AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2021, Vol. 150, No. 4, 633-665
http://dx.doi.org/10.1037/xge0000957

Simple Spans Underestimate Verbal Working Memory Capacity

Pierre Barrouillet and Simon Gorin
Université de Genève

Valérie Camos
Université de Fribourg

Verbal working memory (WM) has been assumed to involve 2 different systems of maintenance, a phonological loop and a central attentional system. Though the capacity estimate for letters of each of these systems is about 4, the maximum number of letters that individuals are able to immediately recall, a measure known as simple span, is not about 8 but 6. We tested the hypothesis that, unaware of the dual structure of their verbal WM, individuals underuse it by trying to verbally rehearse too many items. In order to maximize the use of verbal WM, we designed a new procedure called the maxispan procedure. When performing an immediate serial recall task, participants were invited to cumulatively rehearse a limited number of letters, and to keep rehearsing these letters until the end of the presentation of the list in such a way that the following letters can no longer enter the phonological loop and must be stored in the attentional system. As we expected, in 3 successive experiments, the maxispan procedure resulted in a dramatic increase in spans compared with the traditional simple span procedure, with spans approaching 8 when the to-be-rehearsed letters were presented auditorily and the following letters visually. These results indicate that simple spans, which have been used for more than a century in intelligence tests and are assumed to measure the capacity of short-term memory (STM), actually reflect the complex interplay between different structures and cognitive processes.

Keywords: verbal working memory, simple spans, verbal rehearsal, phonological loop, executive loop

In his theoretical note about the famous Miller's (1956) article "The magical number seven, plus or minus two," Cowan (2015a) noted that one of the most important constraints that science faces is the restriction of topics that are pursued, and he pleaded for returning to research on the topic of the numerical limit of capacity in working memory (WM) introduced by Miller. Sixty years after its publication, Miller's article is still one of the 10 most often cited works in psychology (Ho & Hartley, 2016). Accordingly, every student knows that the capacity of immediate verbal memory is about seven, plus or minus two. As noted by Cowan, Morey, and Chen (2007), even individuals who know very little about psychology are still likely to have heard or read that people can keep in mind about seven items, this number being part of psychological folk wisdoms. The identification of seven as the limit of the human processing system capacity has not only suffused folk psychology, but also spread into other domains of scientific psychology, such as developmental psychology (Pascual-Leone, 1970), educational psychology (Kirschner, Sweller, & Clark, 2006; van Gog & Paas, 2008), and applied psychology (Kareev, 2000). The magical number seven is also evoked in education (Sweller & Chandler, 1994; Sweller, van Marrieboer, & Paas, 1998), management (Van de Ven, 1986), or computing science (Jones, Ravid, & Rafaeli, 2004) to name few of these fields.

Though Miller (1956) launched the modern research on immediate memory, psychologists had surmised long before Miller's article that measuring this type of memory could tell us something about human mind, its functioning (or dysfunction) and development. The measure of immediate verbal memory has indeed played a particular role in the history of psychology. Richardson (2007) mentions that Holmes (1871) was probably the first to assess the immediate recall of lists of digits or letters, a task that Cattel (1890) included in an early mental test. From Binet and his scale for children's intelligence (Binet & Simon, 1904), which was the

Editor's Note. Richard Allen served as the guest editor for this article.—NC

This article was published Online First October 5, 2020.

Pierre Barrouillet and Simon Gorin, Faculté de Psychologie et des Sciences de l'Éducation, Université de Genève; Valérie Camos, Département de Psychologie, Université de Fribourg.

The authors' contributions are reported following the CRediT taxonomy (see https://casrai.org/credit/). Conceptualization: Pierre Barrouillet and Valérie Camos; Formal analysis: Simon Gorin; Investigation: Simon Gorin; Methodology: Pierre Barrouillet, Simon Gorin, and Valérie Camos; Project administration: Simon Gorin; Software: Simon Gorin; Supervision: Pierre Barrouillet and Valérie Camos; Validation: Simon Gorin; Visualization: Simon Gorin; Writing—original draft: Pierre Barrouillet, Simon Gorin, and Valérie Camos.

The authors declare that the research was conducted in the absence of any competing interest that could be construed as a potential conflict of interest and that the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the article. Raw data, scripts for data processing and figures generation, and statistical output files are available via the Open Science Framework https://osf.io/7k52n. Part of these results have been presented at the 59th annual meeting of the Psychonomic Society in New Orleans, November 15–18, 2018.

Correspondence concerning this article should be addressed to Pierre Barrouillet, Faculté de Psychologie et des Sciences de l'Education, Université de Genève, 40 Boulevard du Pont d'Arve, 1205 Genève, Switzerland. E-mail: pierre.barrouillet@unige.ch

first intelligence test and has been the matrix of many similar scales (e.g., the Stanford-Binet test), to the most contemporaneous and widespread psychometric tools (Wechsler, 2014), measuring the number of items that children or adults are able to immediately recall in correct order has been part of virtually all the intelligence scales.[1] This presumed relationship between intelligence and immediate memory span has been the object of continued inquiry, scholars studying individual differences having demonstrated strong correlations between the two constructs (Colom, Flores-Mendoza, Quiroga, & Privado, 2005; Gignac & Weiss, 2015), while brain imaging techniques revealed common cerebral substrates (e.g., Colom, Jung, & Haier, 2007). Thus, measuring individuals' immediate verbal memory is nowadays part of usual psychological and neuropsychological assessment.

However, the digit span task has been used for so long by practitioners, and Miller's (1956) article has been so routinely cited that nobody seems to have questioned the implicit assumption that simple spans[2] accurately reflect the capacity of the mnemonic system they are intended to measure, immediate verbal memory, except for revising this estimation downward (e.g., Cowan, 2001). Contrary to subsequent reappraisals that led to shrink this capacity to smaller and smaller values, the present study, based on a theoretical analysis of what is known about WM structure and functioning, tested the provoking hypothesis that simple span measures actually do not overestimate, but *underestimate* the capacity of verbal WM. As we will see, what this century-old psychological tool measures might be less straightforward than the phrase "simple span" seems to convey.

On Some Reappraisals of Miller's (1956) Estimate

It is worth to note that the limit to seven of immediate verbal memory mainly holds for digits. Dempster (1981) reported that it is lower for letters (about six) and even lower for words (about five), this latter limit depending on the length of these words[3] (Baddeley, Thomson, & Buchanan, 1975; Lewandowsky & Oberauer, 2008). Apart from variations due to the nature of the memoranda, Miller's (1956) article was rapidly followed by a series of proposals assuming that the limit is not seven, but three or four (Broadbent, 1975; Henderson, 1972; Sanders, 1968; Sperling, 1960). Although other authors argued that there is a limit to four in WM capacity (e.g., Halford, 1993; Halford, Wilson, & Phillips, 1998), Cowan (2001) is nowadays considered as the main proponent of this hypothesis, as he defined the conditions under which this pure storage capacity limit can be observed and proposed a theoretical framework accounting for this limitation that is supported by empirical evidence from a variety of domains (see Cowan, 2015b, for a review). Concerning immediate verbal memory, the conditions under which this pure capacity can be observed involve preventing the recoding of the stimuli into larger chunks, limiting long-term memory processes by using over and over the same stimuli that have to be recalled in correct serial order, and by blocking any articulatory rehearsal of the memoranda through the requirement of the concurrent articulation of a syllable. In Cowan's model, the limit to four (or three) would correspond to the capacity of the focus of attention that is able to maintain up to four chunks of information (Cowan, 1999, 2005, 2010). This hypothesis has been verified in the domain of visuospatial WM (Cowan, Blume, & Saults, 2013; Cowan, Saults, & Blume, 2014; Luck & Vogel, 1997), verbal WM (Chen & Cowan, 2005, 2009; Cowan, Rouder, Blume, & Saults, 2012) as well as when visual and verbal items are simultaneously stored (Morey & Cowan, 2004, 2005; Saults & Cowan, 2007).

Explaining Why Verbal Spans Are Higher Than Four

If the pure storage capacity of WM is about three or four chunks as Cowan (2001) assumes, the question is to understand how the span of immediate verbal memory reaches a mean of six or seven units (Dempster, 1981). This latter capacity has often been attributed to the phonological loop as described in Baddeley's multicomponent model of WM (Baddeley, 1986, 2007; Baddeley & Hitch, 1974). This model assumes that verbal information is maintained in a phonological loop consisting in a phonological store and an articulatory loop able to reactivate decayed phonological traces through their articulation. Within this approach, short-term memory (STM) is limited in duration because phonological traces suffer from temporal decay within the phonological store. Individuals appear to be able to immediately recall as many verbal items as they are able to articulate in 2 s (Baddeley et al., 1975). The idea that simple spans reflect the capacity of a single store or system is not limited to Baddeley's phonological loop model. Formal models of the phonological loop have been proposed to account for immediate serial recall (ISR), the paradigm used to measure simple spans (Burgess & Hitch, 1992, 1999, 2006), but several models that sometimes differ strongly from the phonological loop model have nonetheless assumed that ISR is governed by a single system or mechanism for maintaining information in the short term (e.g., Botvinick & Plaut, 2006; Brown, Preece, & Hulme, 2000; Henson, 1998; Lewandowsky & Farrell, 2002; Oberauer & Lewandowsky, 2008; Page & Norris, 1998). This can remain true even when these models deny the existence of a STM system separated from long-term memory (e.g., Nairne, 1990).

However, it might be that simple spans do not reflect the capacity of a single system, but of an assembly of systems involved in verbal WM. This view is in line with the modified version of the multicomponent model proposed by Baddeley (2000) who suggested to add a new component, the episodic buffer, able to maintain verbal information. Thus, seven could correspond to the total number of items maintained in the phono-

[1] In the 1911's version of the Binet and Simon's intelligence metric scale, the preliminary survey aiming at a raw assessment of the intellectual level in order to avoid the subsequent administration of too easy items contained 35% of items involving immediate verbal recall, and 20% concerned repeating digit sequences.

[2] By simple span, we refer to the maximum number of items an individual is able to recall in immediate serial recall tasks in which lists of items presented auditorily or visually must be recalled in correct order. Because simple spans are usually considered as a measure of short-term memory capacity, we use in this article the phrases *immediate verbal memory* and *verbal short-term memory* interchangeably. For sake of simplicity, and because the present article does not aim at addressing the complex question of a possible distinction between short-term and working memory, we follow Cowan's assumption that the simple spans studied by Miller (1956) reflect verbal working memory capacity and consequently use interchangeably the phrases *verbal working memory* and *verbal short-term memory.*

[3] In the present study, we used letters instead of digits because lists of more than nine items were presented, and we wanted to avoid presenting lists with repeated items.

logical loop and the episodic buffer, and offloaded at recall in their order of storage. It is worth to note that, according to Baddeley's model, verbal information such as letters or digits could also be stored in a visual format in the visuospatial sketchpad, even when auditorily presented. In line with this idea, Cowan (1988; see also Zhang & Simon, 1985) has suggested that some chunks of information could be maintained through verbal rehearsal while others could be retained nonverbally. At the very least, the necessity advocated by Cowan (2001) to block verbal rehearsal in order to assess the pure capacity storage of WM presupposes that some items could be maintained through verbal rehearsal outside of the focus of attention.

This hypothesis of the contribution of different systems, already put forward by Craik (1971), is addressed in several contemporary theories of WM. Inspired by Baddeley's multicomponent model, Logie (2011; Logie, Belletier, & Doherty, in press) suggested that multiple systems contribute jointly to WM capacity. A range of domain-specific components allow to sustain the online processing and short-term storage of information, and people may deploy different combinations of components depending on the task characteristics. From a different perspective, Unsworth and Engle (2007) have proposed that a primary memory comparable to the focus of attention in Cowan's model can hold up to four items. However, when more items have to be maintained, as in simple span tasks, some items would be displaced into a secondary memory from which they could be retrieved through a cue-dependent search process (Unsworth, Spillers, & Brewer, 2010). In a similar way, the time-based resource-sharing model (TBRS; Barrouillet & Camos, 2015) assumes that there exist two mechanisms of maintenance for verbal information. Along with a phonological loop in which information is maintained through verbal rehearsal, a central system described as an executive loop can hold items through attentional refreshing (Camos, 2017). This central system is akin to the episodic buffer in Baddeley's multicomponent model and the focus of attention in Cowan's model. Empirical evidence suggests that both systems contribute independently to recall performance in WM span tasks (Camos, Lagner, & Barrouillet, 2009: see Camos, 2015, 2017, for reviews). Thus, the fact that immediate verbal memory span is far higher than the pure storage capacity limit of WM (i.e., four according to Cowan, 2001) can be accounted for by assuming that different stores or systems of maintenance come into play in simple span tasks.

However, this hypothesis raises new questions that have not yet received definite answers. First, if verbal WM involves several systems, for example a rehearsal-based phonological loop and an episodic buffer or an attention-based executive loop, how these systems might jointly contribute to performance remains unclear, as Lewandowsky and Oberauer (2015) recently noted. Is a first segment offloaded from the phonological loop and a subsequent string of items retrieved from the other system? Or, alternatively, do different systems collaborate to maintain the same items? In this case, does each system hold a copy of each of them? Second, whatever the unitary or composite structure of verbal WM, does articulatory rehearsal play a causal role in immediate memory or is it a mere epiphenomenon, as Lewandowsky and Oberauer (2015; see also Souza & Oberauer, 2018) have argued? However, in this latter case, why does the span for immediate memory (i.e., six or seven) exceed the pure storage capacity (i.e., four)? Indeed, if rehearsal is an epiphenomenon, immediate memory for items that can be rehearsed should not exceed pure storage capacity that is estimated under articulatory suppression, that is, when blocking rehearsal, according to Cowan (2005). Beyond these issues is another question, directly related with the problem of the numerical limit of WM capacity. If several systems jointly contribute to verbal WM spans, for example a phonological lop and an attentional system akin to the focus of attention in Cowan's (2005) model, why are spans usually reported in the literature lower than what could be predicted from the capacity of the different systems involved?

It Does Not Add Up: 4 + 4 = 8, Not 6

As we noted above, adults' verbal STM span for letters is about six (Dempster, 1981). However, if we assume that verbal WM spans result from the contribution of a phonological loop and a central attentional system like the executive loop in the TBRS model (Camos, 2015, 2017) or the episodic buffer from the multicomponent model (Baddeley, 2000), and if we simply add up the estimated capacity of these two systems, we will see that the sum exceeds the simple span for letters.

In studying the effect of storage on processing in WM, Vergauwe, Camos, and Barrouillet (2014) had participants performing a Brown-Peterson task in which they were presented with a series of letters to be remembered and asked to perform a parity judgment task on digits presented successively on screen for a period of 12 s before recall. Vergauwe et al. (2014) were interested in the effect of the memory load on the response times (RTs) to the parity task. They observed that, under articulatory suppression preventing verbal rehearsal, RTs to the parity task increased linearly with memory load from one letter onward, suggesting that attention was involved in maintaining the letters. Under articulatory suppression, participants proved able to reliably recall up to four letters in correct order after the 12-s parity task interval, which according to Cowan (2001) corresponds to the capacity of the focus of attention. Interestingly, when the requirement of articulatory suppression was released, it appeared that participants were able to maintain up to four letters without any effect on parity judgment RTs. It is only when more than four letters were to be remembered that RTs increased. This suggests that up to four letters could be maintained in a system that does not rely on attention. Vergauwe et al. (2014) proposed that these four letters were maintained in the phonological loop. Interestingly, when studying rehearsal strategies in ISR, Tan and Ward (2008) observed that what they called the rehearsal set, the set of words cumulatively rehearsed in forward order after the presentation of each stimulus item, tended to be limited to four items. Beyond this limit, participants appeared to refrain from performing cumulative rehearsal on the entire list, or rehearsed incomplete or shorter sequences. Thus, the phonological loop hypothesized by Baddeley (2000) and the TBRS model seems to have approximately the same capacity of four items as the central attentional system theorized as the attentional focus by Cowan (2001) and the executive loop in the TBRS model.

Consequently, if both systems are separate and work jointly for maintaining verbal information in WM as suggested by Camos (2015, 2017; Camos et al., 2009), an optimal combination of the two systems should lead to a span of eight letters, far higher than the simple span for letters traditionally observed, which is six in Dempster's (1981) review. Of course, interference between the two systems during maintenance or output could occur, reducing the total capacity of immediate memory to the span traditionally observed. However, it is also possible that, unaware of the structure and functioning of their WM, people use it in a suboptimal way when performing traditional verbal simple span tasks. Thus, the resulting performance would underestimate the total capacity of WM for verbal material. Because people are mainly aware of verbal rehearsal as an efficient way for maintaining verbal information, it is possible that they try to rehearse too many items, this attempt resulting in unsystematic and incomplete rehearsal as Tan and Ward (2008) observed when more than four items are presented. The number of items rehearsed exceeding the capacity of the phonological loop, attention probably intervenes for holding the extra items between two consecutive articulations. This circulation of items between the two systems, instead of a separate storage, might have deleterious effects resulting in the discrepancy between the theoretical and the observed letter spans.

The Present Study

The aim of the present study was to test the hypothesis that verbal WM has a dual structure that remains underused by individuals in simple span tasks. Thus, instructions leading to a better use of the bipartite structure of verbal WM should lead to recall performance approaching a theoretical optimum corresponding to the addition of the capacities of the two systems, namely eight items for letters. According to the empirical evidences presented above, these instructions should induce the maintenance of two distinct blocks of items into separate systems, one block in the phonological loop through verbal rehearsal and the other in the executive loop through attentional refreshing, avoiding any transfer of items between the two systems. A good way to reach this goal in the context of an ISR task would be to instruct participants to cumulatively rehearse aloud a *limited* number of items that does not exceed the capacity of the phonological loop (about four, say from three to five), and to keep rehearsing them until recall while studying the following items[4] (see Figure 1 for an illustration of this method). This procedure should avoid the storage of an excessive number of items in the phonological loop, the continuous rehearsal of the first items creating an articulatory suppression preventing further storage in this system. Consequently, the following items should enter the executive loop and benefit from an attention-based maintenance without any risk of transfer between the two maintenance systems. At recall, the rehearsed items would be output first before recalling the items retrieved from the attentional system. This method, that we call the *maxispan* procedure, should result in letter spans higher than what is usually observed (i.e., about six) and approaching the theoretical optimum of eight. The prediction that the maxispan procedure should lead to higher letter spans than the traditional simple span tasks in which participants are just instructed to recall the letters in correct order without any specific instructions was tested in three experiments.

Because the maxispan procedure requires encoding items under the articulatory suppression created by rehearsing the first presented items, a first experiment involved the visual presentation of both the to-be-rehearsed and the following letters. Presenting letters visually should avoid the conflict between the following letters auditorily presented and the to-be-rehearsed letters articulated aloud. A second experiment tested the effectiveness of the maxispan method in the least favorable condition by presenting all the letters auditorily, in such a way that the last auditory items were to be encoded while articulating aloud the first segment of the list. Despite this drawback, a maxispan higher than the simple span would represent a strong evidence in favor of the hypothesis of a bipartite verbal WM composed of a phonological loop and an attentional system. Finally, the maxispan procedure was tested in the theoretically optimal condition in which the to-be-rehearsed items were presented auditorily, optimizing their entry in the phonological loop (Baddeley, 1986), whereas the following items were presented visually, facilitating their encoding in a visual format in the executive loop while rehearsing the first segment of the list to be recalled. This last experiment was expected to reveal maxispans approaching the theoretical maximum of eight.

Experiment 1

This first experiment aimed at testing the hypothesis that the maxispan procedure elicits higher letter spans than the simple span procedure. In this experiment, all the letters were presented visually, including those that participants should verbally rehearse in the maxispan procedure.

Method

Experimental design. The experiment relied on a 2 (procedure: simple span vs. maxispan) $\times$ 3 (number of to-be-rehearsed letters: from three to five blue letters) $\times$ 6 (number of following letters: from one to six black letters) mixed repeated-measures design, with procedure as a between-participants factor and the number of blue and black letters as within-participant factors. All the participants were thus exposed to lists of four to 11 letters (i.e., $3 + 1$ to $3 + 6$, and so on until $5 + 6$, for blue and black letters, respectively) with three lists of each type for a total of 54 trials. In the maxispan condition, we manipulated the use of verbal rehearsal and refreshing strategies by requiring participants to rehearse cumulatively the blue letters aloud until recall and to mentally remember the subsequent black letters. In the simple span condition, no specific instructions were given to participants.

Participants. Forty students from the University of Geneva took part in the experiment in exchange of partial course credits. They were tested individually, and alternatively assigned to each of the two procedures, starting with the maxispan group, each group including 20 participants. All the experiments in this study have been approved by the local ethics committee of the Faculty of Psychology and Sciences of Education of the University of Geneva, and all participants gave their informed consent to participate.

