The interaction of word frequency and concreteness in immediate serial recall

LEONIE M. MILLER AND STEVEN ROODENRYS
University of Wollongong, Wollongong, New South Wales, Australia

Word frequency and word concreteness are language attributes that have been shown to independently influence the recall of items in verbal short-term memory (STM). It has been argued that such effects are evidence for the action of long-term memory knowledge on STM traces. However, research to date has not investigated whether these variables interact in serial recall. In two experiments, we examined the behavior of these variables under factorial manipulation and demonstrated that the effect of word frequency is dependent on the level of concreteness of items. Serial recall performance is examined with reference to two explanatory approaches: Walker and Hulme's (1999) dual-redintegration account and language-based models of STM. The data indicate that language-based models are more compatible with the observed effects and challenge the view that frequency and concreteness effects in STM are the products of distinct mechanisms.

Word frequency refers to the level of usage of individual words in written or spoken language. The evidence for a frequency effect in immediate serial recall when words are presented in pure lists (where each item in the list is of the same frequency classification) indicates a clear advantage for high-frequency words (e.g., Allen & Hulme, 2006; Buchner & Erdfelder, 2005; Hulme et al., 1997; Hulme, Stuart, Brown, & Morin, 2003; Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002; Saint-Aubin & LeBlanc, 2005; Stuart & Hulme, 2000). This advantage is maintained across various combinations of presentation (O. C. Watkins & Watkins, 1977) and recall modalities (Hulme et al., 1997; Poirier & Saint-Aubin, 1996), increases with open (rather than closed) pools of words (Roodenrys & Quinlan, 2000), persists when speech rate effects are taken into account (Hulme et al., 1997; Hulme et al., 2003; Majerus & Van der Linden, 2003; Tehan & Humphreys, 1988), and is evident across developmental groups (Majerus & Van der Linden, 2003).

Although specific mechanisms for the effect are not well understood at present (see frequency effects in the recall of mixed lists: Hulme et al., 2003; Morin, Poirier, Fortin, & Hulme, 2006; Saint-Aubin & LeBlanc, 2005) and recent work has highlighted a potential role of stimulus perception (Allen & Hulme, 2006), it is generally acknowledged that the effect arises, at least in part, as a result of the operation of long-term memory (LTM) processes on degraded short-term memory (STM) traces. The redintegration hypothesis (Hulme et al., 1997; Schweikert, 1993), for example, proposes that LTM representations assist recall at the point of retrieval by "cleaning up" remnant phonological traces, through a process that matches partial traces and performs pattern completion. With re-

spect to word frequency, this assistance would appear to be associated with the varying strength of activation of list items as some function of their usage in language and, as such, has been thought to operate on the selection of items in the lexicon as a direct function of use (Hulme et al., 1997; Poirier & Saint-Aubin, 1996; Saint-Aubin & Poirier, 2005), because of prelearned associations between list items (Hulme et al., 2003; Morin et al., 2006; Stuart & Hulme, 2000), or possibly because of differences in available resources for encoding order information, due to differing demands in the encoding of item information (Morin et al., 2006).

Word concreteness is a second variable that has been examined as an LTM influence on short-term serial recall. Highly concrete words refer to material objects, whereas less concrete or abstract words refer to abstract qualities or actions (Walker & Hulme, 1999). The word concreteness effect is the finding that concrete words are recalled better than abstract ones (Allen & Hulme, 2006; Romani, McAlpine, & Martin, 2008; Walker & Hulme, 1999). A related variable, imageability (how easily a mental image of a word can be formed), has been shown to influence recall in a similar way, with high-imageability words demonstrating superior recall, in comparison with low-imageability words (Bourassa & Besner, 1994; Caza & Belleville, 1999; Majerus & Van der Linden, 2003; Tse & Altarriba, 2007). The difference in performance with word sets varying in concreteness is presumed to reflect differential levels of semantic processing in LTM; concrete words have richer semantic representations than do abstract words (Jones, 1985; R. C. Martin & Lesch, 1996; Neath, 1997) and can, therefore, offer stronger supporting activation in the reconstruction of STM traces at retrieval,

---

L. M. Miller, leoniem@uow.edu.au

resulting in better recall (Walker & Hulme, 1999). A second, related explanation is that concrete words have an advantage in that visual features are represented, in addition to verbal ones (Paivio, 1991).

