Est. 1841 | YORK ST JOHN UNIVERSITY

Savill, Nicola ORCID: https://orcid.org/0000-0002-6854-0658, Ellis, Rachel, Brooke, Emma, Koa, Tiffany, Ferguson, Suzie, Rojas-Rodriguez, Elena, Arnold, Dominic, Smallwood, Jonathan and Jefferies, Elizabeth (2018) Keeping It Together: Semantic Coherence Stabilizes Phonological Sequences in Short-Term Memory. Memory & Cognition, 46 (3). pp. 426-437.

Downloaded from: http://ray.yorksj.ac.uk/id/eprint/2681/

The version presented here may differ from the published version or version of record. If you intend to cite from the work you are advised to consult the publisher's version: https://doi.org/10.3758/s13421-017-0775-3

Research at York St John (RaY) is an institutional repository. It supports the principles of open access by making the research outputs of the University available in digital form. Copyright of the items stored in RaY reside with the authors and/or other copyright owners. Users may access full text items free of charge, and may download a copy for private study or non-commercial research. For further reuse terms, see licence terms governing individual outputs. Institutional Repository Policy Statement

RaY

Research at the University of York St John

For more information please contact RaY at ray@yorksj.ac.uk

This is a postprint of a paper to appear in Memory & Cognition

DOI: 10.3758/s13421-017-0775-3

Keeping It Together: Semantic Coherence Stabilises Phonological Sequences in Short-Term Memory

Nicola Savill[1, 2], Rachel Ellis[2], Emma Brooke[2], Tiffany Koa[2], Suzie Ferguson[2], Elena Rojas-Rodriguez[2], Dominic Arnold[2], Jonathan Smallwood[2] & Elizabeth Jefferies[2]

[1]*School of Psychological & Social Sciences, York St John University*

[2]*Department of Psychology, University of York*

Corresponding Author:

Nicola Savill

School of Psychological & Social Sciences,

York St John University,

Lord Mayor's Walk,

York

YO31 7EX

UK

Email: n.savill@yorksj.ac.uk

Tel: +44 (0)1904 876171

Keywords: meaning, phonological binding, verbal short-term memory, semantic coherence, speech

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 found modest and inconsistent benefits of meaning on phonological stability in short-term memory but these studies presented sets of unrelated words. Using a novel design, we examined the immediate recall of sentence-like sequences with coherent meaning, alongside 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, 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, 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.

Introduction

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 contribution of phonological and semantic representations to verbal STM, and the ways in which they interact, is unclear. Semantic manipulations typically have relatively subtle effects on STM compared to phonological manipulations (e.g., Baddeley, 1966; Jefferies, Frankish, & Lambon Ralph, 2006; Majerus & van der Linden, 2003; Savill, Ellis, & Jefferies, 2017) and 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 is 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 which posit direct involvement of lexical-semantic processing/knowledge in STM capacities (e.g., Acheson & MacDonald, 2009; R.C. Martin, Lesch & Bartha, 1999; N. Martin & Gupta, 2004; Knott, Patterson & Hodges, 1997). 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, Graham, & Hodges, 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, characterised 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, Patterson, & Hodges, 1997; Majerus, Norris, & Patterson, 2007; Patterson, Graham, & Hodges, 1994). Although this evidence emphasises the importance of semantic knowledge in STM that 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 – i.e., semantic binding – have produced weak and inconsistent effects, especially when lexical-phonological knowledge is controlled. Jefferies, Frankish and Lambon Ralph (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. While word frequency influenced migrations, there was no clear effect of concreteness (a semantic variable) 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 non-semantic 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 underemphasise the importance of semantic information in maintenance processes in STM and phonological binding for several reasons: (i) studies typically examine performance at the level of whole items, and thus cannot directly examine changes in phoneme migration errors; (ii) it is difficult to experimentally separate effects of conceptual knowledge on the stability of the entire phonological trace from item reconstruction (redintegration); and (iii) studies often use lists of words that are not inherently meaningful and that consequently minimise any advantages that may 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 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 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 provides 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 aged 18-31 years (main experiment M = 21.36, SD = 1.73; replication sample M = 21.13, SD = 3.21), having volunteered and given their informed consent. Twenty-eight participants took part in the main study. This sample size was based on 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 /sneɪz/ sheets /drʌk/ bed /maɪg/"*) (nonwords are indicated with slashes and written using the International Phonetic Alphabet); (4) mixed lists of

words and nonwords, where the words were unrelated (RANDOM MIXED; e.g., *"beat /mɪp/ flag coin /truːk/ /tʃel/"*) and (5) meaningless nonword sequences (NONWORD; e.g., *"/fæmp/ /θiːnd/ /peɪp/ /laːz/ /grɪŋk/ /saʊt/"*). The replication sample was tested on the 26 RANDOM MIXED and 26 SEM MIXED trials.

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 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). 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, Frankish, & Lambon Ralph (2006). We did not manipulate the predictability of word and nonword locations in these lists (unlike Jefferies & Frankish, 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 stimuli Set A was a SEM MIXED trial in Set B, and vice versa, in order 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. Nonwords for both NONWORD lists and RANDOM MIXED lists were created by recombining the phonemes from the words in RANDOM WORD and words from other RANDOM MIXED lists respectively (and were therefore matched for syllable structure). Additional detail regarding stimuli 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, nonwords, and words and nonwords mixed together (or 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 produce item attempts whenever possible, even if unsure. 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 minutes to complete, including short rest breaks every 26 trials, and five initial practice trials per condition (20 minutes for the replication, including a rest break halfway and two practice trials at the start). Responses were digitally recorded 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 allow comparison with most other studies of verbal recall. We used a response-based coding approach (rather than target-based) in order to capture potentially

