Mem Cogn (2016) 44:937–949
DOI 10.3758/s13421-016-0609-8

Phonological similarity in working memory span tasks

Michael Chow[1] · Brooke N. Macnamara[1] · Andrew R. A. Conway[1]

Published online: 5 April 2016

Abstract In a series of four experiments, we explored what conditions are sufficient to produce a phonological similarity facilitation effect in working memory span tasks. By using the same set of memoranda, but differing the secondary-task requirements across experiments, we showed that a phonological similarity facilitation effect is dependent upon the semantic relationship between the memoranda and the secondary-task stimuli, and is robust to changes in the representation, ordering, and pool size of the secondary-task stimuli. These findings are consistent with interference accounts of memory (Brown, Neath, & Chater, *Psychological Review, 114*, 539–576, 2007; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, *Psychonomic Bulletin & Review, 19*, 779–819, 2012), whereby rhyming stimuli provide a form of categorical similarity that allows distractors to be excluded from retrieval at recall.

Keywords Working memory · Short-term memory · Interference/inhibition · Memory retrieval

Traditionally, the *phonological similarity effect* refers to the finding that immediate serial recall is impaired when lists of items are phonologically similar rather than distinct. For example, lists of similar-sounding letters, such as *c*, *b*, *d*, and *v*, are recalled less accurately than lists of dissimilar letters, such as *c*, *r*, *m*, and *k* (Conrad, 1964). The phonological similarity effect is also present when words are used as memoranda, and similarity is operationalized as phoneme overlap—for example, *cat*, *fad*, *pan*, *map*, as compared to *bar*, *kid*, *sun*, *toe* (Baddeley, 1966)—and it remains for similar items even when they are interleaved with dissimilar items (Baddeley, 1968, Exp. 5; Henson, 1996). The phonological similarity effect is such a standard in cognitive psychology that it served as one of the primary motivations for the working memory framework (Baddeley & Hitch, 1974) and is considered a benchmark working memory finding (Guérard, Saint-Aubin, Burns, & Chamberland, 2011).

Phoneme overlap versus rhyme

One prominent explanation of the phonological similarity effect is that when verbal material is stored and maintained in a short-term buffer—the phonological loop—the phonemic similarities of material being rehearsed in that buffer interfere with one another (Hanley & Bakopoulou, 2003). Consistent with this, the phonological similarity effect is not observed when phonological encoding and rehearsal is prevented via articulatory suppression (Larsen & Baddeley, 2003).

In contrast to the negative effect of phoneme overlap on recall, operationalizing phonological similarity using rhyming items—for example, *bat*, *cat*,[1] *hat*, *mat*—can reverse the effect in memory span tasks, such that participants exhibit better recall when presented with phonologically similar as compared to dissimilar lists (Fallon, Groves, & Tehan, 1999;

✉ Michael Chow
machow@princeton.edu

[1] Department of Psychology, Princeton University, Princeton, NJ 08544, USA

[1] The code repository for all analyses and reports is stored at https://osf.io/br854/.

Gupta, Lipinski, & Aktunc, 2005). One explanation for this reversal is that rhyming words provide a category cue that facilitates list retrieval (Gupta et al., 2005; Nairne & Kelley, 1999), thus enhancing recall in the same manner as semantic similarity between memory items on memory span tasks (Huttenlocher & Newcombe, 1976; Saint-Aubin & Poirier, 1999).

According to this view, phonological similarity among list items causes two opposing effects. On the one hand, it has a detrimental effect on recovering an item's serial position within a list. On the other hand, the category cue may have a beneficial effect on overall list recall. An important caveat to this benefit is that the cue must be salient to identifying a list in memory. Accordingly, a rhyming benefit has sometimes been observed in studies that do not repeat memory items across lists, but no rhyming benefit is observed when studies do repeat the memory items across lists (see Gupta et al., 2005, for a review).

Simple span versus complex span tasks

The majority of investigations on phonological similarity effects have required participants to simply store and maintain items in memory—a type of task often referred to as *simple span*. More recent work has extended the investigation of the phonological similarity effect from simple to complex span tasks (Camos, Mora, & Barrouillet, 2013; Lobley, Baddeley, & Gathercole, 2005; Macnamara, Moore, & Conway, 2011; Tehan, Hendry, & Kocinski, 2001). Complex span tasks require participants to store and maintain items in memory while engaging in a concurrent processing activity. The addition of concurrent processing in complex span tasks requires that participants divert attention away from memoranda (Barrouillet, Bernardin, & Camos, 2004; Saito & Miyake, 2004), as well as recall memoranda in the face of interference due to the encoding of irrelevant items within the processing task (Oberauer, 2009). Complex span tasks have been used extensively to study working memory (e.g., Conway et al., 2005).

To be clear, in both simple and complex span tasks, participants view a number of to-be-remembered stimuli before attempting to recall them in the correct serial order. Typically, the stimuli are presented one at a time at a fixed rate (e.g., 1,000 ms per stimulus). The difference between simple span and complex span tasks is that complex span tasks include an additional processing component between the to-be-remembered stimuli. For example, in one common version known as the *reading span task* (RSPAN; Daneman & Carpenter, 1980), the processing component consists of reading a sentence aloud (e.g., "The boy jumped over the fence."). In another version, the *operation span task* (OSPAN; Turner & Engle, 1989), the processing component consists of reading aloud a math problem (e.g., "Is 5 * 3 – 8 = 7?"). Also, in RSPAN and OSPAN, respectively, participants may be required to judge whether each sentence makes sense, or whether each mathematical equation is true.

Phoneme overlap versus rhyme in complex span tasks

The dependence of phonological similarity effects on the way that phonological similarity is operationalized extends to complex span tasks, as well, with explanations following lines similar to those for the effects in simple span (feature overlap: Camos, Mora, & Oberauer, 2011; Lobley et al., 2005; rhyming: Copeland & Radvansky, 2001; Macnamara et al., 2011; see also Unsworth & Engle, 2007, p. 1045). Understandably, researchers who are interested in studying the contribution of the phonological loop using complex span tasks have favored the use of phonological overlap over rhyming, claiming that the use of rhyming memoranda "may reflect processes other than coding in the phonological store" (Lobley et al., 2005, p. 1465).

Although it seems clear that using phonological overlap is preferable when investigating the phonological loop in complex span (e.g., Camos et al., 2013), the use of rhyming memoranda may be useful in clarifying the role of category cueing. Researchers who have used rhyming words to operationalize phonological similarity in complex span tasks have suggested that the rhyming cue facilitates retrieval. For example, Copeland and Radvansky (2001) observed a phonological similarity benefit in an RSPAN task, in which participants memorized the final word of a series of sentences. They claimed that the unique context created by the sentences and a list-rhyming cue produced the phonological similarity benefit. However, Macnamara et al. (2011) observed a phonological similarity benefit even when memory items were presented after each sentence (Exp. 1), as well as when the memory items were contextually unrelated to the preceding sentence (Exp. 2). On the basis of this evidence, Macnamara et al. suggested that performance on complex span tasks is facilitated by a categorical listwise rhyming cue.

