Neuroscience Institute Publications

Western Institute for Neuroscience

2-1-2022

The role of phonological and semantic representations in verbal short-term memory and delayed retention

Theresa Pham
Western University

Lisa M.D. Archibald
Western University, larchiba@uwo.ca

The role of phonological and semantic representations in verbal short-term memory and delayed retention

Theresa Pham[1] · Lisa M. D. Archibald[1]

Abstract

It has been suggested that phonological representations play a central role in verbal short-term memory, but when semantic knowledge has been investigated, it has also been shown to influence verbal short-term memory. Explaining this interaction between verbal short-term memory and the linguistic system has produced different theoretical positions: whether semantic knowledge is used to redintegrate phonological traces or if there is direct activation of both phonological and semantic knowledge upon encountering a word. The present study employed a new paradigm to systematically examine phonological and semantic representations in verbal short-term memory as well as long-term impacts. Across two experiments, a list of words was presented sequentially, followed by a probe word. Participants were to judge whether the probe word rhymed or was synonymous with any items on the list. Delayed memory was also tested. In Experiment 1, we found that immediate performance was better for synonym than rhyme judgements, and this continued to be the case after a brief delay. In Experiment 2, under a fast-encoding, running-span paradigm, we found similar activation of phonological and semantic knowledge. Nevertheless, accuracy was again higher for items probed with the semantic than rhyme cue in the long term. Results showed that indeed there are short-term semantic effects, in addition to phonological effects. Further, semantic processing can occur in a highly automatic and rapid manner, with strong influence on long-term memory. These findings provide a new perspective on viewing verbal short-term memory as operating more dynamically within the context of a complex linguistic system.

Keywords Working memory · Short-term memory · Word recognition · Psycholinguistics

Introduction

Knowing a familiar word entails representing the word form via phonological representations as well as having enriched linguistic knowledge (lexical, semantic, syntactic knowledge) associated with it. Yet, traditional accounts of verbal short-term memory suggest a distinction between language processes involved in short- and long-term memory tasks. In particular, rapid encoding of phonological representations has been attributed to short-term memory processes, whereas activation of associated semantic knowledge has been considered independently related to episodic long-term memory processes or strategy use. In contrast, alternative explanations based on language-based models of verbal short-term memory conceptualize verbal short-term memory as the activation of different levels of linguistic knowledge (phonological, lexical, semantic) within the language system (Majerus, 2013; Schwering & MacDonald, 2020). As such, language-based models allow for direct activation of phonological and semantic information during language processing. Serial recall, the immediate repetition of items in presented order, is one of the most common measures of short-term memory, but is best suited for investigating the role of phonological processing (Campoy & Baddeley, 2008) and rehearsal (Tan & Ward, 2008) in keeping verbal items active. In contrast, there is a relative dearth of paradigms for investigating semantic processing in verbal short-term memory. Thus, the current study aimed to reconcile traditional phonological effects with the growing literature of semantics effects in verbal short-term memory by employing a new paradigm to directly compare the retention of phonological and semantic information in verbal short-term memory as well as long-term impacts.

The notion that verbal short-term memory is largely influenced by phonological effects and minimally by semantic effects comes from a long history. Seminal work by Baddeley (1966) found a greater detrimental effect of

✉ Theresa Pham
   tpham62@uwo.ca

1   School of Communication Sciences and Disorders, University of
    Western Ontario, London, Ontario, Canada

phonological similarity (poorer recall for lists that are phonologically similar as opposed to distinct) than semantic similarity on short-term memory, which was taken as evidence for phonological coding in verbal short-term memory. This finding spurred on a large body of research focusing on phonological properties and rehearsal mechanisms in verbal short-term memory, including the word-length effect (better recall for lists of shorter than longer words) and articulatory suppression (Baddeley et al., 1975). Some researchers continue to use the terms phonological short-term memory and verbal short-term memory interchangeably (e.g., Papagno & Cecchetto, 2019).

However, it is now well established that semantic knowledge influences verbal short-term memory as well. Evidence comes from the findings that words with richer semantic representations are remembered better: the lexicality effect (better recall for words than non-words; Hulme et al., 1991) and the concreteness effect (better recall for concrete than abstract words; Romani et al., 2008; Walker & Hulme, 1999). Neuropsychological studies have been instrumental in demonstrating the interaction between language processing and verbal short-term memory. On verbal short-term memory tasks, some patients present with phonological deficits but intact semantic effects (difficulties repeating words, but not sentences, Baldo et al., 2008; semantic support despite impaired rehearsal, Howard & Nickels, 2005), while other patients have difficulties maintaining semantic but not phonological information (diminished lexicality effects, Jefferies et al., 2005; N. Martin et al., 1996; R. C. Martin et al., 1994). Additionally, behavioural data (Nishiyama, 2014, 2018) and neuroimaging studies (Fiebach et al., 2007) have provided corroborating evidence supporting unique phonological and semantic contributions throughout all stages of a verbal short-term memory task. Explaining this interaction between verbal short-term memory and the linguistic system has produced different theoretical positions, two receiving the most research attention are the redintegration hypothesis (Gathercole et al., 2001; Hulme et al., 1997; Schweikert, 1993) and language-based models (Majerus, 2013; N. Martin et al., 1996; R. C. Martin et al., 1999; Schwering & MacDonald, 2020).

The redintegration hypothesis assumes a two-part process to recall. According to this view, processing verbal information first relies on forming a phonological representation of the item and then semantic knowledge stored in long-term memory is accessed and used at retrieval to reconstruct or "clean up" degraded phonological traces (Gathercole et al., 2001; Hulme et al., 1997; Schweikert, 1993). Some argue that redintegration can also take place during rehearsal (Hulme et al., 1999) or maintenance (Barrouillet & Camos, 2015) and not only at recall. Nevertheless, the distinction between phonological short-term and semantic long-term processes is inherent in this theory, with the redintegration mechanism as a potential

account for the influence of semantic knowledge in immediate recall tasks. More recently however, there has been a shift away from viewing semantic contributions through this stepwise perspective toward a more integrated view of linguistic knowledge and short-term memory.

Language-based models offer a more parsimonious explanation for the interaction between language processing and verbal short-term memory. These current models account for the influence of semantic knowledge on immediate memory by assuming that activation occurs within the linguistic system (N. Martin & Saffran, 1997) or a dedicated buffer for short-term semantic maintenance (semantic short-term memory, R. C. Martin et al. 1999; conceptual short-term memory, Potter, 2012). Despite their differences, language-based models collectively assume that semantic representations are maintained along with phonological representations when verbal items are encountered and processed (for review, see Majerus, 2013, and Schwering & MacDonald, 2020). Motivated by this line of reasoning, we conducted the present study to systematically explore whether phonological and semantic information, while interactive, could have independent effects in verbal short-term memory using novel techniques.

Verbal short-term memory has typically been investigated by using serial recall tasks. Serial recall has been used to study the efficiency of phonological encoding (Campoy & Baddeley, 2008), the effect of list length (Grenfell-Essam & Ward, 2012), and rehearsal in immediate memory (Tan & Ward, 2008). However, serial recall stresses order over item information, and thereby taps linguistic knowledge minimally (Majerus, 2009, 2013). Phonological effects might have been inevitable with serial recall because phonological coding is very effective for storing serial order (Gathercole et al., 2001; Romani et al., 2008). Indeed, with rapid presentation rates, participants can easily encode serial order via phonological processes, while semantic encoding appears to be less optimal (Campoy & Baddeley, 2008; Campoy et al., 2015, Experiment 1). Semantic processing, conversely, then, is more engaged when processing item information and when verbal items are familiar and meaningful. Notably, when semantics have been considered in the methodology, they have been found to impact serial recall (Acheson et al., 2011; Poirier et al., 2015).