Sampling plan. The sampling plan for this and the following experiments was based on a Bayesian open-ended sequential design (Schönbrodt & Wagenmakers, 2018). The decision to stop data collection was conditioned to the results of the main analysis (i.e., comparison of the best span from the maxispan and the simple span procedures, see the Results section for more details). It was planned to first collect a minimum of 20 participants in each group before performing the main analysis. If the results from this

[4] Asking participants to rehearse aloud would allow for a control of the correct execution of this procedure.

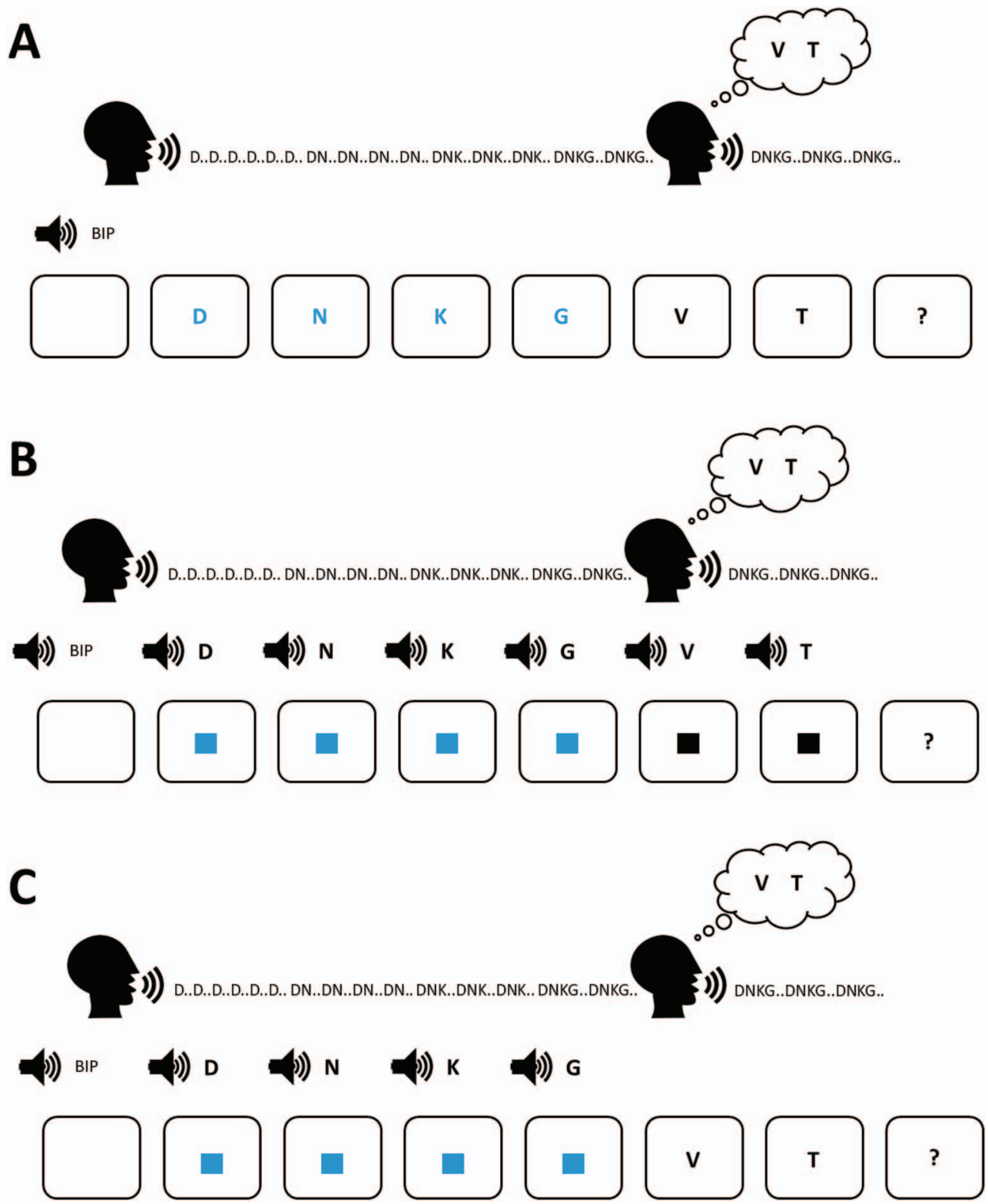

Figure 1. Graphical depiction of the task requirement for participants in the maxispan group in Experiments 1 to 3. As shown on the figure, participants in Experiment 1 had to rehearse the blue letters cumulatively as they appeared on the screen, and to continue rehearsing them during the presentation of the black letters. They were also required to refresh the black letters, but not rehearse them, while continue rehearsing overtly the blue letters. The procedure was the same for the simple span group, except that they received no specific advice concerning the use of rehearsing and refreshing strategies (A). In Experiment 2, the procedure was the same, except that all the letters were presented in an auditory format and blue and black squares were displayed on the screen to indicate which letters were to be rehearsed (B). In Experiment 3, the procedure was also the same, at the difference that the first letters were presented in an auditory format, and accompanied by blue squares on the screen to indicate they should be rehearsed, followed next by black letters visually presented (C). See the online article for the color version of this figure.

analysis provided substantial evidence (Bayes factor or BF >3) for either the null (H_0) or the alternative model (H_1), data collection was stopped. If not, the plan was to collect more participants by batch of five in each group until the desired level of evidence was obtained.

Materials. The material consisted in lists of consonants, ranging from length four to length 11. The lists were constructed from a sample consisting in all the consonants of the Latin alphabet, except W, which is trisyllabic in French, and Y, which is vowel in French (19 remaining consonants). For each list, we randomly selected the number of consonants required to conform to the list length while ensuring that no consonant was repeated within a list. We also controlled that no successive consonants in the lists were alphabetically adjacent (e.g., C followed by D was not allowed), and checked that the same consonant was not presented at the same serial position in two consecutive trials. To avoid any effect associated with the use of a specific set of lists, we created 20 sets of 54 lists for a total of 1,080 lists with the constraint that all the lists were unique within a given set. The same sets of 54 lists were used in both experimental groups (one set per participant) in order to ensure a strict comparability between the two procedures.

Procedure. The task was programmed and controlled with the open source software OpenSesame (Version 3.1.3, Mathôt, Schreij, & Theeuwes, 2012), running on a workstation computer. Each trial started by a 440-Hz tone played during 500 ms to indicate the beginning of the trial. After a 500 ms silent period, the letters were visually displayed one after the other at the center of the screen for 2,000 ms, with 0 ms interstimuli interval (ISI). Note that this rather slow presentation rate was defined in order to allow participants to rehearse aloud all the to-be-rehearsed letters between each letter presentation (maximum of five). The to-be-rehearsed first letters (three, four, or five depending on the condition) were displayed in blue font followed by one to six letters in black font. Consecutive to the presentation of the last letter, a question mark appeared on the screen to prompt participants to recall aloud all the letters in correct serial order. Participants could explicitly skip a serial position if they forgot the letter in this position. The experimenter wrote down the response for later correction and participants had to press the space bar to start the next trial.

The lists were presented in increasing length, from 3 + 1 to 3 + 6, and so on until 5 + 6 letters, for blue and black letters, respectively. For each combination, three lists were presented. No stopping rule was applied and a message appeared on the screen when the number of blue or black letters was about to change.

In the simple span group, participants were simply instructed to remember the letters for further serial recall without paying attention to their color. In the maxispan group, participants were asked to cumulatively rehearse the blue letters aloud during their presentation, and to continue to rehearse them aloud during the entire presentation of the black letters while trying, at the same time, to remember these black letters (Figure 1A). These instructions were displayed on screen along with Figure 1 that was used by the experimenter to explain and demonstrate how the blue letters should be rehearsed. After this demonstration, and before the experimental trials, participants performed three training trials with two blue and two black letters to ensure perfect understanding of the task requirements and the cumulative rehearsal method.

Scoring method. In this series of three experiments, we analyzed in the same way span sizes and serial recall accuracy. The first set of analyses focused on span measures, aiming at determining that the maxispan procedure results in higher spans. Spans were computed separately for each condition of to-be-rehearsed letters (i.e., three, four, and five blue letters) and were obtained with the following formula (Barrouillet, Bernardin, & Camos, 2004; Kemps, De Rammelaere, & Desmet, 2000):

$$\text{span} = base + \frac{N_{correct}}{3}$$

where *base* corresponds to the number of blue letters to be verbally rehearsed and $N_{correct}$ to the number of lists correctly recalled (i.e., all the letters, blue and then black, recalled in correct order). For example, if a participant perfectly recalled 10 lists in the five-blue-letters condition, her span was equal to 5 + (10/3) = 8.33. It is important to note that some participants failed to recall correctly at least three lists in one or several conditions of to-be-rehearsed letters. In that case, using the above formula could lead to span overestimation. For instance, if a given participant recall only one list correctly in the five-blue letters condition, his or her span for that condition of letters to be rehearsed would be 5 + (1/3) = 5.33; taking as granted that the participant could recall correctly five-letter lists anyway, even though it was not tested since the shortest lists in the five to-be-rehearsed letters condition is of six letters. When such situation occurred, we checked whether the baseline level for the given number of to-be-rehearsed letters has been reached in any previous condition (in the previous example, the baseline level was five). If the baseline has been reached before, we granted the baseline to the participant and used the above formula. If not, we set the span to the same level as the highest span in the previous conditions of to-be-rehearsed letters. We also determined the *best span* score by selecting for each participant her highest span score across the three conditions of blue letters to rehearse.

Second, because in the maxispan condition participants were specifically asked to only rehearse a specific subset of letters, we were interested whether the number of black letters influenced the proportion of blue letters recalled at their correct serial position, and vice versa. To reach this, we computed separately for each level of the number of black letters factor the proportion of blue letters recalled at their correct serial position. Next, we computed separately for each level of to-be-rehearsed blue letters factor the proportion of black letters recalled at their correct serial position.

Finally, we analyzed the serial position curves for blue and black letters in the simple span and maxispan conditions. Statistical analyses focused on the serial position curves of the blue letters in order to assess the effect of their cumulative rehearsal in the maxispan condition.

Results

All the statistical analyses were conducted according to a Bayesian inference framework with the open source program JASP (Version 0.9, JASP Team, 2018). For *t* tests, we used default priors consisting in a Cauchy distribution with an r scale of .707. For analyses of variance (ANOVA), the default priors were Cauchy distribution with an r scale of 0.5, 1, and 0.354 for fixed effects, random effects, and covariates, respectively.

Span measures. We compared the best span from the maxispan and the simple span procedures through an undirected Bayesian independent samples *t* test. This analysis provided decisive evidence for a higher mean span in the maxispan than in the simple span group (M = 7.30, SE = 0.13, and M = 6.10, SE = 0.20, respectively), BF_{10} = 1496.25, Cohen's d = 1.60 (see Figure 2).

We next analyzed span scores through a Bayesian mixed ANOVA with the procedure (maxi- vs. simple span) as a between-subjects factor and the number of blue letters to rehearse (three,

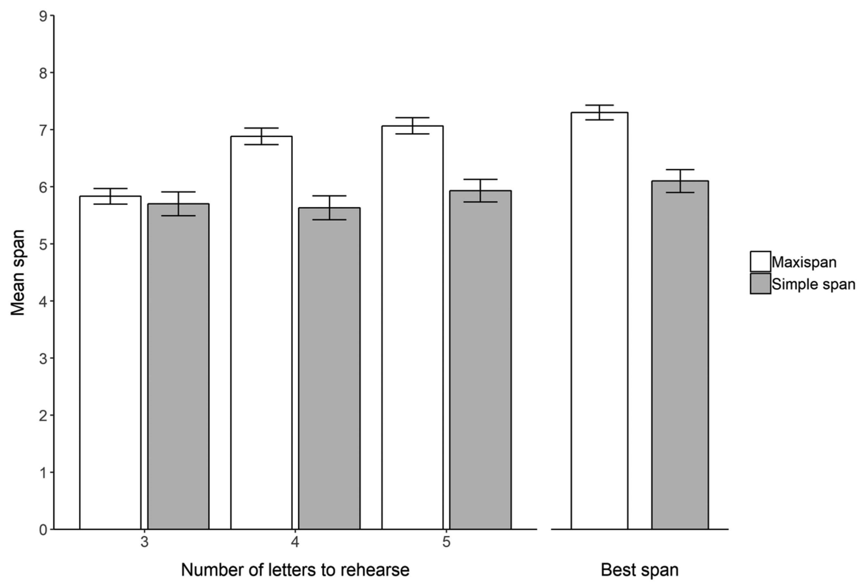

Figure 2. Mean spans for the simple span and the maxispan procedures as a function of the number of blue letters to rehearse in Experiment 1. The right part of the graph represents the best span reached over the three conditions. The bars represent standard error of the mean.

four, or five) as a within-subject factor. The results provided decisive evidence in favor of the full model (BF10 = 1.37E + 10) preferred by a factor of 8959.35 over the second best model. This supported the presence of a main effect of the procedure with higher spans in the maxispan than the simple span group (M = 6.59, SE = 0.11, and M = 5.76, SE = 0.19, respectively), and an effect of the number of blue letters to rehearse that interacted with the procedures (see Figure 2). Given the evidence supporting the interaction, we analyzed each group separately via a Bayesian repeated-measures ANOVA with the number of blue letters to rehearse as a within-subject factor. For the simple span procedure group, we obtained anecdotal evidence in favor of the null model supporting the absence of an effect of the number of blue letters to rehearse on mean spans (M = 5.70, SE = 0.21, M = 5.63, SE = 0.21, and M = 5.93, SE = 0.20 for three, four, and five blue letters, respectively), BF_{01} = 1.27. For the maxispan procedure group, the results yielded decisive evidence in favor of an effect of the number of blue letters to rehearse, mean spans increasing with their number (M = 5.83, SE = 0.14, M = 6.88, SE = 0.15, and M = 7.07, SE = 0.14 for three, four, and five blue letters, respectively), BF_{10} = 1.30E + 7. Post hoc analyses provided decisive evidence for a difference in spans between the three- and four-blue letters conditions, BF_{10} = 5744.31, while we obtained anecdotal evidence in favor of the null model supporting the *absence* of difference between the four- and five-blue letters conditions, BF_{01} = 2.58.

Finally, we also explored the interaction through undirected Bayesian independent samples t tests, comparing for each number of blue letters mean spans from the maxispan and the simple span groups. The results provided anecdotal evidence in favor of an *absence* of difference between the two groups for the three-blue letters, BF_{01} = 2.89, while providing decisive evidence for a difference between the procedures in both the four- and five-blue letters conditions, BF_{10} = 958.35 and 495.43, respectively.

Serial recall accuracy. Beyond the span measures, we were interested in the effect of the maxispan procedure on the recall of the two types of letters presented (see Table 1). Did blue letters benefit from cumulative rehearsal, and black letters suffer from being encoded under concurrent articulation in the maxispan procedure? Was recall of one of these blocks affected by variations of the size of the other? To answer these questions, we first analyzed the proportion of to-be-rehearsed blue letters recalled in correct serial position through the mean of a Bayesian mixed ANOVA with the procedure (maxi- vs. simple span) as a between-subjects factor and the number of black letters (from one to six) as a within-subject factor. This analysis provided decisive evidence in favor of the full model supporting the existence of the two main effects and their interaction (BF10 = 9.69E + 16), this model being favored over the second best model by a factor of 2.04E + 5 (see Figure 3). Blue letters were better recalled in the maxispan than the simple span group (M = .95, SE = .01 and M = .80, SE = .02, respectively), and the proportion of correct recall decreased as the number of black letters increased, this decline being stronger in the simple span procedure group. Thus, we analyzed each group separately via a Bayesian repeated-measures ANOVA with the number of black letters as a within-subject factor. For the simple span procedure group, the results provided decisive evidence for an effect of the number of black letters on the proportion of blue letters correctly recalled (BF_{10} = 5.53E + 7), the proportion of blue letters decreasing progressively as the number of black letters increased. By contrast, there was strong evidence for the *absence* of such an effect in the maxispan procedure group, BF_{01} = 14.33 (see Figure 3).

Next, we analyzed the proportion of black letters recalled in correct serial position via a Bayesian mixed ANOVA with the procedure (maxi- vs. simple span) as a between-subjects factor and the number of blue letters to rehearse (three, four, or five) as a within-subject factor. The results provided decisive evidence in

Table 1
Mean Number of Recall in Correct Serial Position as a Function of the Number of Blue Letters (to Be Rehearsed in the Maxispan Procedure) and Black Following Letters in the Maxispan and Simple Span Procedures in Experiment 1

			Number of black letters						
Procedure	Number of blue letters	Type of letters	1	2	3	4	5	6	*M* rate
Maxispan	3	Blue	3.00	3.00	2.95	2.98	3.00	2.97	.99
		Black	0.97	1.85	2.42	2.33	2.15	2.05	.68
		Total	3.97	4.85	5.37	5.32	5.15	5.02	.80
	4	Blue	3.90	3.97	3.93	3.87	3.78	3.83	.97
		Black	0.97	1.80	2.33	2.57	1.95	2.02	.67
		Total	4.87	5.77	6.27	6.43	5.73	5.58	.81
	5	Blue	4.55	4.47	4.47	4.37	4.50	4.32	.89
		Black	0.78	1.55	1.95	1.70	1.57	1.82	.54
		Total	5.33	6.02	6.42	6.07	6.07	6.13	.73
Simple span	3	Blue	2.95	2.87	2.77	2.38	2.25	2.27	.86
		Black	0.93	1.75	2.20	2.05	1.87	1.75	.62
		Total	3.88	4.62	4.97	4.43	4.12	4.02	.72
	4	Blue	3.62	3.50	3.25	3.43	3.00	3.07	.83
		Black	0.87	1.10	1.63	1.65	1.25	1.07	.47
		Total	4.48	4.60	4.88	5.08	4.25	4.13	.65
	5	Blue	4.03	3.80	3.43	3.32	3.47	3.05	.70
		Black	0.70	0.85	1.10	1.00	0.90	0.82	.34
		Total	4.73	4.65	4.53	4.32	4.37	3.87	.55

favor of the full model supporting the presence of the two main effects and their interaction (BF10 = 6.17E + 16), the model being favored by a factor of 545.10 over the second best model. Black letters were better recalled in the maxispan than the simple span condition (M = .63, SE = .02 and M = .48, SE = .04, respectively), and the proportion of black letters correctly recalled decreased when the number of blue letters to rehearse increased (see Figure 4). In order to analyze the interaction, we ran for each group separately a Bayesian repeated-measures ANOVA with the number of blue letters to rehearse as a within-participant factor. For the simple span procedure group, the results yielded decisive evidence for a decrease in the proportion of correct recall of black letters with more blue letters to be rehearsed (BF_{10} = 7.45E + 8). Post hoc analysis confirmed the progressive decrease by supporting the presence of a difference between the three- and four-blue letters conditions (M = .62, SE = .04, and M = .47, SE = .04, respectively), and between the four- and five-blue letters (M = .34, SE = .05) conditions, BF_{10} = 3831.20 and BF_{10} = 4161.21, respectively.

For the maxispan procedure group, the alternative model supporting the presence of an effect of the number of blue letters to rehearse on the proportion of black letters correctly recalled was

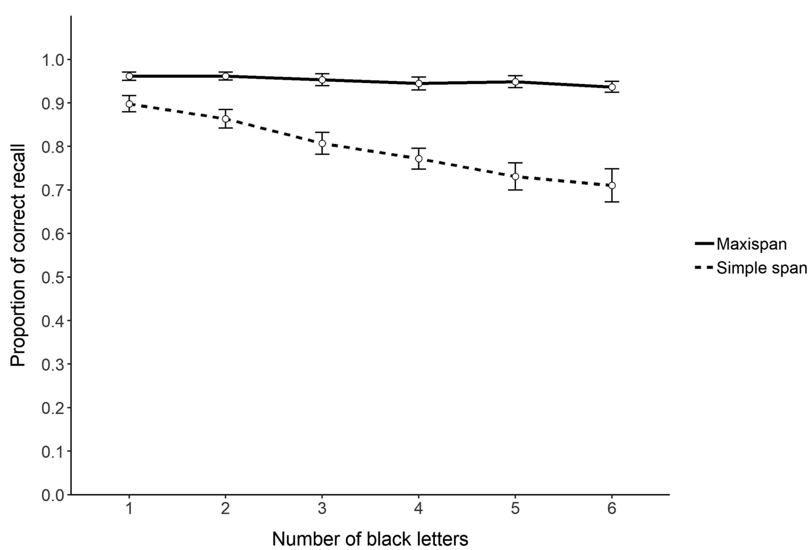

Figure 3. Mean proportion of blue letters correctly recalled at their serial position as a function of the number of following black letters in Experiment 1. Error bars represent standard error of the mean.