relevant errors that were not phonologically-related to the target (e.g., semantically-related errors). The eight types of response were identified as follows: (1) We coded items correctly recalled in position. (2) Item order errors were target items produced in an incorrect list position. Non-target responses containing target phonemes (in the same syllable position), were classified as either (3) phoneme recombination errors, when response phonemes originated from *more than one* target item in the list, or (4) non-recombination yet phonologically-related responses, which were partially correct but only contained 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) an item intrusion from one of the previous six lists, or (8) unrelated. These data were expressed as a percentage of 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 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 analysed. 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; this includes the results of 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 SEM MIXED and RANDOM MIXED conditions with 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 words and 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 categorise 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 non-recombination 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 total word and nonword target phonemes. For nonwords, we report paired sample *t*-tests (Bonferroni-corrected for each comparison with α=.025) comparing (i) nonword recall in RANDOM MIXED lists vs. pure NONWORD lists (to assess the effect of nonwords being presented alongside words) and (ii) nonword recall in SEM MIXED lists vs. 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

[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.

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 α=.017), examining (i) word recall in RANDOM MIXED and pure WORD lists (to assess the effect on random words of being presented alongside nonwords), (ii) word recall in SEM WORD compared with SEM MIXED lists (to assess the effect on semantically-coherent words of being presented alongside nonwords) and (iii) word recall in SEM MIXED lists vs. 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 vs. SEM MIXED lists, and word phoneme recall in RANDOM MIXED lists vs. 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 phonemes, nonword phonemes), i.e., 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 mixed lists tested in both experiments. This analysis included a within-subjects factor of semantic coherence (RANDOM MIXED vs. SEM MIXED) and a between-subjects factor of experiment (MAIN EXP. vs. REPLICATION), assessing responses both at the item and phoneme levels. This analysis is reported in full in the Supplementary Material (Table S3).

Results

Item-level responses

Fig. 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 for the main experiment.

Accuracy: There were 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; 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 presented alongside words than when presented with other nonwords, while words were recalled more poorly in mixed vs. pure word lists[2]. Moreover, in the case 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 mixed lists [RANDOM WORD > RANDOM MIXED] and in mixed lists than 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, Frankish, et al., 2006). The effect of list composition on item order errors for semantically coherent items did

[2] Analyses demonstrating that these mixing effects were not accounted for by differences in the number of nonword targets in the lists can be found in the Supplementary Materials.

not survive correction [SEM WORD ≈ SEM MIXED]. There was no clear effect 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 the pattern as accuracy (but in the opposite direction), with significant changes between *all* conditions (Panel b in Fig. 1, Table 1). Fewest recombination errors were produced for SEM WORD, followed by RANDOM WORD, SEM MIXED, and RANDOM MIXED, with most phoneme recombinations in the NONWORD condition. Thus, 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 non-recombination errors and unrelated errors followed a similar pattern to 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).