The role of the secondary processing component in complex span tasks

Although consistent and robust phonological similarity facilitation effects have been observed for the RSPAN, outcomes have been mixed for the OSPAN, especially when phonological similarity is operationalized by using rhyming words as memoranda. For instance, Tehan et al. (2001, Exp. 2B) found no evidence of improvement when participants simply had to read problems aloud without solving them. However, Tehan et al. (Exp. 2A) found mixed evidence of improved serial-recall accuracy when participants had to verify whether the math problems during the processing component were true (*p*s = .028 and .137). To complicate matters further, in a similar administration in which participants had to verify the math problems, Copeland and Radvansky (2001) observed weak evidence for a phonological similarity *decrement* (*p* = .07).

The relative lack of consistency of a phonological similarity effect in OSPAN as compared to RSPAN raises several possibilities. As Macnamara et al. (2011) mentioned, the processing components in RSPAN and OSPAN tasks differ along several dimensions. First, the OSPAN component imposes the additional requirement that participants verify the solution to the arithmetic sentence rather than simply reading it. Second, Copeland and Radvansky (2001), as well as Macnamara et al., considered the role that sentence reading might play in the effect. However, although participants have been required to read grammatical sentences or operations in previous studies, it is not yet clear whether adhering to a sentence grammar is a necessary factor for producing the phonological similarity benefit.

Alternatively, instead of the processing features (solving problems or reading sentences) playing a role, the characteristics of the distractor items themselves may be sufficient to explain the differential effects of phonological similarity in RSPAN and OSPAN. For example, Macnamara et al. (2011) argued that the phonological similarity benefit on complex span tasks is due to a listwise rhyming cue. However, it seems plausible that such a cue would be beneficial to the extent that it reduces interference from distractor items.

One reason that interference from distractors could be greater in RSPAN is that the contents of the processing component and the memoranda are from a broadly similar class (words). In contrast, in OSPAN the content of the processing component differs categorically from the memoranda (numbers and operators vs. words). Evidence in support of this idea has come from Turner and Engle (1989), who administered four complex span tasks by fully crossing whether words or numbers were used for the memoranda and whether words or numbers were used as the processing content. Recall accuracy was lower when the type of memoranda matched the processing content (words with words or numbers with numbers),[2] suggesting that interference is increased by higher categorical similarity between the memoranda and distractors (see also Conlin, Gathercole, & Adams, 2005).

Furthermore, Oberauer and colleagues (2012) claimed that if the distractors differ categorically from the memory items, they may be excluded as candidates at recall, thereby reducing the interference brought on by those distractors. This notion is also demonstrated in other interference-based models, such as SIMPLE (Neath & Brown, 2006), in which interference at recall is a function of the distance between the items in memory along several feature dimensions.

To simplify, we can consider previous studies of the phonological similarity effect in complex span in terms of a 2 × 2 design (Distractors: words [RSPAN] or numbers/operations [OSPAN] × Phonology: similar or dissimilar). Using this framework, the only condition with high competition at retrieval is the RSPAN with phonologically dissimilar words. That is, the distractor items, which consist of phonologically dissimilar words, are not categorically distinct from the memoranda, which also consist of phonologically dissimilar words, and therefore the items compete at retrieval. Thus, assuming all other things are equal, there should be a reduction in serial accuracy for the phonologically dissimilar condition of RSPAN relative to the other three conditions (i.e., the phonologically similar condition of RSPAN as well as both phonological similarity conditions of OSPAN). Though not conclusive, this reasoning appears consistent with the mean serial-recall accuracies that Copeland and Radvansky (2001) found for OSPAN (similar, 36.67 %; dissimilar, 41.67 %) and RSPAN (similar, 40.00 %; dissimilar, 28.33 %).

Goals of the present experiments

Before we can attribute differences in phonological similarity effects for OSPAN and RSPAN to differences between distractor items (Exp. 4), we must first rule out several confounds pertaining to differences in verification requirements, grammars, and the sizes of the pools of distractor stimuli (Exps. 1–3). Below, we summarize the motivation behind each experiment.

Experiment 1 OSPAN and RSPAN tasks can differ in their verification requirements. Previous investigations of phonological similarity employing the RSPAN task have not required participants to make any judgment about the distractor material (e.g., whether it was a grammatical sentence), whereas several phonological similarity investigations employing OSPAN have required participants to indicate whether each math equation was true or false. These differences could lead to differing cognitive loads, which could moderate the phonological similarity effect. The goal of Experiment 1 was to examine the role of verification requirements in the OSPAN. The design of Experiment 1 allowed us to answer the question of whether making a judgment about the distractor materials in OSPAN is necessary to produce an effect of phonological similarity.

[2] One important caveat in interpreting Turner and Engle (1989) is that participants performed free recall when numbers were used for processing, but serial recall when words were used for processing. Importantly, however, their finding was an interaction between the memoranda and processing task.

Experiment 2 The distractor items in OSPAN and RSPAN adhere to different grammars. The distractor items in OSPAN use arithmetic equations, whereas those in RSPAN use English sentences. These differences could lead to differing predictabilities of the distractor cadence, which could lead to differences in memoranda refreshing. The goal of Experiment 2 was to directly compare the phonological similarity effects in RSPAN and OSPAN, while removing the grammatical structures specific to each task type. The design of Experiment 2 allowed us to answer the question of whether the grammatical structures of the processing components in OSPAN and in RSPAN are necessary for the phonological similarity effect.

Experiment 3 OSPAN and RSPAN distractor stimuli may be drawn from pools that contain different numbers of unique stimuli. OSPAN relies on a limited number of digits and operations, whereas RSPAN relies on a near infinite pool of English words. Using a small number of unique distractor stimuli across an experiment, as in OSPAN, might allow participants to exclude the distractor items at recall. The goal of Experiment 3 was to examine phonological similarity with different pool sizes of distractors. The design of Experiment 3 thus allowed us to answer the question of whether the phonological similarity effect is moderated by the size of the pool from which unique distractors are drawn.

Experiment 4 OSPAN distractor stimuli (numbers and operators) are categorically distinct from the memory items used (words). A categorical cue from rhyming memoranda may be beneficial only when the distractor and memoranda items do not have another distinguishing cue. The goal of Experiment 4 was to examine phonological similarity using a new category of distractor items that would be distinct from the memoranda. The design of Experiment 4 allowed us the answer the question of whether the phonological similarity effect depends upon a distinguishing categorical cue at retrieval.

General method

All experiments followed the same general methodology. Each experimental task manipulated phonological similarity within subjects. All trials within each experimental task consisted of a processing component, such as reading a sentence or arithmetic equation, interleaved with a memory component. The memoranda were consistent across all experiments, but the processing component was manipulated within each experiment and varied across experiments.

Participants

Participants were recruited from Princeton University and the surrounding community. Students recruited from the Psychology Department participated in exchange for partial course credit, and students and community members recruited from the university's paid participant pool were compensated $12/h for their participation.