Although both frequency and concreteness have been examined separately in the context of short-term serial recall, no detailed investigation testing their interaction has been reported. Such information would serve to provide a further constraint on STM models and provide insight into the organization of LTM knowledge components as they relate to STM performance.

A framework in which frequency and concreteness effects, although not their interaction, have been considered is the redintegration proposal of Walker and Hulme (1999). In an examination of the behavior of word concreteness, they compared the differences between frequency and concreteness effects, with particular reference to how they are manifested across serial position. In their experiments, the concreteness effect was found to display a consistent difference in recall for medial positions of the list sequence. The authors compared this pattern with those previously observed for word frequency (e.g., Hulme et al., 1997), where there was a typical increase in the effect across serial positions, except for the last. This interaction was taken as a demonstration that LTM effects operate for the recall of items across the entire list, and not merely for items early in the list, as M. J. Watkins (1977) had suggested. The increase in frequency effect for later items was presumed to reflect the advantage high-frequency items enjoyed with respect to redintegration, where more severely degraded items, argued to be affected by the output of early list items (Cowan, 1992), could be reinstated more successfully, due to their accessibility within phonological LTM. Low-frequency words, by comparison, would become more difficult to reconstruct further into the list, due to the relative difficulties in the pattern completion process arising from reduced access to LTM representations, coupled with less intact STM traces on which to base redintegration. In contrast, the absence of a change in the effect of word concreteness across serial positions was seen as evidence that this variable acts on the degree and type of semantic encoding and is more resistant to interference from output processes than are the phonological traces presumed to underpin the frequency effect.

This difference in serial position interactions for frequency and concreteness motivated Walker and Hulme (1999) to extend Hulme et al.'s (1997) description of redintegration to allow for multiple STM codes and LTM representations operating in parallel, with the frequency and concreteness effects outcomes of separate redintegrative processes. The authors, however, did not elaborate on the specific role short-term semantic encoding would play in the restoration of items, or how this contribution would link with redintegrated output from the phonological system (Romani et al., 2008).

Walker and Hulme (1999) also found that the concreteness effect is maintained in backward recall, whereas for word frequency it is eliminated (Hulme et al., 1997). Backward recall is viewed to be a complex procedure that may involve other retrieval strategies (e.g., visuospatial;

Li & Lewandowksy, 1995); however, Hulme et al. (1997) interpreted the failure to observe a frequency effect under these conditions as reflecting retrieval based on semantic information. The preservation of the concreteness effect in backward recall was seen by Walker and Hulme as a substantiation of this idea.

Lastly, Walker and Hulme (1999) used a matching span task on lists varying in concreteness. This recognition-based procedure delivered a null result, which, it was argued, placed the concreteness effect at the output stage of STM recall. However, subsequent research has thrown into question the sensitivity of matching span and serial recognition procedures in identifying the influence of lexical-semantic variables prior to output (Jefferies, Frankish, & Lambon Ralph, 2006; Monnier & Syssau, 2008; Romani et al., 2008). Since the effects of these variables relate predominantly to the retention of item information (Jefferies et al., 2006; Majerus, 2008), matching span tasks that limit item information requirements are likely to eradicate effects that might otherwise arise. To this end, Romani et al. showed that if an open set of items, instead of a closed one, as in Walker and Hulme, is used, a matching span procedure can produce a concreteness effect, and Monnier and Syssau found effects of another semantic variable, word pleasantness, in two serial recognition experiments. Furthermore, Jefferies et al. employed a novel form of matching span task that tested the sensitivity to change in lists varying in frequency and imageability by altering order information at either the item or the list level. This was achieved, in the first instance, by exchanging phonemes of items to alter item identity or, in the second, by the exchange of whole items to change their order within a list. They found that matching span for frequency and imageability is more sensitive to effects of item change than to those of order change, but importantly, identified the effects of these variables in a task that did not require the overt output of items. On the basis of this evidence, it would appear much less tenable that the location of the concreteness effect—and for that matter, the effect of word frequency—is specifically late stage.

