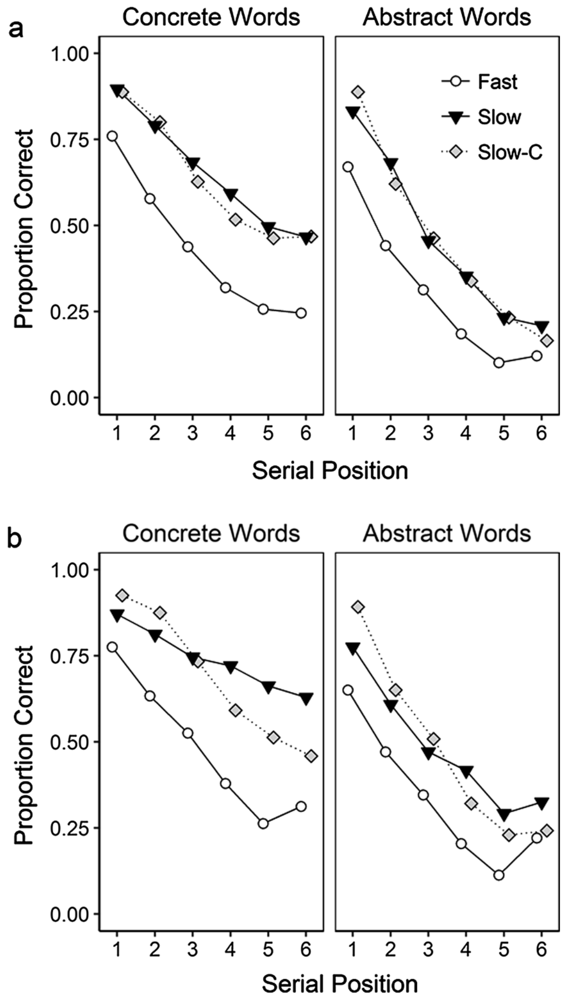

Fig. 3. Serial position curves for the fast and slow presentation rates with unconstrained rehearsal, and with slow presentation rate and the instruction to use cumulative rehearsal (Slow-C condition). Panel a shows data of Experiment 1a (oral recall) and panel b shows data of Experiment 1b (typed recall).

Table 1
Bayes factors quantifying the strength of evidence for the main effects and interactions of the variables manipulated in Experiment 1. Separate sets of analyses were performed on the data of Experiment 1a and Experiment 1b.

Predictor	Condition Contrast			
	Fast vs. Slow		Slow vs. Slow-C	
	E1a	E1b	E1a	E1b
Condition	2753.6	3.7×10^5	0.27	0.45
Concreteness	6524.4	2.9×10^6	20,897	7.58×10^6
Serial Position	5.2×10^{42}	3.6×10^{22}	4.1×10^{47}	1.86×10^{30}
Condition × Concreteness	8.17	2129.8	0.18	1.55
Condition × Serial Position	0.07	234.5	0.02	1.1×10^8
Concreteness × Serial Position	1.29	3.86	7.7×10^6	3.2×10^7
Three-way interaction	0.33	2.25	0.13	0.04

In other regards our results largely replicated those of Tan and Ward (2008): When lists are presented at the – most commonly used – rate of one word per second, people at most rehearse the last presented word. Whereas Tan and Ward (2008) reported that participants engaged in “fixed rehearsal” nearly all the time, participants in Experiment 1 showed this pattern only about half the time. This difference reflects the fact that Tan and Ward instructed participants to always read the presented word aloud, so that merely by following this instruction they were classified as using “fixed rehearsal”, whereas here participants were free to stay entirely silent, and they chose to remain silent on a substantial number of occasions. At a slower presentation rate, participants mostly engaged in cumulative rehearsal, which they increasingly failed to complete as the list became longer, so that towards the end they more often showed partial cumulative rehearsal attempts. Still, when they were free to rehearse as they wished, they used cumulative rehearsal much less than when instructed to always do so.

We found substantial support for the speculation that the benefit of slower presentation rate arises from elaboration: Recall of concrete, highly imageable words benefited more from the slow presentation rate than recall of abstract, less imageable words. Hence the data of Experiment 1 are consistent with the possibility that participants use the long gap between words to create mental images of the memoranda, and that this benefits WM recall. This result is also in line with the results of studies requesting strategy self-report, which has shown better recall in trials in which participants report using elaboration (Bailey, Dunlosky, & Hertzog, 2009, 2014; Bailey, Dunlosky, & Kane, 2008).

3. Experiment 2

Although cumulative rehearsal has been pointed out via theoretical considerations as well as via correlational studies as the strategy that would be the most effective for helping memory for serial order, Experiment 1 showed no beneficial effect of increasing the frequency and length of cumulative rehearsals for serial recall performance. One may wonder, however, whether other types of rehearsal strategies may be beneficial, or whether the effectivity of a rehearsal strategy may change over the course of a trial. A reviewer of this article pointed out that when participants are left to rehearse freely, they mostly rehearse the words cumulatively in the beginning of the list, but then they shift to a fixed rehearsal strategy towards the end of the list. This shift in rehearsal strategies could suggest that different rehearsal strategies are most effective early vs. late in the list. Experiment 1b showed that cumulative rehearsal may increase the accessibility of early list items at the expense of later list items. It is conceivable that a fixed rehearsal strategy is beneficial to later list items, but not for early list items. If this is the case, then the combination of cumulative rehearsal in the beginning of the list with a fixed rehearsal towards the end of the list may yield the desired overall increase in serial recall performance.

The goal of Experiment 2 was to examine whether cumulative and fixed rehearsal strategies are differentially beneficial across the memory list. Experiment 2 serves two specific goals. First, it replicates the conditions from Experiment 1b which yielded the tilting of the serial position curve following a cumulative rehearsal instruction. Second, it implemented a fixed rehearsal instruction. If we observe that a fixed rehearsal strategy benefits later list items whereas cumulative rehearsal benefits early list items, this would lend support to the hypothesis that a mixture of early cumulative and later fixed rehearsal strategies could be beneficial for serial recall.

3.1. Method

A new sample of twenty-four students of the University of Zurich was invited to take part in a single 75-min session for course credit or a reimbursement of 20 Swiss francs. Participants completed the same task as described for the Experiment 1b (i.e., using the same word pool, and having to recall by typing the words) with the following exceptions: all three experimental conditions used a slow pace of presentation of the memoranda (i.e., one word every 5 s, with 0.9 s display and 4.1 s blank). The fast-pace condition was dropped because it did not add information to answer our question. The three slow conditions differed only regarding the rehearsal instructions. The Slow and Slow-C conditions replicated the corresponding conditions in Experiment 1b. The third condition, *Slow-F*, implemented a fixed rehearsal instruction. Participants were instructed to rehearse only the last presented item in each gap, and to do so continuously until the next word appeared. We asked participants to rehearse continuously to prevent them from using any other rehearsal strategy covertly in the meantime. Similarly to Experiment 1b, there were four practice trials and 20 test trials in each experimental condition, which were completed in separate blocks.

3.2. Results

3.2.1. Rehearsal

Participants rehearsed an average of 17.6 words per trial [CI: 13.5, 21.7] in the Slow condition, 36.2 words [33.1, 39.2] in the Slow-C condition, and 28.4 words [25.2, 31.6] in the Slow-F condition. The evidence was strong for an increase in the number of rehearsed words in the Slow-C (BF_{10} = 67541.6) and in the Slow-F (BF_{10} = 75.5) conditions compared to the Slow condition, indicating that the instructions to rehearse increased participant's tendency to rehearse the memoranda.

The next relevant question is whether the patterns of rehearsal differed across the instructed conditions. Fig. 4 presents the frequency of each rehearsal pattern in the Slow, Slow-C, and Slow-F conditions, and an additional analysis in which we filtered the data of the Slow-F condition (explained below). When left to rehearse on their own (Slow condition), a mixture of cumulative rehearsals in the beginning of the list combined with a switch to more fixed rehearsals towards the end was observed, replicating Experiment 1. When participants were instructed to rehearse cumulatively (Slow-C condition), the frequency of cumulative and partial cumulative rehearsal attempts increased throughout the list, again replicating Experiment 1. When participants were instructed to rehearsed in a fixed fashion (Slow-F condition), fixed rehearsals were the most frequent strategy throughout the list. Nevertheless, our automatized rehearsal-strategy coding program[4] detected a substantial amount of cumulative rehearsals in the second position and other rehearsal strategies for the other positions. Closer inspection of the rehearsal data revealed that this was

[4] Implemented in Matlab and available in the OSF.