Materials and procedure

A total of 108 single-syllable nouns were used as memoranda. The words were normed for frequency (Kučera & Francis, 1967), meaningfulness (Toglia & Battig, 1978), familiarity, concreteness, and imageability (pooled; Gilhooly & Logie, 1980; Toglia & Battig, 1978), using the MRC Psycholinguistic Database (Coltheart, 1981). Fifty-four of the words were arranged in 12 lists, ranging in length from three to six memoranda, such that the words within each list were phonologically similar to one another (e.g., *shawl*, *hall*, *doll*). The other 54 words were also arranged in 12 lists, ranging in length from three to six memoranda, such that the words within each list were phonologically dissimilar from one another (e.g., *deck*, *frown*, *sea*). In each case, the 12 lists were composed of three sets of memoranda for each of the four list lengths.

On each trial, participants alternated between the processing and memory components presented on a computer screen, until being prompted to recall as many memoranda as possible in serial order. During the processing component, stimuli were presented on the computer screen to participants. Once the participant had finished reading the presented stimuli aloud, an experimenter advanced the screen to the memory component (experimenter-paced; see Conway et al., 2005). For the memory component, a to-be-remembered word was presented for 1,000 ms and read aloud by the participant. During recall, the participant was prompted with a box in which to type, and asked to enter each word in serial order on a new line. Participants began at the first line, which corresponded to the first serial position, and progressed downward. If they were unable to remember a word, they were instructed to leave the line corresponding to that word blank. Before the beginning of each trial, a ready screen was displayed. All tasks were created in-house using E-Prime.

Scoring

Serial recall For each trial, an item was scored as correct if it was recalled in the correct serial position. The number of correct items was summed across all trials to produce an overall score within each condition for all participants. Thus, a trial with list length six was worth twice as much as a trial with list length three (partial-credit load scoring; Conway et al., 2005).

Each total score was then divided by the maximum score possible to obtain the proportion correct.

Item recall For each trial, an item was scored as correct if it was recalled accurately, regardless of serial position. As in serial recall, the proportion of total items correctly recalled was used.

Order accuracy For each participant, *order accuracy* was defined as the ratio of serial-recall accuracy to item - recall accuracy. This is the proportion of the total items recalled that were also recalled in the correct serial position.

Experiment 1

In Experiment 1, we sought to examine whether the format of the processing task might impact the phonological similarity effect. Half of the participants ($n = 18$) completed a "Standard OSPAN" task (Turner & Engle, 1989). In the Standard OSPAN, participants were tasked with reading aloud basic arithmetic problems, along with a possible solution, and indicating verbally whether that solution was true or false—for example, "Is (2*8) – 9 = 5?" "no." The remaining participants ($n = 18$) completed a "Reading OSPAN" task. In the Reading OSPAN, participants were presented with the same arithmetic problems as the other participants, except that the problems were written out in word form—for example, "Is two times eight minus nine equal to five?" Moreover, the participants in this condition simply needed to read the arithmetic problems aloud, rather than evaluate their correctness. This condition was included in order to make the OSPAN task more similar to previous versions of the RSPAN used to test phonological similarity (e.g., Macnamara et al., 2011), in that both would now require word stimuli to be read aloud, but no judgments about those stimuli to be made. Phonological similarity was manipulated within subjects.

If phonological similarity facilitation disappears when verification of math problems is required, then we should observe phonological similarity facilitation in the Reading OSPAN condition described, but not in the Standard OSPAN condition. However, if phonological similarity facilitation is driven largely by categorical exclusion of the distractor material at retrieval when the distractor material would otherwise cause interference—that is, by reduced competition between the memoranda and distractor material—then neither requiring participants to solve each arithmetic problem nor writing the problem in word form should greatly alter the effect of phonological similarity. This is because the memoranda (words) were categorically distinct from the arithmetic problems (composed of digits and operators), regardless of form.

Results

Serial recall A 2 (Processing Component: standard or reading) × 2 (Phonology: similar or dissimilar) mixed factorial analysis of variance (ANOVA) revealed a significant main effect of processing component, $F(1, 34) = 11.19$, $p = .002$, $\eta_p^2 = .24$, indicating that participants recalled significantly more memoranda when the form of the processing component was spelled out in sentence form without requiring a solution, than when numbers and mathematical symbols were presented and required a solution. There was weak evidence for a main effect of phonological similarity, $F(1, 34) = 3.15$, $p = .084$, $\eta_p^2 = .084$. However, the direction of the effect indicated a phonological similarity *decrement*, rather than facilitation. We observed no evidence for an interaction between task and phonological similarity, $F(1, 34) = 0.16$, $p = .69$, $\eta_p^2 = .004$. See Fig. 1.

Item recall As with serial recall, a 2 × 2 mixed factorial ANOVA revealed a main effect of processing component, $F(1, 34) = 9.30$, $p = .004$, $\eta_p^2 = .21$, such that participants performed better when the processing component was in sentence form and did not require a verbal solution. Unlike with serial recall, however, we found no evidence for a main effect of phonological similarity, $F(1, 34) = 0.001$, $p = .975$, $\eta_p^2 < .001$. There was also no evidence for an interaction between task and phonological similarity, $F(1, 34) = 0.95$, $p = .33$, $\eta_p^2 = .027$.

Order accuracy A 2 × 2 mixed factorial ANOVA revealed a main effect of processing component, $F(1, 34) = 9.97$, $p = .003$, $\eta_p^2 = .22$, as well as evidence for a main effect of phonological similarity, $F(1, 34) = 6.46$, $p = .015$, $\eta_p^2 = .15$, indicating that participants performed better when the memoranda did not rhyme. We found no evidence for an interaction, $F(1, 34) = 0.022$, $p = .88$, $\eta_p^2 < .001$.

Discussion

The results of this experiment replicated Copeland and Radvansky's (2001) findings, in that no serial-recall phonological similarity facilitation emerged in OSPAN, only a weak detriment. Moreover, this null finding persisted after the processing component was expressed in word form and required only that each component be read aloud.

The processing component of the Reading OSPAN differs from RSPAN largely in content only; both tasks may be viewed as RSPAN tasks using different distractor contents. However, the distractor content in OSPAN, regardless of form, follows a systematic ordering that could be moderating the effect. Moreover, the two tasks have not been examined side by side. We conducted Experiment 2 to investigate whether the task-specific distractor material structure (i.e., an

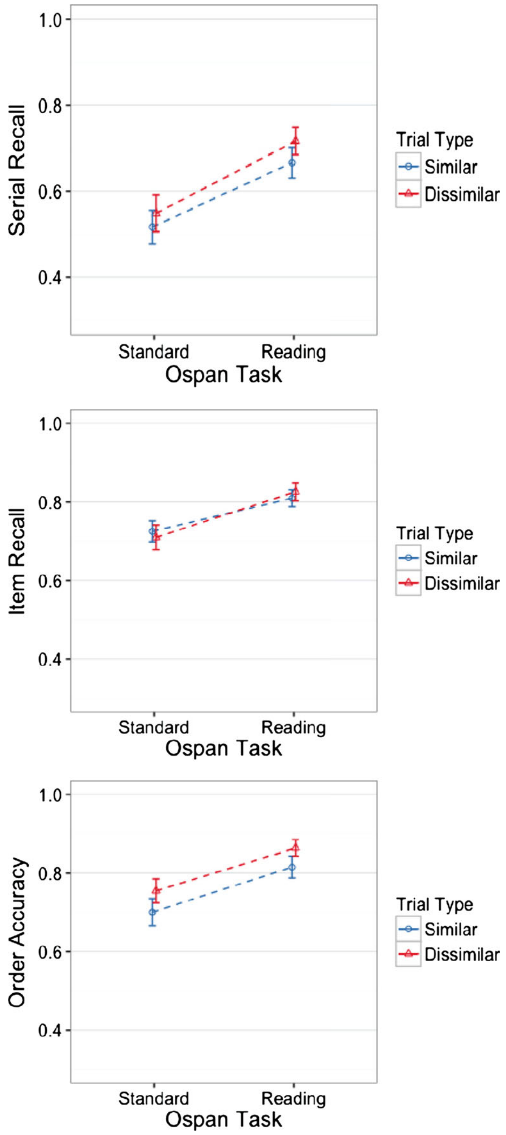

Fig. 1 Serial, item, and order recall accuracies for Experiment 1. "Standard" refers to the typical operation span (OSPAN) task, which uses mathematical problems as the distractor material and requires verification. "Reading" refers to a version of the OSPAN task in which the distractor material was spelled out and verification was not required