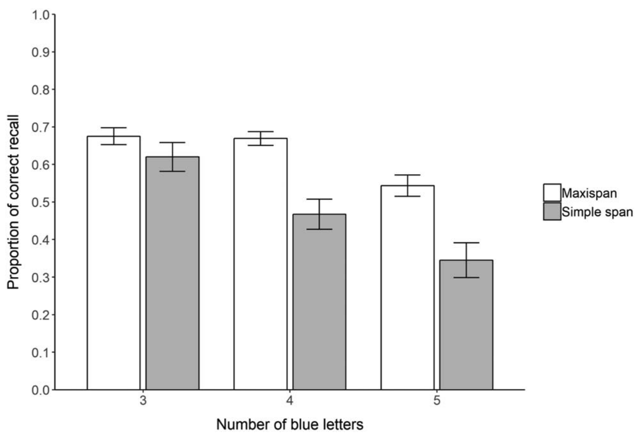

Figure 4. Means proportion of black letters correctly recalled at their serial position as a function of the number of blue letters to rehearse in Experiment 1. Error bars represent standard error of the mean.

also favored with decisive evidence, $BF_{10} = 5.15E + 4$. However, the decrease was not as progressive as it was in the simple span procedure group. Post hoc analyses provided moderate evidence for an *absence* of difference between the 3-and the 4-blue letters conditions ($M = .67$, $SE = .02$ and $M = .67$, $SE = .02$, respectively), $BF_{01} = 4.15$. By contrast, the five-blue letters conditions resulted in a drop in the proportion of black letters recalled ($M = .54$, $SE = .03$) when compared with the four-blue letters condition, $BF_{10} = 1460.26$.

Serial position curves. Another way to look at the results is to express the rate of recall in correct position of each letter as a function of its serial position for the blue and black letters presented in both conditions for each list length (see Figure 5). Serial position curves revealed similarities and differences between the maxispan and the simple span conditions. For both conditions, recall was characterized by a strong primacy effect, but at best a small and sporadic recency effect. When this latter effect occurred, it affected the last item (see e.g., the Condition 3–4 in the simple span procedure or the Condition 4–4 in the maxispan procedure). Overall, across all the list lengths and numbers of blue letters, the difference in the rates of recall between the penultimate and last items was null in the simple span condition, while in the maxispan condition this rate diminished by .02 from the penultimate to the last item.

However, whereas the rate of recall monotonically decreased as serial position increased in the simple span condition, serial position only affected the black letters in the maxispan condition, the quasiperfect recall of the rehearsed letters abolishing any effect of serial position for the blue letters. It is only when five blue letters were to be rehearsed in the maxispan procedure that serial position affected recall of these letters (from .94 to .86 for serial positions 1 and 5, respectively; see Appendix A for statistical analyses). This suggests, in line with the literature previously reviewed, that the capacity of the phonological loop is exceeded with five letters. Nonetheless, these five blue letters were better recalled in the maxispan procedure, and the effect of serial position was stronger in the simple span condition.

Discussion

This first experiment revealed that the maxispan procedure results in letter spans higher than the traditional simple spans by more than one item. This large difference resulted from a better recall of both sets of letters, those that were verbally rehearsed (i.e., the blue letters) and those that followed (the black letters). The better recall of the rehearsed letters in the maxispan procedure provides evidence that verbal rehearsal is beneficial for STM. This is particularly clear when analyzing the effect of the number of the following black letters on the rate of recall of the blue letters rehearsed in the maxispan procedure. Whereas the rate of correct recall of these letters progressively declined as the number of the following black letters increased in the simple span procedure, verbal rehearsal in the maxispan procedure seems to have shielded blue letters, their recall remaining unaffected by the number of black letters subsequently presented. Interestingly, this better recall of the rehearsed letters did not come at the expense of the recall of the following letters. In other words, whereas black letters were encoded and maintained under the concurrent articulation produced by rehearsing aloud the blue letters in the maxispan procedure, they were nonetheless better recalled than in the simple span procedure.

Our interpretation of these findings is that this better recall of both portions of the lists (i.e., blue and black letters) in the maxispan procedure is due to the fact that this procedure directs items toward separate systems of maintenance, and prevents transfer of these items from one system to the other in such a way that both systems achieve their maximum efficiency. Two elements support this interpretation. First, it seems that when the articulatory system is underused, maxispans are not higher than simple spans. It can be seen in Figure 2 that rehearsing three items does not lead to higher maxispans than simple spans, a difference that arises with four letters rehearsed, which corresponds to the hypothesized capacity of the phonological loop for letters. Second, whereas increasing the number of rehearsed letters from three to four had no effect on the recall of black letters in the maxispan procedure, adding a fifth to-be rehearsed letter affected this recall. This suggests that the maintenance of this additional blue letter, which exceeded the capacity of the phonological loop, relied on some capacity previously used for maintaining black letters. Our hypothesis is that the capacity of the phonological loop being exceeded, the supernumerary letter was stored in the central system at the expense of the black letters that this central system maintains, their entrance in the phonological loop being prevented by the continuous articulation of the blue letters. This trade-off from a maintenance system to the other could explain why mean spans in the maxispan procedure were not significantly higher when rehearsing five instead of four blue letters (see Figure 2), and why an effect of serial position appeared on blue letters when five letters were to be rehearsed.

Although the results of this first experiment were in line with our predictions, they could be peculiar to the visual presentation of the letters. Presenting the same letters auditorily would provide a more stringent test of our hypothesis. The auditory presentation should ease the encoding of the to-be-rehearsed letters into the phonological loop (Baddeley, 1986). However, the recall of these letters in the maxispan procedure was already almost perfect in Experiment 1 (rate of correct recall of .99, .97,

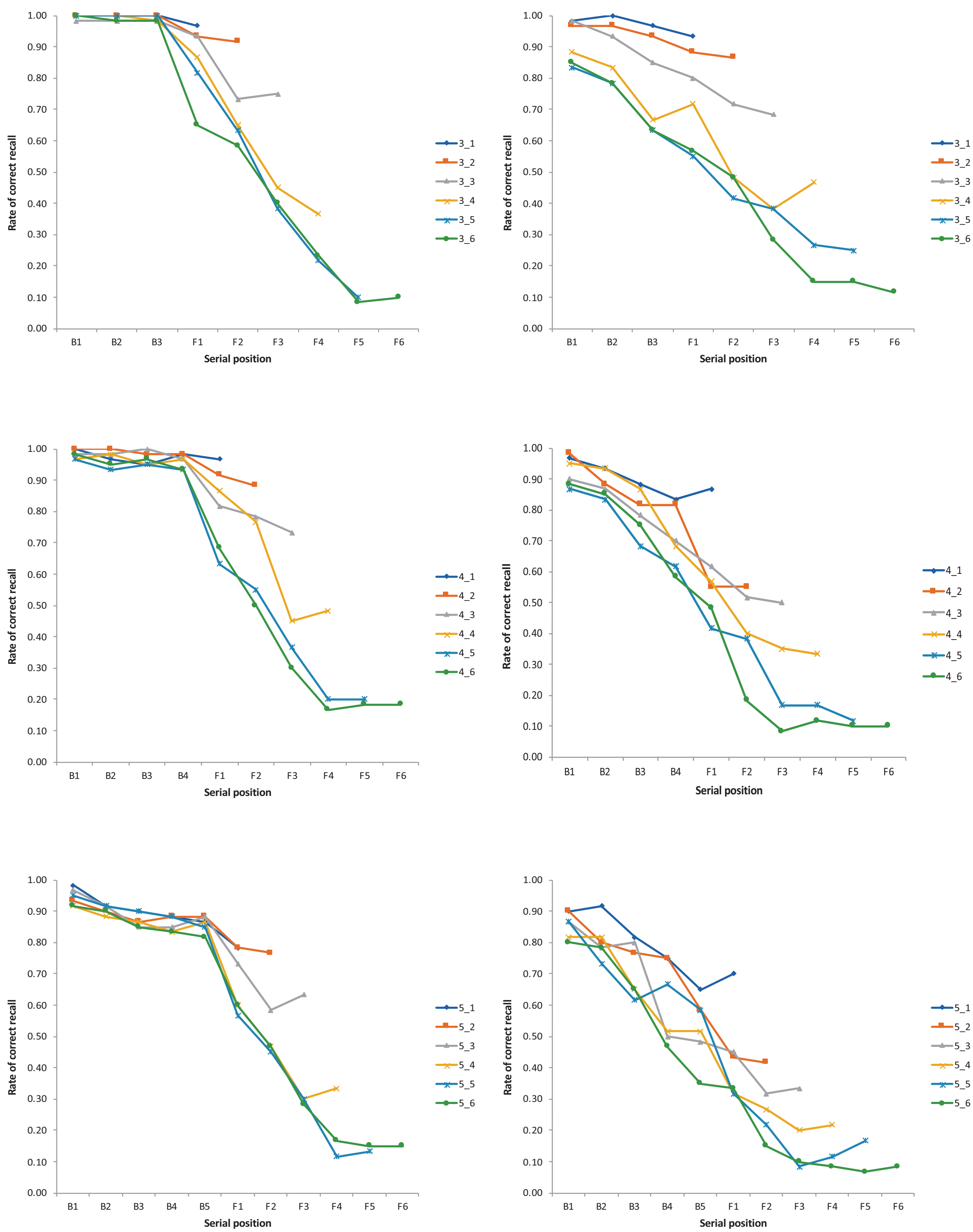

Figure 5. Rate of correct recall of the letters as a function of their serial position and list length in Experiment 1. Left and right panels are for the maxispan and the simple span procedures, respectively. In the x-axes, B and F refer to blue and black letters, respectively. List lengths are expressed as the number of blue and black letters (e.g., 4–3 refers to a series with four blue and three black letters). See the online article for the color version of this figure.

and .89 for three, four, and five rehearsed letters, respectively), leaving little room for improvement. By contrast, articulating aloud these letters should hinder the encoding of the following black letters when presented auditorily. Thus, replicating the superiority of recall performance with the maxispan procedure when the letters are auditorily presented as in the following Experiment 2 would constitute strong evidence for our hypothesis.

Experiment 2

This experiment aimed at replicating the results of the previous experiment when the letters to be recalled are not presented visually, but auditorily. As in Experiment 1, we predicted higher spans from the maxispan procedure.

Method

Participants. Forty participants were recruited on the campus of the University of Geneva to took part in the experiment in exchange of partial course credits. They did not participate in Experiment 1 and were tested individually. We alternatively assigned the participants to one of the two experimental groups, starting with the maxispan group, with 20 participants in each group.

Material and procedure. The design and procedure were similar to Experiment 1 except that the letters were presented auditorily. Auditory consonants were recorded by a native female French speaker and saved as mono wav files with a sampling rate of 44.1 kHz. They were presented through loudspeakers at a rate of one letter every 2,000 ms (durations of these auditory stimuli ranging from 259 ms to 682 ms with a mean of 483 ms). Auditory letter onsets coincided with the presentation on screen of a square for 1,500 ms followed by a blank screen for 500 ms. The letters to be verbally rehearsed were indicated by blue squares, and the others by black squares (Figure 1B).

Results

Span measures. The best span reached by participants in the maxispan and simple span procedures was compared via an undirected Bayesian independent samples t test. The results provided moderate evidence for an effect of procedure, with better spans in the maxispan than in the simple span group ($M = 6.52$, $SE = 0.14$, and $M = 5.82$, $SE = 0.21$, respectively), $BF_{10} = 5.74$, Cohen's $d = 0.88$ (see Figure 6).

We performed a Bayesian mixed ANOVA on spans with the procedure (maxi- vs. simple span) as a between-subjects factor and the number of blue letters to rehearse (three, four, and five) as a within-subject factor. The results provided very strong evidence in favor of the full model with the two main effects and their interaction (BF10 = 3.34E + 9) that was preferred by a factor of 34.60 over the second best model (see Figure 6). This represents very strong support to the presence of an effect of the procedure, characterized by higher spans in the maxispan relative to the simple span group ($M = 5.96$, $SE = 0.60$, and $M = 5.47$, $SE = 0.82$, respectively), as well as an effect of the number of blue letters to rehearse that interacted with the procedure. Given the evidence supporting the interaction, we analyzed the two groups separately through Bayesian repeated-measures ANOVAs with the number of blue letters to rehearse as a within-subject factor. For the simple span group, the results provided very strong evidence for an effect of the number of blue letters on mean spans ($BF_{10} = 31.72$). Whereas post hoc analyses revealed anecdotal evidence for an absence of difference between the three-blue and four-blue letters conditions ($M = 5.25$, $SE = 0.19$, and $M = 5.43$, $SE = 0.20$, respectively), $BF_{01} = 1.64$, we observed moderate evidence

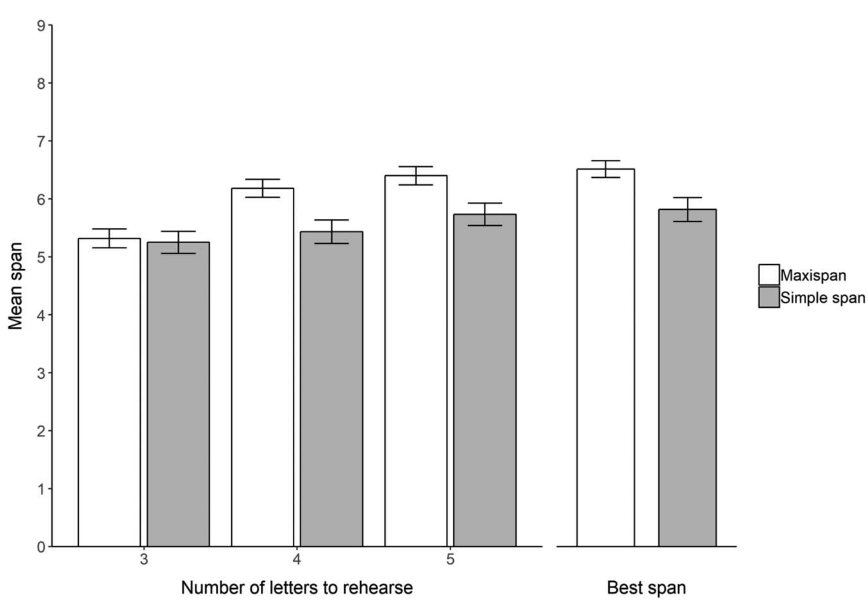

Figure 6. Mean spans for the simple span and the maxispan procedures as a function of the number of blue letters to rehearse in Experiment 2. The right part of the graph represents the best span reached over the three conditions. The bars represent standard error of the mean.

for a difference between the 4-blue and 5-blue letters conditions ($M = 5.73$, $SE = 0.19$), $BF_{10} = 5.84$.

For the maxispan procedure group, the results provided decisive evidence in favor of an effect of the number of blue letters to rehearse on mean spans, $BF_{10} = 1.18E + 6$. Post hoc analyses provided decisive evidence for a difference in spans between the three-blue and four-blue letters conditions ($M = 5.32$, $SE = 0.16$, and $M = 6.18$, $SE = 0.16$, respectively, $BF_{10} = 1204.89$), but anecdotal evidence for the absence of difference in spans between the four- and the five-blue letters ($M = 6.40$, $SE = 0.16$) conditions, $BF_{01} = 1.35$. As in the previous experiment, we compared mean spans from the two procedures for each level of the number of blue letters condition via undirected Bayesian independent samples *t* tests. As in Experiment 1, these analyses revealed moderate evidence for an absence of difference in the three-blue letters condition, $BF_{01} = 3.15$, but moderate evidence for higher spans in the maxispan group with four and five blue letters, $BF_{10} = 7.57$ and 4.63, respectively.

Serial recall accuracy. Rates of recall of both types of letters in their correct serial position in all the experimental conditions can be found in Table 2. We next analyzed the proportion of blue letters correctly recalled as a function of the number of black letters through Bayesian mixed ANOVA with the procedure (maxi- vs. simple span) as a between-subjects factor and the number of black letters (from one to six) as a within-subject factor. The results provided decisive evidence in favor of the full model supporting the presence of the two main effects and their interaction (BF10 = 2.89E + 16), preferred over the second best model by a factor of 779.83 (see Figure 7). We observed higher recall performance in the maxispan compared with the simple span procedure group ($M = .89$, $SE = .01$, and $M = .71$, $SE = .03$, respectively), and a decrease in blue letters recall accuracy as the number of black letters increased. Given the evidence supporting the presence of the interaction, we analyzed each group separately via a Bayesian repeated-measures ANOVA on blue letters recall accuracy with the number of black letters as a within-subject factor. For the simple span procedure group, we obtained decisive evidence in favor of the presence of an effect of the number of black letters ($BF_{10} = 2.83E + 8$), the proportion of blue letters correctly recalled decreasing progressively as the number of black letters increased. For the maxispan procedure group, the same analysis yielded anecdotal evidence in favor of the null model supporting the *absence* of an effect of the number of black letters on blue letters recall accuracy ($BF_{01} = 1.16$).

In a second analysis, we performed a Bayesian mixed ANOVA on recall accuracy for the black letters with the procedure (maxi- vs. simple span) as a between-subjects factor and the number of blue letters to rehearse (three to five) as a within-subject factor. The results indicated that the full model was the best model (BF10 = 4.74E + 20) favored over the second best model by a factor of 3.29, supporting the presence of an effect of procedure with better performance in the maxispan than in the simple span procedure group ($M = .56$, $SE = .02$ and $M = .44$, $SE = .03$, respectively), and an effect of the number of blue letters that interacted with the procedure, this effect being stronger in the simple than in the maxispan procedure (see Figure 8). Due to the interaction, we analyzed the data separately for each group via Bayesian repeated-measures ANOVAs with the number of blue letters as a within-participant factor. For the simple span procedure group, the results provided decisive evidence in favor of an effect of the number of blue letters to rehearse on black letters recall accuracy, $BF_{10} = 9.39E + 10$. Post hoc analyses provided decisive evidence for a difference between the three- and four-blue letters conditions ($M = .57$, $SE = .03$, and $M = .43$, $SE = .04$, respectively), and between the four-blue and five-blue letters ($M = .32$, $SE = .04$) conditions, $BF_{10} = 1.19E + 4$, and $BF_{10} = 1059.58$, respec-

Table 2
Mean Number of Recall in Correct Serial Position as a Function of the Number of Blue Letters (to Be Rehearsed in the Maxispan Procedure) and Black Following Letters in the Maxispan and Simple Span Procedures in Experiment 2

Procedure	Number of blue letters	Type of letters	Number of black letters 1	2	3	4	5	6	*M* rate
Maxispan	3	Blue	2.85	2.90	2.87	2.83	2.92	2.83	0.96
		Black	0.95	1.88	2.13	2.30	1.72	1.83	0.64
		Total	3.80	4.78	5.00	5.13	4.63	4.67	0.76
	4	Blue	3.83	3.78	3.63	3.60	3.60	3.60	0.92
		Black	0.98	1.68	1.98	1.60	1.50	1.83	0.58
		Total	4.82	5.47	5.62	5.20	5.10	5.43	0.74
	5	Blue	4.30	4.23	4.00	3.97	3.75	3.78	0.80
		Black	0.80	1.38	1.23	1.40	1.30	1.07	0.45
		Total	5.10	5.62	5.23	5.37	5.05	4.85	0.64
Simple span	3	Blue	2.87	2.62	2.52	2.32	2.12	2.15	0.81
		Black	0.95	1.72	2.03	1.87	1.25	1.32	0.57
		Total	3.82	4.33	4.55	4.18	3.37	3.47	0.66
	4	Blue	3.27	3.05	2.83	2.73	2.55	2.38	0.70
		Black	0.82	1.32	1.20	1.33	0.87	1.15	0.43
		Total	4.08	4.37	4.03	4.07	3.42	3.53	0.56
	5	Blue	3.95	3.67	3.12	2.85	2.65	2.73	0.63
		Black	0.72	0.93	1.00	0.75	0.55	0.68	0.32
		Total	4.67	4.60	4.12	3.60	3.20	3.42	0.50

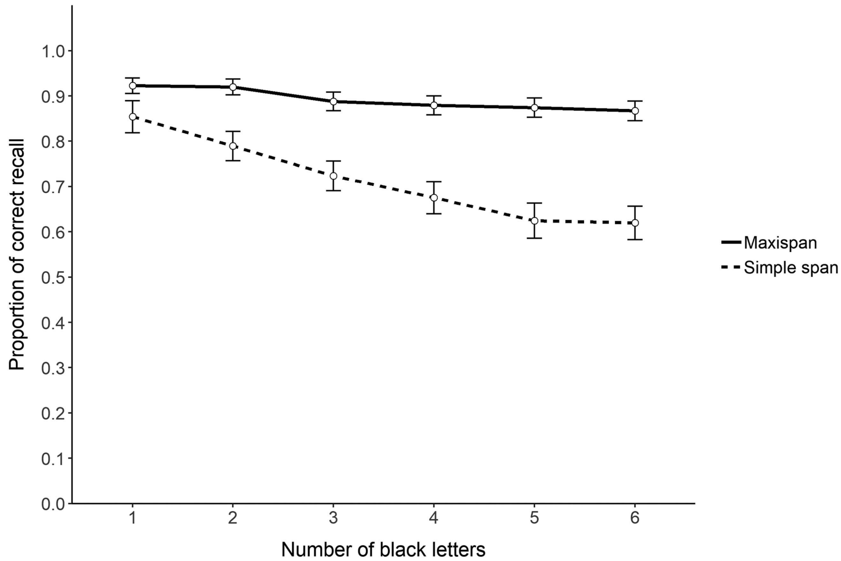

Figure 7. Mean proportion of blue letters correctly recalled at their serial position as a function of the number of following black letters in Experiment 2. Error bars represent standard error of the mean.

tively. For the maxispan procedure group, the results provided decisive evidence in favor of the presence of an effect of the number of blue letters to rehearse on black letters recall accuracy (BF_{10} = 3.32E + 7). Post hoc analyses revealed evidence for a difference between the three-blue and four-blue letters conditions as well as between this latter condition and the five-blue letters condition (M = .64, SE = .02, M = .58, SE = .03, and M = .45, SE = .03 for three, four, and five blue letters, respectively), BF_{10} = 5.25 and BF_{10} = 5529.48, respectively.

