



Keeping it together: Semantic coherence stabilizes phonological sequences in short-term memory

Nicola Savill^{1,2}  · Rachel Ellis² · Emma Brooke² · Tiffany Koa² · Suzie Ferguson² · Elena Rojas-Rodriguez² · Dominic Arnold² · Jonathan Smallwood² · Elizabeth Jefferies²

Published online: 6 December 2017
© Psychonomic Society, Inc. 2017

Abstract

Our ability to hold a sequence of speech sounds in mind, in the correct configuration, supports many aspects of communication, but the contribution of conceptual information to this basic phonological capacity remains controversial. Previous research has shown modest and inconsistent benefits of meaning on phonological stability in short-term memory, but these studies were based on sets of unrelated words. Using a novel design, we examined the immediate recall of sentence-like sequences with coherent meaning, alongside both standard word lists and mixed lists containing words and nonwords. We found, and replicated, substantial effects of coherent meaning on phoneme-level accuracy: The phonemes of both words and nonwords within conceptually coherent sequences were more likely to be produced together and in the correct order. Since nonwords do not exist as items in long-term memory, the semantic enhancement of phoneme-level recall for both item types cannot be explained by a lexically based item reconstruction process employed at the point of retrieval (“redintegration”). Instead, our data show, for naturalistic input, that when meaning emerges from the combination of words, the phonological traces that support language are reinforced by a semantic-binding process that has been largely overlooked by past short-term memory research.

Keywords Meaning · Phonological binding · Verbal short-term memory · Semantic coherence · Speech

The comprehension and production of spoken language depends on maintaining speech sounds in order: “roaring with pain” has a rather different meaning than “pouring with rain.” The retention of these sounds and words in verbal short-term memory (STM) is thought to draw on linguistic representations—particularly the buffering of phonological information between speech perception and production systems (Baddeley, 2012; Jacquemot & Scott, 2006). However, the relative contributions of phonological and semantic representations to verbal STM, and the ways in which they interact, are unclear. Semantic manipulations typically have relatively subtle effects on STM, as compared to phonological

manipulations (e.g., Baddeley, 1966; Jefferies, Frankish, & Lambon Ralph, 2006; Majerus & van der Linden, 2003; Savill, Ellis, & Jefferies, 2017), so the consensus view is that STM is best explained by factors that influence the efficiency of phonological processing. Traditional accounts of STM suggest that the semantic contribution operates at the whole-word level—for example, activated semantic and lexical representations allow people to complete missing pieces of a phonological trace. In line with the notion of STM as an independent short-term phonological store (after Baddeley, 1986), semantic information is thought to influence the availability and selection of words used to reconstruct the phonological trace, but not the integrity of the phonological trace itself, via a retrieval-based process known as “redintegration” (see Poirier & Saint-Aubin, 1995; Saint-Aubin & Poirier, 1999).

The primacy of phonology in STM has been challenged by evidence that conceptual information can directly influence the stability of the phonological trace, from studies of patients with semantic dementia (Patterson, Graham, & Hodges, 1994), and broadly by theoretical perspectives that posit the direct involvement of lexical–semantic processing/knowledge in STM capacities (e.g., Acheson & MacDonald, 2009; Knott, Patterson, & Hodges, 1997; N. Martin & Gupta, 2004; R. C.

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13421-017-0775-3>) contains supplementary material, which is available to authorized users.

✉ Nicola Savill
n.savill@yorksj.ac.uk

¹ School of Psychological & Social Sciences, York St John University, Lord Mayor’s Walk, York YO31 7EX, UK

² Department of Psychology, University of York, York, UK

Martin, Lesch, & Bartha, 1999). One such perspective suggests that semantic information directly impacts the stability of phonological information in short-term memory: According to the “semantic binding” account (Patterson et al., 1994), sequenced speech sounds processed by the phonological system interact with representations of word meaning whenever we comprehend or produce language, allowing conceptual knowledge to scaffold evolving phonological processing in STM. Patients with semantic dementia show progressive degradation of conceptual knowledge but possess relatively preserved language skills, including fluent, well-formed speech and normal digit span (Jefferies, Jones, Bateman, & Lambon Ralph, 2005). When asked to repeat lists of words that they understand poorly, these patients make frequent phonological errors, characterized by phoneme migrations between the items (e.g., “cat, dog” might be recalled as “dat, cog”; Hoffman, Jefferies, Ehsan, Jones, & Lambon Ralph, 2009; Jefferies, Crisp, & Lambon Ralph, 2006; Jefferies, Hoffman, Jones, & Lambon Ralph, 2008; Knott et al., 1997; Majerus, Norris, & Patterson, 2007; Patterson et al., 1994). Although this evidence emphasizes the importance of semantic knowledge in STM and cannot be readily explained by a retrieval-based process, it is controversial, since the neurodegeneration in semantic dementia may affect lexical–phonological as well as semantic knowledge (Papagno, Vernice, & Cecchetto, 2013).

Attempts to identify evidence of more stable phonological processing for meaningful items in healthy participants—that is, semantic binding—have produced weak and inconsistent effects, especially when lexical–phonological knowledge has been controlled. Jefferies et al. (2006) presented unrelated words and nonwords in unpredictable mixed lists, a procedure that elicits significant numbers of phoneme migration errors even for words, providing a paradigm in which the effects of semantic and lexical variables on phonological stability can be tested. Although word frequency influenced migrations, concreteness (a semantic variable) had no clear effect on stability. Consequently, all of the effects in this study could be explained in terms of the contribution of phonological–lexical representations, as opposed to a role for conceptual knowledge. Encoding tasks that require participants to attend to semantic and nonsemantic features of words have revealed fewer phoneme migrations for semantically encoded items, which aligns with the semantic-binding account (Savill, Metcalfe, Ellis, & Jefferies, 2015); however, studies investigating the effect of training new lexical–phonological forms with or without associated meanings have produced conflicting results (Benetello, Cecchetto, & Papagno, 2015; Savill et al., 2017).

It is possible that we currently underemphasize the importance of semantic information in maintenance processes in STM and phonological binding, for several reasons: (1) Studies typically examine performance at the level of whole

items, and thus cannot directly examine changes in phoneme migration errors; (2) it is difficult to experimentally separate the effects of conceptual knowledge on the stability of the entire phonological trace from item reconstruction (redintegration); and (3) studies often use lists of words that are not inherently meaningful and that consequently minimize any advantages that might occur from conceptual retrieval. Using a novel method, we overcame all of these issues. We constructed controlled sets of stimuli that allowed us to quantify phoneme-binding errors for sequences of words that established a coherent overall meaning (like a story) and for more standard lists of random (unrelated) words and nonwords. In two experiments, we compared immediate serial recall (ISR) for these coherent and random word conditions in both pure-word lists and mixed lists containing both words and nonwords. Substantial differences in the phonological stability of coherent and random lists (indexed by different rates of specific errors in which phonemes strayed out of position and incorrectly recombined with other phonemes) would provide converging evidence for semantic binding in healthy individuals. The inclusion of nonwords in mixed lists allowed us to test the contribution of semantic knowledge to the stability of the phonological trace beyond the reconstruction of specific items, since nonwords have no whole-item long-term memory representations to draw upon for support. Therefore, the effect of semantic coherence on nonword recall in mixed lists would provide a key test of the predictions of the semantic-binding hypothesis (and with it, a potential challenge for purely item-based reconstruction explanations of semantic effects in STM).