Another class of model offering an alternative view of the contribution of LTM knowledge in STM performance has a psycholinguistic basis (e.g., N. Martin & Saffran, 1997; R. C. Martin, Lesch, & Bartha, 1999; Romani et al., 2008; see Figure 1). Language-based models have evolved primarily in response to the observed relationships between language processing and STM performance of neuropsychological patients with various forms of memory impairment. As such, this class of model focuses on the mechanisms and capacities responsible for language processing—more specifically, those relating to speech perception and speech production (Majerus, 2008). It has been argued that, in effect, verbal STM is part of the psycholinguistic architecture that provides short-term storage for language-based activities, through the activation and maintenance of representations across multiple levels (phonological, lexical, semantic, and syntactic) within the LTM network. This occurs through interactive activation (N. Martin, 2008; N. Martin & Saffran, 1997), with the degree of activation dependent on the connection strength
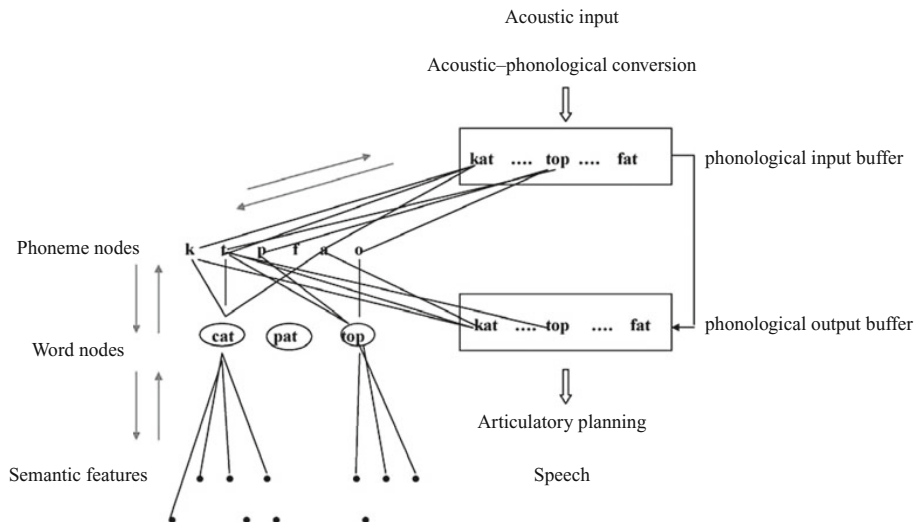
Figure 1. A language-based model of short-term memory. The model includes a placeholder at the phonological level to maintain the serial order of items. From "Concreteness Effects in Different Tasks: Implications for Models of Short-Term Memory," by C. Romani, S. McAlpine, and R. C. Martin, 2008, *Quarterly Journal of Experimental Psychology*, *61*, p. 315. Copyright 2008 by Taylor & Francis Group. Reprinted with permission.

of bidirectional links between representations at adjacent levels, coupled with the rate of decay of activation. The maintenance of activation results from ongoing feedforward and feedback activation to linked nodes at adjacent levels, generating a stable, mutually supporting pattern of activity. The selection of candidates for retrieval is based on the level of activation and the availability of language representations (N. Martin, 2008).