Serial position curves. As in Experiment 1, recall was dominated by a strong primacy effect, but we observed a small recency effect, more pronounced than in Experiment 1, that mainly affected the last item in both the simple span and the maxispan procedures (see Figure 9). Across all the conditions, the rate of recall increased between the penultimate and the last serial position by 6% and 8% in the maxispan and the simple span procedures, respectively. This was probably due to the auditory presentation of the letters that favored retrieval of the last item in sensory store. As in Experiment 1, blue letters remained immune from serial position effect when three letters were to be rehearsed, but a small effect appeared with four letters (rate of correct recall of .94, .93, .92, and .89 from Serial Position 1 to 4). This effect was stronger with five blue letters,

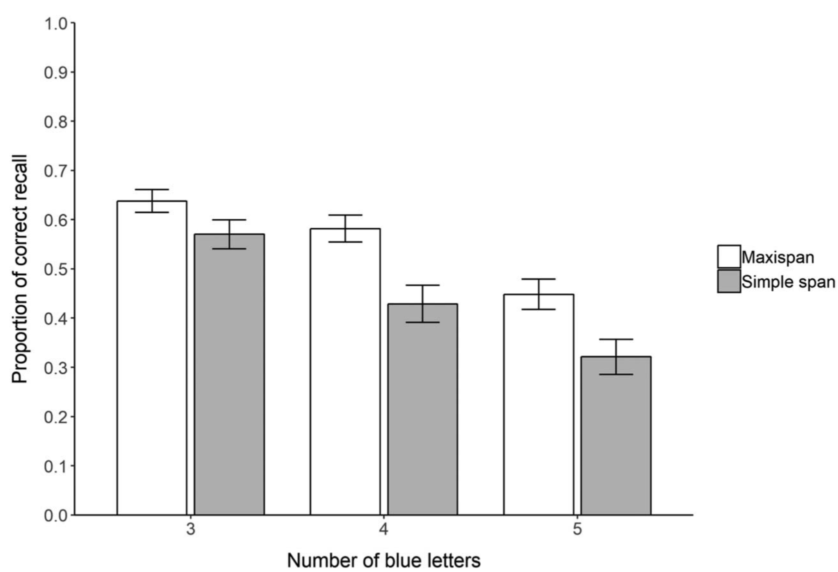

Figure 8. Means proportion of black letters correctly recalled at their serial position as a function of the number of blue letters to rehearse in Experiment 2. Error bars represent standard error of the mean.

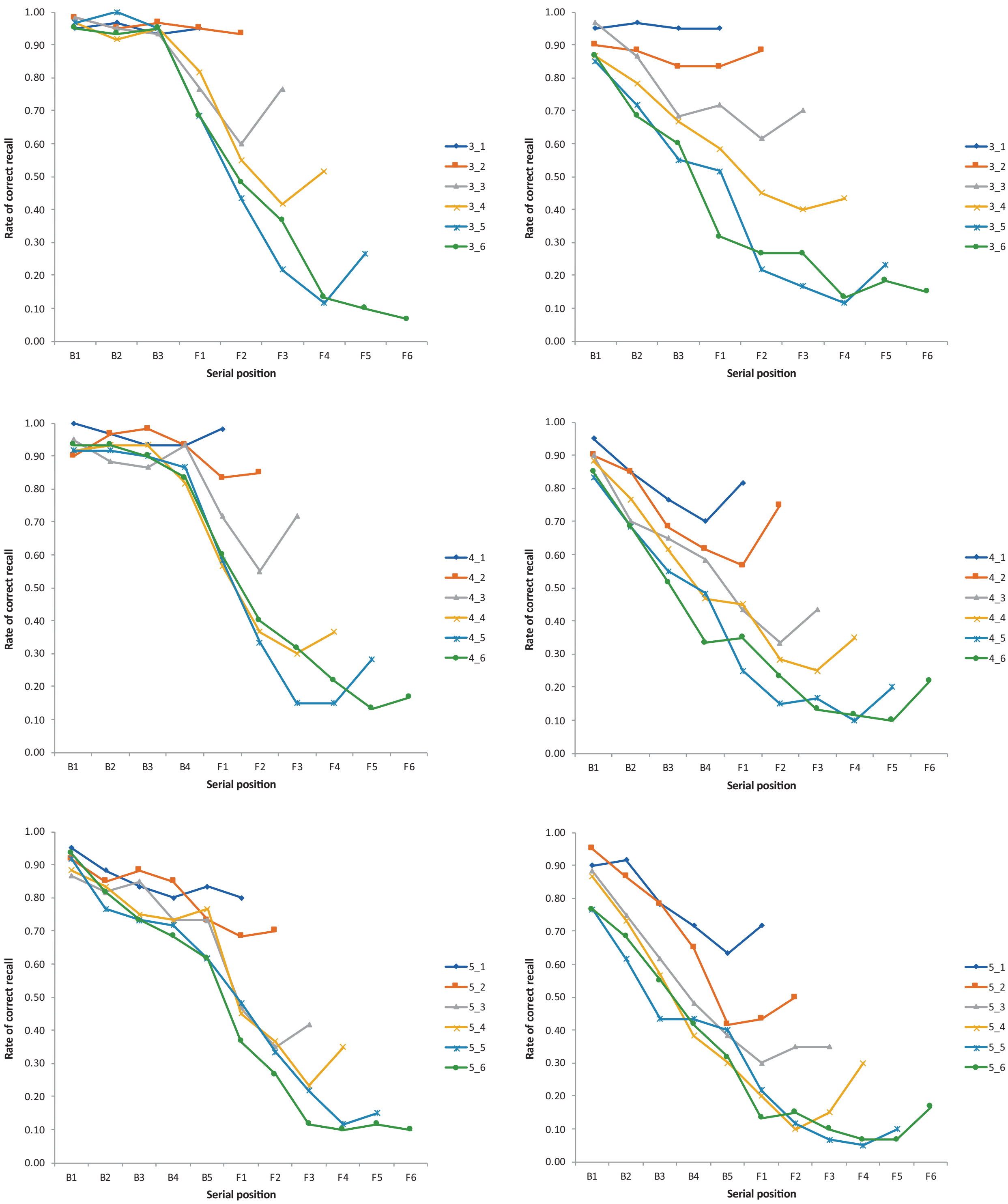

Figure 9. Rate of correct recall of the letters as a function of their serial position and list length in Experiment 2. Left and right panels are for the maxispan and the simple span procedures, respectively. In the x-axes, B and F refer to blue and black letters, respectively. List lengths are expressed as the number of blue and black letters (e.g., 4–3 refers to a series with four blue and three black letters). See the online article for the color version of this figure.

their rate of correct recall decreasing from .91 to .72 between the first and the fifth position. Note that this effect was nonetheless far more pronounced in the simple span procedure (from .77 to .32; see Appendix A for statistical analyses).

Discussion

This experiment, in which all the letters were auditorily presented, replicated the results of Experiment 1. The mean best span was still higher in the maxispan than in the simple span procedure, and the differences between the two procedures as a function of the number of blue letters to be rehearsed in the maxispan procedure replicated very closely those observed in Experiment 1 (compare Figures 2 and 6). Overall, spans were lower in the present experiment, but this decrease in recall performance was more pronounced in the maxispan (mean span across the three conditions of blue letters of 6.59 in Experiment 1, but only 5.96 in Experiment 2) than in the simple span procedure (mean spans of 5.76 and 5.47, respectively). This was due to the fact that, as we expected, the recall of black letters in the maxispan procedure was strongly affected by their auditory presentation, their overall rate of recall dropping from .63 in Experiment 1 to .56, but the same decline was observed in the rate of recall of the blue letters (from .95 to .89). Overall, the auditory presentation seemed detrimental for both procedures. Nonetheless, both blue and black letters were still better recalled in the maxispan procedure, indicating that despite being encoded and maintained under concurrent articulation, auditorily presented black letters benefitted from the segregation between storage systems facilitated by the maxispan procedure, while the recall of blue letters strongly benefitted from articulatory rehearsal.

The next and last experiment of this article aimed at maximizing the difference in recall performance between the maxispan and the simple span procedure. Based on our proposed theoretical framework that distinguishes two systems, we design a maxispan procedure that should facilitate the encoding of the letters in each of these stores assumed to hold information, the phonological loop and the central attentional system.

Experiment 3

The aim of this experiment was to optimize the use of the two storage systems we hypothesized, the phonological and the executive loops, by presenting the memoranda in a way that should facilitate encoding and maintenance. Because phonological information is assumed to directly enter the phonological loop (Baddeley, 1986), the blue letters to be rehearsed in the maxispan procedure were presented auditorily. Because, in the maxispan procedure, the following black letters must be encoded under the concurrent articulation of the blue letters, their visual presentation should facilitate encoding. Not only the auditory output of the articulation of the blue letters should not interfere with the perception of the black letters as it was probably the case when they were presented auditorily in Experiment 2, but the visual presentation of these black letters should induce a multimodal encoding (visual and phonological) appropriate to a storage and maintenance in the executive loop. Indeed, this central system is assumed to maintain any kind of information through attentional refreshing (Barrouillet & Camos, 2015), and richer representations resulting from multimodal encoding should be more resistant to interference and decay as well as easier to retrieve for refreshing. This optimal auditory presentation of blue letters and then visual presentation of black letters should therefore lead to letter spans approaching the theoretical optimum of eight (4 + 4) through the optimal use of the two memory stores involved as we explained in introducing this article.

Method

Participants. We recruited on the campus of the University of Geneva 40 participants who took part in the experiment in exchange of partial course credits. Participants did not participate to the previous experiments, and they were tested individually, by alternatively assigning them to one of the two groups, starting with the maxispan group. Each group was composed of 20 participants.

Material and procedure. This design was similar to the previous experiments, but the procedure was a mix of these experiments with the same auditory presentation of the to-be-rehearsed letters as in Experiment 2 (accompanied by blue squares displayed on screen) and the same visual presentation of the black letters as in Experiment 1 (Figure 1C).

Results

Span measures. Better spans in the maxispan and the simple span groups were submitted to an undirected Bayesian independent samples *t* test. This analysis provided decisive evidence for higher spans in the maxispan than in the simple span procedure, with a difference approaching 1.4 items (M = 7.73, SE = 0.16, and M = 6.37, SE = 0.17, respectively), BF_{10} = 1.34E + 4, Cohen's d = 1.85 (see Figure 10).

Span scores were next analyzed via a Bayesian mixed ANOVA with the procedure (maxi- vs. simple span) as a between-subjects factor and the number of blue letters to rehearse (three to five) as a within-subject factor. The results yielded only anecdotal evidence in favor of the full model (BF10 = 7.26E + 12) supporting the presence of the two main effects as well as their interaction, this model being preferred over the second best model without interaction by a factor of 1.34. In order to untangle the ambiguous evidence, we ran an analysis of *specific effects*, a method averaging evidence for testing the presence of a specific effect across models including the effect, as proposed in JASP. This analysis revealed decisive evidence for an effect of procedure with higher spans in the maxispan than the simple span group (M = 6.96, SE = 0.12, and M = 5.85, SE = 0.15, respectively), $BF_{\text{Inclusion}}$ = 2.73E + 4, decisive evidence for an effect of the number of blue letters with span decreasing as this number increases, $BF_{\text{Inclusion}}$ = 5.98E + 8, as well as moderate evidence for the presence of an interaction between the two factors, $BF_{\text{Inclusion}}$ = 5.36 (see Figure 10). We therefore analyzed span scores from the two groups separately with Bayesian repeated-measures ANOVAs with the number of blue letters to rehearse as a within-subject factor. For the simple span procedure group, the results provided decisive evidence in favor of the presence of an effect of the number of blue letters on span scores, BF_{10} = 131.14. Post hoc analyses yielded anecdotal evidence supporting the absence of a difference between spans from the three-blue and four-blue letters conditions (M = 5.53, SE = 0.17, M = 5.78, SE = 0.17, respectively), BF_{01} = 1.34, but strong evidence supporting a difference between spans from the four-blue to the five-blue

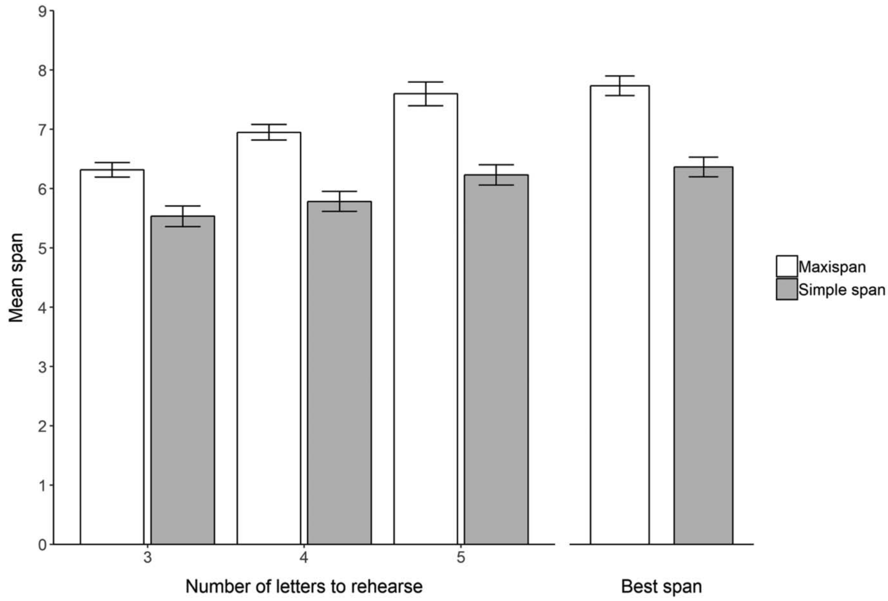

Figure 10. Mean spans for the simple span and the maxispan procedures as a function of the number of blue letters to rehearse in Experiment 3. The right part of the graph represents the best span reached over the three conditions. The bars represent standard error of the mean.

letters condition (M = 6.23, SE = 0.17), BF_{10} = 29.61. For the maxispan procedure group, we obtained decisive evidence in favor of an effect of the number of blue letters to rehearse on span scores, BF_{10} = 6.02E + 5. Post hoc analyses provided evidence for a difference between span scores from the three-blue and four-blue letters conditions, as well from four-blue and five-blue letters conditions (M = 6.32, SE = 0.12, M = 6.95, SE = 0.13 and M = 7.60, SE = 0.02, for three, four, and five blue letters), BF_{10} = 870.99, and BF_{10} = 8.89, respectively.

As in the previous experiments, we conducted complementary undirected Bayesian independent samples t tests comparing mean spans from the maxispan and simple span procedure groups, separately for each number of blue letters. The results provided very strong or decisive evidence for higher spans with the maxispan whatever the number of blue letters, BF_{10} = 42.29, 4301.02, and 2101.41 for three, four, and five blue letters, respectively.

Serial recall accuracy. Rates of recall of both types of letters in their correct serial position in all the experimental conditions can be found in Table 3. We ran a Bayesian mixed ANOVA on the proportion of blue letters recalled at their correct serial position with the procedure (maxi- vs. simple span) as a between-subjects factor and the number of black letters (from one to six) as a within-subject factor. We obtained decisive evidence in favor of the full model (BF10 = 3,71E + 24), preferred over the second best model by a factor of 2.92E + 9. This supports the presence of an effect of procedure characterized by higher recall performance in the maxispan compared with the simple span group (M = .94, SE = .01 and M = .78, SE = .02, respectively), as well as an effect of the number of black letters, with blue letters recall accuracy decreasing as the number of black letters increased, the two factors interacting with each other (see Figure 11). Thus, we analyzed the data separately for each group via Bayesian repeated-measures ANOVAs with the number of black letters as a within-subject factor. For the simple span group, the results provided decisive evidence in favor of an effect of the number of black letters on blue letters recall accuracy (BF_{10} = 7.75E + 11), the proportion of blue letters correctly recalled clearly decreasing as the number of black letters increases. In contrast, for the maxispan group, we obtained strong evidence in favor of the null model, BF_{01} = 11.72, supporting the *absence* of an effect of the number of black letters on blue letters recall accuracy.

In a second step, we performed a Bayesian mixed ANOVA on the proportion of black letters correctly recalled with the procedure as a between-subjects factor and the number of blue letters to rehearse (three to five) as a within-subject factor. We obtained anecdotal evidence for the model containing the two main effects (BF10 = 8.51E + 11), this model being preferred over the full model by a factor of 1.56 (see Figure 12). Given the ambiguous results, we ran an analysis of specific effect, confirming the presence of an effect of the number of blue letters characterized by a decrease of black letters recall accuracy as the number of blue letters to rehearse increased, $BF_{Inclusion}$ = 1.42E + 8, an effect of the procedure with better recall performance in the maxispan than the simple span group (M = 71, SE = 0.2M = .52, SE = .03, respectively), $BF_{Inclusion}$ = 6947.32, as well as some evidence, but still anecdotal, for the presence of an interaction, $BF_{Inclusion}$ = 2.57. We analyzed the data from each group separately through Bayesian repeated-measures ANOVAs with the number of blue letters as a within-subject factor. For the simple span group, we obtained decisive evidence in favor of the presence of an effect of the number of blue letters on black letters recall accuracy, BF_{10} = 6.81E + 4. Post hoc analyses revealed very strong support to the presence of a difference in terms of recall accuracy between the three- and four-blue letters, as well as between four- and five-blue letters conditions (M = .62, SE = .03, M = .52, SE = .03, and M = .42, SE = .04, for three, four, and five blue letters, respec-

Table 3
Mean Number of Recall in Correct Serial Position as a Function of the Number of Blue Letters (to Be Rehearsed in the Maxispan Procedure) and Black Following Letters in the Maxispan and Simple Span Procedures in Experiment 3

			Number of black letters						
Procedure	Number of blue letters	Type of letters	1	2	3	4	5	6	*M* rate
Maxispan	3	Blue	2.92	2.95	2.92	2.97	3.00	2.95	0.98
		Black	1.00	2.00	2.73	2.93	2.60	2.40	0.76
		Total	3.92	4.95	5.65	5.90	5.60	5.35	0.84
	4	Blue	3.83	3.83	3.78	3.73	3.68	3.93	0.95
		Black	1.00	1.93	2.62	2.77	2.48	2.48	0.74
		Total	4.83	5.77	6.40	6.50	6.17	6.42	0.83
	5	Blue	4.63	4.62	4.47	4.38	4.35	4.38	0.89
		Black	0.93	1.75	2.07	2.22	2.37	1.87	0.64
		Total	5.57	6.37	6.53	6.60	6.72	6.25	0.77
Simple span	3	Blue	2.90	2.80	2.65	2.72	2.25	2.20	0.86
		Black	0.97	1.77	2.10	2.20	1.90	1.30	0.62
		Total	3.87	4.57	4.75	4.92	4.15	3.50	0.71
	4	Blue	3.42	3.53	3.20	3.25	2.40	2.72	0.77
		Black	0.90	1.57	1.75	1.73	1.15	1.33	0.53
		Total	4.32	5.10	4.95	4.98	3.55	4.05	0.64
	5	Blue	4.22	3.92	3.53	3.30	3.17	2.95	0.70
		Black	0.83	1.30	1.45	1.13	0.75	0.83	0.42
		Total	5.05	5.22	4.98	4.43	3.92	3.78	0.57

tively), BF_{10} = 65.73 and BF_{10} = 110.15, respectively. For the maxispan group, we also obtained decisive evidence for the presence of an effect of the number of blue letters on black letters recall accuracy, BF_{10} = 238.87. Post hoc analyses provided moderate support for an absence of difference, BF_{01} = 3.10, between the three- and four-blue letters conditions (M = .77, SE = .02, and M = .74, SE = .02, respectively), but strong evidence in favor of a difference between the four- and five5-blue letters (M = .64, SE = .03) conditions, BF_{10} = 29.03.