Method

We examined semantic binding in two independent datasets, allowing us to assess the robustness of the results. The methods of these two experiments are presented together for simplicity. Below we report how we determined our sample sizes, any data exclusions, all manipulations, and all measures in the study.

Participants

All participants, across both experiments, were native British-English-speaking adults with normal hearing, 18–31 years of age (main experiment, $M = 21.36$, $SD = 1.73$; replication sample, $M = 21.13$, $SD = 3.21$), who volunteered and gave their informed consent. Twenty-eight participants took part in the main study. This sample size was based on those in previous research (e.g., Jefferies et al., 2006; Savill et al., 2015) and allowed the four different versions of the task to be fully counterbalanced. One participant’s data were not used due to

an audio recording failure. Twenty-four participants took part in the replication.

Stimuli

In the main experiment, participants were presented with 130 lists of six spoken monosyllabic items in five phonologically controlled experimental conditions (26 lists per condition): (1) semantically meaningful telegraphic word sequences (SEM WORD; e.g., “watch band first live gig stage”); (2) “random” unrelated word sequences (RANDOM WORD; e.g., “lamp seal phase part think ground”); (3) mixed lists of words and nonwords, with the words forming a meaningful sequence (SEM MIXED; e.g., “wash /sne z/ sheets /dr k/ bed /ma g/”) (nonwords are indicated with slashes and written using the International Phonetic Alphabet); (4) mixed lists of words and nonwords in which the words were unrelated (RANDOM MIXED; e.g., “beat /m p/ flag coin /tru k/ /t el/”); and (5) meaningless nonword sequences (NONWORD; e.g., “/fæmp/ /θi nd/ /pe p/ /la z/ / r ŋk/ /sa t/”). The replication sample was tested on the 26 RANDOM MIXED and 26 SEM MIXED trials.

The SEM WORD and SEM MIXED lists were constructed from an initial pool of 52 six-word, semantically coherent sequences (SEM WORD trials). Each list was constructed so that the individual phonemes occurred no more than once at the same syllabic position across the items in the list; this allowed the majority of phoneme migrations to be traced (since migrations largely preserve syllable position; Ellis, 1980). The SEM MIXED trials were created by replacing three items from these lists with nonwords, so that the remaining words were not all in consecutive positions in the list. These nonwords were created by recombining the phonemes from the three words that were replaced (and therefore phonemically matched to the SEM WORD list). The nonwords were otherwise in unpredictable locations, such that, across lists, words and nonwords occurred in each serial position an equal number of times, similar to the mixed-list structures of random words and nonwords in Jefferies et al. (2006). We did not manipulate the predictability of word and nonword locations in these lists (unlike Jefferies, Frankish, & Noble, 2009), and specific mixed-list structures were presented different numbers of times. These lists were then divided into two matched sets of 26 SEM WORD and 26 SEM MIXED trials (to be tested in different participants; i.e., a SEM WORD trial in stimulus Set A was a SEM MIXED trial in Set B, and vice versa, to avoid item-specific effects) on the basis of the lists’ average lexical frequency (SUBTLEX: Van Heuven, Mandera, Keuleers, & Brysbaert, 2014), imageability (Cortese, 2004), and average ratings of the semantic coherence and emotionality of each SEM WORD sequence and SEM MIXED three-word set (the ratings were made by five participants who did not take part in the main study) (see

Table S1 in the supplementary material). The 26 RANDOM MIXED lists and RANDOM WORD lists were constructed from additional words chosen to match the average properties of the SEM lists. The nonwords for both the NONWORD lists and RANDOM MIXED lists were created by recombining, respectively, the phonemes from the words in the RANDOM WORD list and the words from other RANDOM MIXED lists (and were therefore matched for syllable structure). Additional details regarding stimulus construction can be found in the [supplementary materials](#).

The replication retested the mixed trials from one version of the main experiment.

Procedure

Participants wore a headset with an integrated microphone to record spoken responses. They were advised that they would hear six-item lists that consisted of words, of nonwords, and of words and nonwords mixed together (or of only mixtures of words and nonwords, for the replication sample). They were asked to attempt to repeat all six items, in order of presentation, immediately at the end of each list, and to attempt to produce items whenever possible, even if they were unsure. The stimuli were presented at a rate of one item per second. After recalling each list, participants pressed a key to start the next trial. The experiment took approximately 50 min to complete, including short rest breaks every 26 trials, as well as five initial practice trials per condition (20 min for the replication, including a rest break halfway through and two practice trials at the start). Responses were recorded digitally for later coding.

Response coding and analysis

Verbal responses were transcribed phoneme by phoneme. We report two complementary analyses.

The first analysis examined eight types of *responses* at the whole-item level, as a function of list condition. These item-level analyses allowed for comparisons with most other studies of verbal recall. We used a response-based (rather than target-based) coding approach in order to capture potentially relevant errors that were not phonologically related to the target (e.g., semantically related errors). The eight types of responses were identified as follows: (1) We coded items correctly recalled in position. (2) Item order errors were target items produced in incorrect list positions. Non-target responses containing target phonemes (in the same syllable position) were classified as either (3) phoneme recombination errors, when the response phonemes originated from *more than one* target item in the list, or (4) nonrecombination yet phonologically related responses, which were partially correct but contained only the phonemes from one target. When fewer than six responses were given, the missing response was

identified as (5) an omission. The remaining responses, which did not contain target phonemes, were counted as (6) semantically related to the target list (e.g., “mud” when “swamp” was a target word), (7) item intrusions from one of the previous six lists, or (8) unrelated. These data were expressed as a percentage of the total target items.

While it is well-established that sentence-like sequences of words should be recalled more easily than random words (e.g., Brener, 1940; Jefferies, Lambon Ralph, & Baddeley, 2004; Miller & Selfridge, 1950), this analysis identified changes in specific error types across conditions. Savill et al. (2015) used the same item-level coding scheme and found reductions in phoneme recombination errors related to the use of a semantic-encoding strategy (in line with the semantic-binding predictions); we therefore predicted differences in accuracy (correct-in-position responses) and phoneme recombination errors, which directly index the stability of the phonological trace.

Further details of the coding scheme and a worked example of a single trial are provided in the supplemental materials (Table S2). Response types that captured <1% of possible responses were not analyzed. Note that all aggregated ISR response data used for analysis are accessible at <https://goo.gl/KPBVyB>.

In analyses of item-level responses for the main experiment, we computed one-way ANOVAs with five levels: SEM WORD, RANDOM WORD, SEM MIXED, RANDOM MIXED, and NONWORD.¹ Greenhouse–Geisser correction was applied to the degrees of freedom when the sphericity assumption was violated. Bonferroni-corrected pairwise comparisons (at a conservative α of .005, allowing for multiple comparisons) were used to determine which conditions contributed the effect of list condition (Table 1, which includes the results of the key comparisons of SEM WORD vs. RANDOM WORD and SEM MIXED vs. RANDOM MIXED responses). For the replication sample, we compared each response type in the SEM MIXED and RANDOM MIXED conditions using paired *t* tests. These analyses are at the level of complete responses, and therefore cannot examine the retention of target phonemes presented as part of either words or nonwords in mixed lists.