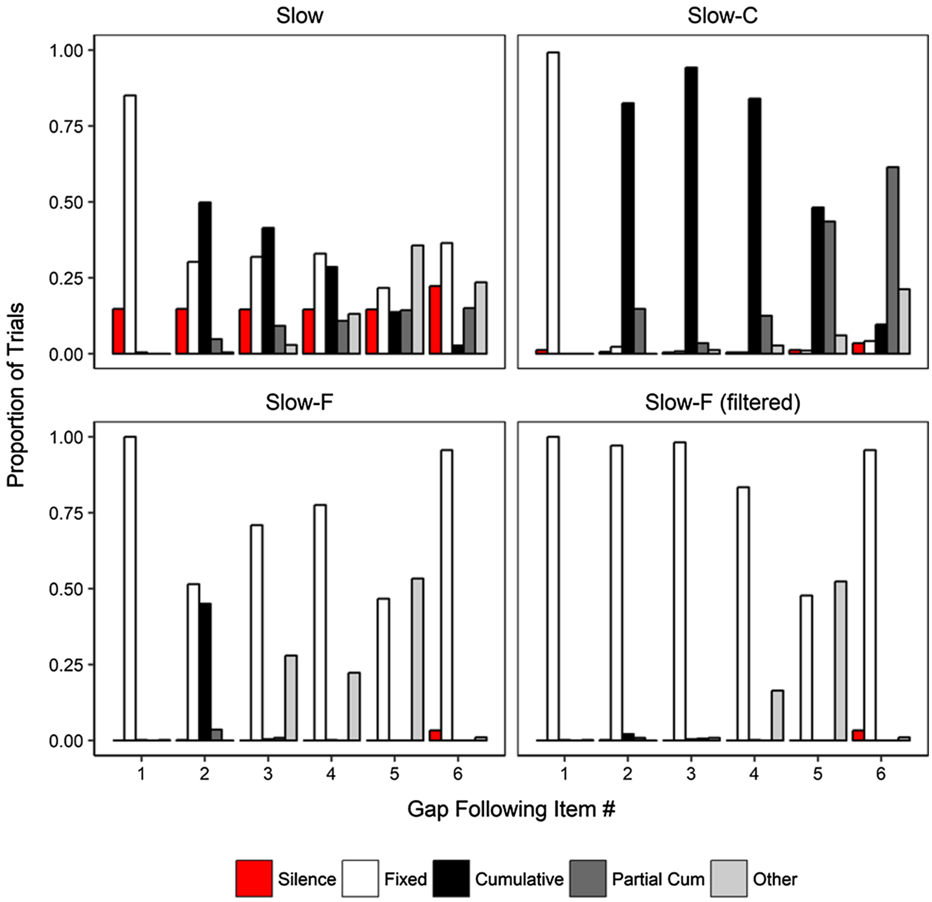

Fig. 4. Relative frequency of each type of rehearsal strategy plotted separately for the three experimental conditions (i.e., Slow, Slow-C, and Slow-F) in Experiment 2. The last panel shows the classification of the rehearsal strategies in the Slow-F condition when the delay in switching between rehearsing the previous word and the current one is taken into account.

mainly due to participants being slightly slow to switch between fixed rehearsals of the preceding word to the fixed rehearsals of the just presented word. Thereby rehearsals of the previously presented word tended to be coded in the next gap. Reanalysis of the data taking into consideration this delay (Slow-F (filtered) panel) shows that participants were indeed mostly rehearsing in a fixed fashion in the Slow-F condition.

In agreement with Fig. 4, the length of cumulative rehearsal differed between the experimental conditions. The average maximum cumulative rehearsal length was 2.14 [CI: 1.59, 2.69] in the Slow condition. This value doubled in the Slow-C condition to $M = 4.38$ [4.04, 4.72], and the difference between the Slow and Slow-C conditions was strongly supported statistically, $BF_{10} = 18{,}325$. Cumulative rehearsals decreased in the Slow-F condition, $M = 0.95$ [0.64, 1.25].[5] This decrease was also strongly supported in comparison to the Slow condition, $BF_{10} = 29.49$.

Finally, Fig. 5 shows the average frequency with which each word was rehearsed within a trial. Participants showed an uneven distribution of rehearsals across the list in the Slow condition, with items from the beginning of the list being rehearsed more often than items from the end of the list. This is typical of a cumulative rehearsal strategy. This pattern was further accentuated under the cumulative rehearsal instructions (Slow-C condition). In contrast, the fixed rehearsal instruction (Slow-F condition) was successful in producing an even distribution of repetitions of the words across the list. Critically, the cumulative rehearsal instruction was particularly successful in increasing the frequency with which words from early list positions (1–3) were rehearsed, whereas the fixed rehearsal instruction was particularly effective in increasing the frequency with which late list positions (4–6) were rehearsed compared to the Slow condition. Altogether, our findings indicate that the instructions to rehearse cumulatively or in a fixed-fashion yielded the corresponding rehearsal patterns.

[5] It is worth noting that fixed rehearsals were assigned a length of 1 for this analysis. Hence an average length of 1 is expected if participants were mostly doing fixed rehearsals.

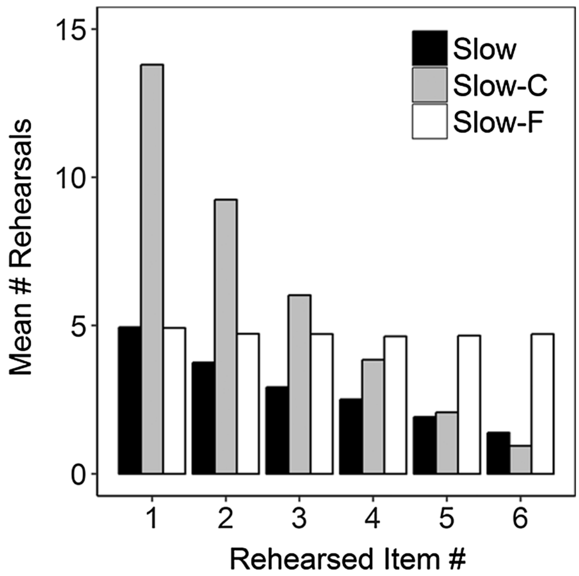

Fig. 5. Average frequency of rehearsing words in each serial position during an entire trial, separately for the three conditions of Experiment 2.

3.2.2. Immediate recall

Fig. 6 presents serial recall performance for concrete and abstract words across rehearsal conditions. We separately contrasted the Slow condition either to the Slow-C or to the Slow-F condition in a three-way BANOVA entering rehearsal condition, concreteness, and serial position as predictors. Table 2 presents the evidence in this analysis for the main effects and interactions.

There was strong evidence for the effects of concreteness, serial position, and also rehearsal in both rehearsal condition contrasts we ran (see Table 2). Of particular interest for our question, the effect of rehearsal obtained here was in the direction of poorer serial recall performance for both instructed rehearsal conditions compared to the baseline condition. Focused one-sided t-tests on the average recall performance across all serial positions showed very strong evidence against better performance in the Slow-C condition compared to the Slow condition, $BF_{10} = 0.05$, and against better performance in Slow-F condition compared to the Slow condition, $BF_{10} = 0.02$.