A further way in which results from serial recall studies have lent themselves to interpretations related to short-term and long-term memory has been in examining phonological effects using short lists (e.g., Campoy & Baddeley, 2008; Tan & Ward, 2008) and semantic effects using long lists (e.g., Kowialiewski & Majerus, 2018a; Nishiyama, 2014). Theorists argue that, given the limits of short-term memory, short list lengths can be kept within the focus of attention (Cowan, 2001), while longer lists would likely involve episodic long-term memory to support recall. Thus, as list length increases, there should be less involvement of short-term

memory processes. Indeed, adults and children spontaneously use rehearsal to aid serial recall of short lists (McGilly & Siegler, 1989; Tan & Ward, 2008), and it has been suggested that this strategy is abandoned at longer list lengths as cognitive load increases (e.g., Baddeley & Larsen, 2007). Rehearsal at short list lengths seems to be reflected in reaction time such that reaction time linearly increases up to about four to six items (serial recall, Vergauwe et al., 2014; recognition task, Rypma & Gabrieli, 2001; Sternberg, 1966). Others report flatter reaction times in recognition tasks, especially with long list lengths (Burrows & Okada, 1975), indicating that memory search may occur in parallel rather than in a serial fashion (Townsend & Fific, 2004; see Cowan, 2001, for theoretical discussion). This latter notion is broadly in line with language-based models, according to which access to both phonological and semantic representations would be available throughout encoding and maintenance, without the need for rehearsal or redintegration processes. Therefore, the current study seeks to better understand whether there is evidence for semantic effects, in addition to phonological effects, in verbal short-term memory when the role of list length and strategies such as rehearsal are considered.

Given the limitations of serial recall paradigms, techniques more suitable for the investigation of phonological and semantic retention capacities are needed. Such tasks should avoid stressing serial order and thereby phonological encoding. Tasks that may be particularly well suited are the rhyme (or homophone) and synonym probe-recognition paradigms (McElree, 1996; Martin, et al., 1994). Item probe-recognition task (primarily testing item information) is in contrast to the serial order or list probe-recognition task, which was designed to primarily test order information and requires serial rehearsal (Henson et al., 2003; Murdock, 1976). Rhyme (or homophone) and synonym probe tasks are typically studied separately. Participants would first process a list of words followed by a probe word. In the rhyme (or homophone) probe task, participants must decide if the probe word rhymes with (or sounds the same as) a word on the list. In the synonym probe task, participants must decide if the probe word is synonymous with a word on the list. Since these two tasks have primarily been used to show a dissociation between phonological and semantic short-term memory in patients (R. C. Martin & He, 2004; R. C. Martin, et al., 1994), it is important to also investigate phonological and semantic mechanisms in healthy adults, and only few studies on this topic exist.

Although addressing a different theoretical question than the current work, an early study by McElree (1996) provided preliminary evidence for phonological and semantic effects in healthy adults. Participants studied a list of five words followed by a recognition probe probing for item, rhyme, or synonym information. Participants answered more quickly or slowly (range = 0.128–3 s) depending on the condition. Results revealed slower retrieval dynamics (a measure of speed-accuracy trade-off) for rhyme and synonym judgements compared to item judgements, except there was a recency effect across all conditions. In turn, rhyme and synonym judgements did not differ. The data were interpreted to indicate sufficient access to phonological and semantic information to enable a comparison with the probe (see also Shulman, 1970). This interpretation, however, should be made with caution due to a very small sample (n = 4), use of closed sets – which could encourage phonological representations, and rhyme and synonym probe stimuli not being perfectly comparable as synonym probes were multisyllabic words (car-automobile) while rhyme probes were shorter words (car-far). Using a modified recognition task paradigm, Nishiyama (2014) investigated the separability of phonological and semantic representations in working memory in healthy adults. Participants studied ten-word lists using either a phonological (rehearsal) or a semantic strategy (focus on the meaning) while completing a concurrent task impairing phonological (articulatory suppression) or semantic processing (finger-tapping). At test, participants had to choose the target word that was a homophone or synonym for one of the items on the list and were tested on all ten words. Results revealed that articulatory suppression impaired homophone judgements, whereas the finger-tapping task impaired synonym judgements, indicating distinct representations in healthy adults. However, task instructions could have favoured phonological or semantic encoding, and the use of long list lengths (ten words) could have involved episodic long-term memory in supporting recall. Further, the influence of semantics was found only in an indirect way – tapping impaired semantic via attentional load (Ruchkin et al., 2003). We addressed these issues by not prompting a particular strategy in order to necessitate engagement of both phonological and semantic processing across both short and long lists lengths.

The running-span task may be another way to study the content of short-term memory (and mitigate involvement of episodic long-term memory) as it minimizes opportunities for strategy use (e.g., Cowan, 2001; Pollack et al., 1959). In a typical task, participants process items presented at a fast presentation rate and then are cued to recall the most recent *n*-items unpredictably (where *n* is the number of items they are asked to recall). The fast presentation rate requires constant updating and serves to necessitate attention to each item, while the unpredictable list length prevents the use of strategic encoding such as the use of rehearsal or grouping (Bunting et al., 2006; Cowan, 2001). Rehearsal is actually detrimental to performance (Hockey, 1973). Similarly, research in visual working memory has shown that conceptual knowledge can be activated rapidly and without rehearsal using the Rapid Serial Visual Presentation procedures (Potter, 2012).

Kowialiewski and Majerus (2018a) developed a novel recognition variant of the running-span task to more closely evaluate the direct activation of semantic knowledge as well as

rapidity of such processing in verbal short-term memory. Participants studied a list of words and nonwords presented at a fast rate of two items/s with list length varying from 11 to 14 items. A probe word appeared after the word list and participants had 1,750 ms to decide if the probe word matched one of the items in the word list. This speeded response further served to prevent redintegration during retrieval. Findings revealed a lexicality effect, with better and faster performance for real words compared to nonwords, even in a task that minimized rehearsal and redintegration processes. The results were interpreted to be consistent with language-based models. Semantic knowledge that was activated when a word (vs. nonword) was encountered served to stabilize phonological representations. Task conditions could, however, have favoured semantic encoding; participants were only being tested on semantic knowledge (lexicality effect) and with long list lengths (11–14 items). Further, item or matching judgements behave differently than judgements based on phonological and semantic information, indicated by McElree (1996). Thus, the current study used a similar but modified procedure that required accessing phonological and semantic information in short-term memory across various list lengths and processing times.

Finally, little is known about how information processed in verbal short-term memory can help or hinder long-term memory encoding. On the one hand, phonological processing or rehearsal can support short-term retention, but phonological information decays rapidly (Baddeley, 2012), making it ineffective for long-term retention (Gallo et al., 2008; Craik & Lockhart, 1972). Semantic processing, on the other hand, is a process that involves deeper processing, leading to encoding of more contextually unique features and making it less susceptible to forgetting (Gallo et al., 2008). Related to the methodology adopted here, the probe word in the probe recognition task may act as a cue itself to re-activate relevant information. Studies in the visual working memory domain have found that when a cue (arrow) is presented after displaying the to-be-remembered items and before the probe item, this reactivates relevant information *already* in working memory and benefits retention for cued items (e.g., Berryhill et al., 2012). For word processing, by consequence of verbal short-term memory being emergent from the language network, probing for reactivation of semantic (vs. phonological) representations may promote encoding into long-term memory.

Taken together, results from the aforementioned studies using novel paradigms to study semantic effects in verbal short-term memory provide complementary data to well-established work on phonological effects (e.g., Kowialiewski & Majerus, 2018a; Nishiyama, 2014). However, past studies on short-term phonological effects differ from these more recent studies on semantic effects in many respects (e.g., serial recall vs. probe recognition; letters, numbers vs. words, sentences; short vs. long lists), making a direct comparison

difficult. Therefore, our aim was to use one common paradigm to minimize methodological differences in our investigation of the retention of phonological and semantic information in verbal short-term memory and long-term impacts.

The present study aimed to address methodological issues with respect to testing of order information, instructions encouraging semantic or phonological encoding, the use of long list lengths favoring semantic encoding, and the possibility of post-list processes and rehearsal strategies. In contrast to Nishiyama (2014, 2018), participants were not instructed in advance on how to encode or maintain the word list. Our approach also extends work by Kowialiewski and Majerus (2018a) by directly tapping semantic (synonym judgement) and phonological information (rhyme judgement). Short and long list lengths were also used. Across two studies, participants studied a list of words that varied in length from three to 11 items. Following the word list, participants were cued to make an immediate rhyme or synonym judgement on a probe word appearing after the cue. Critically, this cue was not known in advance, meaning that participants had to engage in both phonological and semantic processing. After a 10-min delay, participants completed a surprise delayed recognition test to assess long-term retention. In Experiment 1 we used a modified version of the rhyme and synonym probe recognition tasks, and this paradigm was further modified in Experiment 2 by incorporating aspects of the running-span procedure to minimize strategy use and redintegration processes.