The second analysis, at the phoneme level, examined the preservation of target phonemes, split by lexicality, in more detail. Since it was not possible to categorize item-level errors by source lexicality, the purpose of these phoneme-level analyses was to separate word phonemes from nonword phonemes so that we could identify the target sources of recombination errors in the mixed lists. We traced the origins of preserved target phonemes from words and nonwords produced as part

of items correct-in-position, item order errors, recombination errors, and nonrecombination phonological error responses (which were phonologically related to at least one target), to examine the retention of the phonological elements of items presented in mixed lists. The corresponding data for each response type were expressed as percentages of the total word and nonword target phonemes. For nonwords, we report paired-sample *t* tests (Bonferroni-corrected for each comparison with $\alpha = .025$) comparing (1) nonword recall in RANDOM MIXED lists versus pure NONWORD lists (to assess the effect of nonwords being presented alongside words) and (2) nonword recall in SEM MIXED lists versus RANDOM MIXED lists (to examine the effect of semantic coherence on the stability of nonword items). This analysis is important for establishing whether the effects of semantic binding were specific to words or affected the stability of the entire phonological trace. Any effects of semantic coherence for nonwords are unlikely to reflect item-specific reconstruction processes (i.e., redintegration). For phoneme responses traced to word targets in mixed lists, we report three comparisons (Bonferroni-corrected for each comparison, with $\alpha = .017$), examining (1) word recall in RANDOM MIXED and pure WORD lists (to assess the effect on random words of being presented alongside nonwords), (2) word recall in SEM WORD compared with SEM MIXED lists (to assess the effect on semantically coherent words of being presented alongside nonwords) and (3) word recall in SEM MIXED lists versus RANDOM MIXED lists (to examine the effect of semantic coherence in mixed lists). For the replication sample, paired sample *t* tests compared nonword phoneme recall in RANDOM MIXED lists versus SEM MIXED lists, and word phoneme recall in RANDOM MIXED lists versus SEM MIXED lists.

For completeness, in both experiments we also ran 2×2 repeated measures ANOVAs to specifically test the effects of the semantic manipulation on mixed-list performance (RANDOM MIXED, SEM MIXED) according to the lexicality of the source phonemes (word, nonword)—that is, to examine how the semantic effects on recall compared across target types. These analyses are provided in the [supplementary materials](#).

Comparison of main experiment and replication

Finally, we used mixed ANOVAs to assess the stability of the semantic-coherence effects in the mixed lists tested in both experiments. This analysis included the within-subjects factor of semantic coherence (RANDOM MIXED vs. SEM MIXED) and the between-subjects factor of experiment (main experiment vs. replication), assessing responses both at the item and phoneme levels. This analysis is reported in full in the [supplementary material](#) (Table S3).

¹ All analyses were also performed on arcsine-transformed proportions, to better meet parametric assumptions for proportion/percentage data. These produced the same pattern of statistical outcomes.

Table 1 Item-level response categories and pairwise comparisons in the main experiment

ISR Resp.	ANOVA Main Effect of List Condition	Bonferroni-Corrected Pairwise Comparisons				
		Mixing			Semantic Coherence	
		Nonword vs. Random Mixed	Random Word vs. Random Mixed	SEM Word vs. SEM Mixed	Random Word vs. SEM Word	Random Mixed vs. SEM Mixed
Correct	<i>F</i> = 453.13 <i>p</i> < .001 $\eta_p^2 = .95$	− 13.10*** <i>d</i> = − 1.49	17.54*** <i>d</i> = 2.09	24.96*** <i>d</i> = 3.76	− 10.80*** <i>d</i> = − 1.61	− 6.31*** <i>d</i> = − 0.67
IT-ORD	<i>F</i> = 28.20 <i>p</i> < .001 $\eta_p^2 = .52$	− 4.91*** <i>d</i> = − 1.25	5.74*** <i>d</i> = 0.94	2.13, n.s., <i>d</i> = 0.50	0.87, n.s., <i>d</i> = 0.20	− 1.90, n.s., <i>d</i> = − 0.40
RECOMB	<i>F</i> = 364.06 <i>p</i> < .001 $\eta_p^2 = .93$	11.80*** <i>d</i> = 1.82	− 17.84*** <i>d</i> = − 2.49	− 29.07** <i>d</i> = − 4.87	10.54*** <i>d</i> = 1.97	5.14*** <i>d</i> = 0.94
NON-RECOMB	<i>F</i> = 217.31 <i>p</i> < .001 $\eta_p^2 = .89$	6.65*** <i>d</i> = 0.93	− 14.09*** <i>d</i> = − 1.93	− 17.37*** <i>d</i> = − 3.30	7.37*** <i>d</i> = 1.25	2.58, n.s., <i>d</i> = 0.48
OM	<i>F</i> = 2.79 <i>p</i> = .070 $\eta_p^2 = .10$	1.66, n.s., <i>d</i> = 0.19	− 1.78, n.s., <i>d</i> = − 0.18	− 0.95, n.s., <i>d</i> = − 0.22	0.93, n.s., <i>d</i> = 0.19	0.80, n.s., <i>d</i> = 0.10
UNR	<i>F</i> = 25.95 <i>p</i> < .001 $\eta_p^2 = .50$	4.42*** <i>d</i> = 0.69	− 3.49** <i>d</i> = − 0.64	− 5.80*** <i>d</i> = − 1.45	3.16* <i>d</i> = 0.69	− 0.71, n.s., <i>d</i> = − 0.14

The main effects of list condition for each item-level response type are shown in the lefthand column in italics with partial eta-squared estimates of effect size. All *t* values are shown for each planned comparison with asterisks to denote Bonferroni-corrected significance levels, and all comparisons with corrected *p* values less than .1 are further highlighted in bold. Beneath the *t* values are the respective Cohen's *d* measures of effect size. IT-ORD = whole item order errors. RECOMB = responses recombining target phonemes from more than one item. NON-RECOMB = phonologically related errors that did not recombine target phonemes from more than one item; OM = omissions. List intrusions and semantic errors accounted for <1% of responses and were not analyzed. *** Corrected *p* < .001, ** Corrected *p* < .01, * Corrected *p* < .05.

Results

Item-level responses

Figure 1 shows the percentage of ISR responses of each type at the item level for each condition in the main experiment, alongside data from the replication sample. Table 1 reports the ANOVA analyses for each response type in the main experiment.

Accuracy We found substantial effects of list condition on correct-in-position scores (see Table 1). Bonferroni-corrected pairwise comparisons showed significant differences between all five conditions: Recall was most accurate in the SEM WORD condition, followed by RANDOM WORD, SEM MIXED, RANDOM MIXED, and NONWORD conditions (Table 1 and panel a in Fig. 1). The effect of mixing words and nonwords replicated the findings of Jefferies et al. (2006): Nonwords were recalled better when they were presented alongside words than when they were presented with other nonwords, whereas words were recalled more poorly in mixed than in pure

word lists.² Moreover, in the cases of both mixed and pure word lists, the items providing a coherent semantic structure were recalled more accurately (Table 1).

Item errors Phonologically unrelated semantic and list intrusion errors were rare (less than 1% of responses) and did not permit inferential analyses. With the exception of omission errors, all error categories were significantly affected by list condition (see Table 1). Paired comparisons showed that *more* complete items were recalled out of sequence in pure word lists than in mixed lists [RANDOM WORD > RANDOM MIXED] and in mixed lists than in nonword lists [NONWORD < RANDOM MIXED], presumably because when phonological stability was lower, phonemes were more likely to break apart and recombine with the elements of other list items (Jefferies et al., 2006). The effect of list composition on item order errors for semantically coherent items did not survive correction [SEM WORD ≈ SEM MIXED]. There was no clear effect at all of semantic coherence on item order errors [SEM WORD ≈ RANDOM WORD and SEM MIXED ≈ RANDOM MIXED].