Serial position curves. As in the previous experiments, recall was dominated by a strong primacy effect, with no recency effect as in Experiment 1 (the rate of recall decreased by 2% between the penultimate and the last serial position in the maxispan condition, whereas there was no difference in the simple span condition, Figure 13). As in Experiment 1, the analysis revealed no effect of serial position for the rehearsed blue letters in the maxispan procedure, except when their number was five, in which case their rate of recall decreased from Position 1 to 3 (from .93 to .86) and increased from Position 3 to 5 (rate of recall of .90). By contrast, the recall of blue letters in the simple span condition was affected by a clear primacy effect with no recency effect, recall rates decreasing as serial position increased for four and five blue letters

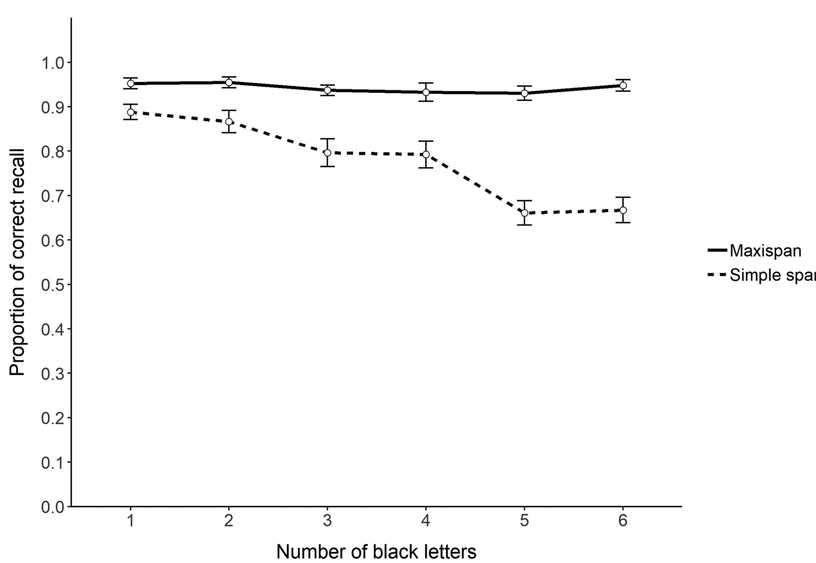

Figure 11. Mean proportion of blue letters correctly recalled at their serial position as a function of the number of following black letters in Experiment 3. Error bars represent standard error of the mean.

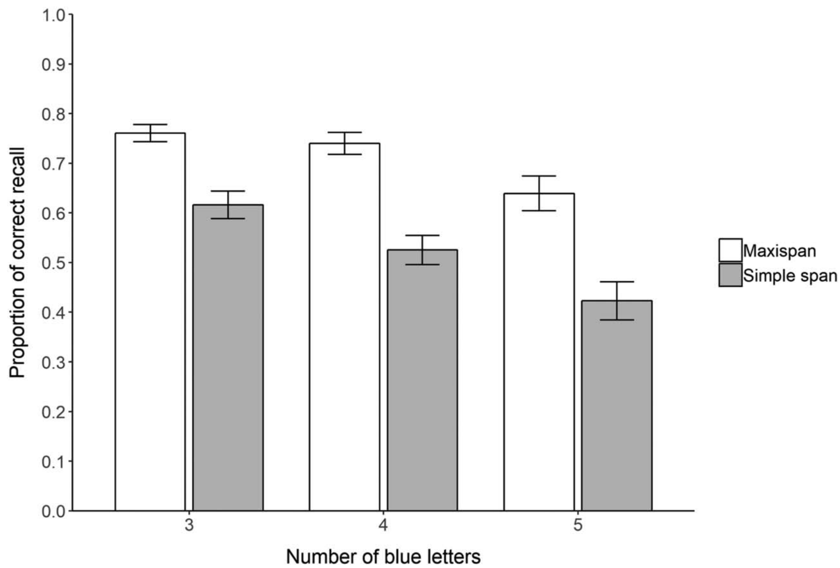

Figure 12. Means proportion of black letters correctly recalled at their serial position as a function of the number of blue letters to rehearse in Experiment 3. Error bars represent standard error of the mean.

(there was only anecdotal evidence for this effect with three blue letters, see Appendix A for statistical analyses and figures).

Discussion

As we predicted, the combined auditory and visual presentations of the memoranda led to a large difference of about 1.4 letters between the maxispan and the simple span procedures, the mean letter span approaching eight (7.73). Ten out of the 20 participants involved in the maxispan condition reached a maxispan of eight or more. Contrary to the previous experiments, the difference between maxispans and simple spans was significant even when three letters were to be rehearsed. The high maxispans observed in the present experiment compared with Experiment 1 were mainly due to a better recall of the black letters, while the blue letters were almost at ceiling in two out of the three conditions of the factor Number of blue letters, with rates of correct recall of blue letters of .98, .95, and .89 for three, four, and five blue letters, respectively. The overall rates of correct recall of black letters for these three conditions were .76, .74 and .64, respectively, which can be compared with those observed in Experiment 1 (.68, .67, and .54). Because black letters were presented in the same way in both experiments (i.e., visually), this suggests that the blue letters interfered more strongly with the following letters when presented visually in Experiment 1 than auditorily in Experiment 3. This suggests in turn that the efficiency of the maxispan procedure relies not only on the storage of the two segments of the letter lists in two different systems, but also in distinct representational formats. The implications of these findings for our understanding of the structure and functioning of verbal WM are addressed in the following General Discussion.

General Discussion

The present study aimed at verifying that theoretically grounded instructions given to participants performing a serial recall task led to a dramatic increase in recall performance. According to a dual-system approach of verbal WM, the maintenance of information is supported by two separate systems, a phonological loop in which items are maintained through articulatory rehearsal, and a central system described as an episodic buffer in Baddeley's (2000) model or an executive loop in the TBRS model (Barrouillet & Camos, 2015). Within this latter model, information would be maintained in the executive loop through attentional refreshing. Because maintenance capacity for letters is approximately of four units in both systems, their optimal use would lead to a letter span approaching eight, while letter span in adults is typically about six (Dempster, 1981). Based on these data, we hypothesized that, unaware of the structure of their WM, individuals misuse it and tend to overload their phonological loop. Thus, we designed the maxispan procedure aiming to limit the number of items entering the phonological loop, the following items being hypothesized to enter the executive loop. In three successive experiments, the maxispan procedure resulted in higher spans compared with a traditional simple span task with no specific instructions. Importantly, the way of presenting the stimuli that was predicted by the theory to produce the highest performance in the maxispan procedure, that is an auditory presentation of the letters to be maintained in the phonological loop and a visual presentation of the letters meant to be attentionally maintained, actually resulted in spans that approached the theoretical value of eight, maximizing the difference with the simple span (about 1.4 items). The present findings have theoretical implications for our understanding of verbal WM and the way ISR is usually interpreted, but also practical implications concerning the interpretation of performance on simple spans tasks routinely used for psychological assessment.

Structure and Functioning of Verbal WM

We hypothesized that the interplay between a phonological loop and an attentional system (the executive loop in the TBRS model)

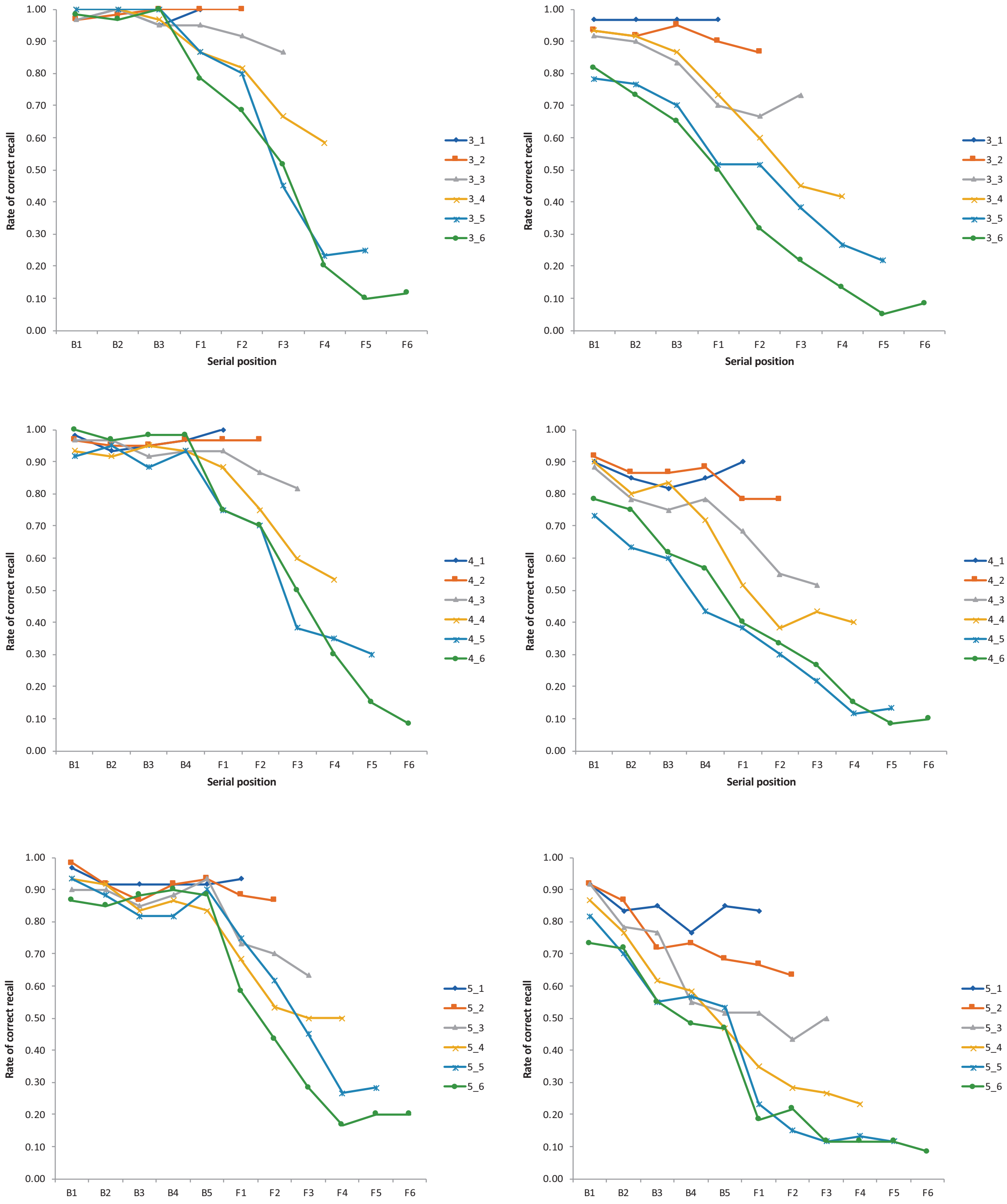

Figure 13. Rate of correct recall of the letters as a function of their serial position and list length in Experiment 3. Left and right panels are for the maxispan and the simple span procedures, respectively. In the x-axes, B and F refer to blue and black letters, respectively. List lengths are expressed as the number of blue and black letters (e.g., 4–3 refers to a series with four blue and 3 black letters). See the online article for the color version of this figure.

contributes to serial recall performance. In addition to the predicted increase of immediate memory spans with the maxispan procedure, several findings strongly support the existence of the dual structure of verbal WM, as well as the independence of the two systems.

First, the clear discontinuity of the serial position curves between the rehearsed and nonrehearsed segments in the maxispan procedure reveals two functionally different systems (Figures 5, 9, and 13). Whereas recall of rehearsed letters, at least up to four items, was virtually perfect and consequently resulted in flat serial position curves, recall of black letters was characterized by steep primacy curves, the rate of correct recall decreasing as serial position increased.

Second, recall of the rehearsed letters in the maxispan procedure remained largely unaffected by the number of subsequent letters to be memorized, whereas the recall rate for the same letters in the simple span procedure progressively decreased as the number of the following "black letters" increased (see Figures 3, 7, and 11). Even in Experiment 2 in which the first letters in the maxispan procedure were rehearsed concurrently to the presentation of the following letters presented auditorily, creating a possible source of interference and confusion, recall of these first rehearsed letters proved to be far more unaffected by the number of the following letters than in the simple span procedure. At the same time, as long as the theoretically grounded four-item capacity of the phonological loop was not exceeded, increasing the number of rehearsed letters from three to four in the maxispan procedure did not greatly affect the rate of recall of the following letters, whereas large drops in performance were observed in the simple span procedure (Figures 4, 8, and 12). This strongly suggests that, in the maxispan procedure, the rehearsed letters and the following letters were stored and maintained in separate systems with minimal interference between them.

Third, the to-be-rehearsed letters were systematically better recalled in the maxispan than in the simple span procedure. This finding demonstrates that, contrary to recent claims (Lewandowsky & Oberauer, 2015), verbal rehearsal is not an epiphenomenon without any impact on WM, but a highly efficient way of maintaining verbal information for further recall. Our results echo previous findings by Rundus (1971) or more recently Tan and Ward (2000) in the domain of free recall showing that items cumulatively rehearsed until the end of the list presentation are better recalled than the other items. In the same way, they support Tan and Ward's (2008) claim that similar mechanisms underpin rehearsal and recall in ISR (see also Burgess & Hitch, 1999). As long as the length of the rehearsed segment does not exceed four and allows for a rehearsal to the end of the list, this cumulative rehearsal leads to virtually perfect serial recall. Together with the evidence of independence of the two systems (e.g., Camos et al., 2009) and the results of the study by Vergauwe et al. (2014) evoked in the Introduction section, our results provide additional evidence for the existence of a phonological loop as Baddeley (1986) described, that is a peripheral system separated from the central attention-based system and endowed with its own system of maintenance (i.e., articulatory rehearsal). However, the capacity of this articulatory system seems to be more limited than it has been traditionally thought. While moving from three to four rehearsed letters had almost no impact on the recall of the black letters in Experiments 1 and 3, this recall began to be impacted when adding a fifth letter to be rehearsed. This suggests that the maintenance of this fifth letter began to impinge on the system used to maintain the black letters. Thus, although verbal rehearsal proves to be an efficient way of maintaining verbal items in WM, verbal STM spans probably do not reflect the capacity of the sole phonological loop.

Fourth, despite the fact that the black letters were encoded under articulatory suppression in the maxispan procedure, they were better recalled than in the simple span procedure. This suggests that they were stored in a system different from the phonological loop in which they could not enter, and that the maintenance of these items in this system is not strongly affected by concurrent articulation. The most plausible hypothesis is that this system is attentional in nature. Indeed, there is empirical evidence that the maintenance of information that cannot be maintained through articulatory rehearsal requires attention. In simple or complex span tasks, when information is maintained under articulatory suppression that prevents storage in the phonological loop, a concurrent attention-demanding task has a detrimental effect on recall performance, suggesting that the maintenance of information that cannot enter the phonological loop requires attention (Camos et al., 2009; Camos, Mora, & Barrouillet, 2013; Camos, Mora, & Oberauer, 2011; Hudjetz & Oberauer, 2007; Mora & Camos, 2013). The other way round, a memory load maintained under articulatory suppression postpones the completion of concurrent attention demanding tasks (Camos et al., 2019; Vergauwe et al., 2014). Considered together with our results, this strongly suggests that nonarticulatory maintenance involves attention, as suggested by several WM models such as the TBRS or the embedded processes model (Cowan, 2005). However, in our study, the number of letters that could be maintained by the mean of the attentional system never reached four, the capacity of the focus of attention hypothesized by Cowan (2001). Indeed, the mean spans in the conditions with three, four, or five letters to rehearse never reached the size that a focus of attention storing four items would have allowed to reach (i.e., seven, eight, and nine, respectively). However, it remains possible that recall performance of the nonrehearsed black letters was reduced by output interference, these letters being recalled after the rehearsed letters. Inverting the order of storage by presenting first the black letters under concurrent articulation of an irrelevant material, and then the letters to be rehearsed would provide us with a more accurate estimate of the capacity of the attentional system.

Overall, our findings confirm the hypothesis that simple spans or WM spans for verbal material reflect the intervention of two systems. These two systems are described in the TBRS model as an executive and a phonological loop. This proposal is akin to that of Baddeley and Hitch (1974) who proposed a WM system acting as a central executive with a storage function, and a supplementary articulatory loop with an estimated capacity of about three items. Indeed, they observed that a memory load of three items had no effect on a concurrent reasoning task that was nonetheless affected by a memory load of six. Nevertheless, the existence of this articulatory loop might be interpreted in two distinct ways. A first hypothesis might assume that the supplementary articulatory loop is a genuine storage system holding items that are separated from those maintained in the more central executive system. The articulatory loop and the central system would have their own mechanism of maintenance. Alternatively, it might be assumed that there is a single store holding verbal items on which two different strategies of maintenance could operate, a strategy of articulatory rehearsal on the one hand and, on the other, a strategy based on attention such as the attentional refreshing hypothesized by Raye, Johnson, Mitchell, Greene, and Johnson (2007) and assumed by the TBRS model as maintaining WM representations in the executive loop (see Camos et al., 2018, for a review).

Different Maintenance Mechanisms or Different Systems?

The view according to which two different strategies would operate on verbal items stored in a unique store has been examined by Lewandowsky and Oberauer (2015). The authors conceptualized "rehearsal as involving the serial retrieval of memoranda, perhaps into a 'focus of attention' (e.g., Oberauer, 2002), followed by the (strengthened) reencoding of each retrieved item" (Lewandowsky & Oberauer, 2015, p. 680). Within this conception, articulatory rehearsal would approximately have the same effect as attentional refreshing. Both mechanisms would retrieve items in some memory store and reactivate them by their reencoding before complete loss. The two mechanisms would operate in parallel, the only difference between them being that rehearsal is assumed to take more time. Whereas attentional refreshing is assumed to be a very fast mechanism reactivating memory traces at a rate of one item every 50 ms (Camos et al., 2019; Vergauwe et al., 2014), articulatory rehearsal is far slower, taking between 250 ms and 500 ms per item (Lewandowsky & Oberauer, 2015; Page & Norris, 1998). Lewandowsky and Oberauer (2015) made several attempts of computational simulations of the parallel functioning of these two maintenance mechanisms on a unique store. These attempts having been unsuccessful, Lewandowsky and Oberauer conclude that articulatory rehearsal is an unworkable solution for the nonexistent problem of decay. However, this failure could simply reveal that the single-store two-strategy hypothesis on which their simulations were based is not adequate to account for WM functioning.

Our results indicate that the hypothesis of two storage systems with their own mechanism of maintenance should be preferred. The fact that increasing the number of rehearsed letters from three to four in the maxispan procedure did not affect the recall of the following black letters, while the recall of the rehearsed letters remained virtually perfect, is particularly enlightening. This means that adding one letter to the lists studied in the three to-be-rehearsed letters condition resulted in an increase of one unit in recall performance. If all the letters of the lists studied in the maxispan procedure were maintained in a single store, adding one item to these lists should have the same effect whatever its nature (i.e., either a blue to-be-rehearsed letter or a black letter). This is not what we observed. For example, in Experiment 1, lists of three blue letters followed by four black letters (for a total of seven letters) elicited the correct recall of 5.32 letters. Adding a to-be-rehearsed blue letter (for a total of eight letters, four blue and four black) resulted in 6.43 letters correctly recalled, whereas adding a black letter (for a total of still eight letters, but three blue and five black) was far from having the same effect with only 5.15 letters correctly recalled. Note that this finding also strongly contradicts the claim that verbal rehearsal is not an efficient mechanism of maintenance as Lewandowsky and Oberauer (2015) asserted. If this was the case, adding a letter to be rehearsed should have no effect on recall performance. Instead of supposing a single store in which two maintenance mechanisms operate in parallel, it seems more appropriate to consider the existence of two systems. This view allows to explain the effect of the addition of a blue letter. Because three letters do not exhaust the capacity of the articulatory loop, it can hold an additional letter without any effect on the functioning of the other, attentional, system.