The main goal was to develop a new methodology that would be more suitable for assessing language representations underlying maintenance of verbal items than paradigms used in earlier studies (e.g., serial recall; Kowialiewski & Majerus, 2018a; Nishiyama, 2014). We drew inspiration from the item probe-recognition and running-span tasks because these paradigms do not stress phonological or semantic maintenance to a greater degree nor should they primarily involve rehearsal or serial order retention (Henson et al., 2003; Murdock, 1976). We hypothesized that the emergence of any phonological and semantic effects would reflect rapid access in short-term memory and further indicate that semantic effects do not require post-encoding reconstruction processes, aligning with language-based models. In contrast, if semantic representations rely on post-list processing, that is, activation of semantic knowledge to reconstruct incomplete phonological representations, then we expect to observe subtle semantic effects across the two experiments. Specifically, with the rapidity of Experiment 2, we expect that semantic knowledge would not be accessed within the allotted time.

A secondary goal was to examine performance across short and long list lengths. There may be an effect of list length in that long list lengths with increased load impair performance regardless of probe. However, at a given list length, we hypothesized comparable phonological and semantic

performance that would indicate direct activation, and thus, the ability to access linguistic knowledge regardless of list length. If, however, phonological and semantic effects reflect short- and long-term memory processes, respectively, then different advantages should emerge across list lengths. Short list lengths could easily be maintained using rehearsal, which would convey an advantage for phonological over semantic probe, but rehearsal would be less efficient as list length increases. This should also be reflected in reaction time analyses; reaction time should increase linearly with increased load (Vergauwe et al., 2014). If semantics only affect episodic long-term memory, then accuracy and reaction times for the semantic relative to phonological probes should be much worse for the short list lengths, but perhaps be better for longer list lengths.

Finally, short- and long-term retention for probed items were examined. The novelty of this task relied on the use of the probe word to reactivate relevant phonological or semantic information that would induce phonological or semantic processing, respectively. Meaningful processing should differentiate words more than focusing on sounds. As such, memory is expected to be better for words probed with a semantic than a phonological cue.

Experiment 1

Experiment 1 served to demonstrate the use of our modified probe recognition paradigm as a verbal short-term memory task that could tap phonological and semantic information. The paradigm was revised to address the limitations of previous work by (a) requiring both phonological and semantic processing (probe type was a within-subject variable), (b) no instructed encoding or maintenance strategy (probing occurred after the word list), and (c) presenting short and long list lengths. Further, a surprise delayed recognition test was added to compare short- and long-term retention.

Method

Participants We recruited 31 participants (15 females; M_{age} = 19.55 years; SD_{age} = 2.01) who were proficient or native English speakers. Three additional participants were excluded because they did not follow directions or understand the instructions. Informed written consent was obtained for all participants. Ethical approval was provided by the University of Western Ontario's research ethics committee.

Materials We selected a total of 350 monosyllabic words (nouns, verbs, adjectives) using the SUBTLEX norms (Brysbaert & New, 2009). Words were of medium lexical frequency, with a mean word frequency of 73.70 per million

(SD = 39.60). They had a mean concreteness rating of 3.70 (SD = 0.93), with 5 being the most concrete (Brysbaert et al., 2014). Lists of three, five, seven, nine, and 11 words were generated by random selection without replacement, so that each word appeared only once during the experiment. There were ten lists at each length; matched for word frequency and concreteness rating. List lengths were presented in an ascending order (trials within list lengths were presented randomly). Within each length, participants made a rhyme judgement on five trials (rhyme probe) and a synonym judgement on five trials (synonym probe). Further, within those five trials, three of them were matching probes (affirmative responses) and two of them were nonmatching probes (negative responses).

For matching trials, one word within each list was paired with a rhyming or a synonymous word using an open set of monosyllabic words as well. Rhyming words were orthographically both similar (e.g., *hat – cat*) and dissimilar (e.g., *note – throat*). A monosyllabic synonymous word (e.g., *hat – cap*) was obtained from the norms of Nelson et al. (1998). The mean forward-associative strength was 0.27 (SD = 0.25).

Procedure Participants were tested individually, seated in front of a 14-in. laptop, using the PsychoPy 3.1 software (Peirce et al., 2019). The paradigm is shown in Fig. 1. Each trial began with a fixation cross at the center of the screen for 1,000 ms. Each word from the word list was presented at the top-center of the screen one at a time for 1,000 ms with an interstimulus interval of 1,000 ms. After the word list, participants received a cue word, "rhymes" or "means", in capitalized letters at the center of the screen, followed by the probe word at the bottom-center of the screen. Participants were instructed to press the key labelled "Yes" or "No" in response to whether the probe word had rhymed or meant something similar to an item on the word list (depending on the cue). For example, in Fig. 1, if the probe word *cap* appears after the cue RHYMES, then the participant must decide if "cap" rhymed with any words from the list. Participants would press "No" since there is no rhyming word on the list. If the cue was MEANS, then participants would press "Yes" because "cap" is a synonym for the word list *hat*. Participants were required to make their response as quickly and as accurately as possible. Participants completed practice trials at list length four. After completing the immediate probe-recognition task, participants performed a nonverbal task for 10 min, then they completed a surprise delayed recognition test.

For the delayed test, the 30 list words that had a matching probe word (i.e., given the match *hat* (list word) – *cap* (probe word), the word *hat* would be tested) and 30 new words were presented individually on the computer. The new words were matched on word frequency and concreteness rating. Participants responded to each word by pressing the key labelled "Old" for old words or "New" for new words. The word was considered "old" if it was a word from the word
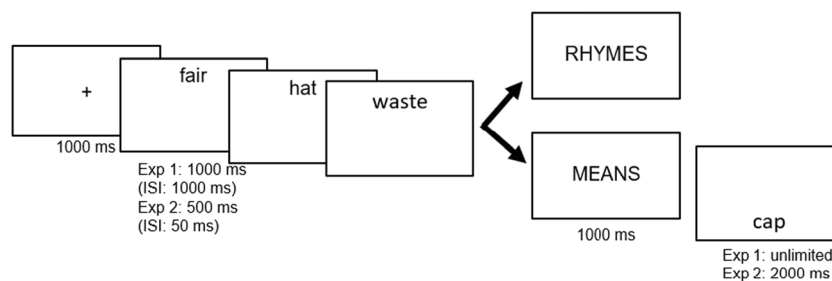
Fig. 1 Schematic representation of the experimental design. After a word list, a single cue appeared. The cue was either "rhymes" or "means". After the cue, a probe word appeared, and participants were instructed to indicate whether the probe word was associated with an item on the list

based on the cue. The timing intervals were the same in both experiments unless otherwise indicated. *Note*: If this was a synonym probe trial, then the word "hat" would be presented in the delayed recognition test

list, while a "new" word meant that it was never presented in the experiment, neither as list nor probe word. The word remained on the screen until participants made a decision.

Data analysis Recognition accuracy was analyzed using d' to remove response bias. Based on signal detection theory, d' is a measure of sensitivity that accounts for false alarms when measuring proportion of hits (Stanislaw & Todorov, 1999). Hits reflect correctly identifying that the probe word rhymed or was synonymous with a word from the list, whereas false alarms were acceptances even when there were no associations. High d' indicates high sensitivity, or more accurate performance (fewer misses or false alarms). The effective limit is 4.65, for a hit rate of .99 and false alarm of .01 (Macmillan & Creelman, 2005). Zero d' indicates a lack of sensitivity, or chance-level performance. The typical range of d' for yes-no paradigms is 0.5–2.5, corresponding to about 60–90% accuracy. We used a Bayesian analysis approach to analyze recognition accuracy (d') and reaction time (Wagenmakers et al., 2018). For each model, Bayes Factor or BF_{10} is used to evaluate the strength of evidence for the alternative model (H_1) against a null model (H_0). If support for the model was ambiguous, we ran an analysis of *specific effects* to untangle this ambiguity and report inclusion BF (BF_{incl}) based on all models (van den Bergh et al., 2020). We used the following classification scheme to interpret BF: BF < 1 provides no evidence, BF between 1 and 3 provides anecdotal evidence, BF between 3 and 10 provides substantial evidence, BF between 10 and 30 provides strong evidence, BF between 30 and 100 provides very strong evidence, and BF > 100 provides decisive evidence (Jeffreys, 1961; Wagenmakers et al., 2018). Delayed data were analyzed using Bayesian Wilcoxon Signed-Rank Test due to the small sample size.[1] Bayesian analyses were conducted using JASP (JASP Team, 2020). Immediate and delayed data were submitted to separate analyses given the differences between the tasks, and visual inspection was used to compare performance.