Phoneme recombination errors were the only error type that showed the same pattern as accuracy (but in the opposite direction), with significant changes between *all* conditions (panel c in Fig. 1, Table 1). The fewest recombination errors

² Analyses demonstrating that these mixing effects were not accounted for by differences in the numbers of nonword targets in the lists can be found in the supplementary materials.

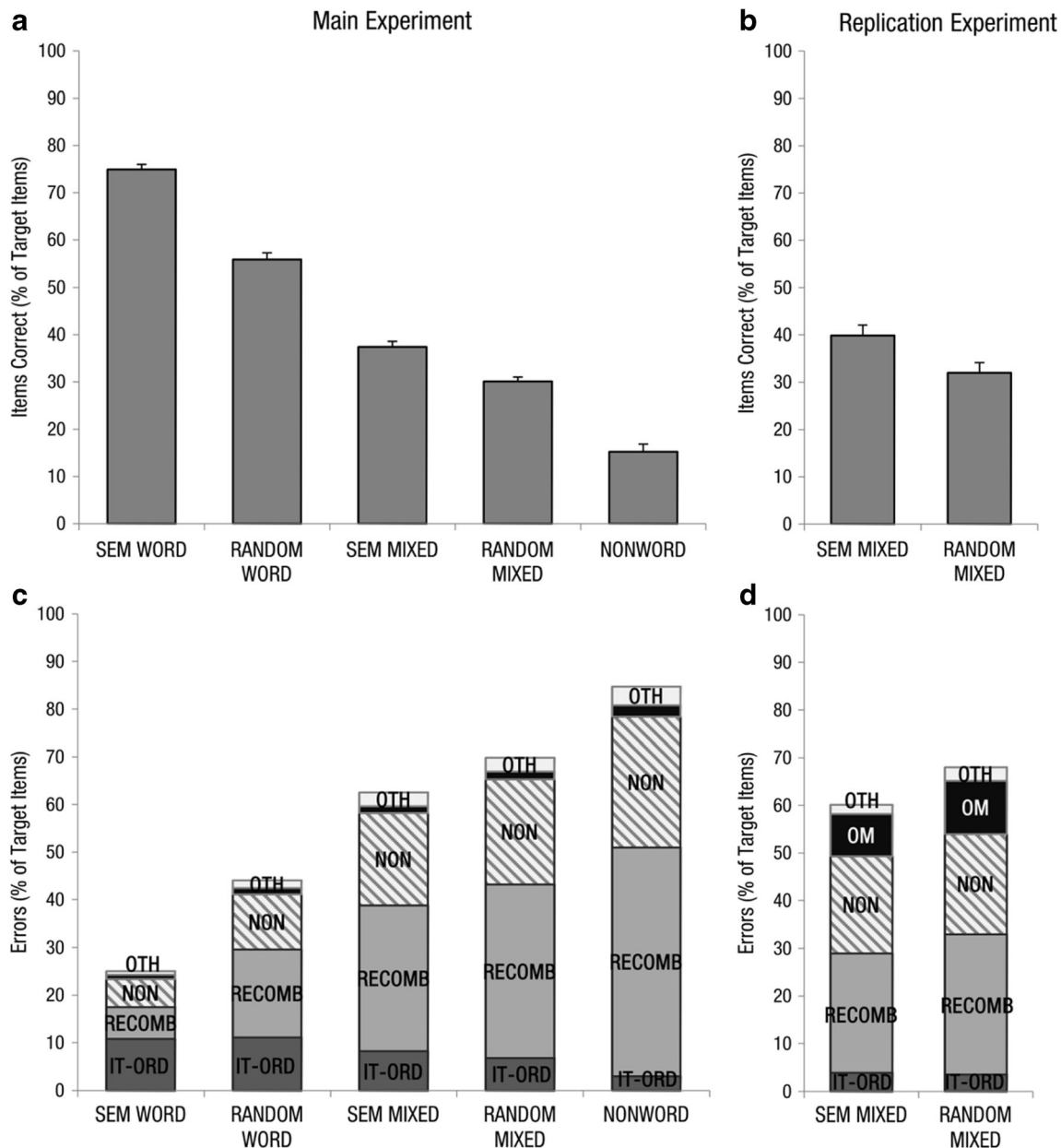


Fig. 1 Item-level response coding in each condition in the main experiment (left) and in the subsequent replication (right). The upper panels show the percentages of items correct in position for each condition (panel **a** = main experiment; panel **b** = replication). Error bars show 95% confidence intervals for a within-subjects design (Cousineau, 2005). The lower panels show the percentages of errors of each type (in stacked bars) for the different conditions (panel **c** = main experiment; panel **d** = replication). N.B. The sum of accurate responses and errors totals 100%

($a + c = 100\%$ and $b + d = 100\%$). IT-ORD = whole-item order errors. RECOMB = responses recombining target phonemes from more than one item. NON = phonologically related errors that did not recombine target phonemes from more than one item; OM = omissions; OTH = all other responses that were phonologically unrelated to the list targets (unrelated errors, list intrusion errors, and semantic errors). In panel **c**, the unlabeled response category marked in black corresponds to omissions.

were produced for SEM WORD, followed by RANDOM WORD, SEM MIXED, and RANDOM MIXED, with the most phoneme recombinations in the NONWORD condition. Thus, the rates of recombination errors were influenced by both the lexical composition of the lists [RANDOM WORD < RANDOM MIXED < NONWORD] and the availability of semantic structure [RANDOM WORD > SEM WORD; RANDOM MIXED > SEM MIXED].

Phonologically related nonrecombination errors and unrelated errors followed a pattern similar to that for recombination errors (in terms of effects of list composition; RANDOM WORD < RANDOM MIXED < NONWORD), but the differences between the mixed-list conditions did not survive Bonferroni correction (i.e., SEM WORD < RANDOM WORD < SEM MIXED ≈ RANDOM MIXED < NONWORD).

Replication sample Despite some differences in the overall frequencies of error types, across conditions, between the sets of participants tested (Fig. 1), the follow-up data replicated the key semantic-binding effects from the main experiment [correct-in-position responses: RANDOM MIXED < SEM MIXED, $t(27) = -5.92, p < .001, d = -0.75$; item order errors: RANDOM MIXED \approx SEM MIXED $t(27) = -0.84, p = .41, d = -0.14$; recombination errors: RANDOM MIXED > SEM MIXED, $t(27) = 5.13, p < .001, d = 0.70$; nonrecombination phonological errors: RANDOM MIXED \approx SEM MIXED, $t(27) = 0.61, p = .55, d = 0.10$], with the exception that omission errors and unrelated errors were also significantly reduced in the SEM MIXED condition relative to the RANDOM MIXED condition [omissions: RANDOM MIXED > SEM MIXED, $t(27) = 2.35, p < .05, d = 0.18$; unrelated: RANDOM MIXED > SEM MIXED, $t(27) = 2.11, p = .05, d = 0.50$]. Importantly, despite the differences in the task structures and participant groups tested, the size of the semantic effect did not differ for any phonologically related response type between the main experiment and the replication (items correct in position, item order errors, recombination errors, and nonrecombination phonological errors were similarly modulated—i.e., null interactions between task and semantic coherence); only the rates of omissions and unrelated responses scaled differently (see the [supplementary material](#) and Fig. 1).