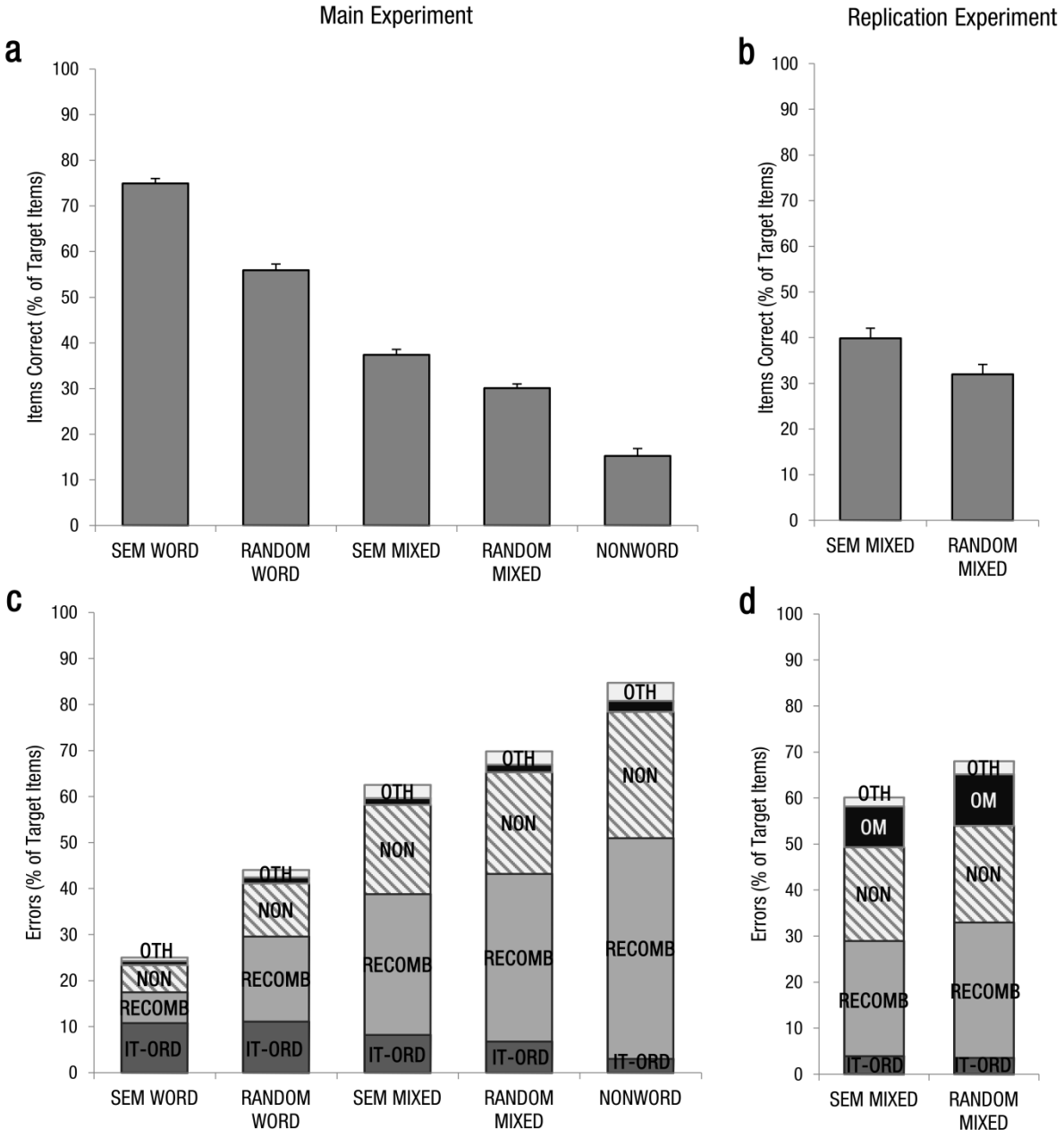

Fig. 1.

Item-level response coding in each condition in the main experiment (left) and in the subsequent replication (right). The upper panels show the percentage of items correct in position for each condition (Panel a = main Experiment; Panel b = replication). Error bars are 95% confidence intervals for a within-subject design (Cousineau, 2005). The lower panels show the proportion 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 unlabelled response category marked in black corresponds to omissions.

Table 1.

Item-level response categories and pairwise comparisons in the Main Experiment

ISR Resp.	*ANOVA*	Bonferroni-corrected pairwise comparisons				
	Main effect of List condition	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	***F=453.13***	**-13.10*** **	**17.54*** **	**24.96*** **	**-10.80*** **	**-6.31*** **
	p<.001					
	η_p^2=.95	***d*=-1.49**	***d*=2.09**	***d*=3.76**	***d*=-1.61**	***d*=-0.67**
IT-ORD	***F=28.20***	**-4.91*** **	**5.74*** **	2.13, *ns*,	0.87, *ns*,	-1.90, *ns*,
	p<.001					
	η_p^2=.52	***d*=-1.25**	***d*=0.94**	d=0.50	d=0.20	d=-0.40
RECOMB	***F=364.06***	**11.80*** **	**-17.84*** **	**-29.07** **	**10.54*** **	**5.14*** **
	p<.001					
	η_p^2=.93	***d*=1.82**	***d*=-2.49**	***d*=-4.87**	***d*=1.97**	***d*=0.94**
NON-RECOMB	***F=217.31***	**6.65*** **	**-14.09*** **	**-17.37*** **	**7.37*** **	2.58, *ns*,
	p<.001					
	η_p^2=.89	***d*=0.93**	***d*=-1.93**	***d*=-3.30**	***d*=1.25**	*d*=0.48
OM	*F=2.79*	1.66, *ns*,	-1.78, *ns*,	-0.95, *ns*,	0.93, *ns*,	0.80, *ns*,
	p=.070					
	η_p^2=.10	*d*=0.19	*d*=-0.18	*d*=-0.22	*d*=0.19	*d*=0.10
UNR	***F=25.95***	**4.42*** **	**-3.49** **	**-5.80*** **	**3.16***	-0.71, *ns*,
	p<.001					
	η_p^2=.50	**d=0.69**	**d=-0.64**	**d=-1.45**	**d=0.69**	d=-0.14

Note. The main effects of list condition for each item-level response type are shown in the left-hand column in italics with partial eta-squared estimates of effect size. *t* values are shown for each planned comparison with asterisks to denote Bonferroni-corrected significance levels and all comparisons with corrected *p* values of less than .1 are further highlighted in bold. Beneath the *t* values are 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 analysed. *** corrected $p < .001$, **corrected $p < .01$, *corrected $p < .05$.

Replication Sample: Despite some differences in the overall frequency 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 ≈ SEM MIXED $t(27) = -0.84$, $p = .41$, $d = -0.14$; Recombination errors: RANDOM MIXED > SEM MIXED, $t(27) = 5.13$, $p < .001$, $d = 0.70$; Non-recombination phonological errors: RANDOM MIXED ≈ 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 compared 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 task structure and participant group tested, the size of the semantic effect did not differ for any phonologically related response type between the main experiment and replication (items correct-in-position, item order errors, recombination errors, and non-recombination phonological errors similarly modulated; i.e., null interactions between task and semantic coherence); only rates of omissions and unrelated responses scaled differently (see Supplementary Material and Fig. 1.)

Phoneme-level responses

Fig. 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 outcome 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