Results and discussion

Preliminary analysis Orthographically dissimilar rhyme pairs may be more difficult than similar pairs so we first verified that this manipulation did not unintentionally affect performance, and indeed there was no evidence for a difference, $BF_{10} = 0.61$. These conditions were collapsed in all remaining analyses.

Accuracy Performance for each probe across list length is depicted in Fig. 2 and descriptive statistics are provided in Table 1. A Bayesian repeated-measures ANOVA with probe and list length as within-subject factors provided anecdotal evidence in favour of the full model containing both main effects and the interaction term ($BF_{10} = 1.03e+8$) preferred by a factor of 1.47 over the second best model containing both main effects ($BF_{10} = 7.03e+7$). We ran an analysis of specific effects in order to untangle the ambiguous evidence. This revealed decisive evidence for an effect of probe with better accuracy for synonym than rhyme probes, $BF_{incl} = 1.57e+4$. There was also decisive evidence for an effect of list length, $BF_{incl} = 3.58e+4$. Follow-up Bayesian t-tests indicated that performance was best at list length 3 compared to all other lengths ($BF_{10} > 29.80$, all cases), anecdotal evidence for 9 versus 11 ($BF_{10} = 1.75$), and no evidence for remaining comparisons ($BF_{10} < 0.34$, all cases). Finally, substantial evidence

Table 1 Proportion of hit responses as a function of probe type and list length and proportions for hits and false alarms for Experiment 1
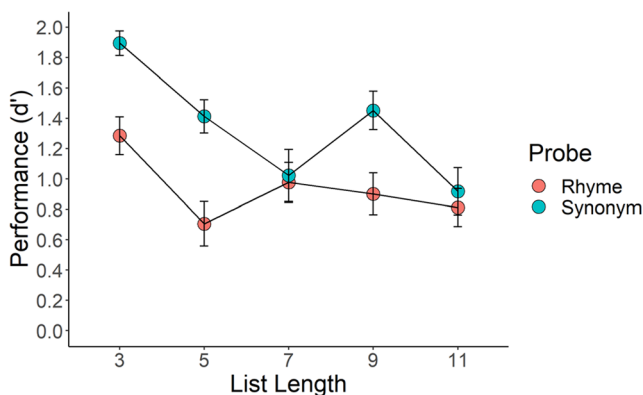
List length	Rhyme	Synonym
3***	0.73 (0.32)	0.94 (0.13)
5**	0.70 (0.30)	0.78 (0.22)
7	0.65 (0.24)	0.73 (0.26)
9**	0.65 (0.21)	0.80 (0.25)
11	0.76 (0.21)	0.67 (0.29)
Hit	0.70 (0.26)	0.78 (0.25)
False alarms	0.24 (0.35)	0.14 (0.37)

Standard deviations are in parenthesis. Probe effect, where Rhyme < Synonym: ** = strong evidence; *** = decisive evidence

---

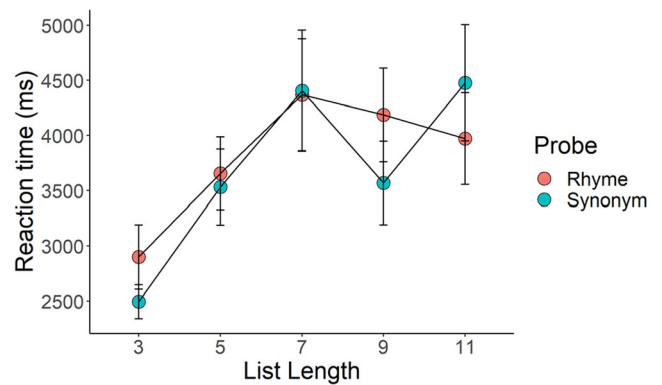[1] Due to experimenter error, only 13 of the 31 participants completed the delayed task in Experiment 1.

supported the interaction between probe and list length, BF_{incl} = 5.48. Follow-up Bayesian t-tests showed better performance for the synonym than rhyme probe at list length 3 (BF_{10} = 115.96) as well as at list lengths 5 and 9 ($BF_{10} > 12.12$); however, no evidence for a difference at list length 7 or 11 ($BF_{10} < 0.22$, both cases). These results indicate that synonym judgements were just as good as – or even better than – rhyme judgements across most list lengths. Additionally, we analyzed each probe separately via a Bayesian repeated-measures ANOVA. There was decisive evidence supporting differences across list lengths for the synonym probe, BF_{10} = 4.96e+4, but no reliable evidence for the rhyme probe, BF_{10} = 1.11. For the synonym probe, list length 3 differed from all other lengths ($BF_{10} > 3.83$, all cases), strong evidence between 9 versus 11 ($BF_{10} = 14.77$), and anecdotal evidence between 5 versus 7, 5 versus 11, and 7 versus 9 ($BF_{10} > 1.96$, all cases).

Reaction time Only correct trials were included in reaction time analyses and data from three participants were removed due to insufficient data for analysis (i.e., no correct trials in a given list length). A Bayesian repeated-measures ANOVA with probe and list length as within-subject factors revealed that the model with the effect of list length only was the best model (BF_{10} = 2.54e+5) favoured over the second best model including both main effects (BF_{10} = 3.92e+4) by a factor of 6.47 (see Fig. 3). Further comparisons using Bayesian t-tests revealed decisive evidence that list length 3 was faster than all other lists ($BF_{10} > 110.32$, all cases), anecdotal evidence for 5 versus 7 and 5 versus 11 ($BF_{10} > 2.50$, both cases), and no support for remaining comparisons ($BF_{10} < 0.48$, all cases).

Delayed data Although the difference in task precludes direct comparison, the immediate data are provided to contrast with delayed performance in Fig. 4. There was substantial support for a difference between recognition of words previously probed by rhyme and synonym cue, BF_{10} = 8.48.
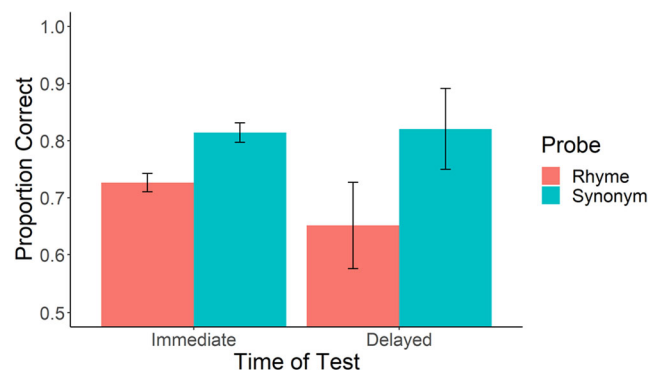

Fig. 3 Mean reaction time across list lengths for both the rhyme and synonym probe in Experiment 1. Error bars are the standard error of the mean

Semantically processed words were recognized better than phonologically processed words.

Using a modified probe recognition paradigm, Experiment 1 demonstrated that verbal items immediately activated phonological and semantic knowledge. Performance was better for synonym than rhyme judgements, and this advantage was consistent until the longest list length. Importantly, words probed semantically had an advantage not only in long list lengths exceeding the capacity of short-term memory (list length 9), but also when items were within the focus of attention (list lengths 3 and 5). Semantic effects in short lists minimize the contribution of episodic long-term memory processes, and addresses a limitation from prior work investigating semantics effects with only long lists (e.g., Kowialiewski & Majerus, 2018a; Nishiyama, 2014). Reaction time did not increase linearly. Instead, in both accuracy and reaction time analyses, there was only decisive evidence for a difference between the shortest list length (3) with longer list lengths; there were no reliable evidence for differences among long list lengths. Finally, delayed data revealed that, similar to immediate recognition, participants were better at remembering words that were probed with a semantic than a phonological cue.