Another way to salvage the hypothesis of a unique system might be to argue that the higher recall performance in the maxispan procedure is due to some chunking or subjective grouping strategy induced by the presentation of the letters in two groups differing by their color (i.e., blue and then black) and their treatment (rehearsing only the first group). Indeed, there is some evidence that participants could spontaneously use a subjective grouping strategy during verbal STM tasks, even in the absence of any grouping cue (e.g., Farrell, 2008; Henson, 1996), this grouping resulting in better recall performance. However, grouping is known to result in specific serial position effects with a scalloped appearance of the curve characteristic of temporal (Hartley, Hurlstone, & Hitch, 2016; Henson, 1996; Ryan, 1969) and perceptual grouping (Frankish, 1985, 1989), primacy and recency effects occurring within each group (e.g., Farrell, 2012; Henson, 1998; Hitch, Burgess, Towse, & Culpin, 1996). It can be verified in Figures 5, 9, and 13 that nothing of this sort occurred in the maxispan procedure or in the simple span condition either. In fact, the peculiar serial position curves produced by grouping have very little to do with the curves elicited by the maxispan procedure with their quasi perfect recall of the rehearsed letters followed by the monotonic primacy effect with a small and rare recency effect affecting recall of the black letters (Figures 5 and 13).

Another signature of the grouping effect is the specific patterns of transposition errors (i.e., recalling a presented item but at the wrong serial position). There is evidence that temporal grouping increases the rate of *interpositions*, the tendency to confuse items that share the same position within their respective group; for example, recalling the second item of a group in the second position of another group (Henson, 1996). If the maxispan procedure had induced the use of a grouping strategy, we should thus have observed in this condition the same effect on transposition errors as observed with temporal grouping, that is an increase of interpositions in the maxispan compared to the simple span condition. By contrast, our hypothesis was that the maxispan procedure results in better recall because items are maintained in two separate systems. This segregated maintenance should not increase, but reduce the rate of between-groups transpositions (between blue and black letters), including interpositions, compared with a simple span procedure.

The analysis of the errors revealed that the mean number of transpositions (i.e., inter- or intragroup) was higher in the simple span than in the maxispan condition (across all the experiments, 79.15 and 50.45, respectively), something not so surprising when considering the better recall performance in the maxispan condition. However, among these transpositions errors, the proportion of between-groups transpositions (from blue to black letters or vice versa) was systematically higher in the simple span than in the maxispan procedure (across all the experiments, 34% and 26%, respectively), a difference that was also observed for the specific type of between-groups transpositions that we defined earlier as interpositions (6% and 3%, respectively; see detailed analyses and statistics in Appendix B).

Thus, neither serial position curves nor patterns of errors corresponded to what might be expected if the increase in recall performance observed in the maxispan procedure was due to a grouping of the items as studied in the ISR literature (e.g., Farrell, 2012; Hitch et al., 1996). On the contrary, the analysis of serial position curves and patterns of errors confirmed that the maxispan procedure elicited higher spans by allowing the segregation of the rehearsed and the following letters in two distinct and separate systems. This view is also supported by evidence that the difference between the two conditions in terms of the proportion of between-groups transpositions disappeared in Experiment 2 where all the letters were presented auditorily. Indeed, black letters presented auditorily could have automatically entered the phonological loop and let a trace that could have increased the rate of between-groups transpositions. In line with this idea, when the format of presentation of letters favored the separated encoding and maintenance into the two systems by presenting blue letters auditorily and black letters visually, as in Experiment 3, the simple span procedure group exhibited its lowest rate of between-groups transpositions.

It is worth to note that previous neurological evidence support the hypothesis that verbal WM does not involve one, but two different systems. Gruber (2001) observed that when articulatory rehearsal can be used in verbal WM tasks, the Broca's area, the left premotor cortex, the cortex along the left intraparietal sulcus, and the right cerebellum are activated. However, when articulatory rehearsal is prevented by articulatory suppression, another network is activated including the anterior prefrontal cortex and the inferior parietal cortex. In the same way, Raye et al. (2007) observed that Broca's area is selectively activated by rehearsal, whereas attentional refreshing selectively involves activation of the dorsolateral prefrontal cortex. Double dissociations in patients have also been observed, with reduced recall performance under articulatory rehearsal while nonarticulatory maintenance remained unimpaired after a brain lesion in the Broca's area, and conversely, diminished memory performance through refreshing but preserved subvocal rehearsal following a lesion in the anterior middle frontal gyrus (Trost & Gruber, 2012).

Thus, both behavioral and neurological evidence converge toward the idea that two systems are involved in immediate verbal memory. One of these systems seems to involve brain areas usually recruited in executive control such as the dorsolateral prefrontal cortex, whereas the other system seems to be based on mechanisms related to speech production and Broca's area. However, the nature of these systems remains to be specified, and several hypotheses have been put forward about their nature, structure, and functioning.

What Is the Nature of the Different Systems?

As we predicted, the maxispan performance seems to reflect the interplay of two separate systems that the TBRS model describes as a phonological loop and an executive loop and Baddeley's model as a phonological loop along with either a central executive with storage capacity (Baddeley & Hitch, 1974) or an episodic buffer (Baddeley, 2000). However, other models have assumed a dual structure of WM and might constitute alternative accounts of our findings. As we mentioned earlier, Unsworth and Engle (2007) have proposed that a primary memory and a secondary memory, which are qualitatively and functionally distinct, jointly contribute to recall in simple and complex span tasks. They assume that primary memory can hold up to four items, a number that echoes what we observed in the recall of blue letters. When more items have to be maintained, as it is the case in simple span tasks and in the present experiments, some items would be displaced into secondary memory (Unsworth et al., 2010). It might be assumed that, in the maxispan procedure, blue letters are held in primary memory while black letters are retrieved from secondary memory. However, this appealing hypothesis does not correspond to what we observed.

First, despite a coincidental similarity in capacity, several findings suggest that what we call the phonological loop is not the primary memory hypothesized by Unsworth and Engle (2007). Primary memory, in their theoretical framework, is the equivalent of the focus of attention in Cowan's (2005) model or the activation buffer in Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, and Usher's (2005) model, and serves "to maintain a distinct number of separate representations active for ongoing processing by means of the continued allocation of attention" (Unsworth & Engle, 2007, p. 106; see Hall, Jarrold, Towse, & Zarandi, 2015, for a similar conception). However, as we mentioned earlier, Baddeley and Hitch (1974) as well as Vergauwe et al. (2014) have shown that the maintenance through articulatory rehearsal of up to four letters has no impact on concurrent attention-demanding activities, suggesting that this maintenance does not depend on the continued allocation of attention. Thus, the mechanism for maintaining verbal information through verbal rehearsal is not the primary memory evoked by Unsworth and Engle's (2007) model.

Second, the secondary memory is described by Unsworth and Engle (2007) as functionally distinct from primary memory and explicitly compared with the long-term store in Atkinson and Shiffrin (1968) modal model. This suggests a more passive maintenance in secondary than in primary memory that does not correspond to what we observed concerning black letters. Passive maintenance of short lists of items has been associated in the literature with serial position curves exhibiting a recency effect instead of the primacy effect that characterizes active maintenance as in ISR. For example, Palladino and Jarrold (2008) compared recall of short lists of the same length (four or five items) in either an ISR task or a running span task (a task in which participants are instructed to recall the *n* last items of a list of unpredictable length presented rapidly). The presentation of short lists in the running span task was intended to induce a maintenance of items without any update, the number of items presented not exceeding the number of items to recall. While standard serial position curves with primacy effect were observed in the ISR task, recall in the running span task was characterized by marked recency with reduced or absent primacy effects. Palladino and Jarrold's (2008) interpretation of this latter finding was that participants in the running span task waited passively for the end of the list before trying to recall the items. The same type of recency-based serial position curves in running span with the same interpretation of passive maintenance was reported by Bunting, Cowan, and Saults (2006). Thus, if the black letters were passively held, their recall should be characterized by recency effects. On the contrary, recall of the nonrehearsed black letters in the maxispan procedure was dominated by marked primacy effects with almost no recency effects (see Figures 5, 9, and

13). This suggests that black letters were not passively, but actively maintained. It is thus doubtful that recalled black letters were retrieved from the secondary memory described by Unsworth and Engle (2007)in which they would have been passively maintained until retrieval and recall.

Note that our results suggest that the black letters were not retrieved from some sensory memory either. It has been assumed that sensory memory stores are the structures from which items are retrieved in running span tasks (Bunting et al., 2006; Cowan, 2005), and we have seen that the serial position curves for black letters do not correspond to the curves observed in running span. Studies on dichotic listening can also inform us about the mnemonics systems involved in the maxispan procedure. In dichotic listening, participants are presented with items in both ears simultaneously, but asked to attend only one stream before recalling the two lists. Results have revealed a pattern of quasi perfect recall for the attended stream assumed to be maintained in primary memory, whereas recall from the unattended hear follows a steep and monotonic recency curve assumed to reflect storage in sensory memory (Broadbent, 1958; Bryden, 1971). It is worth noting that even in Experiment 2 where a recency effect affected the last item (see Figure 9), recall of black letters was nonetheless dominated by a primacy effect. Thus, nonrehearsed black letters in the maxispan procedure do not seem to have been passively held in some secondary or sensory memory.

The serial position curves with marked primacy effect we observed for black letters rather reveal an active maintenance. Because this maintenance was performed under concurrent articulation, it probably relies on attentional allocation. The executive loop is the system devoted to this type of maintenance in the TBRS model, a system closer to the primary than to the secondary memory in Unsworth and Engle's (2007) theory. Thus, our results point toward a dual structure of verbal WM in which, instead of a primary and a secondary memory, a rehearsal-based system and an attention-based system concur to maintaining items. However, although we have thus far characterized the rehearsal-based system as a phonological loop, this apparently simple system has received different descriptions and theorizations that deserve some comments.

The Nature of the Articulatory Loop

There are at least two hypotheses concerning the system that we have described as a phonological loop. The first, developed by Baddeley (1986), assumes that the system comprises a phonological store capable of holding speech-based information, and an articulatory control based on inner speech. Memory traces within the store are assumed to suffer from temporal decay, becoming unretrievable after about 1.5 s or 2 s. This loss can be prevented by the articulatory loop process that reactivates memory traces by reading them off from the store and feeding them back into this store. This system can also convert written material into a phonological code that is passed to and can be registered in the phonological store. Empirical evidence was gathered that memory traces within the store involve codes that are phonological, but not semantic (Salamé & Baddeley, 1982) and not purely auditory because noise does not disrupt verbal maintenance (Salamé & Baddeley, 1987, 1989). The hypothesis that this system stores phonological representations is supported by the fact that it is more difficult to remember phonologically similar than dissimilar items, due to the confusions that occur between phonologically similar items when they are retrieved from the store for reactivation or recall (Baddeley, 1966; Conrad & Hull, 1964). However, Baddeley (1997) noted that the nature of the store, its time characteristics, and how information is retrieved from the store and read by the articulatory system remain unknown. According to this view of the phonological loop, and as Baddeley (1997, p. 59) stressed, "the essence of the phonological loop hypothesis is that memory span will depend on rate of rehearsal, being approximately equivalent to the number of items that can be spoken in two seconds." The capacity of the phonological loop would therefore be about seven items.

More than 10 years before the previously described hypothesis, Baddeley made a slightly different proposal based on the framework suggested by Baddeley and Hitch (1974), that is, before the reconceptualization by Baddeley (1986) of a central executive devoid of any storage capacity. This second hypothesis assumes the existence of an executive WM system supplemented by an articulatory rehearsal loop with a capacity of about three, and not seven, items (Baddeley et al., 1975). This smaller capacity is explained by the fact that the loop is only an auxiliary system. When the access to the loop is prevented, for example by concurrent articulation, memory items can be stored in the executive system that is endowed with storage capacities. The other critical difference relative to the previous hypothesis is the nature of what is stored in the phonological loop. Following Sperling (1963), Baddeley et al. (1975) suggested that the loop could be an output buffer holding the motor program necessary for the verbal production of the memory items. This buffer might be separate from the act of articulation itself, permitting to set up new articulatory programs while current programs are running. These articulatory programs could be set up either by the act of articulation, but also directly through auditory stimulation. Interestingly, more recent studies have indeed established that the perception of speech activates the motor areas involved in speech production (e.g., Wilson, Saygin, Sereno, & Iacoboni, 2004). Such a loop was presented by Baddeley et al. (1975) as a backup system for the immediate retention of phonological material. The authors acknowledged the very tentative aspect of this view and noted that it left underspecified the problem of how such a system maintaining articulatory motor programs would interface with the central executive system.

Do one of these two theoretical views can account for the data we observed with the maxispan procedure? Considering the capacity of the phonological loop estimated by Vergauwe et al. (2014) to four items along with the necessary dual structure of verbal WM lead us to favor the second hypothesis. This approach can be seen as the reverse of that adopted by Baddeley (1986) in order to simplify the theoretical options concerning the maintenance of verbal information in WM. While he suggested operating this simplification by depriving the central executive of its storage function, which would be entirely ensured by the phonological loop, we propose to concentrate the maintenance activities in the central attentional system which, like the episodic buffer, would therefore contain multimodal representations which may as well be phonological, semantic, visual, or spatial. This would avoid the potential theoretical conundrum of a verbal WM in which phonological representations could be stored both in the phonological loop and the episodic buffer. This would also help us understanding why the maintenance of verbal information interferes with visuospatial activities when articulatory rehearsal is unavailable due to concurrent articulation or when the capacity of the articulatory loop is exceeded (Vergauwe et al., 2014). This would also account for the fact that, in the same conditions, verbal and visuospatial maintenance interfere with each other (Uittenhove, Chaabi, Camos, & Barrouillet, 2019) and, more generally, for the processing-storage trade-off that has been observed in virtually all the possible combinations of modalities (see Barrouillet & Camos, 2015, in press, for reviews). In the next section, we present an attempt to describe how the phonological loop, which should be more accurately described as an *articulatory loop*, and the central executive system (the executive loop in the TBRS model) could interact when performing an ISR task. This proposal is clearly speculative and aims only at stimulating further research.

A Tentative Account of the Maxispan and the Simple Span

A verbal ISR task as we used in both the maxispan and the simple span procedures involves the successive presentation, either visually or auditorily, of a series of items. When presented auditorily, we assume that the perception of a letter results in the construction of a representation of this letter in the executive loop. This representation could bind phonological, visual (the visual form of the letter), and semantic (e.g., some idiosyncratic association of the letter with words, names or surnames) features. This representation is subjected to temporal decay as soon as it leaves the focus of attention. Listening the letter also activates the motor program associated with its production, this motor program being stored in the motor program buffer, and available for articulating the letter for maintenance purpose. When presented visually, the letter is read off and converted into a phonological code by the activation of the motor program for articulation. We assume that this articulation is the strategy that individuals spontaneously use when receiving the beginning of the series (Tan & Ward, 2008). The presentation of the following letters creates representations in the executive loop and activates the corresponding motor programs for their articulation. In case of a cumulative rehearsal as we instructed participants to use for the first letters, the motor programs of the successive letters might be concatenated to form an articulatory object akin to the program for articulating a word (e.g., BTK articulated in succession as "beeteekey"). Rehearsing this chunk would lead to the strengthening of its articulatory program and produce a verbal output, the perception of which might contribute to even further reinforce this program by its automatic reactivation.

Due to this loop between the motor program buffer and the articulatory mechanism, the maintenance of at least the first letters is possible through a mere echolalia process, a process that rapidly turns into something akin to a palilalia. Consequently, their corresponding symbolic representations created and stored in the executive loop would become useless for maintenance. This cyclic articulatory production in which perception, and subsequently articulation, directly activates the underpinning motor program for repetition is assumed to require very few attention (Naveh-Benjamin & Jonides, 1984). It is worth noting that it constitutes one of the most basic sensorimotor cognitive process in humans, already present in the first months of life, and described by Baldwin (1894) and Piaget (1936) as primary circular reactions. Infants tend to repeat syllables pertaining to their own repertoire, the output of the articulation triggering its repetition in a behavioral loop. At any moment of a serial recall task, this articulatory production can be used for recall or, if needed, for creating or recreating WM representations in the executive loop by encoding its output.

The results from Vergauwe et al. (2014), but also Baddeley and Hitch (1974) indicate that the number of letters that can be maintained through articulatory rehearsal without any implication of the central attentional system is limited to three or four. This suggests that there is a limit in the complexity of the motor program that can be set up and reactivated by its mere articulation without any support of representations from the central attentional system. The rationale of the maxispan procedure presented here is to exploit the maximum capacity limit of this articulatory loop while blocking its access to subsequent letters. In this way, the first to-be-rehearsed letters simply require the automatic reactivation of a motor program through articulation that protect them from forgetting. Despite the fact that representations of these rehearsed letters are probably constructed at encoding, it is likely that these useless representations rapidly fade away. This would free the executive loop, allowing attention-based maintenance of the following letters, hence the observation of higher spans within the maxispan procedure. To sum up, the rehearsed letters are protected from forgetting by the mechanical reactivation of a motor program, leaving the executive loop free for storing and maintaining additional letters.

So, what goes wrong in the simple span procedure that makes simple spans lower than maxispans? Our hypothesis is that, because encoding letters presented either auditorily or visually results in the automatic activation of the corresponding motor programs for articulation via a kind of affordance process (Jones & Macken, 2018; Macken, Taylor, & Jones, 2014), participants probably rely naturally on rehearsal strategy and try to concatenate motor programs, even though the number of items exceeds the capacity of the buffer. When the articulatory program breaks down due to overload, the rehearsal schedule might become chaotic. In their study of rehearsal in ISR, Tan and Ward (2008) observed that this occurs usually after the fourth presented item. Individuals might keep rehearsing a subset of the letters they tried, but failed, to concatenate, or they might attempt to rehearse letters in middle or late serial positions after having rehearsed the first letters that are passed to the executive loop in order to allow rehearsing the following letters. In any way, a small and suboptimal number of letters are probably maintained in the motor program buffer, and it is possible that the set of letters rehearsed change from the beginning to the end of the presentation of the memory list. This suboptimal use of the phonological loop leads to maintain more items in the executive loop, resulting in its overload. Reactivating these letters through articulatory rehearsal would involve retrieving their representations from the executive loop and the associated motor programs for articulation while memory traces of the others letters decay. Moreover, the coordination of verbal rehearsal and attentional refreshing is probably suboptimal, attention being paid to the letters currently articulated leaving the other items decaying.

Thus, our hypothesis is that simple spans are lower than maxispans because the simple span procedure leads to a misuse of the capacity of the articulatory loop by an initial overload of its motor program buffer. By trying to verbally rehearse too many letters, individuals end up using only a part of the capacity of the motor program buffer as exemplified by Tan and Ward (2008), and they consequently overload their executive loop. This is probably what occurred in Souza and Oberauer (2018) who failed to find any beneficial effect of instructed cumulative rehearsal. Participants were presented with lists of six words with several of these words being multisyllabic, their mean number of characters being 7.8 (see OSF https://osf.io/5q7nd/). Thus, the cumulative rehearsal of these lists of six words necessarily exceeded the capacity of the articulatory system. The longer sequences of cumulative rehearsal observed by Souza and Oberauer (2018)involved about four words (e.g., 4.4 and 3.9 in Experiment 1), which already exceeded the capacity of the articulatory loop, most of these words being multisyllabic. Consequently, the instructed cumulative rehearsal studied by Souza and Oberauer (2018) led to the misuse of the dual-system we described above, hence the apparent ineffectiveness of rehearsal. Our experiments demonstrated that when less items are rehearsed, recall performance strongly increases.

Importantly, this suboptimal use of articulatory rehearsal is nonetheless beneficial compared to its absence of use. Performing a serial recall task under articulatory suppression results in lower spans, even when the articulatory suppression is limited to the maintenance phase (see Doherty et al., 2019, for a recent demonstration). Although it has been argued that the effect of articulatory suppression is due to interference, and not impediment of verbal rehearsal reactivating decaying memory traces (e.g., Lewandowsky & Oberauer, 2015), the interference hypothesis does not stand up to scrutiny (see Camos et al., 2013, for discussion). In some sense, simple spans stand between recall performance under concurrent articulation, when the articulatory loop is unavailable, and maxispans in which this articulatory loop is used at the maximum of its capacities.

What Do Simple Spans Measure?