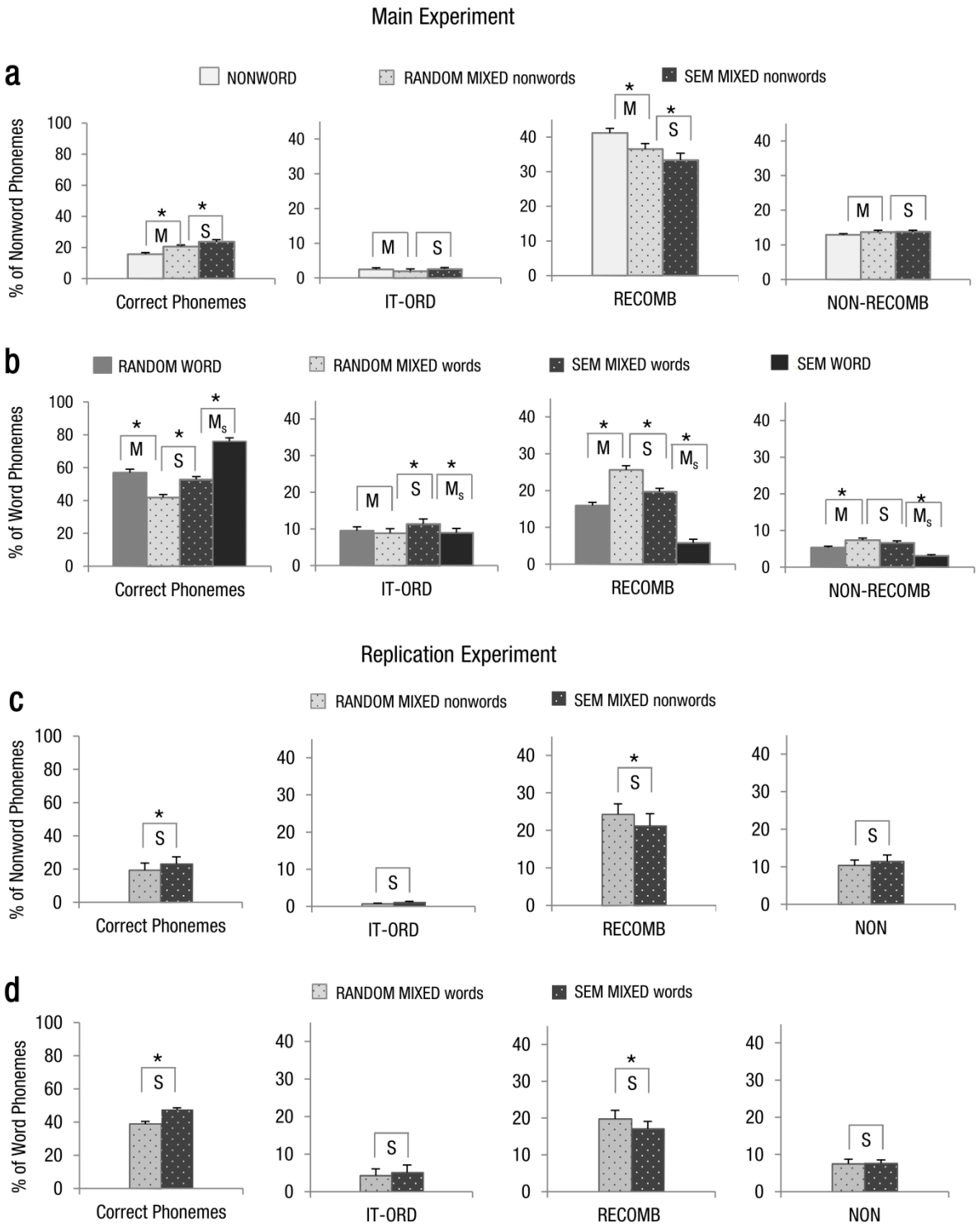

Fig. 2.

Phoneme-level responses in the main experiment and replication experiment. Responses that corresponded with *nonword* target phonemes (a: main experiment; c: replication) are shown above responses corresponding with *word* target phonemes (b: main experiment; d: replication), each split by response type and expressed as a percentage of 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, labelled M) and the effect of semantic coherence (labelled S) are shown. Since not

all target phonemes were produced in the response, the bars for each condition do not total 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 vs. semantically coherent pure lists; S = semantically coherent mixed lists vs. random mixed lists. * denotes significantly different comparisons (p<.05 corrected). Errors bars are 95% confidence intervals for a within-subject design (Cousineau, 2005).

[RANDOM MIXED > NONWORD]. Phonemes from nonwords in mixed lists were also more frequently recalled 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 as opposed to 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; Fig. 2.]. This pattern shows that semantic support from the words affected the stability of the complete phonological trace.

There were no differences between conditions in the rate of nonword phonemes produced as part of whole target items out-of-sequence or as part of partially incorrect non-recombination 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].

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*** **	**-2.88** **	**9.33*** **	**15.41*** **	**-7.20*** **
	***d*=-0.50**	***d*=0.29**	***d*=1.13**	***d*=2.16**	***d*=-0.89**
IT-ORD	0.96, *ns*	-1.35, *ns*	0.81, *ns*	**-2.39***	**-2.39***
	d=0.27	*d*=-0.34	*d*=0.15	***d*=-0.59**	***d*=-0.53**
RECOMB	**5.49*** **	**2.41***	**-10.03*** **	**-18.02*** **	**4.54*** **
	***d*=0.86**	***d*=0.57**	***d*=-1.45**	***d*=-3.19**	***d*=0.97**
NON-RECOMB	-1.61, *ns*	-0.09, *ns*	**-4.77*** **	**-9.08*** **	1.97, *ns*
	d=-0.34	*d*=-0.02	***d*=-0.86**	***d*=-1.69**	*d*=0.43

Note. 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 of less than .05 are further highlighted in bold. Beneath the *t* values are respective Cohen's *d* measures of effect size. *** corrected $p < .001$, **corrected $p < .01$, *corrected $p < .05$. † corrected $p < .1$.

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 ≈ 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 non-recombination errors were relatively frequent for mixed lists [RANDOM MIXED words > RANDOM WORD and SEM MIXED words > SEM WORD]. However, word phoneme non-recombination errors did not vary according to the semantic coherence of the mixed lists [RANDOM MIXED words ≈ SEM MIXED words].

Replication Sample: Table 3 reports the outcome of statistical tests for RANDOM MIXED vs. 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; 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 ≈ SEM MIXED nonwords] or as part of partially incorrect non-recombination responses [Non-recombination errors: RANDOM MIXED nonwords ≈ SEM MIXED nonwords; see Table 3).