The change in rehearsal patterns induced by the instructions led to a change of the serial position curves, as indicated by the interaction of rehearsal x serial position (see Table 2) in both condition contrasts performed: As in the previous experiments with typed recall, cumulative rehearsal led to a steeper primacy effect at the expense of later list items. Fixed rehearsal had the opposite effect, weakening the primacy effect. Yet, despite doubling the frequency of rehearsal of the last three items compared to the control

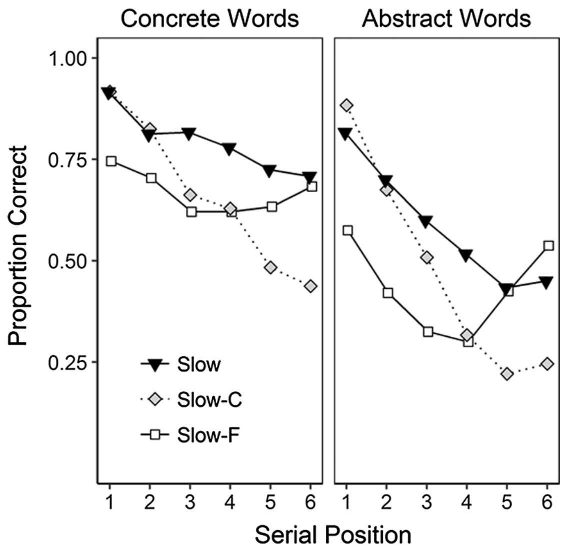

Fig. 6. Serial position curves for the three slow presentation rate conditions in Experiment 2. In the Slow baseline participants were free to rehearse as they wished. In the Slow-C condition participants were instructed to rehearse cumulatively, whereas in the Slow-F condition they were instructed to rehearse only the last presented word (fixed rehearsal strategy).

Table 2
Bayes factors quantifying the strength of evidence for the main effects and interactions of the variables manipulated in Experiment 2.

Predictor	Rehearsal Contrast	
	Slow vs. Slow-C	Slow vs. Slow-F
Rehearsal	57.45	979.79
Concreteness	6.63×10^5	44,388
Serial Position	1.39×10^{33}	1.42×10^8
Rehearsal × Concreteness	0.22	0.31
Rehearsal × Serial Position	6.36×10^{12}	3.96×10^{10}
Concreteness × Serial Position	1.99×10^6	13.88
Three-way interaction	0.08	13.91

condition (Slow), fixed rehearsal did not improve memory for the end of the list, $BF_{10} = 0.08$ ($BF_{10} = 0.02$, 0.11, and 0.48 – for one-tailed t-tests assessing evidence for a rehearsal benefit in recalling the 4th, 5th, and 6th list-item, respectively).

3.3. Discussion

The manipulation of a cumulative rehearsal instruction in Experiment 2 replicated the results observed in Experiment 1b: The increase in cumulative rehearsals tilted the serial position curves towards more primacy and less recency. The only difference between Experiments 1b and 2 is that in the present experiment we obtained substantial evidence for a detrimental effect of cumulative rehearsals on memory (see main effect of rehearsal in Table 2). A fixed rehearsal strategy did not yield a selective increase in performance for items from the end of the list. Hence these results lend little support to the hypothesis that a mixed cumulative-fixed rehearsal strategy could be beneficial to serial recall.

4. Experiment 3

Experiments 1 and 2 showed that neither instructing a cumulative rehearsal strategy nor a fixed rehearsal strategy benefited serial recall performance compared to a condition in which participants were free to rehearse as they wished. This finding is in line with the hypothesis that rehearsal does not benefit serial recall performance.

One objection that has been raised against this conclusion refers to the baseline used in these experiments, namely the slow, free rehearsal condition. Some of our colleagues have argued that the free rehearsal condition is a situation in which rehearsal is occurring in a naturally optimal manner. A related argument is that rehearsing according to an instructed schedule might be cognitively more demanding than rehearsing freely, thereby impeding participant's ability to engage in additional maintenance processes alongside rehearsal. Accordingly, instructing a cumulative rehearsal strategy will yield an effect that can hardly surpass the optimum level achieved in a free-rehearsal baseline. Proponents of this position argue that rehearsal according to a person's freely chosen schedule would certainly be helpful in comparison to a condition in which articulatory rehearsal was blocked. Testing this conjecture was the aim of Experiment 3.

As discussed in the introduction, several studies have imposed the requirement to articulate irrelevant syllabi or words in order to block rehearsal (i.e., an AS procedure). Many of these studies have demonstrated that AS has a detrimental effect on memory performance, which has been considered as one piece of evidence supporting the role of rehearsal in WM. We have raised objections to this conclusion because the articulation of irrelevant information does not only prevent rehearsal but also introduces interference. In addition, most studies implemented AS throughout the WM trial, thereby preventing not only rehearsal but also the encoding of the phonological properties of the memoranda. Several studies have observed that implementing AS during the encoding phase has a more detrimental effect than when AS is implemented only during the retention interval (Chein & Fiez, 2010; Miles, Jones, & Madden, 1991; Toppino & Pisegna, 2005). This shows that in order to selectively block rehearsal but not phonological encoding, one should prompt AS only during the retention interval. In Experiment 3 we did this to create a no-rehearsal baseline condition.

The goal of Experiment 3 was to test for the conjecture that rehearsal following a person's freely chosen schedule yields better performance compared to a baseline where rehearsal is blocked through AS. To attain this goal, we asked participants to complete three conditions. The *Fast* condition, presenting one word per second, replicated the corresponding condition in Experiment 1b. The other two conditions used a slow presentation rate. In the *Slow-AS* condition, participants were required to read aloud the syllabi "babibu" continuously during the inter-word gaps. This procedure allowed us to block rehearsals from occurring during the inter-stimulus interval while not introducing competition with the phonological encoding of the memoranda. In the third condition, *Slow-R*, we yoked the rehearsal each participant in Experiment 3 to the overt rehearsal of a participant in Experiment 1b (Dewar, Brown, & Della Sala, 2011; Tan & Ward, 2000). Each participant in Experiment 3 received the exact same memory lists, and also the same pattern of cumulative rehearsals carried out by their yoked partner during the inter-word gaps in the Slow condition of Experiment 1b. The words rehearsed by the partner were presented one at a time on the computer screen, thereby eliminating the need for participants to retrieve the to-be-rehearsed words from memory. Participants only had to read the words out from the screen as they were presented.

We see four advantages of this yoked-rehearsal procedure. First, it enables us to experimentally induce a rehearsal schedule that matches the one freely chosen by another person in the same situation. If the mixture of rehearsal strategies freely chosen by participants is optimal, then this should result in the best memory performance rehearsal can produce. Second, the requirement to read words from the screen equated the AS and cumulative rehearsal conditions in all regards, except for the nature of the information that was being articulated, namely irrelevant information or the memoranda. Third, because participants only were instructed to read the words from the screen, the Slow-AS and the Slow-R conditions are comparable with respect to their cognitive demands, and hence with regard to participants' opportunity to use other strategies for maintaining information in WM (such as elaboration or attention-based refreshing). This mitigates the risk of conflating the effect of our manipulation of rehearsal with a qualitative shift in the use of maintenance strategies. Fourth, because the rehearsal schedule was presented to participants instead of self-generated, this removes the dependency of successful rehearsal on the ability to remember the list, which otherwise may mask the beneficial effects of rehearsal.

We predict that if rehearsals are beneficial to serial recall, then we should observe that the Slow-R condition in Experiment 3 yields better performance than the Slow-AS condition. Furthermore, the comparison of the Fast and the Slow-AS conditions allowed us to assess whether the beneficial effect of a slow presentation rate can be observed even when rehearsal is blocked. If slow presentation rates only benefits performance due to providing more opportunities for rehearsal, then blocking rehearsal should eliminate or even reverse the beneficial effects of slower rates. If we assume decay, the longer retention interval in the Slow-AS condition should render memory even worse in that condition than in the Fast condition. The same prediction can be made on the assumption that articulatory suppression in the Slow-AS condition impairs memory.