Fig. 2 Participants' accuracy (d′) across list lengths for both the rhyme and synonym probes in Experiment 1. Error bars are the standard error of d′

Fig. 4 Mean proportion correct for rhyme/synonym probe across short- and long-term retention in Experiment 1. Error bars are the standard error of the mean

Of note, performance on rhyme probes at list length 3 may not have been at ceiling because vigilance could be low in tasks that seem trivially easy (Thomson et al., 2015); practice trials using list length 4 were at ceiling (rhyme: $M = .90$, $SD = 0.20$; synonym: $M = 0.87$, $SD = 0.22$). An alternative explanation is that participants found the synonym task more challenging (e.g., it is less clear whether two words are synonymous than rhyme with each other) and thus chose to focus more attention on semantic access and maintenance given the relatively slow presentation rate, causing a decrement in performance for rhyme probes. Another potential indication of a trade-off between task difficulty and accuracy comes from inspecting the performance patterns across lists 7, 9, and 11 for synonym probes. Although there was no evidence or, at best, anecdotal evidence for differences in accuracy or reaction time in direct comparisons of these lengths, visual inspection of the data revealed somewhat slower but more accurate responses for list length 9, which could reflect a shift to increased vigilance, and a relative "giving up" of the task with list length 11.

One limitation of Experiment 1 is that the timing intervals could potentially introduce confounding effects. In our initial conceptualization of this study, we were drawing on traditional serial recall tasks where a presentation rate of one item/s is common (e.g., Baddeley, 1966; Nishiyama, 2014; Poirier et al., 2015). However, longer times could have given participants time to make the strategic choice to focus attention on semantic relations as well as time for elaborative processing and engaging episodic long-term memory. Probe effects in delayed retention too could be attributed to stabilization from tapping long-term memory. We addressed this limitation in Experiment 2 by using a fast-encoding, running-span procedure to prevent rehearsal and grouping strategies (Bunting et al., 2006; Cowan, 2001). Words were presented at a rate of about two items/s (500 ms on, 50 ms off). Further, we started at list length 4 so that the task did not seem trivial and only used orthographically similar rhyme pairs. In this way, Experiment 2 was designed to investigate whether access to phonological and semantic information in verbal short-term memory occurs rapidly, in the absence of strategy use and redintegration processes. If short-term effects arise by consequence of the linguistic system directly supporting verbal short-term memory maintenance rather than being attributed to redintegration processes, then we should expect similar phonological and semantic effects, even under conditions of fast encoding.

Experiment 2

In Experiment 2, we re-examined phonological and semantic processing in verbal short-term memory by incorporating a running-span procedure. This meant manipulating the

presentation rate of the probe recognition task to be very fast and ending lists unpredictably to minimize opportunities for strategy use and redintegration. Experiment 2 was revised by (a) presenting words visually and auditorily at a rate of one every 550 ms, (b) using list lengths 4, 6, 8, and 10, and (c) selecting only orthographically similar rhyme pairs.

Method

Participants We recruited 30 new participants (19 females; $M_{age} = 18.33$ years; $SD_{age} = 0.61$). None had participated in Experiment 1.

Materials We selected 280 monosyllabic words from Experiment 1 ($M_{word\ frequency} = 74.90$ per million, $SD_{word\ frequency} = 37$; $M_{concreteness} = 3.76$, $SD_{concreteness} = 0.95$). From this set, words were further divided into ten trials for list lengths 4, 6, 8, and 10. Trials were presented in random order so that the list ended unpredictably. Trials varied by probe type (i.e., rhyme or synonym) and whether there was a match or not in the same way as was done in Experiment 1. Rhyme pairs were orthographically similar (e.g., *hat – cat*). The mean forward-associative strength for synonym pairs was 0.45 ($SD = 0.23$).

To control the duration in this experiment, audio recordings of each word were recorded by the first author, a native female English speaker. Audacity was used to remove background noise and adjust the duration of each word to 500 ms without altering the pitch.

Procedure In this experiment, the running-span procedure was incorporated into the probe recognition task (from Experiment 1). Stimuli were presented both visually and auditorily. A fixation cross appeared at the center of the screen for 1,000 ms to begin each trial. Words from the list were presented sequentially for 500 ms with an interstimulus interval of 50 ms. After the word list, participants were cued to make a yes-no decision on the probe word to indicate whether it had rhymed or was synonymous with a word on the list. Participants had 2,000 ms to respond, further limiting the use of strategies and redintegration processes (Experiment 1 had no time restrictions). When participants did not respond in time, the word "FASTER" would appear on the screen reminding them to respond faster, and the trial was not repeated. Participants completed a practice before the experiment. After the immediate task and a 10-min delay, participants performed a surprise delayed recognition test.

For the delayed test, the 24 old and 24 new words were presented both visually and auditorily, one at a time, in the center of the computer screen. For each word, participants had to determine whether the word was old or new.

Table 2 Proportion of hit responses as a function of probe type and list length and proportions for hits and false alarms for Experiment 2

List length	Rhyme	Synonym
4	0.83 (0.25)	0.72 (0.31)
6*	0.86 (0.19)	0.73 (0.25)
8*	0.70 (0.28)	0.75 (0.27)
10	0.75 (0.25)	0.64 (0.27)
HitFalse alarms	0.79 (0.25)	0.71 (0.28)
	0.21 (0.32)	0.18 (0.31)

Standard deviations are in parenthesis. Probe effect: * = anecdotal evidence

Data analysis Accuracy (d′) and reaction time were analyzed in the same way as in Experiment 1.

Results and discussion

Accuracy Performance accuracy on the immediate test is displayed in Fig. 5 and provided in Table 2. Data from one participant was removed due to insufficient data for analysis. We further excluded 26 rhyme trials (2.24% of the data) and 40 synonym trials (3.45%) where participants did not respond within the allotted time. A Bayesian repeated-measures ANOVA on recognition accuracy scores (d′) with probe and list length as within-subject factors provided anecdotal evidence in favour of the full model ($BF_{10} = 4.75e+5$) preferred by a factor of 1.48 over the second best model containing the list length effect only ($BF_{10} = 3.20e+5$). Given the ambiguous results, we ran an analysis of specific effect. There was no reliable evidence for an effect of probe, $BF_{incl} = 1.19$, suggesting that there was no evidence for the probe effect, nor evidence to state that there is no effect.[2] There was a list length effect, $BF_{incl} = 4.73e+5$, characterized by decisive support for a difference between 4 versus 8 and 6 versus 8 ($BF_{10} > 832.41$, both cases), strong support for 4 versus 10 ($BF_{10} = 13.44$), and substantial support for 6 versus 10 ($BF_{10} = 7.40$). Finally, there was substantial evidence in favour of the interaction, $BF_{incl} = 4.72$. Follow-up comparisons using Bayesian t-tests indicated only anecdotal evidence for a probe effect at list lengths 6 and 8 ($BF_{10} > 1.27$, both cases) and no evidence at list lengths 4 and 10 ($BF_{10} < 0.84$, both cases). This indicates that, at minimum, immediate recognition of semantic information was just as good as recognition of phonological information across all list lengths. Instead, the interaction was due to decisive evidence supporting differences across list lengths for rhyme probes, $BF_{10} = 1.44e+7$, but no evidence

for synonym probes, $BF_{10} = 0.49$. A comparison for rhyme judgements across list lengths revealed decisive support for a difference between 4 versus 8 and 6 versus 8 ($BF_{10} > 6850.63$, both cases), substantial support for 6 versus 10 and 8 versus 10 ($BF_{10} > 4.71$, both cases), and anecdotal support for 4 versus 10 ($BF_{10} = 2.95$).
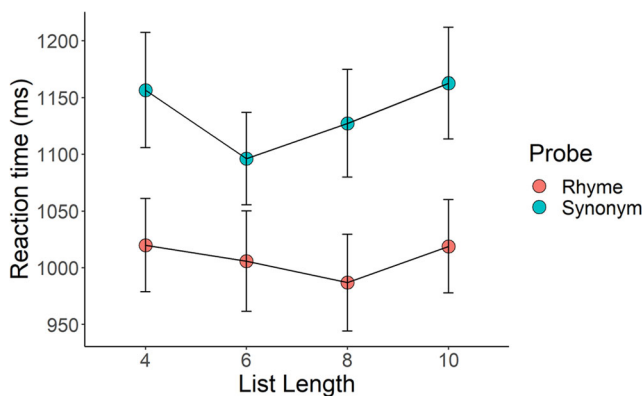
Reaction time Only correct trials were included in reaction time analyses. Data from two participants were removed due to insufficient data for analysis. We conducted a Bayesian repeated-measures ANOVA to assess reaction time performance (Fig. 6). The best model included the main effect of probe only ($BF_{10} = 7.66e+6$), which was 12.50 times more likely than the second best model with the two main effects ($BF_{10} = 6.13e+5$). Participants were faster when making rhyme ($M = 1007.85$ ms, $SD = 221.41$) than synonym judgements ($M = 1135.65$ ms, $SD = 247.93$).