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	**-2.80***	-1.23, *ns*	**3.07****	-1.46, ns
vs. SEM MIXED	***d*=-0.40**	*d*=-0.33	***d*=0.42**	*d*=-0.27
Word phonemes:				
RANDOM MIXED	**-5.04****	-1.34, *ns*	**3.02****	-0.12, *ns*
vs. SEM MIXED	***d*=-0.86**	*d*=-0.17	***d*=0.52**	*d*=-0.04

Note. 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 of less than .05 are further highlighted in bold. *** $p < .001$, ** $p < .01$, * $p < .05$.

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 phonemes, nonword phonemes) (see Supplementary Materials). These analyses confirmed that, while semantic influences on recall accuracy were stronger for words than nonwords [interactions of semantic manipulation and lexicality], the semantic effects of nonword recall were not carried by 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 effect on each response type was similar between the sets, and showed that semantic coherence improves the *overall* recall of word and nonword target phonemes (details in Supplementary Materials, Table S3).

Discussion

We have 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 stabilising effect of semantic coherence on phoneme order extends to meaningless nonwords when these are mixed with 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 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 utilised 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 under-emphasised the semantic contribution to phonological coherence. For unrelated words lacking an over-arching 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 under-emphasises 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 et al., 1997; Hulme, Maughan, & Brown, Gordon, 1991; 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

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

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 – where 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. By 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, while 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 are not 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 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 compared 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 stabilises ongoing phonological processing in STM. We cannot determine if it is specifically the conceptual coherence available from word combinations (e.g., the word '*stage*' in '*gig 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 comprised 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 current 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 stabilise 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 stabilises STM; more stable phonological sequencing could manifest at recall through facilitated phonological encoding, strengthened phonological maintenance at rehearsal or through guiding production at recall – or across these stages. Combining our task measures with an online measure of STM, such as event-related potentials, could for example determine if 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.

Acknowledgements

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

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: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:10.1037/a0017679

Baddeley, A. D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *The Quarterly Journal of Experimental Psychology*, *18*, 302–309. doi: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:10.1146/annurev-psych-120710-100422

Benetello, A., Cecchetto, C., & Papagno, C. (2015). When meaning is useless. *Memory*, *23*, 1001–1012. doi:10.1080/09658211.2014.945939

Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology*, *26*, 467–482. doi:10.1037/h0061096

Cortese, M. J. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers*, *36*, 384–387. doi: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: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:10.1016/0749-596X(86)90018-5

Dell, G. S. (1986). A spreading-activation theory of retreival in sentence production. *Psychological Review*, *93*, 283–321. doi: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: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:10.1016/j.neuropsychologia.2008.12.001

Hulme, C., Maughan, S., & Brown, Gordon, 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:10.1016/0749-596X(91)90032-F

Hulme, C., Roodenrys, S., Schweickert, R., Brown, Gordon, 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: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: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: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: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: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: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: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: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: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:10.1177/0963721415620776

Majerus, S. (2013). Language repetition and short-term memory: an integrative framework. *Frontiers in Human Neuroscience, 7*, 357. doi: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: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: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: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 & Language, 41,* 3–29 10.1006/jmla.1999.2637

Miller, G. A., & Selfridge, J. A. (1950). Verbal Context and the Recall of Meaningful Material. *The American Journal of Psychology, 63,* 176–185. doi: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: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: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. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 48,* 384–404. doi: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: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: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. *The Behavioral and Brain Sciences, 26,* 709-28-77. doi: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: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: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: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:10.3758/BF03202729

Stuart, G., & Hulme, C. (2000). The effects of word co-occurance 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: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: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: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:10.1037/0278-7393.25.5.1256

Supplementary Material

Table S1.

Matching of psycholinguistic variables across conditions

		SEM WORD		SEM MIXED		RANDOM MIXED	RANDOM WORD	NONWORD
		Set A	Set B	Set A	Set B	Sets A & B	Sets A & B	Sets A & B
Word Items								
Lexical Frequency	*M*	4.63	4.58	4.37	4.47	4.45	4.60	N/A
	SD	0.85	0.86	1.68	1.52	0.74	0.77	
Imageability	*M*	4.86	4.85	5.14	5.12	5.16	4.80	N/A
	SD	0.80	0.81	1.68	1.52	1.28	1.37	
Coherence Rating	*M*	5.91	5.88	6.23	6.15	Not Collected	Not Collected	N/A
	SD	0.75	0.74	1.22	1.11			
Affect Rating	*M*	2.92	2.49	2.24	2.58	Not Collected	Not Collected	N/A
	SD	0.80	0.82	1.35	1.25			
Grammatical class								
Noun[a]	%	78.85	80.77	85.90	85.90	82.05	84.62	N/A
Verb	%	11.54	8.97	8.97	10.26	10.26	8.33	N/A
Adjective	%	7.69	8.33	5.13	3.85	7.69	6.41	N/A
Other	%	1.92	1.92	0.00	0.00	0.00	0.64	N/A
All items								
Item phonemes	*M*	3.27	3.33	3.32	3.27	3.28	3.31	3.31
	SD	0.51	0.53	0.53	0.51	0.53	0.57	0.57
List Biphon.Prob.	*M*	0.009	0.010	0.010	0.009	0.009	0.010	0.009
	SD	0.003	0.003	0.003	0.003	0.002	0.003	0.003

Note. [a] includes nouns that may also fall into other word classes (like 'watch' and 'light'). Lexical frequency refers to Zipf values taken from SUBTLEX-UK (Van Heuven et al., 2014; these values are on a scale from 1 to 7 where 1=lowest frequency and 7=highest frequency) and Imageability refers to values taken from Cortese (2004; these values correspond to ratings of monosyllabic words on a scale from 1 to 7 where 1=lowest imageability rating and 7=highest imageability rating). The sequences were rated for semantic coherence and affect on scales of 1 to 7 (semantic coherence: 1=not coherent, 7=very coherent; affect: 1=not emotional; 7=very emotional). List Biphon.Prob.= Summed biphone probabilities averaged by list.