4.1. Method

4.1.1. Participants, materials, and procedure

A new sample of 24 students was invited to take part in a 60-min session in exchange of course credit or 15 CHF. Participants completed three experimental conditions. In the Fast condition, a list of six words was presented at a rate of one word per second (0.9 s onscreen, 0.1 off-screen). All words were printed in black. In the other two conditions, the memoranda were presented at a rate of one word every 5 s (0.9 onscreen, 4.1 off-screen). Participants were instructed that the memory words will also appear in black, and that they should remember them in their order of presentation. In the interval between two memory words, words printed in red appeared in the middle of the screen. Participants were instructed to read the red words aloud into a microphone attached to a headset (and their verbal responses were recorded), but they did not need to try to remember them. In the Slow-AS condition, participants were instructed that the printed red words will always be the three-syllable utterance "babibu", and they should read this repeatedly aloud, once each time this character string re-appeared on the screen. "Babibu" was flashed six times during the inter-stimulus interval to indicate that participants should articulate it six times. In the Slow-R condition, the red words were the rehearsals performed by one of the participants in the Slow condition of Experiment 1b (to which the present participant was yoked). Because there was variability in the rehearsals carried out by participants in Experiment 1b, we edited their cumulative rehearsals in the following ways. If participants rehearsed more than 6 words in the 4.1 s inter-stimulus interval, only the first 6 rehearsed words were presented to the yoked participant. This provided a relatively comfortable rate of reading, while at the same time providing the possibility for rehearsal of the entire list during the gap. When the yoked participant in Experiment 1b rehearsed less than 6 words, we filled up the remaining time with repetitions of "babibu". This ensured that participants were not using the remaining inter-word time to rehearse on their own. In cases where the yoked rehearsal schedule included other words than the list words, these words were presented for reading. Hence, the cumulative rehearsals implemented in this condition were not error free, but they were representative of what participants do when attempting to rehearse by themselves. In this condition, participants were instructed that the red words would mostly constitute repetitions of the memoranda, but that occasionally they may be requested to read a word that was not part of the memory list, or to repeat "babibu" aloud.

The pool of words and the recall procedure was identical to the one implemented in Experiment 1b. Participants completed four practice trials and 20 test trials in each experimental condition, and these conditions were completed across different blocks.

4.2. Results

Given that both the rehearsal and the AS procedure was controlled by visually displaying which words participants uttered, the only relevant dependent variable in Experiment 3 was the proportion of correctly recalled words in each condition as a function of serial position. Fig. 7 presents the relevant data, and Table 3 presents the evidence obtained in a BANOVA entering condition, material, and serial position as predictors.

Our main interest was in comparing performance between the two slow conditions: the Slow-AS condition in which rehearsal was blocked through AS, and the Slow-R condition in which participants carried out a rehearsal schedule freely chosen by another person. Fig. 7 shows that these two conditions yielded similar performance levels. Overall, there was weak evidence against an effect of condition in the BANOVA: The BF for the condition effect in Table 3 shows that the Null is preferred by a factor of about 2. Condition did not enter in any interactions with the other variables, serial position or material. As for the other experiments reported here, we performed a more focused test of the directional hypothesis that rehearsal should yield better performance than AS by performing a one-sided *t*-test on the overall recall accuracy in the task. This analysis showed substantial evidence against rehearsal yielding better performance than AS, $BF_{10} = 0.09$.

Our second analysis assessed whether the effect of presentation rate vanishes or reverses when the inter-stimulus interval is filled

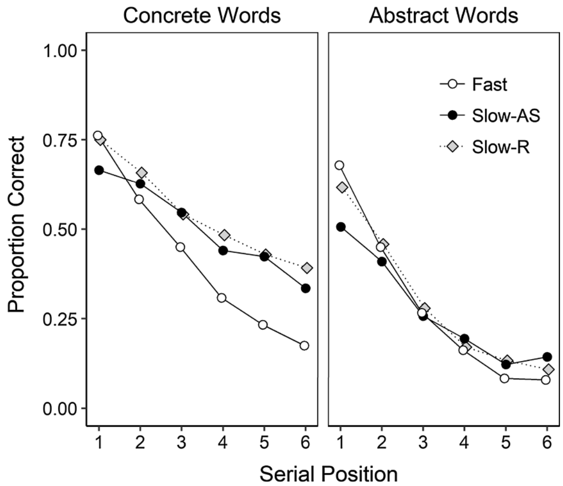

Fig. 7. Serial position curves for the three conditions in Experiment 3. In the Fast condition, words were presented every 1 s. In the Slow-AS condition, one new memory item was presented every 5 s, and in between the memoranda, participants repeated "babibu". Finally, in the Slow-R condition the inter-stimulus interval was filled with repetitions of the memoranda, which constituted the free rehearsals of a yoked participant in the Slow condition of Experiment 1b.

Table 3
Bayes factors quantifying the strength of evidence in the data for including the main effects and interactions of the variables manipulated in Experiment 3.

Predictor	Condition Contrast	
	Slow-AS vs. Slow-R	Fast vs. Slow-AS
Condition	0.49	0.002
Concreteness	10605.56	17211.05
Serial Position	1.85×10^{32}	3.76×10^{31}
Condition × Concreteness	0.18	221.54
Condition × Serial Position	0.31	5.70×10^{6}
Concreteness × Serial Position	67.04	0.52
Three-way interaction	0.08	0.02

with AS. For that, we contrasted the Fast and Slow-AS conditions. Fig. 7 shows that the Slow-AS condition yielded better performance than the Fast condition when the word-lists consisted of concrete, highly imageable words, but not for abstract, low imageable words. Accordingly, the evidence presented in Table 3 indicates strong evidence against a main effect of condition, coupled with strong evidence for an interaction between condition and concreteness. We followed up this analysis with two one-tailed t-tests comparing overall recall performance between these conditions separately for each word-list type. There was strong evidence for better performance in the Slow-AS than the Fast condition for concrete words, $BF_{10} = 9.09$; conversely, the evidence was substantial against better performance in the Slow-AS than Fast condition for abstract words, $BF_{10} = 0.26$. These results shows that the beneficial effect of a slower presentation rate can also be observed when rehearsal is blocked, but this effect occurs only for concrete, highly imageable words.

4.3. Discussion

Experiment 3 controlled for the articulations participants carried out during the interval between presentations of two memory words. This allowed us to test whether repeated articulation of the memory list yields a benefit compared to articulating irrelevant information. Experiment 3 showed that rehearsing the list according to the schedule chosen by a yoked participant in a corresponding condition added no benefit to memory performance compared to saying "babibu" continuously. This result clearly goes against the hypothesis that rehearsal is beneficial to memory for serial order.

Articulatory suppression during the inter-word gaps did not impair memory relative to the Fast condition. In part, this absence of an AS effect could be explained by the fact that – in contrast to most other AS experiments – we asked for AS only during the inter-word gaps, not during encoding of the list items. Still, previous studies in which AS was performed only in the retention interval substantially impaired serial recall (Chein & Fiez, 2010; Lewandowsky et al., 2010; Miles et al., 1991; Oberauer & Lewandowsky, 2008; Toppino & Pisegna, 2005). One difference between those earlier studies and the present experiment is that the earlier studies tested serial recall of consonants, whereas here we tested serial recall of words. Based on that observation, we propose the following explanation for why we found no detrimental effect of AS in the present experiment: Articulation of irrelevant verbal material during the retention interval adds phonological representations to WM that interfere with the phonological representations of the memoranda. Memory for letters must rely nearly exclusively on phonological representations, whereas memory for words can rely more strongly on lexical and semantic representations that suffer no interference from the articulation of meaningless syllable strings. Moreover, the Slow-AS condition involved much longer inter-word intervals that participants could use to elaborate the words, thereby improving their non-phonological representations even further. The letters that served as memoranda in previous studies did not benefit much from this additional time because they are harder to enrich through elaboration than words. Therefore, elaboration during the longer inter-word intervals in the Slow-AS condition compared to the Fast condition could be sufficient to fully compensate for the interfering effect of AS when the memoranda are words, but not when they are letters.