Within this approach, two prevailing views exist regarding the degree to which verbal STM is separate from the LTM knowledge store. The first view holds that STM comprises that subset of LTM (the language processor) that is currently activated (N. Martin & Saffran, 1997), whereas the second proposes separate STM storage in the form of buffers, in addition to activated LTM representations (R. C. Martin et al., 1999; Romani et al., 2008). In the latter case, buffers record representations of items in a sequence at encoding and integrate their contents with representations in the LTM knowledge store via the action of feedback, continually updating representations that feedback from LTM in return. Both the model of R. C. Martin et al. and that of Romani et al. contain three layers of representation—namely, phonological, lexical, and semantic levels. Lexical nodes are connected to semantic nodes, comprising semantic features, and are also connected to phonological representations, in the form of phoneme nodes (Majerus, Lekeu, Van der Linden, & Salmon, 2001). In addition, on the basis of neuropsychological evidence, phonological representations are linked to either input or output phonological buffers that reflect whether they have been derived from speech perception or speech production processes, respectively.

The model of R. C. Martin et al. (1999) proposes an additional lexical-semantic buffer that houses representa-

tions of the lexical-semantic properties of items in serial order. During word recall, candidates for output are selected from this buffer. In contrast, Romani et al. (2008) have proposed the removal of the lexical-semantic buffer and the inclusion of a placeholder operating at the phonological level (N. Martin & Saffran, 1997) to reflect the observed positive association between the maintenance of serial order and the integrity of phonological representations. Order is maintained through the buffered phonological representations that exist for speech perception or speech production. The former is involved in the conversion of acoustic information into phonological representations, whereas the output buffer is responsible for the transformation of phonological representations into an articulatory code. The "virtual" lexical-semantic buffer in this model comprises the activated lexical-semantic representations in the LTM knowledge structure. The encoding order of lexical-semantic representations is directed from the buffered phonological representations in the placeholder.

Romani et al. (2008) have recently examined the influence of concreteness across a range of STM tasks, the results of which have motivated the language-based model described above. They replicated the concreteness effect found in serial recall by Walker and Hulme (1999) and reported a concreteness × serial position interaction consistent with these authors' results—namely, a concreteness effect that is relatively constant across serial positions, except for the last two, where it is absent. Romani et al. observed this pattern to be similar to that in other reports of lexical-semantic effects that are more prominent in the primacy than in the recency portions of the curve (Kintsch & Buschke, 1969; O. C. Watkins & Watkins, 1977). In particular, O. C. Watkins and Watkins found that the

frequency effect was diminished for recency positions. Romani et al. disagreed with Hulme et al.'s (1997) interpretation of the frequency × position interaction, citing the assumption of a widening effect due to disruption of phonological traces from output sources as ad hoc. They stated that another possibility, put forward by Hulme et al. (2003), is that the frequency and concreteness effects are due to different processes and that concreteness depends on redintegration, whereas frequency relies on a different form of co-occurrence mechanism. However, their favored position was that the typical pattern observed for lexical-semantic effects is a narrowing in the last two positions in the list, and they maintained that this attenuation is due to masking of the effect from a strong phonological record.

As was highlighted earlier, Romani et al. (2008) demonstrated that failure to produce a concreteness effect in matching span is a function of task design. They compared performance between open and closed sets of items under control and articulatory suppression conditions. In control conditions, the closed sets replicated the outcome observed by Walker and Hulme (1999); however, the open sets produced a facilitative effect of concreteness. A reduction in the difference in item information between concrete and abstract words, due to repeated presentation, was considered responsible for the lack of an effect with the closed sets. Furthermore, the addition of articulatory suppression to the task yielded results that were negatively offset to those in the control condition; there was no effect of concreteness with open sets and a reversed concreteness effect with closed sets. These results were taken to indicate that under suppression, additional phonological input serves to confuse the phonological record, disrupting the order of items and, therefore, undermining any item identity advantage concreteness can provide. In the case of open sets, the concreteness advantage is nullified. When closed sets are used, order information is so weak that item and order information can become decoupled, an event argued to harm concrete words more than abstract ones. Concrete words will be more likely than abstract words to produce item information that overrides order information, since the latter will experience less strong lexical-semantic activation, and will encourage participants to use available phonological information, however weak. It should be noted that the performance data of the suppression conditions in this study included sizable proportions of participants with near-chance performance, although all the results were replicated with second samples of participants.