operation statement "grammar") was the source of the phonological similarity effect.

Experiment 2

In this experiment, we sought to examine whether the structure of the processing task might impact the phonological similarity effect. Half of the participants ($n = 20$) completed a "Scrambled OSPAN" task, in which the participants read the same spelled-out equations used in Experiment 1 (Reading OSPAN). However, the order of the words in each equation was randomized. Thus, the processing component contained the same words as the prior condition, but the words were no longer presented in an arithmetic structure. The remaining participants ($n = 17$) completed a "Scrambled RSPAN" task, in which the participants read the same RSPAN sentences used in Experiment 1 of Macnamara et al. (2011), but the order of the words in each sentence was randomized. Thus, the processing component contained the same words as a prior condition, but the words were no longer presented in a sentence structure. For all conditions, the order of the shuffled words was fixed across participants, and phonological similarity was manipulated within subjects. One participant was dropped from the Scrambled RSPAN condition due to computer errors.

Results

Serial recall A 2 × 2 mixed factorial ANOVA revealed a significant main effect of processing component, $F(1, 34) = 7.37$, $p = .010$, $\eta_p^2 = .17$. The effect of similarity was qualified by a significant interaction between processing component and similarity, $F(1, 34) = 3.20$, $p = .014$, $\eta_p^2 = .086$. Simple-effects analyses examining the effect of phonological similarity within each task provided strong evidence for phonological similarity facilitation when the processing component consisted of a scrambled sentence, $F(1, 15) = 9.05$, $p = .008$, $\eta_p^2 = .37$, but no evidence for an effect when the processing component consisted of a scrambled arithmetic equation, $F(1, 19) = 0.15$, $p = .69$, $\eta_p^2 = .008$. See Fig. 2.

Item recall A 2 × 2 mixed factorial ANOVA revealed a significant main effect of processing component, $F(1, 34) = 8.08$, $p = .007$, $\eta_p^2 = .19$. As in serial recall, the relationship between processing component and phonological similarity was qualified by a significant interaction, $F(1, 34) = 13.35$, $p < .001$, $\eta_p^2 = .28$. As in the simple-effects analyses for serial recall, we found strong evidence for phonological similarity facilitation when the processing component consisted of a scrambled sentence, $F(1, 15) = 37.7$, $p < .001$, $\eta_p^2 = .71$, but no evidence of any phonological similarity effect when the processing component consisted of a scrambled arithmetic equation, $F(1, 19) = 1.19$, $p = .28$, $\eta_p^2 = .059$.

Order accuracy A 2 × 2 mixed factorial ANOVA revealed a significant main effect of processing component, $F(1, 34) = 4.21$, $p = .047$, $\eta_p^2 = .11$. Unlike in serial and item recall, here we found no evidence for an interaction, $F(1, 34) = 0.37$, $p = .54$, $\eta_p^2 = .01$. In addition, there was only weak evidence for a main effect of phonological similarity, $F(1, 34) = 2.91$, $p = .097$, $\eta_p^2 = .109$.

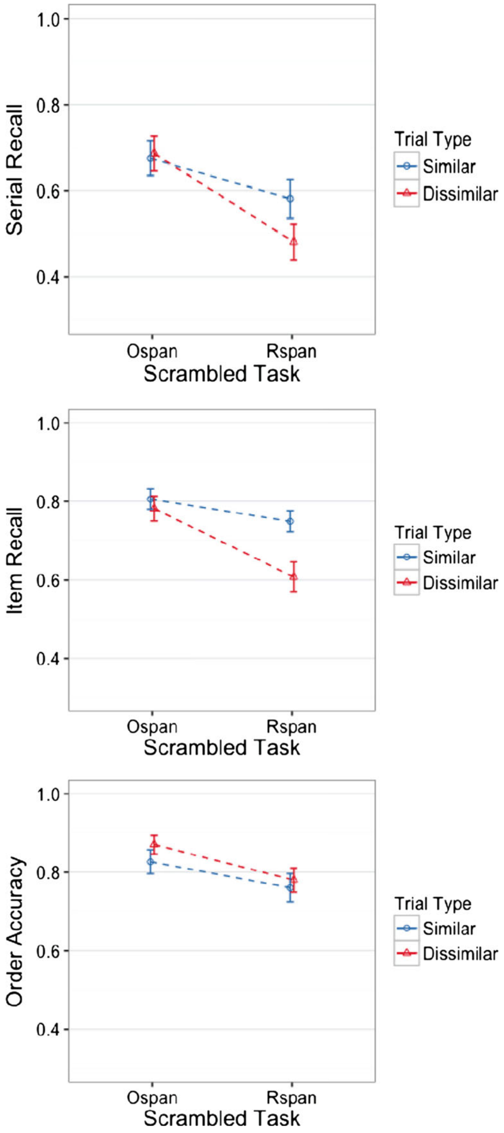

Fig. 2 Serial, item, and order recall accuracies for Experiment 2. All distractor stimuli were presented using random ordering (i.e., were scrambled) within the processing component

Discussion

Recall was enhanced in the phonologically similar condition relative to the dissimilar condition in the Scrambled RSPAN task, but no phonological similarity effect was observed in the Scrambled OSPAN task. This is consistent with the hypothesis that increased interference in RSPAN, due to the lack of categorical distinctness, drives the phonological similarity benefit. In other words, these results suggest that phonological similarity facilitation is driven largely by the categorical exclusion of the distractor material at retrieval when the material would otherwise cause interference.

Although phonological similarity facilitation was observed in the unstructured RSPAN but not in the unstructured OSPAN, one other potential confound is that the distractor items are drawn from a small pool for the OSPAN (nine digits and four operators) and from a large pool for the RSPAN (English words). We conducted Experiment 3 to investigate whether the pool size of the distractor materials was the source of the phonological similarity effect.