Support for the assumption that participants used the slow presentation rate to elaborate the words comes from our finding that serial recall was even better in the Slow-AS condition than the Fast condition, but only for concrete (and highly imageable), not for abstract (less imageable) words. This interaction is what we should expect if the benefit of slow presentation rate is due to elaboration, and abstract words are difficult to elaborate. The same interaction between presentation rate and concreteness has also been observed in Experiment 1a and 1b. In contrast to Experiments 1a and 1b, however, in Experiment 3 the benefit of the slow rate vanished for abstract words. This result might be interpreted as indicating that abstract words do require rehearsal to benefit from additional inter-word time. This is not a satisfactory explanation because it is incongruent with the observation that the Slow-R and Slow-AS condition did not differ for abstract word-lists, and this should have been the case if rehearsal was the mechanism behind the benefit of slow rates for this type of material. Alternatively, it may be that elaboration of abstract words depends more on verbal representations (e.g., forming sentences to link the list words) than processing of concrete words – because the latter can also rely on visual representations, as proposed by the dual-coding theory (Paivio, Walsh, & Bons, 1994). For example, Paivio et al. (1994) observed that recall of abstract word-pairs required more verbal relational integration than recall of concrete word-pairs. Accordingly, it is possible that the reading demand in both the Slow-R and the Slow-AS condition prevented the use of additional verbal processes that were responsible for the benefit of slower rates for abstract words. This assumption could also explain the finding in Experiments 1 and 3 that concrete words benefit more from slower rates: Verbal processing supports memory performance for both concrete and abstract words, but only concrete words can benefit from additional visual processing. Removing the verbal support eliminates the benefit of slower rates for abstract words and reduces the benefit obtained for concrete words. However, concrete words can still benefit from the activation of visual representations, and hence there is still a benefit under slow rates even under suppression for these materials. This hypothesis is also congruent with the observation that slower presentation rates benefit memory for visual materials (Ricker & Hardman, 2017). Considering the complete pattern of effects obtained here, we argue that our findings are inconsistent with the assumption that increased rehearsal opportunities cause the benefit of slower presentation rates.

Altogether, the data of Experiment 3 provide further evidence against the assumption that rehearsal helps serial recall. The results question two hypotheses. First, the hypothesis that rehearsal according to a spontaneously chosen schedule benefits memory compared to a condition in which rehearsal is impossible was not supported by the data. Second, the benefit of slow presentation rates is best explained as arising from a combination of verbal and visual elaboration, not from rehearsal.

5. Joint analyses across experiments

5.1. Effects of rehearsal on free recall score

In serial recall tasks, part of the errors people make are of confusing items (i.e., order errors) and part of these errors are related to omissions or extra-list recalls (i.e., item-memory errors). Rehearsal may increase the chance that participants recall list-items overall (item-memory benefit), but have little impact on order memory. Because we only computed a serial recall score, the data we presented so far does not allow one to assess for the possibility of an item-memory benefit as a function of rehearsal. An item-memory benefit seems plausible given that previous studies have found evidence for a beneficial effect of rehearsal on free recall (Laming, 2008; Rundus, 1971; Tan & Ward, 2000). To assess for this possibility we compute a free recall score for the data of all experiments by

Table 4
Mean free recall accuracy (and 95% within-subjects confidence intervals) in each condition of Experiments 1–3.

	Fast		Slow		Slow-C		Slow-F		Slow-AS	
Concrete words										
E1a	0.72	[0.68, 0.76]	0.85	[0.82, 0.89]	0.83	[0.80, 0.86]				
E1b	0.72	[0.69, 0.75]	0.89	[0.85, 0.92]	0.82	[0.80, 0.84]				
E2			0.90	[0.87, 0.94]	0.81	[0.77, 0.85]	0.83	[0.79, 0.87]		
E3	0.63	[0.58, 0.69]	0.77	[0.74, 0.81]					0.66	[0.60, 0.71]
Abstract words										
E1a	0.59	[0.56, 0.63]	0.72	[0.69, 0.76]	0.73	[0.68, 0.77]				
E1b	0.57	[0.54, 0.60]	0.69	[0.66, 0.72]	0.65	[0.62, 0.68]				
E2			0.76	[0.72, 0.79]	0.69	[0.65, 0.73]	0.65	[0.60, 0.71]		
E3	0.52	[0.48, 0.56]	0.58	[0.54, 0.63]					0.45	[0.40, 0.50]

Note. In Experiment 3, we entered the data of the Slow-R condition in the Slow condition column, given that in this condition participants were rehearsing according to the rehearsal patterns displayed by participants in the Slow condition of Experiment 1b.

calculating how often list items were recalled across any serial position. Table 4 presents the data of each experiment separately for concrete and abstract words as well as the different experimental conditions.

As shown in Table 4, the comparison of the free recall accuracy between the Slow and Slow-C conditions (Experiments 1a, 1b, and 2) does not indicate an advantage for item memory when participants rehearsed cumulatively. If anything, in two experiments (1a and 2) participants recalled list-items less often in the Slow-C than in the Slow condition. A one-tailed *t*-test comparing overall accuracy (collapsed over list-type) yielded strong support for worse item memory in the Slow-C condition compared to the Slow condition in Experiment 1b (BF_{10} = 27.7) and in Experiment 2 (BF_{10} = 1144.9). Only in Experiment 1a, item memory was similar between these conditions, BF_{10} = 0.13. These results mimic the ones found for the serial recall score in these experiments. For the comparison between rehearsing vs. performing articulatory suppression in Experiment 3, however, we do see an item-memory advantage for rehearsal. A one-tailed *t*-test comparing overall accuracy between these conditions yielded very strong support for worse item-memory in the Slow-AS than in the Slow-R condition (BF_{10} = 1268.8).

All in all, this data indicates that rehearsal can be beneficial to item memory if it is compared to a no-rehearsal baseline (Experiment 3). This is line with the findings reviewed above which provided evidence for a role of rehearsal in free recall tasks (Laming, 2008; Rundus, 1971; Tan & Ward, 2000). Taken together with the serial recall data, these results indicate that rehearsal may increase item accessibility, but it does not improve order memory. Order memory is considered as the principal way of assessing WM functioning, and most models of WM focus on explaining performance in serial recall tasks. All in all, we can only conclude that the role of rehearsal in WM tasks is, at best, constrained to boosting item memory.

5.2. Revisiting the correlation of rehearsal with recall

Tan and Ward (2008) reported a strong positive across-subjects correlation between mean correct serial recall and mean length of cumulative rehearsal in their medium and slow presentation-rate conditions. We also assessed this correlation for each condition of Experiments 1a, 1b, and 2. Our aim was to pool the data across experiments to obtain a more robust estimate of this correlation than can be obtained from the small samples in each individual experiment. We also included the correlation estimates from the study by Tan and Ward (2008), as well as the correlations obtained from an unpublished study from our lab that replicated Experiment 1 using auditory presentation of the words (hereafter referred to as the E auditory).[6] We used the metafor package (Viechtbauer, 2010) implemented in R to estimate the overall correlation using a multi-level random-effects model that assumes that estimates vary from study to study, while taking into account the dependency in some of the estimates (in our case, estimates coming from the same sample of participants).

Fig. 8 shows a forest plot with the correlation coefficients (and 95% CIs around these estimates) for the study/conditions included in the analysis, and the meta-analytic estimate of the correlation coefficient across studies. The overall model shows positive relation between rehearsal and recall. The correlations were, however, highly variable across individual experiments and conditions, reflected in substantial estimates of heterogeneity between data sets (I^2 = 61.2%, p = .001). As we only have a small number of studies in this sample, it is still too early to try to robustly estimate what moderates the size of the correlation between cumulative rehearsal and serial recall.

6. General discussion

Does rehearsal help immediate serial recall? Many theories of working memory for verbal materials assume that it does. In particular, theories assuming decay of representations in WM invariably assign a beneficial role to rehearsal (Anderson, Bothell, Lebiere, & Matessa, 1998; Baddeley, Thomson, & Buchanan, 1975; Camos et al., 2009; Cowan, 1999; Kieras, Meyer, Mueller, & Seymour, 1999). Correlational evidence is consistent with this assumption. Our experiments also replicate the positive correlation between rehearsal – in particular cumulative rehearsal – and recall success in most (albeit not all) instances, as shown in Fig. 8.

Yet, when we experimentally increased the degree to which people engaged in cumulative rehearsal, the net effect on memory was not beneficial. Table 5 summarizes the evidence gathered across all experiments for the effect of rehearsal on serial recall. Across all experiments reported here, we found substantial evidence *against* a beneficial effect of cumulative rehearsal on WM.

The results presented in Table 5 matches the finding from strategy self-report studies that rehearsal is an ineffective maintenance strategy (Bailey et al., 2011; Dunlosky & Kane, 2007). Hence, we can only conclude that the positive correlation between cumulative rehearsal and recall must be driven by a third unobserved variable that affects both serial order memory and the spontaneous occurrence of rehearsals (Brown & Hulme, 1995; Nairne, 2002).