Delayed data Figure 7 contrasts phonological and semantic performance at immediate and delayed testing. There was substantial evidence for a difference between words probed semantically and phonologically, $BF_{10} = 3.17$. Delayed recognition again showed an advantage for semantically processed words, even though immediate recognition was similar.

Our aim in Experiment 2 was to use a running-span probe task to investigate whether short-term phonological and semantic effects were evident in a task that prevented strategic encoding and redintegration processes. Indeed, we found that there appears to be similar access to both phonological and semantic information as participants were equally accurate at making rhyme and synonym judgements, indicated by no evidence for a probe effect (unlike Experiment 1). Further support came from demonstrating that there was no evidence or, at best, anecdotal evidence in support of a difference between probe type at any given list length. Therefore, at minimum, our results imply that semantic information was activated as readily and rapidly as phonological information in verbal
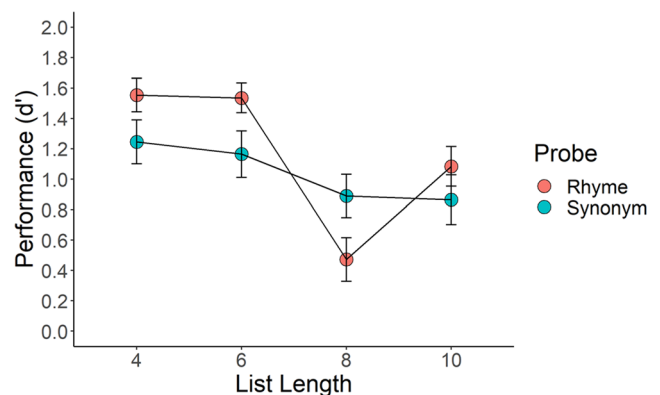

Fig. 5 Participants' accuracy (d′) across list lengths for both the rhyme and synonym probe in Experiment 2. Error bars are the standard error of d′

---

[2] As an alternative, and to untangle this ambiguity further, we computed inclusion probabilities for "matched" models only (van den Bergh et al., 2020) and found no evidence for the main effect of probe, $BF_{incl} = 0.32$. In fact, there is 3.13 times more evidence for the null model than a model including the effect of probe.

Fig. 6 Mean reaction time across list lengths for both the rhyme and synonym probe in Experiment 2. Error bars are the standard error of the mean.

short-term memory, consistent with language-based models. Interestingly, despite no immediate advantage of probing with a rhyme or synonym cue, there was substantial evidence supporting better memory for words processed semantically than phonologically after a brief delay.

Two somewhat surprising results were (a) substantial evidence in support for a difference between list lengths 8 and 10 in the rhyme probe condition and (b) that there was a main effect of probe in the reaction time data. First, improved accuracy between list lengths 8 and 10 could be attributed to inadvertently probing for rhymes across more positions in list length 8 (serial positions 1, 4–8 were probed) than list length 10 (serial positions 5–8 were probed), resulting in more degraded representations in list length 8. Since we did not systematically vary serial position, further analyzes would not be appropriate. But in Experiment 1 there was also some evidence that performance increased after a dip between list lengths 7 and 9 for synonym pairs, indicating a potential shift in processing. We elaborate on this point in the *General discussion*. Second, although synonym judgements were associated with slower responses, this did not correspond to more accurate responses, suggesting it was not due to a speed-accuracy trade-off. Instead, it likely takes longer for
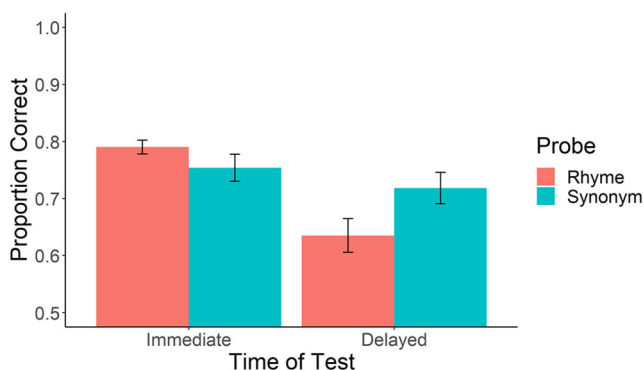
Fig. 7 Mean proportion correct for rhyme/synonym probe across short- and long-term retention in Experiment 2. Error bars are the standard error of the mean

participants to determine if words are synonymous with each other than if they had rhymed because synonym judgements are inherently more difficult.

General discussion

The purpose of the study was to use our modified verbal short-term memory task to tap phonological and semantic information as well as to compare short- and long-term retention. Importantly, the methodology adopted across two experiments, particularly the running-span version of Experiment 2, avoids an emphasis on order information and the employment of rehearsal, both of which favor phonological coding. In Experiment 1, accuracy was better for synonym than rhyme judgements. Further, words probed with a synonym (vs. rhyme) cue were better retained after a delay. However, given the slower presentation rate in Experiment 1, findings of semantic effects could have been attributed to contributions from long-term memory via redintegration rather than immediate co-activation of linguistic knowledge. Experiment 2 was designed to rule out this possibility by preventing redintegration and strategic processes. Indeed, in Experiment 2, under a running-span task, similar activation of phonological and semantic knowledge was indicated by accuracy being equally good for both the rhyme and synonym probes, confirmed by no evidence for a probe effect. Despite initial similar levels of activation, delayed retention was better for words probed with a semantic than phonological cue. Taken together, phonological and semantic effects were evident across both studies suggesting that linguistic knowledge was rapidly and readily available to support word list retention.

Phonological and semantic representations in short-term memory

This study used a modified probe recognition task coupled with the running-span task in Experiment 2 to assess phonological and semantic information across list lengths. Results revealed automatic and rapid access to phonological and semantic representations in a task that minimized redintegration and strategic processes, with slow encoding leading to an immediate semantic advantage. As discussed, synonym judgements are more difficult than rhyme judgments even without short-term memory demands, leading participants to focus their attention on semantic information when they had time to make such strategic choices (Experiment 1) and resulting in longer reaction times (Experiment 2). Further, semantic effects in particular do not require post-encoding processes such as redintegration, consistent with prior work demonstrating the influence of semantics in verbal short-term memory (e.g., Kowialiewski & Majerus, 2018a, 2018b). Although we cannot completely rule out the existence of redintegration

processes, it is difficult to reconcile the rapidity of semantic processing and lack of a phonological advantage with redintegration theories. However, it could be the case that more traditional serial recall tasks or certain procedures requiring serial order and phonological processing may benefit from redintegration. Nonetheless, our study more broadly implies that different levels of representations (phonological and semantic knowledge have been the focus here) influence verbal short-term memory and support retention of verbal information.

What is particularly novel in these results is that we demonstrated access to phonological and semantic representations using a relatively novel recognition variant of the running-span task. Procedures used in prior work including serial recall, short list lengths, and task instructions could have encouraged phonological representations and minimized semantic access (e.g., Nishiyama, 2014; Tan & Ward, 2008; but see McElree, 1996). It was therefore important to adopt a paradigm that would capture verbal short-term memory processes in the absence of rehearsal and redintegration (Kowialiewski & Majerus, 2018a), strategic encoding (Bunting et al., 2006; Cowan, 2001), and without primarily relying on serial order retention (Henson et al., 2003; Murdock, 1976). Results revealed that the same verbal item can engage both phonological and semantic processing when the task does not emphasize certain codes, even under fast encoding conditions. Given that this methodological approach may be better suited for tapping phonological and semantic information simultaneously, this work needs to be replicated and extended to further understand phonological and semantic access and maintenance. Future work can directly compare well-established phonological (e.g., phonological similarity, word length) and semantic effects (e.g., lexicality, concreteness). For instance, Kowialiewski and Majerus (2018b) studied a variety of semantic effects (lexicality, frequency, semantic similarity, and imageability) and found that most of these effects, except imageability, continued to emerge even in a running-span task. Phonological effects, however, were not investigated. Nishiyama (2013) studied both phonological and semantic effects using word frequency and word imageability, respectively. Although results were interpreted to support the separability of phonological and semantic representations, using word frequency to index phonology limits this interpretation because word frequency is typically viewed as a semantic variable.