Experiment 3

In this experiment, we sought to examine whether the number of unique distractor items in the processing component (i.e., the size of the pool of potential distractors) might impact the phonological similarity effect. All of the unique words used in the sentences of the reading span task from Experiment 2 were extracted (583 words in total).[3] Word pools consisting of 10, 15, 30, 60, and 350 words were created by sampling randomly without replacement from the 583 unique words. The pool size was then manipulated between groups. Two sets of stimuli were generated at each pool size, for generality, but participants (n = 59) were assigned to only one set of processing component stimuli for the entire experiment. Three of the participants, two from the 30 and one from the 60 pool size conditions, were dropped due to computer errors. The remaining numbers of participants in the 10-, 15-, 30-, 60-, and 350-item pool size conditions were 12, 10, 11, 9, and 14, respectively. For the processing component, the participants read "sentences" consisting of ten words presented on the screen in an incoherent sentence form—that is, the first word was capitalized and a period followed the final word, as had been the case in the previous experiments—that were drawn at random without replacement from one of the pools. For example, participants in the 10-item pool size condition were presented with the same ten words (in a random order) each time they experienced the processing component, whereas participants in the 350-item pool size condition viewed ten words randomly drawn from the pool of 350 words, and therefore were unlikely to view many redundant words while they experienced the processing component.

Results and discussion

Serial recall Analyses were conducted using a 5 (Pool Size) × 2 (Phonological Similarity) mixed factorial ANOVA. The model revealed very strong evidence for an effect of phonological similarity, $F(1, 51) = 38.8$, $p < .001$, $\eta_p^2 = .43$, but no evidence emerged of a main effect of pool size, $F(4, 51) = 1.08$, $p = .37$, $\eta_p^2 = .07$, or of an interaction, $F(4, 51) = 1.58$, $p = .19$, $\eta_p^2 = .11$. See Fig. 3.

[3] Short prepositions and high-frequency articles such as "a," "at," and "the" were not included.

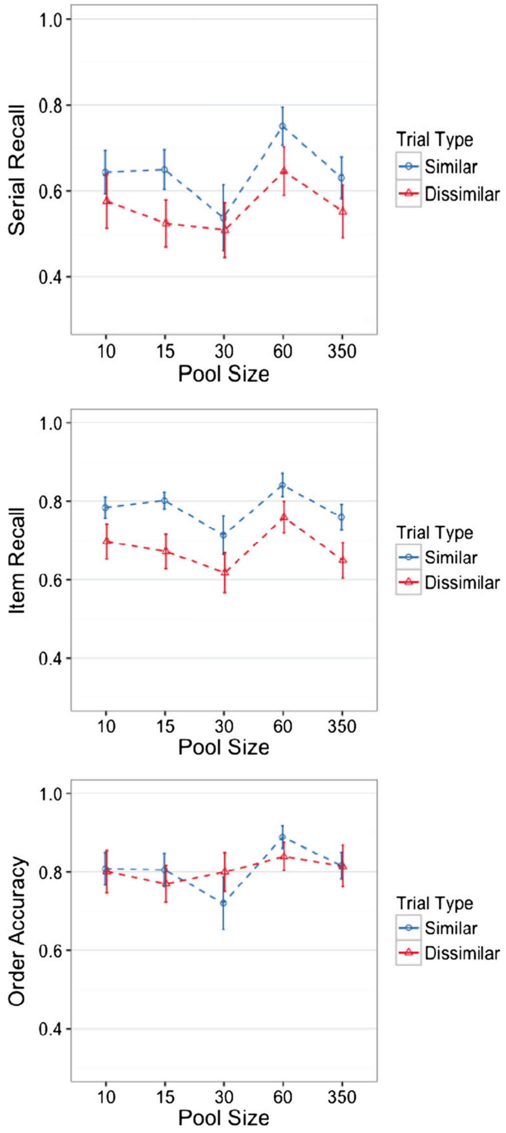

Fig. 3 Serial, item, and order recall accuracies for Experiment 3. Distractor stimuli were sampled without replacement from word pools of varying sizes

Item recall A similar analysis conducted on item recall accuracy also revealed a strong and significant main effect of phonological similarity, $F(1, 51) = 66.85$, $p < .001$, $\eta_p^2 = .56$. Similarly, we observed no evidence for a significant main effect of pool size, $F(4, 51) = 1.54$, $p = .20$, $\eta_p^2 = .10$, or for a significant interaction, $F(4, 51) = 0.44$, $p = .77$, $\eta_p^2 = .03$.

Order accuracy A similar analysis conducted on order recall accuracy showed no evidence for a main effect of phonological similarity, $F(1, 51) = 0.001$, $p = .97$, $\eta_p^2 < .001$. Moreover, we found no evidence for a significant main effect of pool size, $F(4, 51) = 0.63$, $p = .64$, $\eta_p^2 = .04$, and only weak evidence for a significant interaction, $F(4, 51) = 2.37$, $p = .064$, $\eta_p^2 = .15$. The weak evidence for an interaction may have been driven by a slight decrease in serial and (consequently) order accuracy at a pool size of 30.

Discussion

The presence of a phonological similarity enhancement did not depend on the pool size of the processing component. These results suggest that the phonological similarity effects observed previously in RSPAN tasks but not in OSPAN tasks were not due to the differences in the numbers of unique distractors between the two tasks. The results of Experiments 1–3 suggest that differences between distractor materials in OSPAN and RSPAN regarding verification, structure, and pool size were not driving the phonological similarity facilitation found in the RSPAN tasks. We conducted Experiment 4 to investigate the hypothesis that increased interference in the RSPAN condition might drive the phonological similarity benefit through the categorical exclusion of the distractor material at retrieval when the material would otherwise cause interference.

Experiment 4

In this experiment, we examined whether using a new set of categorically related distractors for the processing component, which are also fairly distinct from the memoranda, would eliminate the phonological similarity effect in complex span. In the first task ($n = 17$), participants read "sentences" that consisted of the same ten lexical items in a randomized order, as the participants had in the smallest pool size condition of Experiment 3. However, the "sentences" consisted of 10 three-syllable common names (e.g., *Jonathan*).

To examine whether the lack of phonological similarity enhancement in the operation span tasks was due to the fairly small number of syllables in the processing components consisting of operations (e.g., *2*, *+*, *is*), we also employed a second task. In the second task ($n = 19$), the same procedure was used as in the first task of this experiment, but with shortened, one- or two-syllable versions of those common names (e.g., *Jon*). Two participants were dropped from the shortened-names task, one due to disengagement noted by the experimenter, and another due to computer errors.

If the phonological similarity benefit occurs in RSPAN tasks because of the increased interference of categorically similar distractor material, then the systematic relationship between names and their categorical distinction from the memoranda should reduce distractor interference at recall. If this is the case, we should not observe a phonological similarity effect in either task.