One possibility to salvage the notion that rehearsal helps serial recall is to argue that people, when left to rehearse freely, engage in an optimal amount of rehearsal. Instructing them to rehearse more than they otherwise would merely pushes rehearsal beyond its optimal level, thereby yielding no added benefit. This assumption rests on the premise that in the free-rehearsal conditions every individual calibrates their extent of cumulative rehearsal just right. Our experiments, and that of Tan and Ward (2008), have demonstrated substantial variability in how much people engage in cumulative rehearsal, and that this variability correlates with their serial-recall performance. If every person rehearsed just as much as is optimal for them, then this correlation cannot reflect the presumed causal effect of rehearsal on recall performance. If the correlation reflected the causal effect, it would imply that those individuals who engage in relatively little cumulative rehearsal recalled the lists relatively poorly because they did not rehearse

[6] A summary of the results of this experiment together with the data and analysis script is available at the OSF.

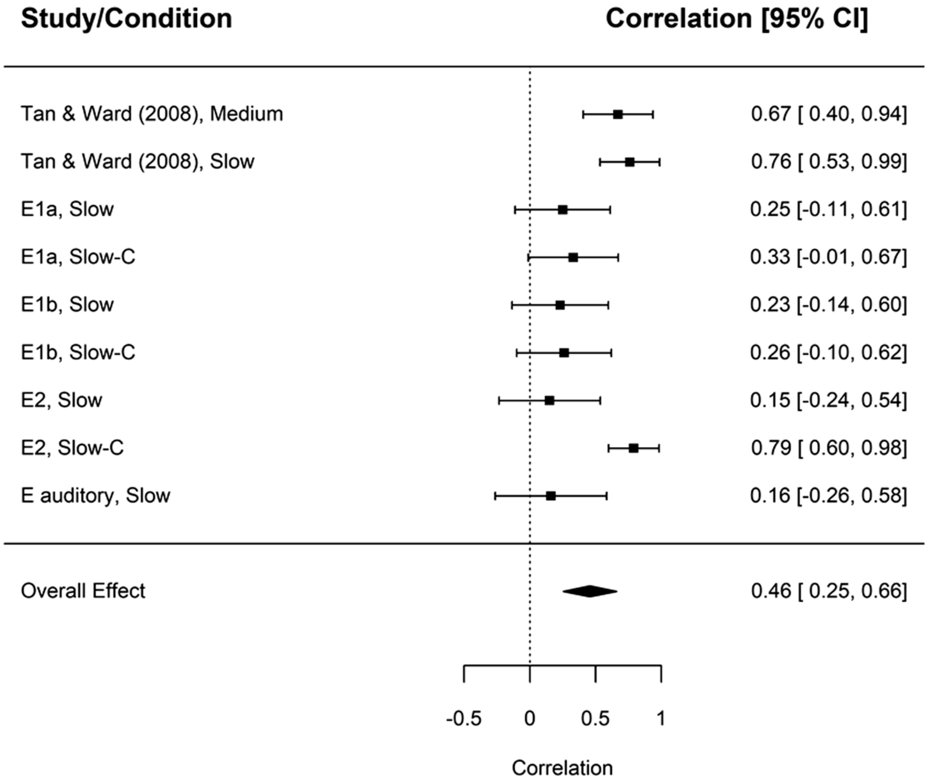

Fig. 8. Forest plot showing the correlation coefficients between mean correct serial recall and mean length of rehearsal sequence across the 11 conditions spanning the studies of Tan and Ward (2008) and the present series of experiments. E = Experiment.

Table 5
Summary of the evidence in favor of a benefit (BF_{10}) of rehearsal on serial recall across all experiments reported here. For convenience, the same evidence is also presented as Bayes factors in favor of the null hypothesis (i.e., $BF_{01} = 1/BF_{10}$).

Experiment	Condition contrast	BF_{10}	BF_{01}
E1a	Slow < Slow-C	0.14	7.14
E1b	Slow < Slow-C	0.10	10
E2	Slow < Slow-C	0.05	20
E2	Slow < Slow-F	0.02	50
E3	Slow-AS < Slow-C	0.09	11.11

Note. The evidence stems from one-sided t-tests performed on overall recall performance averaged across stimulus materials and serial position.

enough. If that were the case, then increasing cumulative rehearsal in these individuals should increase their recall performance. This, in turn, should lift mean accuracy in the entire group, contrary to what we observed.

One criticism that can be raised against this argument is that we are relying on a group effect to determine whether rehearsal is beneficial. Some may argue that increasing cumulative rehearsal could be beneficial to some individuals, but costly to others, thereby yielding a null effect. If that is the case, the effect of the rehearsal instruction on memory should be moderated by two variables: The degree to which participants spontaneously rehearse, and the degree to which they actually increased their cumulative-rehearsal behavior as a function of our instruction. Individuals who spontaneously show relatively little cumulative rehearsal, and those who increase their cumulative rehearsal behavior more strongly in response to the instruction, should show the largest memory improvement through the instruction.

To assess these possibilities, we pooled together the data of Experiments 1 and 2 which contained both a free rehearsal condition (Slow condition) and a cumulative rehearsal instruction condition (Slow-C condition). Fig. 9A plots the average length of cumulative rehearsals spontaneously performed by a given individual in the Slow condition against their own length of cumulative rehearsals achieved in the Slow-C condition. This panel shows that there was substantial variability in the length of spontaneous cumulative rehearsals, and that instructing participants to rehearse cumulatively generally increased the length of cumulative rehearsals they performed (i.e., most of the dots are above the diagonal).

The other two panels in Fig. 9 show the change in serial recall performance between the two slow conditions (i.e., Slow-C

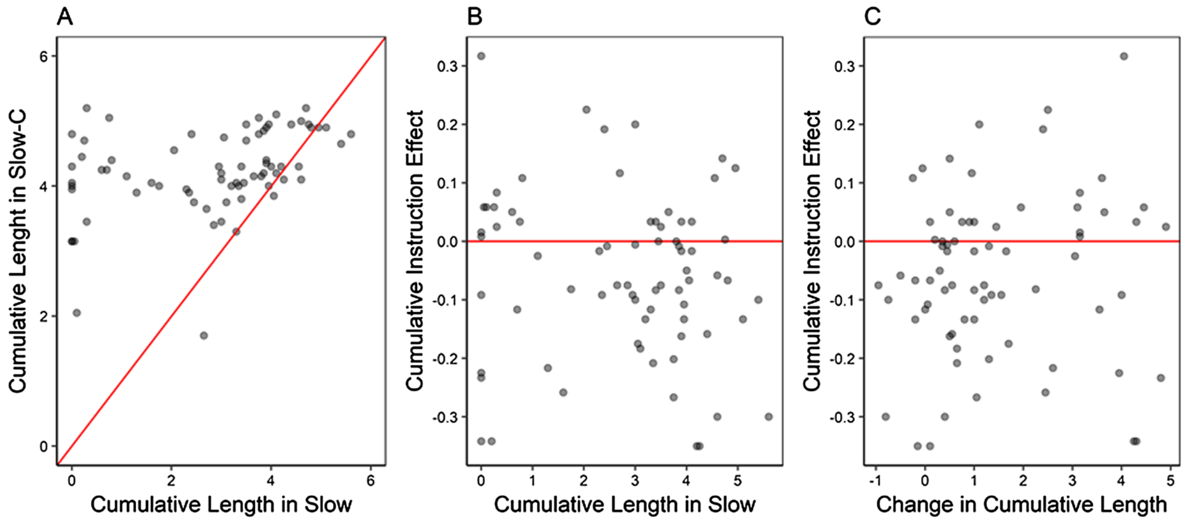

Fig. 9. Panel A: Average length of cumulative rehearsals in the Slow-C condition plotted against the average length of cumulative rehearsals performed in the Slow condition for each individual. Panel B: Change in serial recall accuracy between the Slow and Slow-C conditions plotted as a function of average length of cumulative rehearsal in the Slow condition for each individual. Panel C: Change in serial recall accuracy as a function of the cumulative rehearsal instruction as a function of the change in average cumulative rehearsal length between the Slow and Slow-C condition.

accuracy – Slow accuracy) as a function of the length of cumulative rehearsals in the Slow condition (Fig. 9B), and also as a function of the change in length of cumulative rehearsals between the Slow and Slow-C conditions (Fig. 9C). Both panels show the same pattern: The effect of the rehearsal instruction was centered on zero with a slight tendency for decreases in performance. The variation in the effect of the rehearsal instruction on memory cannot be accounted for by the degree to which participants spontaneously performed cumulative rehearsals ($r = -0.13$, $BF_{10} = 0.17$), nor by how much the cumulative rehearsal instruction motivated them to cumulatively rehearse the lists more ($r = 0.10$, $BF_{10} = 0.14$), despite the fact that the instruction vastly increased participants' tendency to rehearse cumulatively.