Phoneme-level responses

Figure 2 shows the respective percentages of word and nonword target phonemes recalled as part of each response category and for each condition in both experiments. Table 2 reports the outcomes of statistical tests for the main experiment.

Nonword phonemes Phonemes from nonwords were more likely to be correctly recalled as part of a complete item in position when they were presented alongside words in mixed lists [RANDOM MIXED > NONWORD]. Phonemes from nonwords in mixed lists were also recalled more frequently as part of a complete item when the words in the sequence were semantically coherent [RANDOM MIXED < SEM MIXED].

Nonword phonemes were *less* likely to be produced as part of a recombination response in mixed than in pure nonword lists [RANDOM MIXED < NONWORD]. Importantly, they were also less likely to migrate and recombine with other target phonemes when presented alongside coherent words [RANDOM MIXED > SEM MIXED nonword]; see Fig. 2. This pattern shows that semantic support from the words affected the stability of the complete phonological trace.

We found no differences between conditions in the rates of nonword phonemes being produced as part of whole target

items out of sequence or as part of partially incorrect nonrecombination responses (see Table 2).

Word phonemes Phonemes from words were less likely to be recalled as part of a complete item in position when they were presented with nonwords in mixed lists [RANDOM MIXED < RANDOM WORD and SEM MIXED < SEM WORD]. They were also more likely to be produced as part of a complete item in position when the words formed a meaningful sequence [RANDOM MIXED words < SEM MIXED words].

List composition (mixed vs. pure) did not influence the percentage of word phonemes produced as part of whole-item order errors when the words did not form a coherent sequence [RANDOM MIXED \approx RANDOM WORD words]. There was a tendency for these responses to *increase* when the target words formed a meaningful sequence [RANDOM MIXED words < SEM MIXED words and SEM WORD < SEM MIXED word], which again might reflect a tendency of phonemes to migrate together, and not break apart, when phonological stability was higher.

Word phonemes were more likely to migrate and recombine with other target phonemes when in mixed lists, relative to pure word lists [RANDOM MIXED words > RANDOM WORD and SEM MIXED words > SEM WORD]. Word phonemes were also less likely to migrate and recombine with other target phonemes in mixed lists when the words formed a meaningful sequence [RANDOM MIXED words > SEM MIXED words].

Word phonemes produced as part of nonrecombination errors were relatively frequent for mixed lists [RANDOM MIXED words > RANDOM WORD and SEM MIXED words > SEM WORD]. However, word phoneme nonrecombination errors did not vary according to the semantic coherence of the mixed lists [RANDOM MIXED words \approx SEM MIXED words].

Replication sample Table 3 reports the outcome of statistical tests for RANDOM MIXED versus SEM MIXED comparisons. The effects of semantic binding were fully replicated: Both word and nonword phonemes were likely to be produced as part of a complete target item in position when the words in the sequence were semantically coherent. Both word and nonword phonemes were again less likely to migrate and recombine with other target phonemes when the words were semantically coherent [recombinations: RANDOM MIXED > SEM MIXED; see Table 3]. As in the main experiment, semantic coherence in the mixed lists did not significantly influence nonword and word phonemes produced as part of whole target items out of sequence [item order errors: RANDOM MIXED nonwords \approx SEM MIXED nonwords] or as part of partially incorrect nonrecombination responses [nonrecombination errors: RANDOM MIXED nonwords \approx SEM MIXED nonwords; see Table 3].

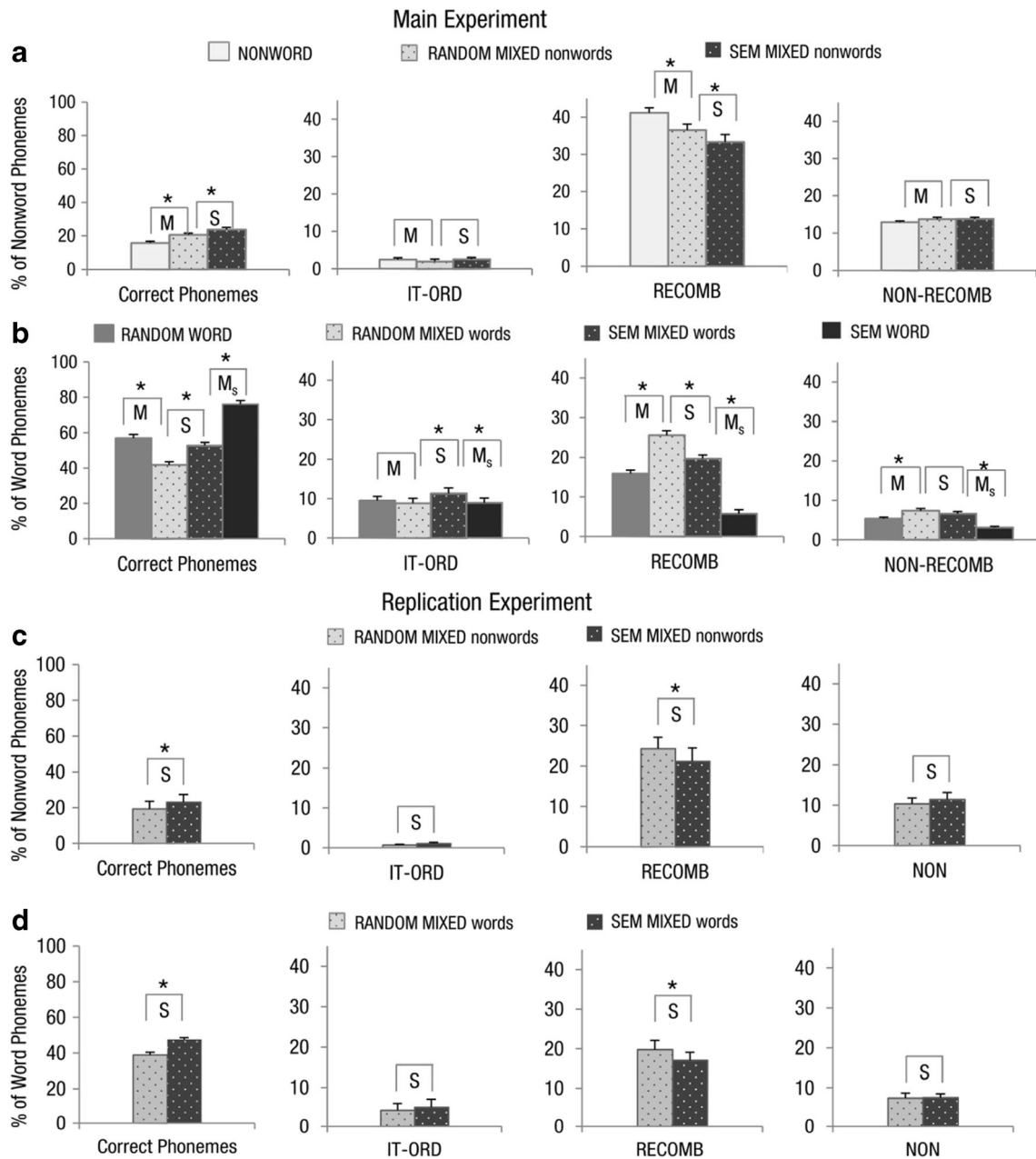


Fig. 2 Phoneme-level responses in the main experiment and the replication experiment. Responses that corresponded with *nonword* target phonemes (**a**: main experiment; **c**: replication) are shown above the responses corresponding with *word* target phonemes (**b**: main experiment; **d**: replication), each split by response type and expressed as a percentage of the total nonword or word target phonemes, respectively. The results of Bonferroni-corrected *t* tests examining the effect of list composition (i.e., mixing of words and nonwords, labeled M) and the effect of semantic coherence (labeled S) are shown. Since not all target phonemes were produced in the response, the bars for each condition do not total to 100%. Correct Phonemes = response phonemes