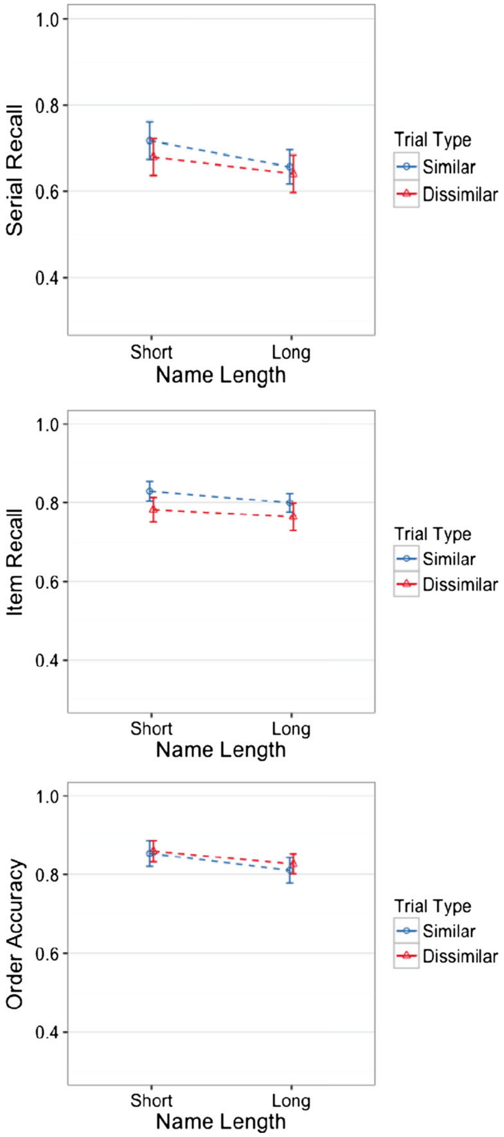

Fig. 4 Serial, item, and order recall accuracies for Experiment 4. All distractor stimuli consisted of either short or long names

Results

Serial recall A 2 × 2 mixed factorial ANOVA did not reveal any significant effects. We found no evidence for a main effect of task, $F(1, 32) = 0.75$, $p = .39$, $\eta_p^2 = .022$; no evidence for a main effect of phonological similarity, $F(1, 32) = 2.51$, $p = .12$, $\eta_p^2 = .072$; and no evidence for an interaction between the two, $F(1, 32) = 0.40$, $p = .53$, $\eta_p^2 = .012$. See Fig. 4.

Item recall A similar analysis showed no evidence for a main effect of task, $F(1, 32) = 0.37$, $p = .54$, $\eta_p^2 = .011$. However, it did reveal strong evidence for a main effect of phonological similarity, $F(1, 32) = 11.44$, $p = .001$, $\eta_p^2 = .26$. Finally, no evidence of an interaction was apparent, $F(1, 32) = 0.19$, $p = .65$, $\eta_p^2 < .006$.

Order accuracy A similar analysis failed to find evidence for any effects or interactions in order accuracy (lowest $p = .32$).

Discussion

The failure to find evidence of phonological similarity facilitation in serial recall is consistent with the hypothesis that the categorical exclusion of distractor items in both similar and dissimilar conditions removes or greatly weakens the phonological similarity facilitation effect. One important qualification is that item recall was still enhanced by phonological similarity. This may have been due to the manipulation weakening, but not providing the same level of interference, as in the OSPAN task. Moreover, it appears that the lack of an effect in serial recall was not due to the relatively short utterances required in the OSPAN task.

Calculating effect sizes across the previous experiments

In order to obtain more accurate measures of effect size, as well as to summarize previous findings, we applied a random-effects model to the present set of experiments, as well as to Experiment 2 of Macnamara et al. (2011), in which they examined phonological similarity in complex span tasks. The Standard OSPAN task from Experiment 1 of the present set of experiments was excluded, because it had the unique requirement that participants verify each math equation, and exhibited markedly lower overall performance than in the other OSPAN tasks. Each task was labeled either *high-interference* or *low-interference*. High-interference tasks were those in which the content of the processing component was categorically similar to the memoranda (RSPAN tasks; phonological similarity facilitation expected). Low-interference tasks were those in which the content of the processing task was categorically distinct (OSPAN tasks/Exp. 4 tasks; phonological similarity effects not expected). The model contained random effects for both task-level variance (s^2_{task}) and subject-level variance (s^2_{subj}). Fixed effects were specified for both levels of phonological similarity at each level of interference.

One issue with reports of Cohen's *d* as a measure of effect size is that it will differ depending on whether a within-subjects (repeated measures) or between-subjects design is used (Lakens, 2013). For our designs, we manipulated phonological similarity within subjects and processing components between subjects. Since each contains meaningful information, we report both types of Cohen's *d*. For each processing component interference group, Cohen's d_{within} was calculated by dividing the increase in serial-recall performance from the phonologically dissimilar to the phonologically similar condition by $\sqrt{2^* s^2_{\text{resid}}}$, where s^2_{resid} is the estimated residual variance for the model. Note that for a simple, dependent *t*-test, this formula returns equivalent Cohen's

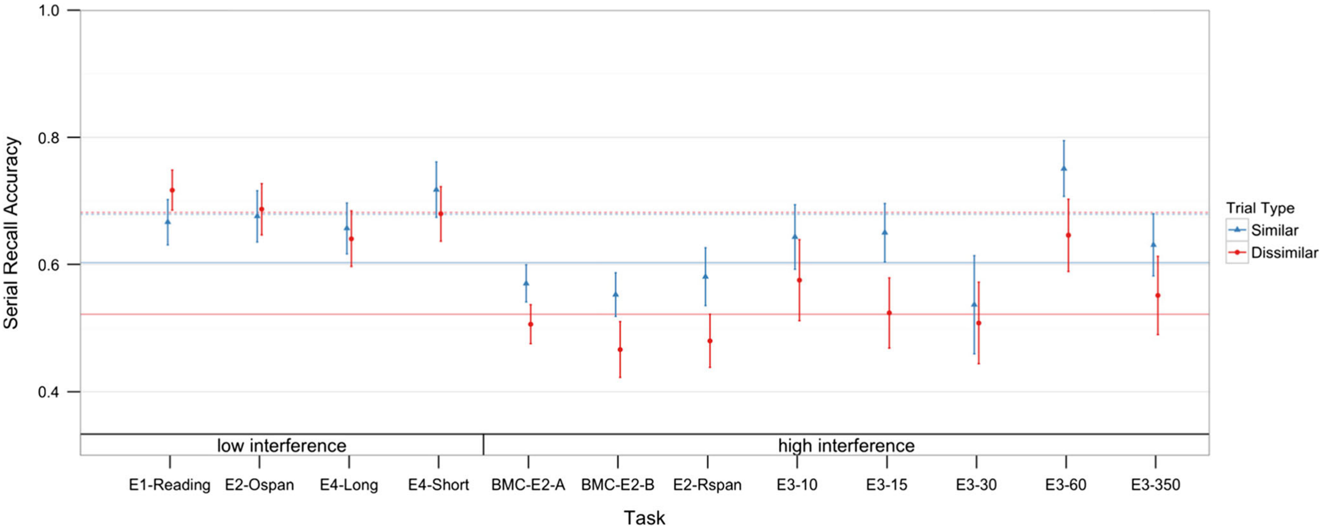

Fig. 5 Serial-recall means for each task. Fixed-effect estimates are plotted as horizontal lines and colored according to phonological similarity. The lines for high interference are solid, whereas those for low interference are dotted. BMC-E2-A and BMC-E2-B are the two tasks used in the second experiment of Macnamara et al. (2011)

d estimates. However, since this effect size estimate is only applicable to within-subjects designs, we also calculated Cohen's d_{betw}, by dividing the increase in serial-recall performance by $\sqrt{s^2_{\mathrm{subj}} + 2{*}s^2_{\mathrm{resid}}}$. Finally, 95 % confidence intervals for all estimates were derived through parametric bootstrapping.