The evidence we provide here against the beneficial effect of rehearsal does not only rely on the comparison between the free-rehearsal and instructed-rehearsal conditions. Experiment 3 provides a further strong test of the effect of rehearsal on WM performance by comparing the rehearsal condition against an AS condition. This experiment also addresses the possibility that rehearsal schedules reflecting the typical strategy mixture generated freely by participants might be more helpful than a pure implementation of the instructed strategy. If rehearsal helps WM, even if only slightly, this comparison should have allowed us to detect this benefit. We still found no evidence for a benefit.

Recently, Jarrold (2017) suggested that rehearsal may be helpful if participants rehearse within their capacity limit. For example, if people can only remember three items, rehearsal of 3-item lists would be beneficial, whereas rehearsal of 4-item lists would yield no benefit, and possibly even a cost, because when rehearsing beyond one's capacity one risks introducing errors into the rehearsed list. This is a valid argument but it does not help to explain our findings: When we instructed participants to rehearse cumulatively, they did so more often, and their cumulative rehearsals were longer, on average. Cumulative rehearsal is by definition error-free rehearsal – any overt rehearsal sequence including an error would not be classified as cumulative. When participants were instructed to rehearse cumulatively, they produced more cumulative rehearsals (see Fig. 9A). These observations rule out the possibility that by instructing cumulative rehearsal we pushed participants to rehearse beyond their capacity – rather, this instruction pushed them towards more fully realizing their capacity. If Jarrold (2017) is correct in assuming that rehearsal is beneficial as long as it stays within a person's capacity, then the increased cumulative rehearsal induced by instructions should have been beneficial in the present experiments.

Rehearsal and recall correlate positively not only between individuals, but also between presentation-rate conditions: Increasing the temporal gaps between presented list items (i.e., slowing the presentation rate) increased the degree to which people used cumulative rehearsal, and increased recall accuracy (Tan & Ward, 2008). Does this correlation reflect a causal effect of rehearsal on recall? Probably not, for two reasons. First, the slow-rate benefit was of about equal size for all serial positions despite the fact that items from the beginning of the list were rehearsed much more often than those at the end. In contrast, experimentally increasing cumulative rehearsal led to a selective increase of the primacy part of the list when written recall was required (Experiments 1b and 2). Second, slow presentations rates have been found to benefit serial recall even under AS (see Longoni, Richardson, & Aiello, 1993; but see Grenfell-Essam, Ward, & Tan, 2013 for free recall), and we replicated this effect in Experiment 3, albeit only for concrete, highly imageable words. Furthermore, slower presentation rates are also beneficial for recall of continuously varying visual stimuli that are difficult to rehearse (Ricker & Hardman, 2017).

What, if not rehearsal, could cause the benefit of slow presentation rates? One possible explanation of this effect is that people use elaborative rehearsal: It is possible that the long inter-item interval favored the use of elaboration or imagery. Self-report studies (Bailey et al., 2008, 2009, 2014) suggest that elaboration is a more effective maintenance strategy than rehearsal. Here we found tentative evidence for the elaboration explanation in Experiments 1 and 3: The benefit of slower presentation rate was more pronounced for concrete, highly imageable words that are relatively easy to elaborate compared to abstract, low imageable words. We readily admit that this evidence is not conclusive, because the slow-presentation benefit could be larger for concrete words for other reasons than by facilitating elaboration.[7] We note, however, that the availability of these alternative explanations for the presentation-rate benefit severely weakens the case for attributing this benefit to articulatory rehearsal.

Rehearsal is not an epiphenomenon: At least with typed recall, it leads to the tilting of the serial-position curve towards more primacy and less recency. One explanation of this effect is that rehearsal increases the absolute strength of rehearsed items in memory in proportion to how often they are rehearsed. With more cumulative rehearsal, items at the beginning of the list gain in strength more than items at the end of the list. Most computational models of serial recall assume that the probability of recalling an item is a function of its relative strength among competing recall candidates at the time of test, a process sometimes referred to as competitive queuing (Bullock, 2004; Burgess & Hitch, 2005; Lewandowsky & Farrell, 2008; Page & Norris, 1998). The effect of cumulative rehearsal could be to bias this competition in favor of items in the primacy portion of the list at the expense of items in the recency portion. This effect was observed in the simulations of rehearsal by Lewandowsky and Oberauer (2015): Rehearsal increased primarily the items at the beginning of the list, sometimes rendering their strength overbearing, so that even when a later list item was cued for recall by its serial position, it lost the competition to an earlier list item. Why this effect of rehearsal on the serial position curve was observed only with typed but not with spoken recall is a question we must leave to future research.

7. Conclusion

We investigated the putative link between articulatory rehearsal and WM performance. Across four experiments, we controlled the occurrence of rehearsal by instructing people to rehearse using a cumulative or fixed strategy, and we compared these conditions to a free-rehearsal baseline (Experiments 1a, 1b, and 2) or to an AS baseline (Experiment 3). In all experiments, we found that rehearsal did not improve serial order memory. Our results show therefore that articulatory rehearsal is an ineffective strategy to improve WM. These results call into question one of the main roles assigned to rehearsal in time-based decay models, which is to protect representations from forgetting.

References

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language, 38*, 341–380.
Baddeley, A. (1986). *Working memory.* Clarendon Press/Oxford University Press.
Baddeley, A., & Lewis, V. (1984). When does rapid presentation enhance digit span? *Bulletin of the Psychonomic Society, 22*, 403–405.
Baddeley, A., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575–589.
Bailey, H., Dunlosky, J., & Hertzog, C. (2009). Does differential strategy use account for age-related deficits in working-memory performance? *Psychology and Aging, 24*, 82–92.
Bailey, H., Dunlosky, J., & Hertzog, C. (2014). Does strategy training reduce age-related deficits in working memory? *Gerontology, 60*, 346–356.
Bailey, H., Dunlosky, J., & Kane, M. J. (2008). Why does working memory span predict complex cognition? Testing the strategy affordance hypothesis. *Memory & Cognition, 36*, 1383–1390.
Bailey, H., Dunlosky, J., & Kane, M. J. (2011). Contribution of strategy use to performance on complex and simple span tasks. *Memory & Cognition, 39*, 447–461.
Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.
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*, 796–808.
Bhatarah, P., Ward, G., Smith, J., & Hayes, L. (2009). Examining the relationship between free recall and immediate serial recall: Similar patterns of rehearsal and similar effects of word length, presentation rate, and articulatory suppression. *Memory & Cognition, 37*, 689–713.
Bourassa, D. C., & Besner, D. (1994). Beyond the articulatory loop: A semantic contribution to serial order recall of subspan lists. *Psychonomic Bulletin & Review, 1*, 122–125.
Brodie, D. A., & Murdock, B. B. (1977). Effect of presentation time on nominal and functional serial-position curves of free recall. *Journal of Verbal Learning and Verbal Behavior, 16*, 185–200.
Brown, G. D. A., & Hulme, C. (1995). Modeling item length effects in memory span: No rehearsal needed? *Journal of Memory and Language, New York, 34*, 594–621.
Brown, G. D. A., Sala, S. D., Foster, J. K., & Vousden, J. I. (2007). Amnesia, rehearsal, and temporal distinctiveness models of recall. *Psychonomic Bulletin & Review, 14*, 256–260.
Bullock, D. (2004). Adaptive neural models of queuing and timing in fluent action. *Trends in Cognitive Sciences, 8*, 426–433.
Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*, 551–581.
Burgess, N., & Hitch, G. (2005). Computational models of working memory: Putting long-term memory into context. *Trends in Cognitive Sciences, 9*, 535–541.
Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language, 61*, 457–469.
Camos, V., Mora, G., & Barrouillet, P. (2013). Phonological similarity effect in complex span task. *The Quarterly Journal of Experimental Psychology, 66*, 1927–1950.
Camos, V., Mora, G., & Oberauer, K. (2011). Adaptive choice between articulatory rehearsal and attentional refreshing in verbal working memory. *Memory & Cognition, 39*, 231–244.
Chein, J. M., & Fiez, J. A. (2010). Evaluating models of working memory through the effects of concurrent irrelevant information. *Journal of Experimental Psychology. General, 139*, 117–137.
Cowan, N. (1999). An embedded-process model of working memory. In A. Miyake, & P. Shah (Eds.). *Models of working memory: Mechanisms of active maintenance and*