Finally, Romani et al. (2008) compared performance in order reconstruction and free recall under control and articulatory suppression conditions to test the size of the concreteness effect in tasks with differing emphases on order and item information and, more generally, to test the proposition that order information relies on the integrity of phonological representations, whereas item identity is influenced by the strength of lexical-semantic codes. The effect of concreteness was greater in free recall than in order reconstruction, reflecting differences in task demands. Concreteness effects were found in the order reconstruction tasks in both control and suppression conditions, once

more implying that lexical-semantic effects operate from the time of presentation onward. Free recall under suppression was less influenced by order information than in the control condition, as indicated by a measure of seriation, supporting the view that serial order is associated with the integrity of phonological representations.

Accordingly, from the perspective of Romani et al. (2008), semantic effects arise because bidirectional links exist between lexical-phonological representations of words and their semantic representations in the long-term knowledge store, and these are activated from the time of presentation onward. It is claimed that words with richer semantic representations will provide stronger feedback to lexical-phonological representations, which, in turn, will supply feedback activation to the buffered phonological representations. The consequent stabilization should benefit concrete words over abstract ones and should be more evident in circumstances when phonological representations might otherwise be weakened. Furthermore, word frequency is assumed to affect the activation levels of the lexical nodes (R. C. Martin et al., 1999); high-frequency items have stronger lexical activations, and these serve to deliver stronger feedback to the connected representations at the phonological level.

The aim of the following experiments was to test the nature of the LTM contributions of word frequency and word concreteness to STM recall performance when these variables are manipulated within the same experiment. Both frameworks presented here can accommodate the possibility of a frequency × concreteness interaction, although the arguments for its existence are quite different. Language-based models explicitly incorporate the interaction of frequency and concreteness via the structure of long-term knowledge and continual access to long-term representations throughout the encoding, retention, and retrieval of items (N. Martin, 2008). In contrast, within the dual-redintegration model of Walker and Hulme (1999), an interaction might be generated from the increasing numbers of items that can be successfully redintegrated from both semantically and phonologically based processes when frequency and concreteness are greater; recall of an item presumably requires successful redintegration from a single system, and high-frequency words that are also highly concrete would produce redundant information when both systems yield the same output. Should no interaction arise, this would be harder to explain within the language-based view, given the specific links between these types of representations. The dual-redintegration approach might have greater scope to accommodate such a finding, in terms of a "one-shot" process, where an item could be redintegrated by one of the phonological or semantic mechanisms, but not by both.

From a qualitative perspective, these approaches might anticipate different forms of serial position interactions for each of the variables involved. The Walker and Hulme (1999) model relies on a qualitative difference between the interactions, with serial position for frequency and word concreteness reflecting the different processes that produce them, whereas Romani et al. (2008) suggested that interactions of lexical-semantic variables with serial posi-

tions should be of the same form—that is, marked by a reduction in effect for the recency positions of the curve.

The first experiment was run to investigate the variations in recall performance of lists of words in a factorial manipulation of word frequency and word concreteness. Items were visually presented, with the specific aim of examining the behavior of these lexical-semantic variables in a context in which, arguably, phonological information is weaker, relative to auditory presentation (Tolan & Tehan, 1999), and possibly less likely to mask lexical-semantic effects (Romani et al., 2008).

EXPERIMENT 1

Method

Participants. Forty University of Wollongong undergraduate and postgraduate students (6 males and 34 females) either participated for course credit or volunteered to participate. The mean age of the sample was 21.7 years ($SD = 6.5$ years), and all the participants had English as a first language.