Additional details on stimuli construction

All list items contained between two phonemes (CVs where C=consonant, V=vowel; e.g., 'tea' /ti/) and five phonemes (CCVCCs and CCCVCs e.g., 'stunt' /stʌnt/ and 'scream' /skrim/) and over 98% of the individual items were unique across lists (the few that were repeated were presented no more than twice across all lists). These stimuli were recorded by a female British English speaker and each item sound file was edited in Praat to 0.75 seconds in length.

ISR lists were arranged pseudo-randomly into four test versions in the main experiment, which together accommodated the two sets of SEM WORD and SEM MIXED trials and allowed counterbalancing of two opposite trial orders. The lists' average phonotactic properties did not differ between conditions in any version (summed phoneme and biphone positional probabilities, see Vitevitch & Luce, 2004).

Response coding details

When fewer than six responses were given on a trial, whole item omissions were positioned within the transcript in a way that minimised the error score (for example, if five responses were produced that largely corresponded with the second through to the sixth target items, the omission would be placed in the first response position). Transcriptions used CELEX DISC notation (Baayen, Piepenbrock, & van Rijn, 1995). Responses were categorised from the transcription at item and phoneme levels using a version of the item coding scheme reported in Savill et al. (2015), adapted to accommodate non-CVC monosyllables. That is, similar to the methods in Savill et al. (2015), a phoneme response was considered a target phoneme out of position if it was not in the correct serial position in the list but corresponded to a target phoneme in the same *relative syllable position*, i.e., when both target and response phoneme were at the onset (i.e., any consonant in an

onset cluster), vowel nucleus, or coda positions (i.e., any consonant in a final cluster) of the syllable. A worked example of the coding of a single trial is provided in Table S2.

Table S2.

Example coding for a single trial.

Example target list	"teen, quorl, shop, glack, dress, mim"
Target list phonetic transcription[a]	tin kwɔl ʃɒp glæk drɛs mɪm
Lexicality of target items	word, nonword, word, nonword, word, nonword
Example verbal recall response	"teen, bowl, vard, meck, dress, shop"
Response phonetic transcription [a] (where bold=target item, green font=phoneme in correct position, red font=migrated phoneme, black font=non-match)	**tin** bəʊl vɑd mɛk **drɛs ʃɒp**
Item response coding	CIP, NON-RECOMB, UNR, RECOMB, CIP, IT-ORD = 2 CIP, 1 IT-ORD, 1 RECOMB, 1 NON-RECOMB & 1 UNR
Tracing lexicality of recalled target phonemes	Correct (CIP) items=7 word phonemes, 0 nonword phonemes IT-ORD errors=3 word phonemes, 0 nonword phonemes RECOMB errors=1 word phoneme (repeated and out of position), 2 nonword phonemes (1 correct and 1 out of position) NON-RECOMB errors=0 word phonemes, 1 nonword phonemes (in correct position)

Note. [a] Transcription are shown using the International Phonetic Alphabet for illustration only. Transcriptions used CELEX DISC notation. Key: CIP: Item in correct position. 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

Transcriptions were completed by five coders while a sixth independently second-coded two data sets transcribed by each coder to assess inter-rater reliability. Reliability ranged from Kappa=0.78 (p <.001), 95% CI (0.735, 0.833) to Kappa=0.88 (p <.001), 95% CI (0.843, 0.917). Following Landis & Koch (1977), this would classify four of the coders (with Kappa > 0.81) as providing 'almost perfect' agreement and the poorest as providing 'substantial agreement' (statistical outcomes remained the same excluding datasets coded by the least consistent coder). For the replication sample, two coders transcribed responses, and a data set was second-coded to assess inter-rater reliability, producing a Kappa rating of 0.85 (p <.001), 95% CI (0.806, 0.900).

Comparison of mixed list performance in the Main Experiment and Replication Experiment:

As might be expected from testing separate sets of participants in different task contexts, the two groups differed in terms of their overall tendencies to produce certain types of errors: Specifically, the participants in the Main Experiment tended to produce more recombination errors (RECOMB) and order errors (IT-ORD) than those in the follow-up experiment but fewer omission errors (OM) (i.e., significant between-group effects; Table S3). Despite these gross differences between tasks, the effects of semantic coherence were similar between experiments. The only differences between results, i.e., interactions between semantic coherence and participant group, emerged for item omissions and unrelated item errors. This reflected a greater overall tendency of the participants in the replication sample to produce omission errors, which in turn showed sensitivity to semantic coherence that did not emerge in the Main Experiment. Whereas unrelated errors were overall reduced in the replication sample, and fewer in the SEM MIXED than the RANDOM MIXED condition.

At the phoneme level, the participants in the Main Experiment tended to be more successful at correctly recalling nonwords than the mixed list-only participants but they also produced more nonword recombination and nonword non-recombination errors, and fewer nonword order errors.

These same participants also tended to produce more word order errors and recombination errors. Importantly however, there were no significant semantic coherence by group interactions when responses were broken down according to whether the target was a word or nonword (Table S3).

Table S3.