[7] There have been few attempts to experimentally investigate the effect of elaboration on WM performance. One study from our lab (Bartsch, Singmann, & Oberauer, 2018) found no convincing evidence for a WM benefit when participants were told to elaborate some items of the memory list, although elaboration improved episodic long-term memory. This study used, however, a mixture of concrete and abstract words, and the time to elaborate was shorter (2 s) than the one provided here. We further note that the benefit of elaboration, although not credible in the WM conditions in the study of Bartsch et al. (2018), was in the direction of a benefit.

executive control (pp. 62–101). New York, NY: Cambridge University Press.
Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–114.
Cowan, N. (2011). The focus of attention as observed in visual working memory tasks: Making sense of competing claims. *Neuropsychologia, 49*, 1401–1406.
Daily, L. Z., Lovett, M. C., & Reder, L. M. (2001). Modeling individual differences in working memory performance: A source activation account. *Cognitive Science, 25*, 315–353.
Dewar, M., Brown, G. D. A., & Della Sala, S. (2011). Restoring primacy in amnesic free recall: Evidence for the recency theory of primacy. *Cognitive Neuropsychology, 28*, 386–396.
Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *The Quarterly Journal of Experimental Psychology, 60*, 1227–1245.
Estes, W. K. (1991). *On types of item coding and source of recall in short-term memory. Relating theory and data: Essays on human memory in honor of Bennet B. Murdock*. Hilsdale, NJ: Erlbaum.
Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review, 119*, 223–271.
Grenfell-Essam, R., Ward, G., & Tan, L. (2013). The role of rehearsal on the output order of immediate free recall of short and long lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 317–347.
Gupta, P., & MacWhinney, B. (1995). Is the articulatory loop articulatory of auditory? Reexamining the effects of concurrent articulation on immediate serial recall. *Journal of Memory and Language; New York, 34*, 63–88.
Guttentag, R. E., Ornstein, P. A., & Siemens, L. (1987). Children's spontaneous rehearsal: Transitions in strategy acquisition. *Cognitive Development, 2*, 307–326.
Hellyer, S. (1962). Supplementary report: Frequency of stimulus presentation and short-term decrement in recall. *Journal of Experimental Psychology, 64*, 650.
Hudjetz, A., & Oberauer, K. (2007). The effects of processing time and processing rate on forgetting in working memory: Testing four models of the complex span paradigm. *Memory & Cognition, 35*, 1675–1684.
Humphreys, M. S., Maguire, A. M., McFarlane, K. A., Burt, J. S., Bolland, S. W., Murray, K. L., & Dunn, R. (2010). Using maintenance rehearsal to explore recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 147–159.
Jarrold, C. (2017). Working out how working memory works: Evidence from typical and atypical development. *The Quarterly Journal of Experimental Psychology, 70*, 1747–1767.
Kaakinen, J. K., & Hyönä, J. (2007). Strategy use in the reading span test: An analysis of eye movements and reported encoding strategies. *Memory, 15*, 634–646.
Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*, 773–795.
Kieras, D. E., Meyer, D. E., Mueller, S., & Seymour, T. (1999). Insights into working memory from the perspective of the EPIC architecture for modeling skilled perceptual-motor and cognitive human performance. In A. Miyake, & P. Shah (Eds.). *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 183–223). Cambridge: Cambridge University Press.
Laming, D. (2008). An improved algorithm for predicting free recalls. *Cognitive Psychology, 57*, 179–219.
Lewandowsky, S., Duncan, M., & Brown, G. D. A. (2004). Time does not cause forgetting in short-term serial recall. *Psychonomic Bulletin & Review, 11*, 771–790.
Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. *Psychology of Learning and Motivation, 49*, 1–48.
Lewandowsky, S., Geiger, S. M., Morrell, D. B., & Oberauer, K. (2010). Turning simple span into complex span: Time for decay or interference from distractors? *Journal of Experimental Psychology: Learning, Memory & Cognition, 36*, 958–978.
Lewandowsky, S., Geiger, S. M., & Oberauer, K. (2008). Interference-based forgetting in verbal short-term memory. *Journal of Memory and Language, 59*, 200–222.
Lewandowsky, S., & Oberauer, K. (2015). Rehearsal in serial recall: An unworkable solution to the nonexistent problem of decay. *Psychological Review, 122*, 674–699.
Longoni, A. M., Richardson, J. T. E., & Aiello, A. (1993). Articulatory rehearsal and phonological storage in working memory. *Memory & Cognition, 21*, 11–22.
McFarlane, K. A., & Humphreys, M. S. (2012). Maintenance rehearsal: The key to the role attention plays in storage and forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 1001–1018.
Miles, C., Jones, D. M., & Madden, C. A. (1991). Locus of the irrelevant speech effect in short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 578–584.
Mora, G., & Camos, V. (2013). Two systems of maintenance in verbal working memory: Evidence from the word length effect. *PLoS ONE, 8*, e70026.
Mora, G., & Camos, V. (2015). Dissociating rehearsal and refreshing in the maintenance of verbal information in 8-year-old children. *Developmental Psychology, 6*, 11.
Morey, R. D., & Rouder, J. N. (2014). BayesFactor: Computation of Bayes factors for common designs (Version 0.9.9). Retrieved from <http://CRAN.R-project.org/package=BayesFactor>.
Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*, 251–269.
Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53*, 53–81.
Nishiyama, R., & Ukita, J. (2013). Articulatory rehearsal is more than refreshing memory traces. *Experimental Psychology (Formerly Zeitschrift Für Experimentelle Psychologie), 1*, 1–9.
Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review, 115*, 544–576.
Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review, 19*, 779–819.
Page, M., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105*, 761–781.
Paivio, A., Walsh, M., & Bons, T. (1994). Concreteness effects on memory: When and why? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 20*, 1196–1204.
Palmer, S. E., & Ornstein, P. A. (1971). Role of rehearsal strategy in serial probed recall. *Journal of Experimental Psychology, 88*, 60–66.
Phaf, R. H., & Wolters, G. (1993). Attentional Shifts in Maintenance Rehearsal. *The American Journal of Psychology, 106*, 353–382.
R core team (2014). *R: A language and environment for statistical computing.* Retrieved fromVienna: Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>.
Ricker, T. J., & Hardman, K. O. (2017). The nature of short-term consolidation in visual working memory. *Journal of Experimental Psychology: General, 146*, 1551–1573.
Romani, C., Mcalpine, S., & Martin, R. C. (2008). Concreteness effects in different tasks: Implications for models of short-term memory. *The Quarterly Journal of Experimental Psychology, 61*, 292–323.
Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89*, 63–77.
Rundus, D., & Atkinson, R. C. (1970). Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior, 9*, 99–105.
Schwibbe, M. H. (n.d.). Der Semantische Atlas. Retrieved from <http://kulturkontor-goe.de/semat/semat.htm>.
Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1589–1625.
Tan, L., & Ward, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin & Review, 15*, 535–542.
Toppino, T. C., & Pisegna, A. (2005). Articulatory suppression and the irrelevantspeech effect in short-term memory: Does the locus of suppression matter? *Psychonomic Bulletin & Review, 12*, 374–379.
Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language, 49*, 446–468.
Vallar, G., & Baddeley, A. D. (1982). Short-term forgetting and the articulatory loop. *The Quarterly Journal of Experimental Psychology Section A, 34*, 53–60.
Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software,* 1–48.