Materials. Four stimulus sets of 12 words each were selected such that a factorial manipulation of word frequency and concreteness was achieved (namely, word sets contained items that were low frequency and low concreteness [LFLC], low frequency and high concreteness [LFHC], high frequency and low concreteness [HFLC], or high frequency and high concreteness [HFHC]). Word frequency ratings were derived from the Celex database (Baayen, Piepenbrock, & Van Rijn, 1993). Frequency ratings were the composite of database entries for the same orthography and across those entries with the same word identification number, so that, for example, the frequency counts for *bird* and *birds* were combined. It is argued that these frequency counts are more reflective of any frequency-based effects in serial recall, given the unrelatedness of list items. Concreteness ratings were sourced from the MRC Psycholinguistic Database (Coltheart, 1981). The sets were matched on phonological neighborhood using values calculated from the Celex database, the number of phonemes, and number of letters of set items. In addition, the stimulus sets were matched on phonological similarity. Mueller, Seymour, Kieras, and Meyer (2003) have developed a formal methodology for quantifying phonological dissimilarity and, hence, similarity between pairs of words (PSIMETRICA). The method involves the decomposition of words by examining syllable structure and phoneme clusters. Since the words in the present experiment were monosyllabic, effects of syllable structure will not be considered here. In intrasyllabic analysis, the vowel nucleus is identified first and defines phoneme clusters that constitute the onset component preceding it and the coda following it. Corresponding clusters of phonemes (onset, nucleus, and coda) are aligned between word pairs, and the phonological features of these elements are compared, resulting in three dissimilarity measures. A dissimilarity profile for a set of words is defined to be the average of these dimensions across all pairwise comparisons in the set. A monosyllabic version of the methodology of PSIMETRICA was written into an Excel Visual Basic program to determine dissimilarity ratings on onset, nucleus, and coda dimensions. Profiles of average dissimilarity for each word in the set, when compared with the others, were obtained. The values lie on a scale of zero to one, where zero indicates identical phonology and numbers closer to one indicate greater dissimilarity.

A MANOVA was performed on the word sets (Wilks's lambda, $\Lambda = .005$, $p < .001$), the analysis demonstrating that (1) high-frequency sets differed significantly from low-frequency sets [$F(3,44) = 37.69$, $MS_e = 1,877.083$, $p < .001$; Tukey's HSD homogeneous subset analysis identified high- and low-frequency conditions as significantly different], and (2) high-concreteness sets differed significantly from low-concreteness sets [$F(3,44) = 565.98$, $MS_e = 689.233$, $p < .001$; Tukey's HSD homogeneous subset analysis identified high- and low-concreteness conditions as significantly

different]. No sets differed significantly from the others with respect to phonological neighborhood [$F(3,44) = 0.76$, $MS_e = 207.087$, $p = .522$], number of phonemes [$F(3,44) = 0.21$, $MS_e = 0.261$, $p = .887$], number of letters [$F(3,44) = 1.66$, $MS_e = 0.402$, $p = .189$], or any of the phonological similarity measures [onset, $F(3,44) = 1.66$, $MS_e = 0.002$, $p = .189$; nucleus, $F(3,44) = 1.70$, $MS_e = 0.002$, $p = .181$; coda, $F(3,44) = 0.79$, $MS_e = 0.002$, $p = .504$]. The word sets and word attributes are presented in the Appendix.

Pseudorandom six-word lists were generated for each condition and each participant, so that no word appeared twice in the same list and each word appeared once in each serial position across the set of lists. The order of conditions (four sets of 12 lists) was factorially counterbalanced across participants to minimize additional influences from learning and practice effects.