that formed part of a correct item, in the correct position. IT-ORD = response phonemes that formed part of an item order error. RECOMB = phonemes that formed part of a recombination response incorporating phonemes from more than one target. NON-RECOMB = phonemes produced as part of a response that was phonologically related to the target but did not include phonemes from more than one item. M = mixed vs. pure lists; M_S = semantically coherent mixed versus semantically coherent pure lists; S = semantically coherent mixed lists versus random mixed lists. *Significantly different comparisons (*p* < .05 corrected). Errors bars show 95% confidence intervals for a within-subjects design (Cousineau, 2005).

For a complete picture, repeated measures ANOVAs were run to assess the semantic effects in mixed lists (RANDOM MIXED, SEM MIXED) according to the lexicality of the source phonemes (word, nonword) (see the [supplementary](#)

[materials](#)). These analyses confirmed that, whereas semantic influences on recall accuracy were stronger for words than for nonwords (interactions of semantic manipulation and lexicality), the semantic effects of nonword recall were not carried by

Table 2 Phoneme-level response categories and pairwise comparisons in the main experiment, split by lexicality for mixed lists

ISR Resp.	Nonword Analyses		Word Analyses		
	Mixing	Semantic Coherence	Mixing		Semantic Coherence
	NONWORD vs. RANDOM MIXED Nonword	RANDOM MIXED Nonword vs. SEM MIXED Nonword	RANDOM WORD vs. RANDOM MIXED Word	SEM WORD vs. SEM MIXED Word	RANDOM MIXED Word vs. SEM MIXED Word
Correct	-5.35^{***} <i>d</i> = -0.50	-2.88^{**} <i>d</i> = 0.29	9.33^{***} <i>d</i> = 1.13	15.41^{***} <i>d</i> = 2.16	-7.20^{***} <i>d</i> = -0.89
IT-ORD	0.96, n.s. <i>d</i> = 0.27	-1.35, n.s. <i>d</i> = -0.34	0.81, n.s. <i>d</i> = 0.15	-2.39[*] <i>d</i> = -0.59	-2.39[*] <i>d</i> = -0.53
RECOMB	5.49^{***} <i>d</i> = 0.86	2.41[*] <i>d</i> = 0.57	-10.03^{***} <i>d</i> = -1.45	-18.02^{***} <i>d</i> = -3.19	4.54^{***} <i>d</i> = 0.97
NON-RECOMB	-1.61, n.s. <i>d</i> = -0.34	-0.09, n.s. <i>d</i> = -0.02	-4.77^{***} <i>d</i> = -0.86	-9.08^{***} <i>d</i> = -1.69	1.97, n.s. <i>d</i> = 0.43

All *t* values are shown for each mixed list comparison in the main experiment with asterisks to denote Bonferroni-corrected significance levels and all comparisons with corrected *p* values less than .05 are further highlighted in bold. Beneath the *t* values are the respective Cohen's *d* measures of effect size. *** Corrected *p* < .001, ** Corrected *p* < .01, * Corrected *p* < .05.

the recall of word items in either experiment (main effects of the semantic manipulation).

Analyses that directly compared the main experiment and replication results at the phoneme level confirmed that the coherence effects on each response type were similar between the sets and showed that semantic coherence improves the *overall* recall of word and nonword target phonemes (details in the supplementary materials, Table S3).

Discussion

We demonstrated, in two independent datasets, that semantic knowledge improves the coherence of linguistic information in short-term memory at the phonological level. Phonemes are more likely to be recalled together in the correct configuration, rather than recombined with the elements of other list items, when target words are presented within a meaningful sequence.

This stabilizing effect of semantic coherence on phoneme order extends to meaningless nonwords when these are mixed with the words, suggesting that semantic binding of phonology influences the stability of the entire phonological trace, and that semantic-binding effects cannot be fully explained in terms of the reconstruction of familiar items from lexical knowledge (since nonwords cannot be reconstructed in the same way). These findings have important theoretical and practical implications for our understanding of STM and language processing, since they point to an alternative mechanistic account of the semantic contribution to verbal STM and indicate that ongoing interactions between semantic and phonological representations are crucial to the ability to maintain a sequence of phonemes verbatim, at least for naturalistic input.

Semantic effects on phonological coherence in STM have been observed before, most clearly in patients with semantic dementia; however, studies of healthy participants have not found convincing evidence for semantic binding effects in

Table 3 Pairwise comparisons for phoneme-level responses in the replication sample, split by lexicality for mixed lists

Paired Comparison	ISR Response			
	Correct	IT-ORD	RECOMB	NON-RECOMB
Nonword phonemes:				
RANDOM MIXED vs. SEM MIXED	-2.80[*] <i>d</i> = -0.40	-1.23, n.s. <i>d</i> = -0.33	3.07^{**} <i>d</i> = 0.42	-1.46, n.s. <i>d</i> = -0.27
Word phonemes:				
RANDOM MIXED vs. SEM MIXED	-5.04^{**} <i>d</i> = -0.86	-1.34, n.s. <i>d</i> = -0.17	3.02^{**} <i>d</i> = 0.52	-0.12, n.s. <i>d</i> = -0.04

All *t* values are shown for each mixed list comparison in the replication experiment, with Cohen's *d* as a measure of effect size. Asterisks denote significance levels, and all comparisons with *p* values less than .05 are further highlighted in bold. ** *p* < .01, * *p* < .05.

STM—with small effect sizes and conflicting conclusions across studies. Consequently, the proposal that semantic information can directly influence the stability of the phonological trace remains highly controversial, especially since the frequent phoneme migration errors produced in the recall of patients with semantic dementia could potentially index neurodegeneration spreading beyond the conceptual system (e.g., a loss of phonological–lexical knowledge that prevents reconstruction of the STM trace; Papagno et al., 2013).