The means of each task are plotted, with the model's prediction for each interference level–phonological similarity combination, in Fig. 5. See Table 1 for the fixed-effect and random-effect estimates. As expected, serial-recall accuracy in the low-interference conditions (OSPAN/Exp. 4 tasks) did not differ between the dissimilar and similar conditions (d_{within} = –0.024, 95 % CI [–0.25, 0.20]; d_{betw} = –0.014, 95 % CI [–0.15, 0.12]). However, in the high-interference conditions (RSPAN), the predicted serial-recall accuracy increased from the dissimilar to the similar condition (d_{within} = 0.69, 95 % CI [0.49, 0.89]; d_{betw} = 0.41, 95 % CI [0.30, 0.53]), but the similar-condition mean still fell below that of the low-interference conditions. In other words, high-interference conditions led to worse serial recall, though phonological similarity somewhat protected against greater interference. Interestingly, the model predicted little to no between-task variability, s_{task} = 0, 95 % CI [0, .045], but substantial between-subjects variability, s_{subj} =.15, 95 % CI [.13, .17].

Table 1 Multilevel model estimated serial-accuracy effects

Fixed Effects		Beta	Lower Bound	Upper Bound
Low dissimilar		.668	.625	.708
Low similar		.668	.621	.707
High dissimilar		.507	.474	.542
High similar		.585	.552	.622
Random Effects	*n*	*SD*	Lower Bound	Upper Bound
Subject (intercept)	184	0.155	.133	.172
Task (intercept)	12	0	.000	.042
Residual		0.081	.073	.090

"Lower bound" and "upper bound" refer to the bounds of the 95 % confidence intervals derived using a parametric bootstrap. *SD* = standard deviation

The same model was also run for item and order accuracies. Overall, the low-interference condition exhibited higher item and order accuracies than did the high-interference condition. Within the low-interference condition, we observed a minor increase in item accuracy due to similarity (d_{within} = 0.24, 95 % CI [0.012, 0.47]; d_{betw} = 0.15, 95 % CI [0.008, 0.31]), but this was offset by a minor decrease in order accuracy due to similarity (d_{within} = –0.25, 95 % CI [–0.49, –0.026]; d_{betw} = –0.18, 95 % CI [–0.35, 0.019]). Within the high-interference condition, there was a large increase in item accuracy due to similarity (d_{within} = 1.17, 95 % CI [0.95, 1.4]; d_{betw} = 0.76, 95 % CI [0.63, 0.90]) and no evidence for a difference in order accuracy (d_{within} = –0.006, 95 % CI [–0.19, 0.18]; d_{betw} = –0.004, 95 % CI [–0.14, 0.13]). Although the increase in item recall due to similarity is consistent with previous findings, it is surprising that similarity did not cause a decrease in order accuracy for both levels of interference, as would be expected by several interference theories of working memory (Brown et al., 2007; Oberauer et al., 2012).

Prior-list, distractor, and other intrusions

In addition to serial, item, and order accuracies, the recall data across experiments were scored according to prior-list, distractor, and other intrusions. A *prior-list intrusion* occurs when a word from a previous memory list is recalled for the current list. A *distractor intrusion* occurs when a distractor item is recalled. Finally, all other incorrect recalls that were not omissions were classified as *other intrusions*.

Prior list intrusions (PLIs) were analyzed using the multilevel model from the previous section. It should be noted that the average proportion of PLIs in any given condition was low (the proportion of PLIs in all tasks was approximately .02 or less). The phonologically similar trials showed virtually no PLIs across the low-interference (proportion PLIs = .0002, 95 % CI [–.002, .003]) and high-interference (proportion PLIs = .0008, 95 % CI [–.001, .003]) tasks. This is consistent with the notion that a rhyming cue distinguishes the memoranda on a trial from those on previous trials. On the other hand, the phonologically dissimilar trials had higher proportions of PLIs in the low-interference (proportion PLIs = .008, 95 % CI [.005, .011]) and in the high-interference (proportion PLIs = .015, 95 % CI [.013, .017]) tasks, with strong evidence for a greater proportion of PLIs in the high-interference tasks (high minus low PLIs = .007; $t = 3.917$, $p = .001$).

Distractor intrusions (DIs) were examined for the conditions in Experiment 3 with pool sizes 10, 15, and 30 (33 participants total), as well as for both names-as-distractor conditions in Experiment 4 (34 participants total). These conditions were chosen because they all used a small pool of distractors, so participants were exposed to all distractors in the pool either during the practice stage or early in the experiment. Accordingly, recalling any item from the distractor pool was considered a DI. Although DIs were extremely rare, each condition of Experiment 3 had either 19 or 20 total cases of DIs on dissimilar trials, and two cases of DIs on similar trials. Each of these results is highly unlikely under the null hypothesis that observed DIs had equal probabilities of being labeled as similar or dissimilar trials (binomial distribution; highest $p < .001$). In the names-as-distractors conditions of Experiment 4, there were no cases of DIs. Overall, this is consistent with the idea that the dissimilar condition of Experiment 3 was unique, in that participants did not necessarily categorically exclude the distractors as recall candidates.

Other intrusions (OIs) were analyzed using the multilevel model from the previous section. In low-interference tasks, OIs were less common for both similar trials (proportion OIs = .10, 95 % CI [.083, .11]) and dissimilar trials (proportion OIs = .09, 95 % CI [.079, .11]), with no evidence for a difference between similar and dissimilar trials within those tasks (d_{within} = .066, 95 % CI [–.16, .30]; d_{betw} = .041, 95 % CI [–.10, .18]). However, high-interference tasks had more OIs, across both similar (proportion OIs = .13, 95 % CI [.12, .15]) and dissimilar (proportion OIs = .16, 95 % CI [.14, .17]) trials, with more OIs in dissimilar than in similar trials within those tasks (d_{within} = .42, 95 % CI [.62, .23]; d_{betw} = .26, 95 % CI [.38, .14]).

General discussion

Although the source of the differential effects of phonological similarity on RSPAN and OSPAN tasks is complicated by a number of factors—including additional processing requirements (e.g., verifying a solution to a math problem), interdistractor dependencies (e.g., reading a grammatical sentence), and distractor pool size—the experiments above demonstrate the minimal conditions under which phonological similarity facilitation occurs. Moreover, under the simplest of conditions, in which ten items are reused across trials as distractors, manipulating which items are used as distractors abolishes the phonological similarity effect.

The present experiments addressed two possible explanations given by Macnamara et al. (2011) for the difference between the effects of phonological similarity in RSPAN and OSPAN tasks. The first was that the requirement that participants solve the math problems in the OSPAN abolishes the phonological similarity facilitation effect. In Experiment 1, we found no evidence for a phonological similarity effect, regardless of whether or not participants were required to solve math problems. A second possible explanation given by Macnamara et al. was that sentence reading produces the phonological similarity facilitation effect. In Experiment 2, we scrambled the sentences in both the OSPAN and RSPAN into an unstructured form. Likewise, in Experiment 3 we scrambled the sentences of the RSPAN into unstructured forms. We continued to observe a phonological similarity effect in the RSPAN tasks and not in the OSPAN tasks. These results demonstrate that reading grammatical sentences is not necessary for the phonological similarity effect.