Outcome of Mixed ANOVAs comparing SEM MIXED and RANDOM MIXED responses in the Main Experiment and the Replication (shown at the item level and when split by the lexicality of target phonemes)

ISR Resp.	Coherence Effect	Participant Group/Task Effect	Coherence × Group/Task
Item Level			
Correct	F=74.88, p<.001, η_p^2=.60	*ns*	*ns*
IT-ORD	F=3.99, p=.051, η_p^2=.08	F=19.66, p<.001, η_p^2=.29	*ns*
RECOMB	F=49.96, p<.001, η_p^2=.51	F=12.34, p<.001, η_p^2=.20	*ns*
NON-RECOMB	F=5.18, p=.027, η_p^2=.10	*ns*	*ns*
OM	F=6.73, p=.012, η_p^2=.12	F=12.23, p<.001, η_p^2=.20	F=4.09, p=.049, η_p^2=.08
UNR	*ns*	*ns*	F=4.61, p=.037, η_p^2=.09
Phoneme Level			
Nonword Phonemes recalled:			
Correct	F=17.90, p<.001, η_p^2=.27	F=36.79, p<.001, η_p^2=.43	*ns*
IT-ORD	*ns*	F=5.28, p=.03, η_p^2=.10	*ns*
RECOMB	F=13.77, p=.001, η_p^2=.22	F=53.91, p<.001, η_p^2=.52	*ns*
NON-RECOMB	*ns*	F=10.40, p=.002, η_p^2=.18	*ns*
Word Phonemes recalled:			
Correct	F=68.61, p<.001, η_p^2=.58	*ns*	*ns*
IT-ORD	F=7.32, p=.009, η_p^2=.13	F=23.27, p<.001, η_p^2=.32	*ns*
RECOMB	F=28.30, p<.001, η_p^2=.37	F=8.72, p=.005, η_p^2=.15	*ns*
NON-RECOMB	*ns*	*ns*	*ns*

Note. df = (1, 49). *ns* = non-significant effects (p>.05). η_p^2 = Partial eta-squared estimates of effect size

Follow-up analyses interrogating the mixing effects on nonword phonemes

We ran several different analyses to interrogate the improvements in nonword recall in the mixed lists.

First, to address whether the general benefit for nonword recall when presented with words in mixed lists (i.e., compared to nonword-only lists) could be explained by fewer opportunities for phonemes migrations between the nonword positions in the mixed lists than the NONWORD condition (i.e., due to fewer nonword items, which would have meant that any migrating phonemes had to travel further), we re-examined the relevant data in the main experiment (the replication did not contain the relevant conditions).

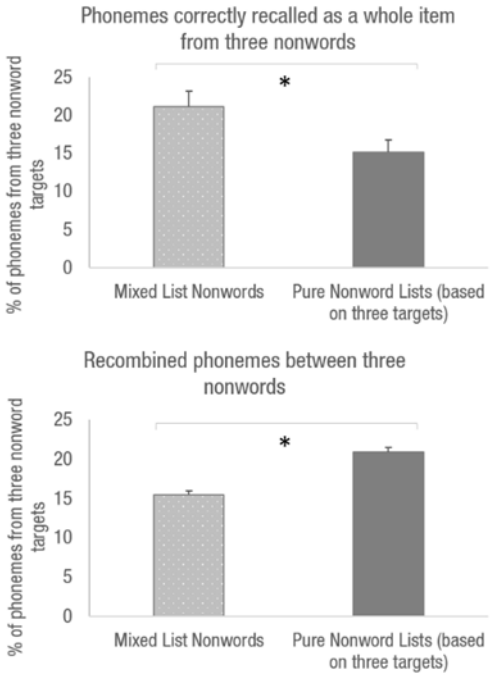

Fig. S1. Analyses constrained to three nonword targets in pure nonword lists and mixed lists in the main experiment. Bars show percentage of the target nonword phonemes in all mixed lists and equivalent targets in pure nonword lists. Errors bars are 95% confidence intervals for a within-subject design (Cousineau, 2005).

Analyses of nonwords in the pure nonword lists were constrained to three target positions that corresponded to nonwords in the most common mixed list structures tested (nonwords in positions 1, 3 and 5; in positions 2, 4, and 6; in positions 1, 2 and 5; and positions 3, 4, and 6). We recalculated phoneme recombinations that arose only from phonemes from the three nonword target positions. Recall averaged across these different three-nonword structures was compared with the average recall of nonwords in mixed lists. The resulting data are displayed in Figure S1. These additional analyses demonstrated that, even when nonword recombination 'opportunities' in the pure lists were constrained to match mixed list levels, the mixing effect on nonword recall – better whole item recall than for nonwords in pure lists [$t(27) = 6.78$, $p < .001$, $d = 0.63$] and fewer recombinations [$t(27) = -8.14$, $p < .001$, $d = -1.82$] – remained. Thus, the stabilising influence from words in mixed lists cannot be explained by fewer available opportunities to migrate between nonword positions.

Next, we considered whether the semantic manipulation of the mixed lists benefitted the overall recall of word and nonword target phonemes, over and above improvements in recall that could be attributed to a reduction in phoneme recombination errors. This is relevant to interpretations of the semantic effects on nonword recall. To do this, we compared the percentages of target word phonemes and target nonword phonemes that were recalled irrespective of the response's serial position within the list (but respecting syllabic position within an item; i.e., target phonemes recalled in a correct or migrated position). Thus, phoneme movements (e.g., those affecting recombinations) would not directly affect this measure (and so any changes in this measure for nonwords could not be accounted for by changes in the phonological stability of word items). Figure S2 displays the respective percentages in the Main Experiment and Replication Sample. Paired t-tests confirmed that both word and nonword phonemes were successfully recalled at an increased rate in the SEM MIXED condition compared to the RANDOM MIXED condition in both experiments [Main Experiment: Word Phonemes, $t(27) = 6.34$, $p < .001$, $d = 0.95$; Nonword Phonemes, $t(27) = 3.32$, $p = .003$, $d = 0.36$; Replication Sample: Word Phonemes, $t(23) = 5.38$, $p <$

.001, d = 0.68; Nonword Phonemes, $t(23) = 2.38$, $p = .026$, d = 0.27]. These data show that improvements in nonword recall linked to the presentation of more meaningful word sequences cannot be entirely explained by reduced opportunities for phoneme recombinations when nonwords are mixed with words.