As we noted in introducing this article, immediate memory tasks have been part of intelligence tests from their very inception (Binet & Simon, 1904). But what do verbal simple spans measure? Miller (1956) surmised that STM spans might reflect the capacity of the channel of information that human mind constitutes, but subsequent appraisals reduced this capacity from seven to four (Cowan, 2015a). We have also seen that they do not measure the capacity of the sole phonological loop as Baddeley (1986) suggested. Immediate verbal memory seems to reflect the interplay between two systems, an articulatory loop and a central attentional system described as an executive loop in the TBRS model. However, the maxispan procedure has revealed that simple spans do not reflect the maximal capacity of this dual-system. Our results suggest that verbal simple spans reflect instead the way individuals use a system the structure of which they are unaware of. Performance on these simple span tasks certainly depends on the capacity of the two systems involved. The capacity of these systems might depend on both the capacity of their store (i.e., the motor program buffer for the articulatory loop and the episodic buffer of the executive loop, see Barrouillet & Camos, in press, on this latter point) and the efficiency of their maintenance mechanisms (verbal rehearsal and attentional refreshing, respectively). However, simple spans necessarily depend also on the way individuals are able to coordinate the functioning of the two systems. This strategic aspect of simple spans, which is neutralized in both the maxispan procedure and the span tasks performed under articulatory suppression, introduces a cognitive component that is not mnemonic in nature. It remains to be established whether the maxispan that neutralizes the strategic use of verbal WM is or is not a better predictor of higher cognition and intelligence than simple spans. On the one hand, it might be that the strategic component is key in the predictive power of the capacity of STM, as measured by simple spans. People able to optimize their cognitive functioning in memory tasks might be able of this optimization in other domains, outperforming their peers in tasks assumed to assess high level cognition. On the other hand, this strategic component might blur the relation existing between STM and high level cognition. The maxispan procedure, through a better assessment of the real capacity of the different components of verbal WM on which several high level cognitive processes rely, would in this case be a better predictor of high level cognition than simple spans. Future studies should provide us with the answer to this question.

Conclusion

In some sense, Miller (1956) erred in supposing that verbal STM spans inform us about the capacity of human mind considered as a channel of information. Our results indicate that the number of letters that can be immediately recalled after presentation would not reflect the capacity of a single short-term buffer, but the complex interplay between two storage systems endowed with their own mechanism of maintenance. It is even probable that the mental objects held by these mechanisms differ in nature, the articulatory loop holding motor programs whereas an attentional central system might maintain symbolic representations. Thus, apparently simple behaviors such as the immediate reproduction of a list of verbal items might involve a variety of mental structures and processes. This means that one should be cautious in establishing immediate equivalence between tasks on the one hand and structures or processes on the other. Performance in a seemingly simple task does not necessarily reflect the capacity of a single structure or the efficiency of a single process, but can result from the complex interplay of different structures and processes. Simple spans prove not that simple and measuring STM capacity proves more complex than expected.

This is not to say that the maxispan procedure might be preferred to the venerable simple span task in psychological and neuropsychological assessments. Simple spans, and more precisely the digit span, would have not been used for more than a century with children, adults, and patients if they were not useful. However, beyond the empirical interest and practical usefulness of a measuring tool, one of the aims of the scientific inquiry is to clarify what this tool precisely measures in order to better understand its diagnostic and prognostic power. The complexity of the processes underpinning the apparently simple behavior of immediately repeating a list of verbal items probably explains why simple spans have sometimes been considered as poorly diagnostic of school achievement and intelligence (e.g., Estes, 1981; Glasser & Zimmerman, 1967; Wechsler, 1958), some authors having even assumed, as Gignac and Weiss (2015) recall, that it is not a test of memory span at all, but instead a test of elementary attention (Hannay, Howieson, Loring, Fischer, & Lezak, 2004; Hebben & Milberg, 2009; Rapaport, Schafer, & Gill, 1945; Sbordone & Saul, 2000). Beyond constituting an empirical test of a theoretical framework concerning verbal WM, the present study must be considered as an attempt to further our understanding of what one of the oldest and most popular psychological test measures.

Context of the Research

The maxispan paradigm introduced in the present study came to our mind some years ago (March 2015) when considering that there was a discrepancy between the simple span for letters usually reported in the literature (about six) and what our TBRS model predicted. If there are two maintenance systems with a capacity of four letters each, as the TBRS model assumes, the simple span for letters should be eight and not six. The initial intuition was that a good way to use verbal rehearsal would be to not exceed the capacity of the phonological loop by rehearsing only four items, even in longer lists. Pilot studies confirmed our intuition: The maxispan was higher than the simple span. At a more general level, this research program is intended to explore the nature of the cognitive objects and processes on which WM and immediate memory are based, with the guiding idea that immediate memory is not necessarily, or not uniquely, based on symbolic representations, but could also involve the sensorimotor schemes described by Piaget as well as what Jones and Macken (2018, p. 351) describe as "perceptual objects that may then be apprehended and manipulated by bodily effector systems."

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 89–195). New York, NY: Academic Press.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *The Quarterly Journal of Experimental Psychology, 18,* 362–365. http://dx.doi.org/10.1080/14640746608400055

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Clarendon Press/New York, NY: Oxford University Press.

Baddeley, A. D. (1997). *Human memory: Theory and practice*. Hove, UK: Psychology Press.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423. http://dx.doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D. (2007). *Working memory, thought, and action*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780198528012.001.0001

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–89). San Diego, CA: Academic Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 575–589. http://dx.doi.org/10.1016/S0022-5371(75)80045-4

Baldwin, J. M. (1894). *Mental development in the child and the race*. New York, NY: Macmillan.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133,* 83–100. http://dx.doi.org/10.1037/0096-3445.133.1.83

Barrouillet, P., & Camos, V. (2015). *Working memory: Loss and reconstruction*. Hove, UK: Psychology Press.

Barrouillet, P., & Camos, V. (in press). The time-based resource-sharing model of working memory. In R. H. Logie, V. Camos, & N. Cowan (Eds.), *Working memory: State of the science*. Oxford, UK: Oxford University Press.

Binet, A., & Simon, T. (1904). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux [New methods for the diagnosis of the intellectual level of subnormals]. *L'Année Psychologique, 11,* 191–244. http://dx.doi.org/10.3406/psy.1904.3675

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review, 113,* 201–233. http://dx.doi.org/10.1037/0033-295X.113.2.201

Broadbent, D. E. (1958). *Perception and communication*. London, UK: Pergamon Press. http://dx.doi.org/10.1037/10037-000

Broadbent, D. E. (1975). The magic number seven after fifteen years. In A. Kennedy & A. Wilkes (Eds.), *Studies in long term memory* (pp. 3–18). Oxford, UK: Wiley.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107,* 127–181. http://dx.doi.org/10.1037/0033-295X.107.1.127

Bryden, M. P. (1971). Attentional strategies and short-term memory in dichotic listening. *Cognitive Psychology, 2,* 99–116. http://dx.doi.org/10.1016/0010-0285(71)90004-1

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

Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language, 31,* 429–460. http://dx.doi.org/10.1016/0749-596X(92)90022-P

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. http://dx.doi.org/10.1037/0033-295X.106.3.551

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

Camos, V. (2015). Storing verbal information in working memory. *Current Directions in Psychological Science, 24,* 440–445. http://dx.doi.org/10.1177/0963721415606630

Camos, V. (2017). Domain-specific versus domain-general maintenance in working memory: Reconciliation within the time-based resource sharing model. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 67, pp. 135–171). Cambridge, MA: Academic Press.

Camos, V., Johnson, M., Loaiza, V., Portrat, S., Souza, A., & Vergauwe, E. (2018). What is attentional refreshing in working memory? *Annals of the New York Academy of Sciences, 1424,* 19–32. http://dx.doi.org/10.1111/nyas.13616

Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language, 61,* 457–469. http://dx.doi.org/10.1016/j.jml.2009.06.002

Camos, V., Mora, G., & Barrouillet, P. (2013). Phonological similarity effect in complex span task. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 66,* 1927–1950. http://dx.doi.org/10.1080/17470218.2013.768275

Camos, V., Mora, G., & Oberauer, K. (2011). Adaptive choice between articulatory rehearsal and attentional refreshing in verbal working memory. *Memory & Cognition, 39,* 231–244. http://dx.doi.org/10.3758/s13421-010-0011-x

Camos, V., Mora, G., Oftinger, A-L., Mariz Elsig, S., Schneider, P., & Vergauwe, E., (2019). Does long-term memory affect refreshing in verbal working memory? *Journal of Experimental Psychology: Learning, Memory and Cognition, 45,* 1664–1682. http://dx.doi.org/10.1037/xlm0000657

Cattel, J. M. (1890). Mental tests and measurements. *Mind, 15,* 373–381. http://dx.doi.org/10.1093/mind/os-XV.59.373

Chen, Z., & Cowan, N. (2005). Chunk limits and length limits in immediate recall: A reconciliation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1235–1249. http://dx.doi.org/10.1037/0278-7393.31.6.1235

Chen, Z., & Cowan, N. (2009). Core verbal working-memory capacity: The limit in words retained without covert articulation. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 62,* 1420–1429. http://dx.doi.org/10.1080/17470210802453977

Colom, R., Flores-Mendoza, C., Quiroga, M. Á., & Privado, J. (2005). Working memory and general intelligence: The role of short-term storage. *Personality and Individual Differences, 39,* 1005–1014. http://dx.doi.org/10.1016/j.paid.2005.03.020

Colom, R., Jung, R. E., & Haier, R. J. (2007). General intelligence and memory span: Evidence for a common neuroanatomic framework. *Cognitive Neuropsychology, 24,* 867–878. http://dx.doi.org/10.1080/02643290701781557

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55,* 429–432. http://dx.doi.org/10.1111/j.2044-8295.1964.tb00928.x

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104,* 163–191. http://dx.doi.org/10.1037/0033-2909.104.2.163

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139174909.006

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–114. http://dx.doi.org/10.1017/S0140525X01003922

Cowan, N. (2005). *Working memory capacity*. New York, NY: Psychology Press.

Cowan, N. (2010). The magical mystery four. *Current Directions in Psychological Science, 19,* 51–57. http://dx.doi.org/10.1177/0963721409359277

Cowan, N. (2015a). George Miller's magical number of immediate memory in retrospect: Observations on the faltering progression of science. *Psychological Review, 122,* 536–541. http://dx.doi.org/10.1037/a0039035

Cowan, N. (2015b). Sensational memorability: Working memory for things we see, hear, feel, or somehow sense. In P. Jolicoeur, C. Levebre, & J. Martinez-Trujillo (Eds.), *Mechanisms of sensory working memory: Attention and perfomance XXV* (pp. 5–22). London, UK: Academic Press. http://dx.doi.org/10.1016/B978-0-12-801371-7.00002-8

Cowan, N., Blume, C. L., & Saults, J. S. (2013). Attention to attributes and objects in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 731–747. http://dx.doi.org/10.1037/a0029687

Cowan, N., Morey, C. C., & Chen, Z. (2007). The legend of the magical number seven. In S. Della Sala (Ed.), *Tall tales about the mind & brain: Separating fact from fiction* (pp. 45–59). Oxford, UK: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780198568773.003.0005

Cowan, N., Rouder, J. N., Blume, C. L., & Saults, J. S. (2012). Models of verbal working memory capacity: What does it take to make them work? *Psychological Review, 119,* 480–499. http://dx.doi.org/10.1037/a0027791

Cowan, N., Saults, J. S., & Blume, C. L. (2014). Central and peripheral components of working memory storage. *Journal of Experimental Psychology: General, 143,* 1806–1836. http://dx.doi.org/10.1037/a0036814

Craik, F. I. (1971). Primary memory. *British Medical Bulletin, 27,* 232–236. http://dx.doi.org/10.1093/oxfordjournals.bmb.a070859

Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review, 112,* 3–42. http://dx.doi.org/10.1037/0033-295X.112.1.3

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin, 89,* 63–100. http://dx.doi.org/10.1037/0033-2909.89.1.63

Doherty, J. M., Belletier, C., Rhodes, S., Jaroslawska, A., Barrouillet, P., Camos, V., . . . Logie, R. H. (2019). Dual-task costs in working memory: An adversarial collaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45,* 1529–1551. http://dx.doi.org/10.1037/xlm0000668

Estes, W. K. (1981). Intelligence and learning. In M. P. Friedman, J. P. Das, & N. O'Connor (Eds.), *Intelligence and learning* (pp. 3–23). New York, NY: Plenum Press. http://dx.doi.org/10.1007/978-1-4684-1083-9_2

Farrell, S. (2008). Multiple roles for time in short-term memory: Evidence from serial recall of order and timing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 128–145. http://dx.doi.org/10.1037/0278-7393.34.1.128

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

Frankish, C. R. (1985). Modality-specific grouping effects in short-term memory. *Journal of Memory and Language, 24,* 200–209. http://dx.doi.org/10.1016/0749-596X(85)90024-5

Frankish, C. (1989). Perceptual organization and precategorical acoustic storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 469–479. http://dx.doi.org/10.1037/0278-7393.15.3.469

Gignac, G. E., & Weiss, L. G. (2015). Digit Span is (mostly) related linearly to general intelligence: Every extra bit of span counts. *Psychological Assessment, 27,* 1312–1323. http://dx.doi.org/10.1037/pas0000105

Glasser, A. J., & Zimmerman, I. L. (1967). *Clinical interpretation of the Wechsler Intelligence Scale for Children (WISC)*. New York, NY: Grune & Stratton.

Gruber, O. (2001). Effects of domain-specific interference on brain activation associated with verbal working memory task performance. *Cerebral Cortex, 11,* 1047–1055. http://dx.doi.org/10.1093/cercor/11.11.1047

Halford, G. S. (1993). *Children's understanding*. Hillsdale, NJ: Erlbaum.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21,* 803–831. http://dx.doi.org/10.1017/S0140525X98001769

Hall, D., Jarrold, C., Towse, J. N., & Zarandi, A. L. (2015). The developmental influence of primary memory capacity on working memory and academic achievement. *Developmental Psychology, 51,* 1131–1147. http://dx.doi.org/10.1037/a0039464

Hannay, H. J., Howieson, D. B., Loring, D. W., Fischer, J. S., & Lezak, M. D. (2004). Neuropathology for neuropsychologists. In M. D. Lezak, D. B. Howieson, & D. W. Loring (Eds.), *Neuropsychological assessment* (4th ed., pp. 157–285). New York, NY: Oxford University Press.

Hartley, T., Hurlstone, M. J., & Hitch, G. J. (2016). Effects of rhythm on memory for spoken sequences: A model and tests of its stimulus-driven mechanism. *Cognitive Psychology, 87,* 135–178. http://dx.doi.org/10.1016/j.cogpsych.2016.05.001

Hebben, N., & Milberg, W. (2009). *Essentials of neuropsychological assessment*. Hoboken, NJ: Wiley.

Henderson, L. (1972). Spatial and verbal codes and the capacity of STM. *The Quarterly Journal of Experimental Psychology, 24,* 485–495. http://dx.doi.org/10.1080/14640747208400308

Henson, R. N. A. (1996). *Short-term memory for serial order* (Unpublished doctoral thesis). Cambridge University, Cambridge, UK. Retrieved from http://www.mrc-cbu.cam.ac.uk/people/rik.henson/personal/thesis/

Henson, R. N. A. (1998). Short-term memory for serial order: The Start-End Model. *Cognitive Psychology, 36,* 73–137. http://dx.doi.org/10.1006/cogp.1998.0685

Henson, R. N. A., & Burgess, N. (1997). Representations of serial order. In J. A. Bullinaria, D. W. Glasspool, & G. Houghton (Eds.), *4th Neural Computation and Psychology Workshop* (pp. 283–300). London, UK: Springer.

Hitch, G. J., Burgess, N., Towse, J. N., & Culpin, V. (1996). Temporal grouping effects in immediate recall: A working memory analysis. *The Quarterly Journal of Experimental Psychology Section A, 49,* 116–139. http://dx.doi.org/10.1080/713755609

Ho, Y.-S., & Hartley, J. (2016). Classic articles in psychology in the *Science Citation Index Expanded*: A bibliometric analysis. *British Journal of Psychology, 107,* 768–780. http://dx.doi.org/10.1111/bjop.12163

Holmes, O. W. (1871). *Mechanism in thought and morals: An address delivered before the Phi Beta Kappa Society of Harvard University, June 29, 1870, with notes and afterthoughts*. Boston, MA: Osgood. http://dx.doi.org/10.1037/12291-001

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. http://dx.doi.org/10.3758/BF03193501

JASP Team. (2018). JASP (version 0.9) [Computer software]. Retrieved from https://jasp-stats.org/

Jones, D. M., & Macken, B. (2018). In the beginning was the deed: Verbal short-term memory as object-oriented action. *Current Directions in Psychological Science, 27,* 351–356. http://dx.doi.org/10.1177/0963721418765796

Jones, Q., Ravid, G., & Rafaeli, S. (2004). Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information Systems Research, 15,* 194–210. http://dx.doi.org/10.1287/isre.1040.0023

Kareev, Y. (2000). Seven (indeed, plus or minus two) and the detection of correlations. *Psychological Review, 107,* 397–402. http://dx.doi.org/10.1037/0033-295X.107.2.397

Kemps, E., De Rammelaere, S., & Desmet, T. (2000). The development of working memory: Exploring the complementarity of two models. *Journal of Experimental Child Psychology, 77,* 89–109. http://dx.doi.org/10.1006/jecp.2000.2589

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41,* 75–86. http://dx.doi.org/10.1207/s15326985ep4102_1

Lewandowsky, S., & Farrell, S. (2002). Computational models of working memory. In L. Nadel, D. Chalmers, P. Culicover, R. Goldstone, & B. French (Eds.), *Encyclopedia of cognitive science* (pp. 578–583). London, UK: Macmillan.

Lewandowsky, S., & Oberauer, K. (2008). The word-length effect provides no evidence for decay in short-term memory. *Psychonomic Bulletin & Review, 15,* 875–888. http://dx.doi.org/10.3758/PBR.15.5.875

Lewandowsky, S., & Oberauer, K. (2015). Rehearsal in serial recall: An unworkable solution to the nonexistent problem of decay. *Psychological Review, 122,* 674–699. http://dx.doi.org/10.1037/a0039684

Logie, R. H. (2011). The functional organization and capacity limits of working memory. *Current Directions in Psychological Science, 20,* 240–245. http://dx.doi.org/10.1177/0963721411415340

Logie, R. H., Belletier, C., & Doherty, J. M. (in press). Integrating theories of working memory. In R. H. Logie, V. Camos, & N. Cowan (Eds.), *Working memory: State of the science*. Oxford, UK: Oxford University Press.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281. http://dx.doi.org/10.1038/36846

Macken, B., Taylor, J. C., & Jones, D. M. (2014). Language and short-term memory: The role of perceptual-motor affordance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1257–1270. http://dx.doi.org/10.1037/a0036845

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44,* 314–324. http://dx.doi.org/10.3758/s13428-011-0168-7

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97. http://dx.doi.org/10.1037/h0043158

Mora, G., & Camos, V. (2013). Two systems of maintenance in verbal working memory: Evidence from the word length effect. *PLoS ONE, 8,* e70026. http://dx.doi.org/10.1371/journal.pone.0070026

Morey, C. C., & Cowan, N. (2004). When visual and verbal memories compete: Evidence of cross-domain limits in working memory. *Psychonomic Bulletin & Review, 11,* 296–301. http://dx.doi.org/10.3758/BF03196573

Morey, C. C., & Cowan, N. (2005). When do visual and verbal memories conflict? The importance of working-memory load and retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 703–713. http://dx.doi.org/10.1037/0278-7393.31.4.703

Nairne, J. S. (1990). Similarity and long-term memory for order. *Journal of Memory and Language, 29,* 733–746. http://dx.doi.org/10.1016/0749-596X(90)90046-3

Naveh-Benjamin, M., & Jonides, J. (1984). Maintenance rehearsal: A two-component analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 369–385. http://dx.doi.org/10.1037/0278-7393.10.3.369

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 411–421. http://dx.doi.org/10.1037/0278-7393.28.3.411

Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review, 115,* 544–576. http://dx.doi.org/10.1037/0033-295X.115.3.544

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105,* 761–781. http://dx.doi.org/10.1037/0033-295X.105.4.761-781

Palladino, P., & Jarrold, C. (2008). Do updating tasks involve updating? Evidence from comparisons with immediate serial recall. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 61,* 392–399. http://dx.doi.org/10.1080/17470210701664989

Pascual-Leone, J. A. (1970). A mathematical model for the transition rule in Piaget's developmental stage. *Acta Psychologica, 32,* 301–345. http://dx.doi.org/10.1016/0001-6918(70)90108-3

Piaget, J. (1936). *La naissance de l'intelligence chez l'enfant* [The orign of intelligence in the child]. Neuchâtel, Switzerland: Delachaux et Niestlé.