Moreover, a third possible explanation is that the phonological similarity effect depends on the size of the pool from which distractors are drawn. However, in Experiment 3 we manipulated the sizes of the distractor word pools, and observed a consistent phonological similarity benefit that was not moderated by pool size.

The present results provide preliminary support for the notion that rhyming serves as a categorical cue, and that this cue is beneficial to the degree that the memoranda and distractor items are not already distinguishable from one another. Thus, using classes of distractor stimuli that were readily distinguishable from the memoranda—numbers and operators in Experiments 1 and 2, and names in Experiment 4—failed to produce evidence of a phonological similarity benefit. This relationship was clarified by considering the data across all experiments. Item recall accuracy performance was best in low-interference tasks; in high-interference tasks, recall was higher when rhyming memoranda were presented, relative to nonrhyming memoranda. This is consistent with the notion that a categorical cue from rhyming memoranda allows participants to exclude the distractors as candidates at recall.

However, one surprising result of the present study was that order errors remained unaffected by phonological similarity for high-interference tasks, whereas low-interference tasks exhibited only a slight decrease in order accuracy for similar items. Although this is consistent with the findings of

Macnamara et al. (2011), it seems at odds with the trade-off of increased positional confusions, as well as item accuracy, as a result of similarity that is anticipated by interference models of memory such as the SOB-CS model (Oberauer et al., 2012) or SIMPLE (Brown et al., 2007).

Conclusion

Three potential confounds with regard to the distractor component of complex span tasks—verification requirements, the grammars by which distractors are arranged, and the size of the stimulus pools from which they are selected—have previously made the cause of the phonological similarity facilitation effect for rhyming memoranda unclear. The present set of experiments suggests that these confounds are not responsible for the phonological similarity facilitation effect, but that phonological similarity facilitation occurs when the distractor items and the memoranda are not otherwise categorically distinct. If the distractor material and the memoranda are drawn from distinct categories (e.g., words and numbers/operators), competition at retrieval is relatively low, and phonological similarity is unnecessary to further differentiate the memoranda in memory. In cases of otherwise high competition, phonological similarity serves to differentiate the memoranda from distractor items at recall. These sets of experiments provide some insight concerning the importance of the role of semantic similarity between memory items and distractors in working memory.

References

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

Baddeley, A. D. (1968). How does acoustic similarity influence short-term memory? *Quarterly Journal of Experimental Psychology, 20,* 249–264. doi:10.1080/14640746808400159

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4,* 829–839. doi:10.1038/nrn1201

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York, NY: Academic Press. doi:10.1016/S0079-7421(08)60452-1

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. doi:10.1037/0096-3445.133.1.83

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114,* 539–576. doi:10.1037/0033-295X.114.3.539

Camos, V., Mora, G., & Barrouillet, P. (2013). Phonological similarity effect in complex span task. *Quarterly Journal of Experimental Psychology, 66,* 1927–1950. doi: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. doi:10.3758/s13421-010-0011-x

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33A,* 497–505. doi:10.1080/14640748108400805

Conlin, J. A., Gathercole, S. E., & Adams, J. W. (2005). Children's working memory: Investigating performance limitations in complex span tasks. *Journal of Experimental Child Psychology, 90,* 303–317. doi:10.1016/j.jecp.2004.12.001

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55,* 75–84.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12,* 769–786. doi:10.3758/BF03196772

Copeland, D. E., & Radvansky, G. A. (2001). Phonological similarity in working memory. *Memory & Cognition, 29,* 774–776. doi:10.3758/BF03200480

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466. doi:10.1016/S0022-5371(80)90312-6

Fallon, A. B., Groves, K., & Tehan, G. (1999). Phonological similarity and trace degradation in the serial recall task: When CAT helps RAT, but not MAN. *International Journal of Psychology, 34,* 301–307. doi:10.1080/002075999399602

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation, 12,* 395–427. doi:10.3758/BF03201693

Guérard, K., Saint-Aubin, J., Burns, S. C., & Chamberland, C. (2011). Revisiting backward recall and benchmark memory effects: A reply to Bireta et al. (2010). *Memory & Cognition*. doi:10.3758/s13421-011-0156-2

Gupta, P., Lipinski, J., & Aktunc, E. (2005). Reexamining the phonological similarity effect in immediate serial recall: The roles of type of similarity, category cuing, and item recall. *Memory & Cognition, 33,* 1001–1016. doi:10.3758/BF03193208

Hanley, J. R., & Bakopoulou, E. (2003). Irrelevant speech, articulatory suppression, and phonological similarity: A test of the phonological loop model and the feature model. *Psychonomic Bulletin & Review, 10,* 435–444. doi:10.3758/BF03196503

Henson, R. N. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology, 49A,* 80–115.

Huttenlocher, J., & Newcombe, N. (1976). Semantic effects on ordered recall. *Journal of Verbal Learning and Verbal Behavior, 15,* 387–399. doi:10.1016/S0022-5371(76)90034-7

Kučera, H., & Francis, N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t* tests and ANOVAs. *Frontiers in Psychology, 4,* 863. doi:10.3389/fpsyg.2013.00863

Larsen, J. D., & Baddeley, A. (2003). Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: Do they have a common source? *Quarterly Journal of Experimental Psychology, 56A,* 1249–1268. doi:10.1080/02724980244000765

Lobley, K. J., Baddeley, A. D., & Gathercole, S. E. (2005). Phonological similarity effects in verbal complex span. *Quarterly Journal of Experimental Psychology, 58A,* 1462–1478. doi:10.1080/02724980443000700

Macnamara, B. N., Moore, A. B., & Conway, A. R. A. (2011). Phonological similarity effects in simple and complex span tasks.

Memory & Cognition, 39, 1174–1186. doi:10.3758/s13421-011-0100-5

Nairne, J. S., & Kelley, M. R. (1999). Reversing the phonological similarity effect. *Memory & Cognition, 27,* 45–53. doi:10.3758/BF03201212

Neath, I., & Brown, G. D. A. (2006). SIMPLE: Further applications of a local distinctiveness model of memory. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 46, pp. 201–243). San Diego, CA: Elsevier Academic Press. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0079742106460060

Oberauer, K. (2009). Interference between storage and processing in working memory: Feature overwriting, not similarity-based competition. *Memory & Cognition, 37,* 346–357. doi:10.3758/MC.37.3.346

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

Saint-Aubin, J., & Poirier, M. (1999). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *Quarterly Journal of Experimental Psychology, 52A,* 367–394. doi:10.1080/713755814

Saito, S., & Miyake, A. (2004). On the nature of forgetting and the processing–storage relationship in reading span performance. *Journal of Memory and Language, 50,* 425–443. doi:10.1016/j.jml.2003.12.003

Tehan, G., Hendry, L., & Kocinski, D. (2001). Word length and phonological similarity effects in simple, complex and delayed serial recall tasks: Implications for working memory. *Memory, 9,* 333–348.

Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127–154. doi:10.1016/0749-596X(89)90040-5

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. doi:10.1037/0033-2909.133.6.1038