Procedure. All the participants were tested individually. The total time to complete the experiment was approximately half an hour. Testing was controlled via the DMDX program (Forster & Forster, 1999) run on an IBM-compatible computer and commenced after two practice trials. Initiation of each trial occurred when the participant pressed the space bar of the keyboard. The program would then present each word in the trial at a rate of one word per second. Words were presented in the center of a black screen and were in white, 40-point Times New Roman font. After the sixth word, a recall prompt appeared ("?????") indicating that the participants should commence recall. Spoken recall was according to strict serial recall criteria; that is, (1) words were recalled in order of presentation, (2) if a word could not be recalled, the participant would indicate by saying "pass," and (3) previous items were not to be recalled after the participant had moved on to successive items in the list.

Results

For each participant and each condition, recall was scored and collapsed across trials to provide the number of correct items by serial position. Correct recall was scored using a strict criterion; namely, responses were considered correct if a word was recalled in the position corresponding to its serial order in presentation. The mean number of correctly recalled items by serial position and condition is shown in Figure 2. Performance collapsed across serial positions was greatest for the HFHC condition ($M = .670$, $SD = .149$), then the HFLC condition ($M = .644$, $SD = $
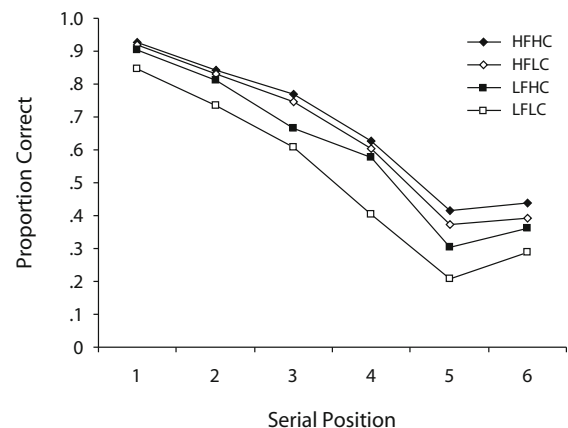

Figure 2. Serial recall of words as a function of frequency and concreteness with visual presentation and spoken recall. HFHC, high frequency, high concreteness; HFLC, high frequency, low concreteness; LFHC, low frequency, high concreteness; LFLC, low frequency, low concreteness.

.148), the LFHC condition ($M = .602$, $SD = .151$), and finally, the LFLC condition ($M = .514$, $SD = .130$).

Serial recall. An alpha level of .05 was applied to the following statistical tests. A $2 \times 2 \times 6$ (frequency $\times$ concreteness $\times$ serial position) repeated measures ANOVA was conducted on the serial recall data. All the main effects were significant [frequency, $F(1,39) = 62.80$, $MS_e = 0.037$, $p < .001$; concreteness, $F(1,39) = 27.80$, $MS_e = 0.028$, $p < .001$; serial position, $F(5,195) = 160.67$, $MS_e = 0.110$, $p < .001$]. Thus, high-frequency words were recalled better than low-frequency words, concrete words were recalled better than abstract words, and the performance across serial position revealed a typical pattern for visually presented material: a decline in performance from the first item through to the fifth item, followed by a modest increase at the last position. The frequency $\times$ concreteness interaction was found to be significant [$F(1,39) = 10.77$, $MS_e = 0.022$, $p = .002$].

The frequency effect was smaller for concrete words than for abstract words, or alternatively, the concreteness effect was smaller for high-frequency words than for low-frequency words. The frequency $\times$ serial position interaction was also significant [$F(5,195) = 3.81$, $MS_e = 0.018$, $p = .006$], as was the concreteness $\times$ serial position interaction [$F(5,195) = 2.66$, $MS_e = 0.011$, $p = .035$]. These results appear to be driven by a widening of the respective effects for the latter serial positions, together with the possibility of a ceiling effect operating on the first position. The three-way interaction was nonsignificant [$F(5,195) = 1.76$, $MS_e = 0.013$, $p = .122$].