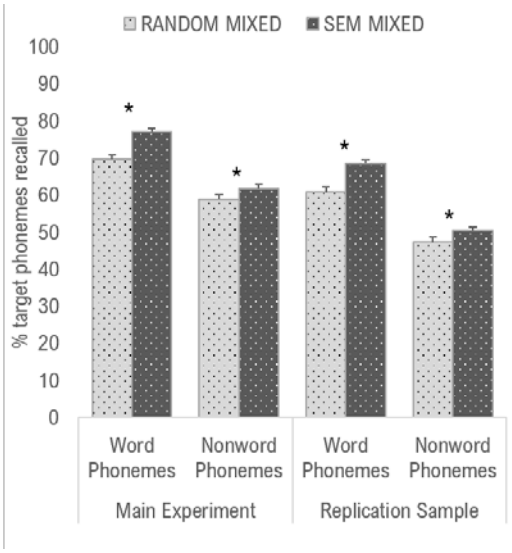

Fig. S2. Target word and nonword phonemes recalled from the mixed list conditions in the main experiment and replication experiment. Errors bars are 95% confidence intervals for a within-subject design (Cousineau, 2005).

Finally, to provide statistical information regarding how the semantic effects on recall for nonwords in mixed lists compared with words, we ran separate 2 × 2 repeated measures ANOVAs for target phonemes recalled overall (a target based analysis) and each corresponding phonologically-related response category in the main experiment and replication sample. These tested the effects of the semantic manipulation on mixed list phoneme recall (RANDOM MIXED, SEM MIXED) according to the lexicality of the source target (word phonemes, nonword phonemes).

Robust main effects of the semantic manipulation in both datasets confirmed that the semantic benefits were present for both word and nonword targets [semantic support increased the percentages of both word and nonword phonemes that were recalled, Main Experiment: $F(1, 27) = 180.02$, $p < .001$,

η_p^2= .87; Replication Sample: $F(1, 23) = 123.25$, $p < .001$, η_p^2= .84. This corresponded with increases in correct items in position, Main Experiment: $F(1, 26) = 26.05$, $p < .001$, η_p^2= .50; Replication Sample: $F(1, 23) = 27.85$, $p < .001$, η_p^2= .55, and decreases in phoneme recombination errors, Main Experiment: $F(1, 26) = 17.75$, $p < .001$, η_p^2= .41; Replication Sample: $F(1, 23) = 14.18$, $p = .001$, η_p^2= .38. Order errors also increased in Main Experiment: $F(1, 26) = 7.22$, $p = .01$, η_p^2= .22; Replication Sample: p = .16; non-recombination errors were not affected in either task, Main Experiment: p = .19; Replication Sample: p = .24]. This is relevant because, as expected, word target phonemes tended to be better recalled than nonword target phonemes [main effects of target lexicality on the percentages of phonemes recalled overall, Main Experiment: $F(1, 27) = 39.05$, $p < .001$, η_p^2= .59; Replication Sample: $F(1, 23)$ = 32.04, $p < .001$, η_p^2= .58, corresponding to more whole words than nonwords in the correct position, Main Experiment: $F(1, 26) = 310.20$, $p < .001$, η_p^2= .92; Replication Sample: p = .16; or incorrect position, Main Experiment: $F(1, 26) = 210.45$, $p < .001$, η_p^2= .89; Replication Sample: $F(1, 23)$ = 210.45, $p < .001$, η_p^2= .89; and fewer recombination errors, Main Experiment: $F(1, 26) = 279.99$, $p <$ .001, η_p^2= .92; Replication Sample: $F(1, 23) = 7.04$, $p = .014$, η_p^2= .23, and non-recombination errors, Main Experiment: $F(1, 26) = 130.10$, $p < .001$, η_p^2 = .83; Replication Sample: $F(1, 26) = 21.23$, $p <$ .001, η_p^2 = .48]. Furthermore, as expected, words tended to be associated with larger magnitude semantic effects on their recall than nonwords [interactions between the semantic manipulation and target lexicality in phonemes recalled, Main Experiment: $F(1, 27) = 11.47$, $p = .002$, η_p^2= .30; Replication Sample: $F(1, 23) = 4.89$, $p = .037$, η_p^2= .18, relating to larger effects on whole items in the correct position, Main Experiment: $F(1, 26) = 35.05$, $p < .001$, η_p^2= .57; Replication Sample: $F(1, 23)$ = 123.15, $p < .001$, η_p^2= .84, and a larger reduction in recombination errors in the Main Experiment: $F(1,$ 26) = 4.62, $p = .04$, η_p^2= .15; Replication Sample: p = .73. There were no differences in the magnitude of semantic effects between words and nonwords for order errors, Main Experiment: p = .69, Replication Sample: p = .83, or non-recombination errors, Main Experiment: p = .12; p = .58]. Thus, the

information regarding interaction effects provided by these analyses confirm – as expected – that the effects of the semantic manipulation on recall accuracy were stronger for words than nonwords.

Supplementary Material References

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1995). The CELEX Lexical Database [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

Cortese, M. J. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers, 36*, 384–387. doi:10.3758/BF03195585

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:10.1016/j.jml.2005.08.001

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33, 159–174. doi:10.2307/2529310

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:10.1016/j.jml.2015.06.003

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:10.1080/17470218.2013.850521

Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. Behavior Research Methods, Instruments, & Computers, 36, 481–487. doi:10.3758/BF03195594