Phonological and semantic effects across list lengths

A secondary goal was to investigate phonological and semantic effects across list lengths. Prior work often used short lists to examine phonological effects (e.g., Tan & Ward, 2008) and long lists to examine semantic effects (e.g., Kowialiewski & Majerus, 2018a; Nishiyama, 2014), with some arguing that

long list lengths likely tap episodic long-term memory. We found that although there was an overall effect of list length in that accuracy decreased with longer lengths, synonym judgements were for the most part better than phonological judgements in Experiment 1 and performance was comparable at any list length in Experiment 2. Importantly, semantic maintenance was observed not only at long list lengths (replicating prior work; Kowialiewski & Majerus, 2018a; Nishiyama, 2014), but also short list lengths, contrasting with the dominant view of semantics only affecting episodic long-term memory. McElree (1996) also found comparable rhyme and synonym judgements using five-word lists. Furthermore, reaction time did not increase linearly with list lengths (except a difference between list length 3 vs. longer lengths). This finding, coupled with the lack of a phonological advantage even at short lists, suggests that rehearsal or phonological processes was not the primary mechanism being used to process the word list. In fact, there was a semantic advantage in Experiment 1, while rehearsal was prevented in Experiment 2. Instead, the results can be accommodated by the predictions of language-based models. That is, upon hearing a word, maintenance would rely on the direct activation of phonological and semantic representations simultaneously within the linguistic system. It would follow that participants can automatically and rapidly activate phonological and semantic knowledge needed for processing, without the need to rehearse and this activation is unaffected by list length manipulations (see also Potter, 2012).

Nevertheless, substantial support for the interaction in both experiments suggests a potential shift in processing with increased load (i.e., longer list lengths). In particular, after an initial decrease, accuracy increased between list lengths 7–8 and 9–10, but it was not due to a speed-accuracy trade-off. Instead, there could have been a trade-off between allocating attentional processes to semantic (and phonological) access and accommodating the increased load. As a result, this likely minimized the semantic advantage at length 7 in Experiment 1 and caused a decrement to rhyme judgements at length 8 in Experiment 2 (which is in line with Experiment 1). The subsequent increase at lengths 9 and 10, respectively, might be due to paying more attention when the task got even harder, leading to better performance. However, we need to interpret this trend with caution as there was only anecdotal support for a difference between 7 and 9 for synonym whereas substantial evidence supported a difference between 8 and 10 for rhyme. Increased performance for rhyming pairs may be due to a number of factors (e.g., serial position probed, visual plus auditory presentation). Additionally, we did not measure individual differences in attention allocation which further limits our interpretations. Future studies, for instance, could evaluate general attentional demands on performance by using a dual-task that primarily imposes an attentional load while not chiefly tapping phonological or semantic processing. Results

would shed light on the role that attentional processes play in list length findings and its interrelationships with language processing in verbal short-term memory more broadly.

Semantic processing facilitates retention

We also examined how probing for phonological and semantic information already in verbal short-term memory benefited delayed retention. Although linguistic knowledge may be directly activated in short-term memory, after a brief delay, focusing on semantic (vs. phonological) information led to better long-term retention. Interestingly, this was even the case in Experiment 2 despite no initial probe advantage. To understand how short-term processing impacts subsequent memory, we integrate ideas from language-based models and from the distinctiveness hypothesis.

According to language-based models, different levels of linguistic representations (including phonological and semantic) are actively maintained in verbal short-term memory (Majerus, 2013; N. Martin et al., 1996; R. C. Martin et al., 1999; Schwering & MacDonald, 2020). On the other hand, creating distinctive representations via immediate semantic processing allows participants to encode more unique features associated with studied words (Gallo et al., 2008). Through these frameworks, when cued to make a synonym judgement, semantic processing is associated with richer and more distinctive representations at the semantic level, which serves to stabilize phonological representations through mutual interactions between the different levels (e.g., Majerus, 2013; N. Martin & Saffran, 1997), and subsequently enhance memory (i.e., less susceptible to forgetting). In contrast, information activated at the semantic level by a rhyme cue will be less rich given the lack of distinctiveness of rhyming pairs that share a limited number of surface features. Thus, phonological processing contributes less to short-term maintenance and would be suboptimal for promoting retention. The results for delayed performance are an interesting adjunct to the short-term memory findings, indicating that semantic processing during short-term memory benefitted retention.

Conclusion

In sum, it is clear that verbal short-term memory does not operate in isolation, but within the context of a complex linguistic system. In particular, when a word is encountered and processed, verbal short-term memory has access to and relies on both phonological and semantic representations for maintenance and retention. Further, focusing on semantic information in short-term memory leads to better long-term memory, even if there seems to be no immediate advantage. Our novel finding of separable phonological and semantic effects on both short- and long-term retention contrast with the dominant view that phonological and semantic effects reflect short- and long-term memory processes, respectively. More broadly, the results are consistent with viewing verbal short-term memory as an emergent property of the language processing system with rapid access to available word-related knowledge.

References

Acheson, D. J., MacDonald, M. C., & Postle, B. R. (2011). The effect of concurrent semantic categorization on delayed serial recall. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 37*(1), 44-59. https://doi.org/10.1037/a0021205

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63(1),* 1-29. https://doi.org/10.1146/annurev-psych-120710-100422

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

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 575-589. https://doi.org/10.1016/S0022-5371(75)80045-4

Baddeley, A. D., & Larsen, J. (2007). The phonological loop unmasked? A comment on the evidence for a "perceptual-gestural" alternative. *The Quarterly Journal of Experimental Psychology, 60(4),* 497–504. https://doi.org/10.1080/17470210601147572

Baldo, J., Klostermann, E., & Dronkers, N. (2008). It's either a cook or a baker: Patients with conduction aphasia get the gist but lose the trace. *Brain and Language, 105(2),* 134–140. https://doi.org/10.1016/j.bandl.2007.12.007

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

Berryhill, M. E., Richmond, L. L., Shay, C. S., & Olson, I. R. (2012). Shifting attention among working memory representations: Testing cue type, awareness, and strategic control. *The Quarterly Journal of Experimental Psychology, 65(3),* 426-438. https://doi.org/10.1080/17470218.2011.604786

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41(4),* 977-990. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46(3),* 904-911. https://doi.org/10.3758/s13428-013-0403-5

Bunting, M., Cowan, N., & Scott Saults, J. (2006). How does running memory span work? *Quarterly Journal of Experimental Psychology, 59(10),* 1691–1700. https://doi.org/10.1080/17470210600848402

Burrows, D., & Okada, R. (1975). Memory retrieval from long and short lists. *Science, 188(4192),* 1031–1033. https://doi.org/10.1126/science.188.4192.1031

Campoy, G., & Baddeley, A. D. (2008). Phonological and semantic strategies in immediate serial recall. *Memory*, *16(4)*, 329–340. https://doi.org/10.1080/09658210701867302

Campoy, G., Castellà, J., Provencio, V., Hitch, G. J., & Baddeley, A. D. (2015). Automatic semantic encoding in verbal short-term memory: Evidence from the concreteness effect. *The Quarterly Journal of Experimental Psychology, 68(4)*, 759-778. https://doi.org/10.1080/17470218.2014.966248

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences, 24(1)*, 87–114. https://doi.org/10.1017/S0140525X01003922

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684

Fiebach, C., Friederici, A., Smith, E., & Swinney, D. (2007). Lateral inferotemporal cortex maintains conceptual-semantic representations in verbal working memory. *Journal of Cognitive Neuroscience, 19(12)*, 2035–2049. https://doi.org/10.1162/jocn.2007.19.12.2035

Gallo, D. A., Meadow, N. G., Johnson, E. L., & Foster, K. T. (2008). Deep levels of processing elicit a distinctiveness heuristic: Evidence from the criterial recollection task. *Journal of Memory and Language, 58(4)*, 1095–1111. https://doi.org/10.1016/j.jml.2007.12.001

Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology Section A, 54(1)*, 1-30. https://doi.org/10.1080/02724980042000002

Grenfell-Essam, R., & Ward, G. (2012). Examining the relationship between free recall and immediate serial recall: The role of list length, strategy use, and test expectancy. *Journal of Memory and Language, 67(1)*, 106–148. https://doi.org/10.1016/j.jml.2012.04.004

Henson, R., Hartley, T., Burgess, N., Hitch, G., & Flude, B. (2003). Selective interference with verbal short-term memory for serial order information: A new paradigm and tests of a timing-signal hypothesis. *The Quarterly Journal of Experimental Psychology Section A, 56(8)*, 1307-1334. https://doi.org/10.1080/02724980244000747

Hockey, R. (1973). Rate of presentation in running memory and direct manipulation of input-processing strategies. *Quarterly Journal of Experimental Psychology, 25(1)*, 104-111. https://doi.org/10.1080/14640747308400328

Howard, D., & Nickels, L. (2005). Separating input and output phonology: semantic, phonological, and orthographic effects in short-term memory impairment. *Cognitive Neuropsychology, 22(1)*, 42–77. https://doi.org/10.1080/02643290342000582

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(6)*, 685-701. https://doi.org/10.1016/0749-596X(91)90032-F

Hulme, C., Newton, P., Cowan, N., Stuart, G., & Brown, G. (1999). Think before you speak: Pauses, memory search, and trace redintegration processes in verbal memory span. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 25(2)*, 447–463. https://doi.org/10.1037/0278-7393.25.2.447

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(5)*, 1217.