Rapaport, D., Schafer, R., & Gill, M. (1945). *Diagnostic psychological testing*. Chicago, IL: The Year Book Publishers. http://dx.doi.org/10.1037/10834-000

Raye, C. L., Johnson, M. K., Mitchell, K. J., Greene, E. J., & Johnson, M. R. (2007). Refreshing: A minimal executive function. *Cortex, 43,* 135–145. http://dx.doi.org/10.1016/S0010-9452(08)70451-9

Richardson, J. T. E. (2007). Measures of short-term memory: A historical review. *Cortex, 43,* 635–650. http://dx.doi.org/10.1016/S0010-9452(08)70493-3

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89,* 63–77. http://dx.doi.org/10.1037/h0031185

Ryan, J. (1969). Temporal grouping, rehearsal and short-term memory. *The Quarterly Journal of Experimental Psychology, 21,* 148–155. http://dx.doi.org/10.1080/14640746908400207

Salamé, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21,* 150–164. http://dx.doi.org/10.1016/S0022-5371(82)90521-7

Salamé, P., & Baddeley, A. (1987). Noise, unattended speech and short-term memory. *Ergonomics, 30,* 1185–1194. http://dx.doi.org/10.1080/00140138708966007

Salamé, P., & Baddeley, A. D. (1989). Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology Section A, 41,* 107–122. http://dx.doi.org/10.1080/14640748908402355

Sanders, A. F. (1968). Short term memory for spatial positions. *Nederlands Tijdschrift voor de Psychologie en Haar Grensgebieden, 23,* 1–15.

Saults, J. S., & Cowan, N. (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology: General, 136,* 663–684. http://dx.doi.org/10.1037/0096-3445.136.4.663

Sbordone, R. J., & Saul, R. E. (2000). *Neuropsychology for health care professionals and attorneys*. Boca Raton, FL: CRC Press.

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review, 25,* 128–142. http://dx.doi.org/10.3758/s13423-017-1230-y

Souza, A. S., & Oberauer, K. (2018). Does articulatory rehearsal help immediate serial recall? *Cognitive Psychology, 107,* 1–21. http://dx.doi.org/10.1016/j.cogpsych.2018.09.002

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74,* 1–29. http://dx.doi.org/10.1037/h0093759

Sperling, G. (1963). A model for visual memory tasks. *Human Factors, 5,* 19–31. http://dx.doi.org/10.1177/001872086300500103

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12,* 185–233. http://dx.doi.org/10.1207/s1532690xci1203_1

Sweller, J., van Marrieboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10,* 251–296. http://dx.doi.org/10.1023/A:1022193728205

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. http://dx.doi.org/10.1037/0278-7393.26.6.1589

Tan, L., & Ward, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin & Review, 15,* 535–542. http://dx.doi.org/10.3758/PBR.15.3.535

Trost, S., & Gruber, O. (2012). Evidence for a double dissociation of articulatory rehearsal and non-articulatory maintenance of phonological information in human verbal working memory. *Neuropsychobiology, 65,* 133–140. http://dx.doi.org/10.1159/000332335

Uittenhove, K., Chaabi, L., & Camos, V., & Barrouillet, P. (2019). Is working memory storage intrinsically domain-specific? *Journal of Experimental Psychology: General, 148,* 2027–2057. http://dx.doi.org/10.1037/xge0000566

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin, 133,* 1038–1066. http://dx.doi.org/10.1037/0033-2909.133.6.1038

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2010). The contributions of primary and secondary memory to working memory capacity: An individual differences analysis of immediate free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 240–247. http://dx.doi.org/10.1037/a0017739

Van de Ven, A. H. (1986). Central problems in the management of innovation. *Management Science, 32,* 590–607. http://dx.doi.org/10.1287/mnsc.32.5.590

van Gog, T., & Paas, F. (2008). Instruction efficiency: Revisiting the original construct in educational research. *Educational Psychologist, 43,* 16–26. http://dx.doi.org/10.1080/00461520701756248

Vergauwe, E., Camos, V., & Barrouillet, P. (2014). The impact of storage on processing: How is information maintained in working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1072–1095. http://dx.doi.org/10.1037/a0035779

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Baltimore, MD: The Williams & Wilkins Company. http://dx.doi.org/10.1037/11167-000

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children–Fifth Edition technical and interpretive manual*. San Antonio, TX: NCS Pearson.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience, 7,* 701–702. http://dx.doi.org/10.1038/nn1263

Zhang, G. J., & Simon, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory & Cognition, 13,* 193–201. http://dx.doi.org/10.3758/BF03197681

(*Appendices follow*)

Appendix A

Serial Position Curve Analysis

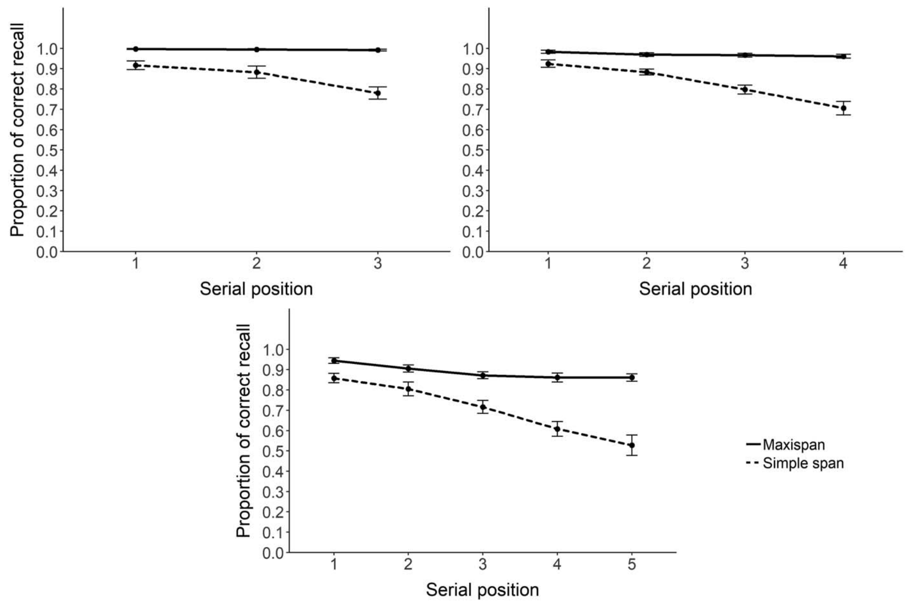

Figure A1. Means and standard errors for the proportion of blue letters correctly in Experiment 1, as a function of serial position. Top left panel: serial position curve for the three-blue letters condition. Top right panel: serial position curve for the four-blue letters condition. Bottom panel: serial position curve for the five-blue letters condition.

Experiment 1

We analyzed, separately for each condition of letters to rehearse (three to five), recall accuracy for the blue letters as a function of serial position via Bayesian mixed ANOVAs with the procedure (maxi vs. simple span) as a between-subject factor and the serial position (from one to three, one to four, and one to five, for each condition of letters to rehearse, respectively) as a within-subject factor. Consistently across the three analyses, the full model containing the two main effects and their interaction was preferred with decisive evidence over the second best model (BF_{10} = 6.46E + 4, BF_{10} = 1.59E + 9, and BF_{10} = 4.89E + 6, for the three-blue, four-blue, and five-blue letters to rehearse conditions, respectively). In addition to the already reported result that recall accuracy is higher in the maxispan than the simple span procedure group (see the main text), we also have evidence that blue letters recall accuracy decreased as the serial position increased and that serial position interacted with the procedure (Figure A1). We thus analyzed the data separately for each group, and each condition of blue letters, through the mean of Bayesian repeated-measures ANOVA with the serial position as a within-subject factor (from one to three, one to four, and one to five, for each condition of letters to rehearse, respectively). For the simple span procedure group, the results provided decisive evidence for an effect of serial position on blue letters recall accuracy (BF_{10} = 2.25E + 4, 8.47E + 9, and 1.13E + 11, for the three-blue, four-blue, and five-blue letters to rehearse conditions, respectively). As one can see in Figure A1, this effect characterized by a clear decrease in recall performance as serial position increased, supporting the presence of a primacy effect, but the absence of an effect of recency. For the maxispan procedure group, we observed anecdotal evidence in favor of the null model supporting the absence of an effect of serial position on blue letters recall accuracy for the 3-blue (BF_{01} = 2.45) and the four-blue letters to rehearse condition (BF_{01} = 1.40). In contrast, in the five-blue letters to rehearse condition, we obtained decisive evidence for the presence of an effect of a serial position (BF_{10} = 5.31E + 4). The difference between serial positions 1 (M = .94, SE = .01) and 2 (M = .91, SE = .02), and Serial Positions 2 and 3 (M = .87, SE = .02) was supported by evidence in post hoc analyses (BF_{10} = 38.30 and BF_{10} = 5.05, respectively). However, the comparison between Serial Positions 3 and 4 (.86, SE = .02), and Serial Positions 4 and 5 (M = .86, SE = .02), provided support of an absence of difference (BF_{01} = 2.38 and BF_{01} = 4.30, respectively).

(*Appendices continue*)

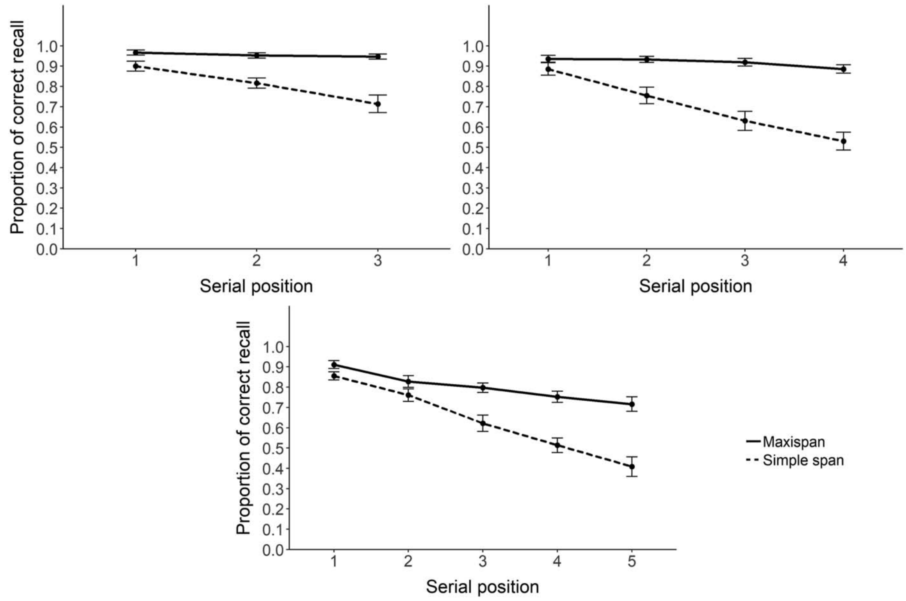

Figure A2. Means and standard errors for the proportion of blue letters correctly in Experiment 2, as a function of serial position. Top left panel: serial position curve for the three-blue letters condition. Top right panel: serial position curve for the four-blue letters condition. Bottom panel: serial position curve for the five-blue letters condition.

Experiment 2

Separately for each condition of letters to rehearse (three to five), we analyzed the blue letters recall accuracy as a function of serial position through the mean of Bayesian mixed ANOVAs with the procedure (maxi vs. simple span) as a between-subject factor and the serial position (from one to three, one to four, and one to five, for each condition of letters to rehearse, respectively) as a within-subject factor. For the three analyses, the results provided decisive evidence in favor of the full model supporting the presence of the two main effects and their interaction, preferred with decisive evidence over the second best model (BF_{10} = 642.44, BF_{10} = 6.47E + 10, and BF_{10} = 2.92E + 7, for the three-blue, four-blue, and five-blue letters to rehearse conditions, respectively). In addition to the already reported effect of procedure, this result provides evidence for an effect of serial position characterized by a decrease of the blue letters recall accuracy as serial position increased, the serial position interacting with the procedure (Figure A2). Given the evidence in favor of the interaction, we analyzed the data separately for each group, and for each condition of blue letters, via Bayesian repeated-measures ANOVAs with the serial position as a within-subject factor (from one to three, one to four, and one to five, for each condition of letters to rehearse, respectively). For the simple span procedure group, the results provided for each analysis decisive evidence in favor of the alternative model supporting the presence of an effect of serial position on blue letters recall accuracy (BF_{10} = 8373.73, BF_{10} = 3.32E + 12, and BF_{10} = 1.64E + 20, for the three-blue, four-blue, and five-blue letters to rehearse conditions, respectively). As shown in Figure A2, it appears clearly that recall accuracy decreased as serial position increased, marking the presence of an effect of primacy, but the absence of a recency effect. For the maxispan procedure group, the results provided moderate evidence in favor of the null model supporting the absence of an effect of serial position for the 3-blue letters to rehearse condition (BF_{01} = 3.87). However, we obtained moderate and decisive evidence in favor of the presence of an effect of serial position on blue letters recall accuracy for the four-blue (BF_{10} = 5.56) and five-blue (BF_{10} = 5.87E+8) letters to rehearse condition, respectively. Post hoc analysis in the 4-blue letters condition revealed an absence of difference between Serial Positions 1 (M = .94, SE = .02) and 2 (M = .93, SE = .01), and Serial Positions 2 and 3 (M = .92, SE = .02), supported by moderate (BF_{01} = 4.24) and anecdotal (BF_{01} = 2.77) evidence, respectively. We also observed ambiguous evidence for an absence of difference (BF_{01} = 1.01) when comparing Serial Positions 3 and 4 (M = .89, SE = .02). However, we obtained moderate evidence for a difference between Serial Positions 1 and 4 (BF_{10} = 6.84) and Serial Positions 2 and 4 (BF_{01} = 4.74). For the five-blue letters condition, it is clear in Figure A2 that recall accuracy decreased progressively as serial position increased.

(*Appendices continue*)

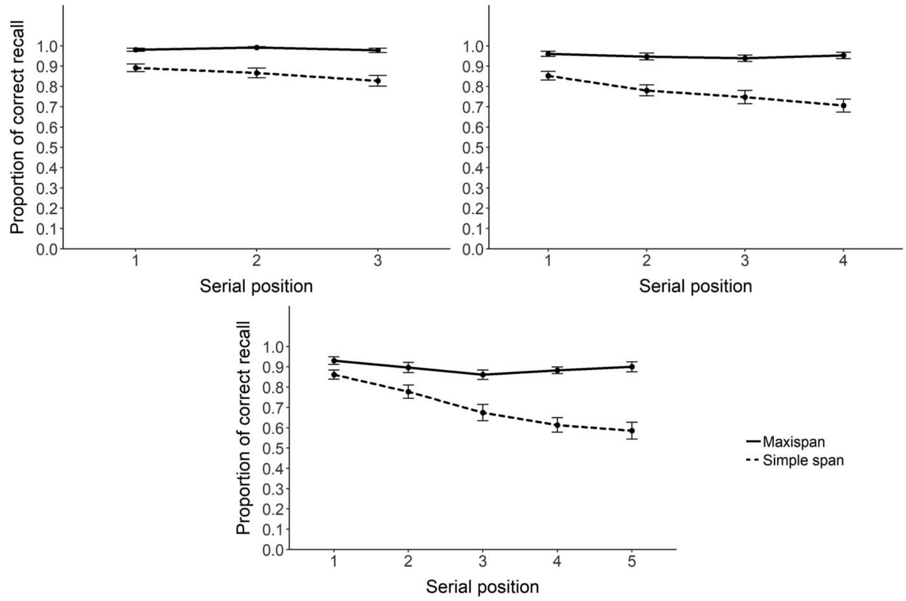

Figure A3. Means and standard errors for the proportion of blue letters correctly in Experiment 3, as a function of serial position. Top left panel: serial position curve for the three-blue letters condition. Top right panel: serial position curve for the four-blue letters condition. Bottom panel: serial position curve for the five-blue letters condition.

Experiment 3

We analyzed recall accuracy for the blue letters as a function of serial position in the same way as in the previous experiments. The results provided evidence in favor of the full model supporting the presence of the two main effects with their interaction, preferred over the second best model by a factor of 1606.08 and 2.23E + 8 in the four-blue and five-blue letters to rehearse condition. In the three-blue letters condition, the results were ambiguous, providing similar level of evidence to the full model, the model with the two main effects, and the model with only the effect of group. An analysis of specific effect provided some evidence, nonetheless anecdotal for the presence of an interaction term ($BF_{Inclusion}$ = 2.20) and an effect of serial position ($BF_{Inclusion}$ = 1.68), but decisive evidence for the effect of group ($BF_{Inclusion}$ = 5.71E + 4). These results added to the already reported effect of procedure on blue letters recall accuracy, evidence for an effect of serial position with performance decreasing as serial position increase, that interacted with the procedure; the pattern of effect is overall consistent across the three blue letters conditions (Figure A3). Given the overall evidence for the presence of an interaction, we analyzed the data separately for each group, and for each condition of blue letters, via Bayesian repeated-measures ANOVAs with the serial position as a within-subject factor (from one to three, one to four, and one to five, for each condition of letters to rehearse, respectively). For the simple span procedure group, the analyses favored to a decisive level the alternative model supporting the presence of an effect of serial position in the four-blue and five-blue letters to rehearse conditions that is characterized by a clear primacy effect, but no effect of recency (BF_{10} = 3.41E + 4 and BF_{10} = 4.35E + 11, respectively). For the three-blue letters to rehearse condition, the evidence in favor of an effect of serial position was only anecdotal (BF_{10} = 1.62). For the maxispan procedure group, the results provided moderate evidence in favor of the null model supporting the absence of an effect of serial position on blue letters recall accuracy in the three-blue and four-blue letters to rehearse conditions (BF_{01} = 3.17 and BF_{01} = 4.53, respectively). In contrast, in the five-blue letters to rehearse condition, we obtained moderate evidence in favor of the alternative model supporting the presence of an effect of serial position on blue letters recall accuracy (BF_{10} = 6.28). Post hoc analyses yielded strong (BF_{10} = 21.00) and moderate (BF_{10} = 4.75) evidence for a difference in terms of recall accuracy between Serial Positions 1 (M = .93, SE = .02) and 3 (M = .86, SE = .02), and Serial Positions 1 and 4 (M = .88, SE = .02), respectively. There was also evidence for a difference, but only to an anecdotal level, BF_{10} = 1.28 and BF_{10} = 2.04, between Serial Positions 2 (M = .90, SE = .02) and 3, and Serial Positions 3 and 5 (M = .90, SE = .02), respectively.

(*Appendices continue*)

Appendix B

Transposition Error Analyses

For theoretical reasons exposed in the General Discussion, we focused on transposition errors distribution patterns. We first computed for each participant and across all the trials the number of transposition errors (i.e., recalling a presented item but at the wrong serial position). However, if a letter presented in the list was recalled at the correct position, and recalled a second time but at the wrong serial position (e.g., recalling *D—N—K—G—V—K* instead of *D—N—K—G—V—T*), the erroneous recall was considered as a repetition error and not as a transposition. After that, we computed among all the transpositions errors the proportion of *between-group transpositions* (i.e., recalling a blue letter to a position where a black letter was presented, or vice versa). We next determined among all the transpositions errors the proportion of relative *interposition* errors for the first and last group positions (i.e., recalling the first or the last blue letter at the position of the first or the last black letter, respectively). We chose to focus on relative interpositions for the first and last letters of each group for two reasons. First, it is known that when groups of items are of unequal sizes, items tend to be transposed following a relative, rather than a positional coding scheme (Henson, 1998; Henson & Burgess, 1997). Second, according to the start-end model (Henson, 1998), whatever the difference of size between two group of items, the first and last positions keep the same relative code in the different group size.

Experiment 1

We compared the maxispan and the simple span procedure conditions regarding their proportions of between-group transposition and interpositions through the mean of undirected Bayesian independent samples *t*-tests. For between-group transposition errors, the analysis yielded decisive evidence for a difference between the two conditions (BF_{10} = 255.48, Cohen's $d = -1.39$), with lower proportions of between-group transpositions with the maxispan (M = .25, SE = .02) than the simple span procedure (M = .35, SE = .02). For interposition errors, the results supported to a decisive level the presence of a difference between conditions (BF_{10} = 2384.23, Cohen's $d = -1.65$) characterized by a lower proportion of interposition errors in the maxispan (M = .02, SE = .01) than in the simple span condition (M = .06, SE = .01).

Experiment 2

As far as Experiment 2 was concerned, for between-group transposition errors, the analysis provided anecdotal evidence for an absence of difference (BF_{01} = 2.01, Cohen's $d = -0.35$) between the two conditions (M = .33, SE = .02 and M = .36, SE = .04, respectively). For interposition errors, the results yielded very strong evidence for a difference between the two conditions (BF_{10} = 95.07, Cohen's $d = -1.27$) with fewer proportion of interposition errors in the maxispan (M = .03, SE = .004) than in the simple span condition (M = .06, SE = .004).

Experiment 3

For between-group transposition errors, the analysis yielded strong evidence for a difference between conditions (BF_{10} = 10.25, Cohen's $d = -0.97$), the maxispan (M = .20, SE = .02) procedure eliciting proportionally less between-group transpositions than the simple span procedure did (M = .29, SE = .02). For interposition errors, the results provided very strong evidence for a difference between the two conditions (BF_{10} = 37.74, Cohen's $d = -1.15$), with lower proportion of interposition errors in the maxispan (M = .03, SE = .005) than in the simple span procedure (M = .06, SE = .01).

Received January 20, 2020
Revision received May 27, 2020
Accepted June 5, 2020 ■