We overcame several methodological limitations in the literature to provide converging evidence in healthy participants consistent with studies of semantic dementia. Namely, we utilized phoneme-level as well as item-level scoring, which allowed us to trace phoneme migrations; we employed mixed lists including both words and nonwords, allowing us to investigate effects of semantic binding across an entire list (i.e., for words and also for nonwords that cannot be reconstructed from lexical knowledge); and, crucially, we presented meaningful story-like sequences, as well as standard lists of unconnected words that are more typically used. Our results suggest that previous research has underemphasized the semantic contribution to phonological coherence. For unrelated words lacking an overarching meaning, STM may draw more strongly on phonological than semantic processes; however, for more naturalistic and meaningful materials, the contribution of semantic information to phoneme binding is increased. Thus, our study demonstrates that the experimental paradigm commonly used in this field systematically underemphasizes the role of meaning in binding phonemes together.³

These observations have clear theoretical implications for our understanding of STM. Strikingly different architectures have been proposed to explain the impact of semantic knowledge on ISR—with one account largely drawing on studies of semantic dementia, and another based on research with healthy volunteers. In line with the strong effects of phonological manipulations in the ISR performance of healthy individuals, redintegration accounts assume that phonological maintenance occurs in isolation from lexical and semantic processing (Hulme, Maughan, & Brown, 1991; Hulme et al., 1997; Schweickert, 1993). The suggestion that semantic knowledge influences recall through the restriction of candidate lexical representations used in the trace reconstruction process (Poirier & Saint-Aubin, 1995; Saint-Aubin & Poirier, 1999) holds that conceptual manipulations should largely influence recall at the level of whole-item accuracy and, in mixed lists, redintegration should be constrained to portions of the phonological trace that contain familiar words. However, neither of these predictions accord with our results. While strategic redintegration is likely to have contributed to performance, particularly when participants could encode which

list positions were words and which were nonwords (Jefferies, Frankish, & Noble, 2009), such mechanisms do not offer a ready account for the enhanced recall of *nonwords* in mixed lists. They also seem unlikely to explain the effect of the semantic coherence of the words on nonword recall. A conceivable indirect, lexically driven explanation of the semantic effect on nonword performance would be if the improvements to nonword recall were a consequence of fewer opportunities to incorrectly assign loose nonword phonemes with word phonemes—in which nonword phonemes available at the point of recall may then be more likely to be recalled correctly (effectively by default). Such an explanation cannot account for the overall semantically related increases in nonword target phonemes recalled that we observed, however. In contrast, our findings are compatible with the view that continual interactions between phonological and semantic representations are fundamental to the maintenance of a sequence of phonemes. According to this view, semantic coherence strengthens the stability of the entire phonological sequence, providing an explanation for why verbatim recall of unfamiliar nonwords is limited to a very small number of items, whereas long sentences can be repeated without error.

Effects of sentence structure and semantic coherence on ISR are well-established (e.g., Brener, 1940; Miller & Selfridge, 1950)—syntactic structures support the reproduction of words in order, and participants are better able to reproduce target words when meaning is constrained by other items. However, our results show that this process of semantic binding goes beyond reconstruction based on gist (cf. Potter & Lombardi, 1990) because conceptual knowledge influences phonological stability at the level of individual phonemes. Thus, these data are highly compatible with the predictions of the semantic binding hypothesis (Patterson et al., 1994) and broader theoretical accounts that couch short-term memory in terms of activations of the underlying language system (e.g., Acheson & MacDonald, 2009; MacDonald, 2016; Majerus, 2013).

The present data do not present irrefutable evidence for the semantic-binding account, however, since there may be alternative explanations for the semantic improvements in phonological stability. The reductions in phoneme recombination errors that we observed generally corresponded to an increase in whole word or whole nonword items correct. Thus, a plausible explanation could be one of resource allocation: People might allocate more attentional resources to nonwords when they are mixed with words than to nonword-only lists, and in such a way that the available attention for nonwords may be further increased when the word memoranda are semantically coherent and easier to encode; this attentionally enhanced encoding and/or maintenance may contribute to their better recall.

Nevertheless, our results allow us to conclude that the availability of a coherent meaning across a sequence of words stabilizes ongoing phonological processing in STM. We cannot determine whether it is specifically the conceptual coherence available from word combinations (e.g., the word “stage” in “gig

³ Our task instructions to produce item attempts even when unsure might have contributed to differences in phonological integrity at the subitem level by discouraging response omissions in cases of uncertainty.

stage” has a more specific meaning in combination than alone), the linguistic co-occurrence of these words that created meaning, or the combined influence of these factors that produced the long-term support driving stronger STM performance. Previous studies have separately manipulated the semantic support for individual words (e.g., the imageability of the items; Acheson, Postle, & MacDonald, 2010; Romani, McAlpine, & Martin, 2008; Tse & Altarriba, 2007; Walker & Hulme, 1999) and the extent to which items co-occur (Stuart & Hulme, 2000) and these factors both influence short-term memory (although these investigations have not examined the stability of phonological processing as in this study). Nevertheless, in more naturalistic language these factors interact and the context in which a word is used strongly constrains its meaning. In the present study we showed that the additional support from long-term message-level meaning was not restricted to the recall of the constituent words but extended to the recall of unfamiliar nonword stimuli when these were embedded within meaningful sequences. This provides strong evidence that support from long-term representations has a dynamic influence over all of the phonological content in STM, since these items cannot otherwise benefit from long-term retrieval strategies (i.e., semantic knowledge helps to stabilize STM at a sub-item level across the whole-trace, and is not purely item-based; more compatible with language-based explanations of STM than redintegration accounts).

One limitation of our recall-based measures is that they do not allow us to infer the stage of processing at which semantic information stabilizes STM; more stable phonological sequencing could manifest at recall through facilitated phonological encoding, strengthened phonological maintenance at rehearsal, or production being guided at recall—or through combined effects across these stages. Combining our task measures with an online measure of STM, such as event-related potentials, could, for example, determine whether nonwords within coherent mixed lists show an encoding advantage (cf. Ruchkin, Grafman, Cameron, & Berndt, 2003).

Our aim was to expand upon existing evidence that semantic effects in short-term memory go beyond lexical representations of individual words. Our resultant demonstrations of semantic binding effects across the overall phonological trace have important real world applications. Our capacity to repeat verbatim long sequences that we hear—especially when some words in the sequence are completely unfamiliar—may support ongoing comprehension, at least in some circumstances. Having a stable representation of the meaning of the words we are planning to say aloud may help us to avoid speech errors in which phoneme segments from one item split off and recombine with other segments (Dell, 1986). Semantic binding mechanisms occurring at the level of a phoneme are also likely to assist the production of complex sentences that span several seconds, and allow us to learn about the appropriate use of words in context, both in our own language and in the process of learning a new one (e.g., Daneman & Green, 1986). These effects are likely to have a

larger influence on real-world language tasks than has hitherto been appreciated.

References

- Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin*, *135*, 50–68. doi:<https://doi.org/10.1037/a0014411>.
- Acheson, D. J., Postle, B. R., & MacDonald, M. C. (2010). The interaction of concreteness and phonological similarity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 17–36. doi:<https://doi.org/10.1037/a0017679>
- Baddeley, A. D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology*, *18*, 302–309. doi:<https://doi.org/10.1080/14640746608400047>
- Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Clarendon Press.
- Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29. doi:<https://doi.org/10.1146/annurev-psych-120710-100422>
- Benetello, A., Cecchetto, C., & Papagno, C. (2015). When meaning is useless. *Memory*, *23*, 1001–1012. doi:<https://doi.org/10.1080/09658211.2014.945939>
- Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology*, *26*, 467–482. doi:<https://doi.org/10.1037/h0061096>
- Cortese, M. J. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers*, *36*, 384–387. doi:<https://doi.org/10.3758/BF03195585>
- Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson’s method. *Tutorials in Quantitative Methods for Psychology*, *1*, 42–45. doi:<https://doi.org/10.20982/tqmp.01.1.p042>
- Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*, 1–18. doi:[https://doi.org/10.1016/0749-596X\(86\)90018-5](https://doi.org/10.1016/0749-596X(86)90018-5)
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321. doi:<https://doi.org/10.1037/0033-295X.93.3.283>
- Ellis, A. W. (1980). Errors in speech and short-term memory: The effects of phonemic similarity and syllable position. *Journal of Verbal Learning and Verbal Behavior*, *19*, 624–634. doi:[https://doi.org/10.1016/S0022-5371\(80\)90672-6](https://doi.org/10.1016/S0022-5371(80)90672-6)
- Hoffman, P., Jefferies, E., Ehsan, S., Jones, R. W., & Lambon Ralph, M. A. (2009). Semantic memory is key to binding phonology: Converging evidence from immediate serial recall in semantic dementia and healthy participants. *Neuropsychologia*, *47*, 747–760. doi:<https://doi.org/10.1016/j.neuropsychologia.2008.12.001>
- Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, *30*, 685–701. doi:[https://doi.org/10.1016/0749-596X\(91\)90032-F](https://doi.org/10.1016/0749-596X(91)90032-F)
- Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1217–1232. doi:<https://doi.org/10.1037/0278-7393.23.5.1217>
- Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in*