JASP Team. (2020). JASP (Version 0.14.1) [Computer software].

Jefferies, E., Jones, R., Bateman, D., & Lambon Ralph, M. (2005). A semantic contribution to nonword recall? Evidence for intact phonological processes in semantic dementia. *Cognitive Neuropsychology, 22(2)*, 183–212. https://doi.org/10.1080/02643290442000068

Jeffreys, H. (1961). *The theory of probability.* Oxford University Press.

Kowialiewski, B., & Majerus, S. (2018a). Testing the redintegration hypothesis by a single probe recognition paradigm. *Memory, 26(9)*, 1256-1264. https://doi.org/10.1080/09658211.2018.1448420

Kowialiewski, B., & Majerus, S. (2018b). The non-strategic nature of linguistic long-term memory effects in verbal short-term memory. *Journal of Memory and Language, 101*, 64-83. https://doi.org/10.1016/j.jml.2018.03.005

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide (2nd ed.).* Lawrence Erlbaum Associates Publishers.

Majerus, S. (2009). Verbal short-term memory and temporary activation of language representations: The importance of distinguishing item and order information. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 244–276). Psychology Press.

Majerus, S. (2013). Language repetition and short-term memory: an integrative framework. *Frontiers in Human Neuroscience, 7*, 357. https://doi.org/10.3389/fnhum.2013.00357

Martin, N., & Saffran, E. M. (1997). Language and auditory–verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology, 14*, 641–682. https://doi.org/10.1080/026432997381402

Martin, N., Saffran, E. M., & Dell, G. S. (1996). Recovery in deep dysphasia: Evidence for a relation between auditory–verbal STM capacity and lexical errors in repetition. *Brain and Language, 52(1)*, 83–113.

Martin, R. C., & He, T. (2004). Semantic short-term memory and its role in sentence processing: A replication. *Brain and Language, 89(1)*, 76-82. https://doi.org/10.1016/S0093-934X(03)00300-6

Martin, R. C., Lesch, M., & Bartha, M. (1999). Independence of Input and Output Phonology in Word Processing and Short-Term Memory. *Journal of Memory and Language, 41(1)*, 3–29. https://doi.org/10.1006/jmla.1999.2637

Martin, R. C., Shelton, J. R., & Yaffee, L. S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language, 33(1)*, 83-111. https://doi.org/10.1006/jmla.1994.1005

McElree, B. (1996). Accessing short-term memory with semantic and phonological information: A time-course analysis. *Memory & Cognition, 24(2)*, 173-187. https://doi.org/10.3758/BF03200879

McGilly, K., & Siegler, R. S. (1989). How children choose among serial recall strategies. *Child Development, 60, 172–182.* https://doi.org/10.2307/1131083

Murdock, B. B. (1976). Item and order information in short-term serial memory. *Journal of Experimental Psychology: General, 105(2)*, 191–216. https://doi.org/10.1037/0096-3445.105.2.191

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms.* http://www.usf.edu/FreeAssociation/

Nishiyama, R. (2013). Dissociative contributions of semantic and lexical-phonological information to immediate recognition. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 39(2)*, 642-648. https://doi.org/10.1037/a0029160

Nishiyama, R. (2014). Active maintenance of semantic representations. *Psychonomic Bulletin & Review, 21(6)*, 1583-1589. https://doi.org/10.3758/s13423-014-0618-1

Nishiyama, R. (2018). Separability of active semantic and phonological maintenance in verbal working memory. *PloS One, 13(3)*, e0193808. https://doi.org/10.1371/journal.pone.0193808

Papagno, C., & Cecchetto, C. (2019). Is STM involved in sentence comprehension? *Cortex, 112*, 80–90. https://doi.org/10.1016/j.cortex.2018.08.028

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2:

Experiments in behavior made easy. *Behavior Research Methods, 51(1),* 195-203, https://doi.org/10.3758/s13428-018-01193-y

Poirier, M., Saint-Aubin, J., Mair, A., Tehan, G., & Tolan, A. (2015). Order recall in verbal short-term memory: The role of semantic networks. *Memory & Cognition, 43(3),* 489-499. https://doi.org/10.3758/s13421-014-0470-6

Pollack, I., Johnson, I. B., & Knaff, P. R. (1959). Running memory span. *Journal of Experimental Psychology, 57,* 137–146.

Potter, M. C. (2012). Conceptual short term memory in perception and thought. *Frontiers in Psychology, 3,* 113–113. https://doi.org/10.3389/fpsyg.2012.00113

Romani, C., Mcalpine, S., & Martin, R. C. (2008). Concreteness effects in different tasks: Implications for models of short-term memory. *The Quarterly Journal of Experimental Psychology, 61(2),* 292-323. 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. *The Behavioral and Brain Sciences, 26(6),* 709-728. https://doi.org/10.1017/S0140525X03000165

Rypma, B., & Gabrieli, J. D. (2001). Functional neuroimaging of short-term memory: The neural mechanisms of mental storage. *The Behavioral and Brain Sciences, 24(1),* 143–144. https://doi.org/10.1017/S0140525X01503921

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21(2),* 168-175. https://doi.org/10.3758/BF03202729

Schwering, S. C. & Macdonald. M. C. (2020). Verbal working memory as emergent from language comprehension and production. *Frontiers in Human Neuroscience, 14,* 68. https://doi.org/10.3389/fnhum.2020.00068

Shulman, H. G. (1970). Encoding and retention of semantic and phonemic information in short term memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 499–508. https://doi.org/10.1016/S0022-5371(70)80093-7

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31(1),* 137-149. https://doi.org/10.3758/BF03207704

Sternberg, S. (1966). High speed scanning in human memory. *Science, 153,* 652–654. https://doi.org/10.1126/science.153.3736.652

Tan, L. & Ward, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin & Review, 15(3),* 535-542. https://doi.org/10.3758/PBR.15.3.535

Thomson, D., Besner, D., & Smilek, D. (2015). A resource-control account of sustained attention: Evidence from mind-wandering and vigilance paradigms. *Perspectives on Psychological Science, 10(1),* 82–96. https://doi.org/10.1177/1745691614556681

Townsend, J. T., & Fifić, M. (2004). Parallel versus serial processing and individual differences in high-speed search in human memory. *Perception & Psychophysics, 66(6),* 953–962. https://doi.org/10.3758/BF03194987

van den Bergh, D., van Doorn, J., Marsman, M., Draws, T., van Kesteren, E., Derks, K .. & Wagenmakers, E. (2020). A Tutorial on Conducting and Interpreting a Bayesian ANOVA in JASP. *L'Année Psychologique, 120,* 73–96. https://doi.org/10.3917/anpsy1.201.0073

Vergauwe, E., Camos, V., & Barrouillet, P. (2014). The impact of storage on processing: How is information maintained in working memory? *Journal of Experimental Psychology. Learning, Memory, and Cognition, 40(4),* 1072–1095. https://doi.org/10.1037/a0035779

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., …Meerhoff, F. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin and Review, 25,* 58–76. https://doi.org/10.3758/s13423-017-1323-7

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(5),* 1256-1271. https://doi.org/10.1037/0278-7393.25.5.1256