- Cognitive Sciences*, 10, 480–486. doi:<https://doi.org/10.1016/j.tics.2006.09.002>
- Jefferies, E., Crisp, J., & Lambon Ralph, M. A. (2006). The impact of phonological or semantic impairment on delayed auditory repetition: Evidence from stroke aphasia and semantic dementia. *Aphasiology*, 20, 963–992. doi:<https://doi.org/10.1080/02687030600739398>
- Jefferies, E., Frankish, C. R., & Lambon Ralph, M. A. (2006). Lexical and semantic binding in verbal short-term memory. *Journal of Memory and Language*, 54, 81–98. doi:<https://doi.org/10.1016/j.jml.2005.08.001>
- Jefferies, E., Frankish, C. R., & Noble, K. (2009). Lexical coherence in short-term memory: Strategic reconstruction or “semantic glue”? *Quarterly Journal of Experimental Psychology*, 62, 1967–1982. doi:<https://doi.org/10.1080/17470210802697672>
- Jefferies, E., Hoffman, P., Jones, R., & Lambon Ralph, M. A. (2008). The impact of semantic impairment on verbal short-term memory in stroke aphasia and semantic dementia: A comparative study. *Journal of Memory and Language*, 58, 66–87. doi:<https://doi.org/10.1016/j.jml.2007.06.004>
- Jefferies, E., Jones, R. W., Bateman, D., & Lambon Ralph, M. A. (2005). A semantic contribution to nonword recall? Evidence for intact phonological processes in semantic dementia. *Cognitive Neuropsychology*, 22, 183–212. doi:<https://doi.org/10.1080/02643290442000068>
- Jefferies, E., Lambon Ralph, M. A., & Baddeley, A. D. (2004). Automatic and controlled processing in sentence recall: The role of long-term and working memory. *Journal of Memory and Language*, 51, 623–643. doi:<https://doi.org/10.1016/j.jml.2004.07.005>
- Knott, R., Patterson, K. E., & Hodges, J. (1997). Lexical and semantic binding effects in short-term memory: Evidence from semantic dementia. *Cognitive Neuropsychology*, 14, 1165–1216. doi:<https://doi.org/10.1080/026432997381303>
- MacDonald, M. C. (2016). Speak, act, remember: The language-production basis of serial order and maintenance in verbal memory. *Current Directions in Psychological Science*, 25, 47–53. doi:<https://doi.org/10.1177/0963721415620776>
- Majerus, S. (2013). Language repetition and short-term memory: An integrative framework. *Frontiers in Human Neuroscience*, 7, 357. doi:<https://doi.org/10.3389/fnhum.2013.00357>
- Majerus, S., Norris, D. G., & Patterson, K. E. (2007). What does a patient with semantic dementia remember in verbal short-term memory? Order and sound but not words. *Cognitive Neuropsychology*, 24, 131–151. doi:<https://doi.org/10.1080/02643290600989376>
- Majerus, S., & van der Linden, M. (2003). Long-term memory effects on verbal short-term memory: A replication study. *British Journal of Developmental Psychology*, 21, 303–310. doi:<https://doi.org/10.1348/026151003765264101>
- Martin, N., & Gupta, P. (2004). Processing and verbal short-term memory: Evidence from associations and dissociations. *Cognitive Neuropsychology*, 21, 213–228. doi:<https://doi.org/10.1080/02643290342000447>
- Martin, R. C., Lesch, M. F., Bartha, M. C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*, 41, 3–29 <https://doi.org/10.1006/jmla.1999.2637>
- Miller, G. A., & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *American Journal of Psychology*, 63, 176–185. doi:<https://doi.org/10.2307/1418920>
- Papagno, C., Vernice, M., & Cecchetto, C. (2013). Phonology without semantics? Good enough for verbal short-term memory: Evidence from a patient with semantic dementia. *Cortex*, 49, 626–636. doi:<https://doi.org/10.1016/j.cortex.2012.04.015>
- Patterson, K. E., Graham, N., & Hodges, J. R. (1994). The impact of semantic memory loss on phonological representations. *Journal of Cognitive Neuroscience*, 6, 57–69. doi:<https://doi.org/10.1162/jocn.1994.6.1.57>
- Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *Quarterly Journal of Experimental Psychology*, 48A, 384–404. doi:<https://doi.org/10.1080/14640749508401396>
- Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, 29, 633–654. doi:[https://doi.org/10.1016/0749-596X\(90\)90042-X](https://doi.org/10.1016/0749-596X(90)90042-X)
- Romani, C., McAlpine, S., & Martin, R. C. (2008). Concreteness effects in different tasks: Implications for models of short-term memory. *Quarterly Journal of Experimental Psychology*, 61, 292–323. doi:<https://doi.org/10.1080/17470210601147747>
- Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain Sciences*, 26, 709–728, disc. 728–777. doi:<https://doi.org/10.1017/S0140525X03000165>
- Saint-Aubin, J., & Poirier, M. (1999). The influence of long-term memory factors on immediate serial recall: An item and order analysis. *International Journal of Psychology*, 34, 347–352. doi:<https://doi.org/10.1080/002075999399675>
- Savill, N., Ellis, A. W., & Jefferies, E. (2017). Newly-acquired words are more phonologically robust in verbal short-term memory when they have associated semantic representations. *Neuropsychologia*, 98, 85–97. doi:<https://doi.org/10.1016/j.neuropsychologia.2016.03.006>
- Savill, N., Metcalfe, T., Ellis, A. W., & Jefferies, E. (2015). Semantic categorisation of a word supports its phonological integrity in verbal short-term memory. *Journal of Memory and Language*, 84, 128–138. doi:<https://doi.org/10.1016/j.jml.2015.06.003>
- Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration. *Memory & Cognition*, 21, 168–175. doi:<https://doi.org/10.3758/BF03202729>
- Stuart, G., & Hulme, C. (2000). The effects of word co-occurrence on short-term memory: Associative links in long-term memory affect short-term memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 796–802. doi:<https://doi.org/10.1037/0278-7393.26.3.796>
- Tse, C.-S., & Altarriba, J. (2007). Testing the associative-link hypothesis in immediate serial recall: Evidence from word frequency and word imageability effects. *Memory*, 15, 675–90. doi:<https://doi.org/10.1080/09658210701467186>
- Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, 67, 1176–1190. doi:<https://doi.org/10.1080/17470218.2013.850521>
- Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 1256–1271. doi:<https://doi.org/10.1037/0278-7393.25.5.1256>

Author note

This study was supported by a European Research Council grant (SEMBIND-283530). We thank James Davey for research assistance.