Serial position interactions. The influence of the ceiling effect on the interactions with serial position was investigated by reanalyzing the data for the last five serial positions. Whereas the frequency $\times$ position interaction remained significant [$F(4,156) = 2.50$, $MS_e = 0.015$, $p = .045$], the concreteness by position interaction was marginal [$F(4,156) = 2.12$, $MS_e = 0.010$, $p = .081$], suggesting that performance in the first position artificially contributed to the results in the full analysis. With this in mind, the differences between effects for the primacy and recency portions of the curves were compared. The average difference in the frequency effect for the primacy positions ($M = .079$, $SD = .076$) was, after Bonferroni correction for multiple comparisons ($\alpha = .025$), not significantly different from that for the recency positions

($M = .119$, $SD = .120$) [$t(39) = -2.03$, $p = .049$]. In the case of word concreteness, the average difference in the primacy positions ($M = .039$, $SD = .077$) was smaller than the average difference in effect for recency positions ($M = .075$, $SD = .090$) [$t(39) = -2.35$, $p = .024$]. It appears unlikely in this instance, despite limitations to the magnitude of differences in the primacy portion of the curve, that effects for either variable would be larger in this region than in the recency portion of the curve.

Item error analysis. Error analysis was performed on the data to further examine the impact that word frequency and concreteness may have on the recall of items. The responses for each trial were classified as correct (the correct item in the correct serial position), an order error (the item recalled corresponded to a word that was presented elsewhere in the trial), or an item error (the item recalled did not match any item presented in the trial or was a repetition of a previously recalled list item). Nonrepetition item errors were further broken down according to the classification suggested by Allen and Hulme (2006), each error being an omission (when participants said "pass" to an item or indicated that they could not recall an item in that position), an intraset intrusion (ISI; participants recalled an item from within the current experimental set but outside the list presented), an intraexperiment intrusion (IEI; participants recalled an item from one of the other experimental sets), an extraexperiment intrusion (EEI; participants recalled an item that did not correspond to any item within any of the experimental sets), or a phonological approximation (PA; participants recalled an item that approximated a list item by at least 50% of the presented item's phonemes). The proportion of errors of each type, collapsed across serial position and participants, is given in Table 1, together with the proportion of items correct.

The proportions of order errors were conditionalized by dividing the total number of order errors by the number of items correctly recalled, regardless of order. This procedure avoids confounding different levels of order memory with differing levels of item memory (Murdock, 1976; Poirier & Saint-Aubin, 1996; Saint-Aubin & Poirier, 1999). The resultant conditionalized rates were .125 for HFHC, .116 for HFLC, .134 for LFHC, and .142 for LFLC. A 2×2 repeated measures ANOVA showed that neither frequency [$F(1,39) = 2.14$, $MS_e = 0.006$, $p = .152$], nor concreteness [$F(1,39) = 0.00$, $MS_e = 0.005$,

Table 1
Proportions of Items Correctly Recalled and Proportions of
Different Error Categories by Condition in Experiment 1

Condition	Correct	Order	Errors						
						Item Errors			
			Total	Rep	Omis	EEI	ISI	IEI	PA
HFHC	.670	.090	.240	.003	.213	.000	.023	.000	.001
HFLC	.645	.080	.275	.004	.244	.002	.017	.004	.005
LFHC	.603	.086	.311	.005	.279	.002	.019	.002	.005
LFLC	.515	.080	.405	.002	.342	.003	.019	.026	.012

Note—Rep, repetitions; Omis, omissions; EEI, extraexperimental intrusions; ISI, intraset intrusions; IEI, intraexperimental intrusions; PA, phonological approximations; HFHC, high frequency, high concreteness; HFLC, high frequency, low concreteness; LFHC, low frequency, high concreteness; LFLC, low frequency, low concreteness.

$p = .982$], nor the frequency × concreteness interaction [$F(1,39) = 0.96$, $MS_e = 0.003$, $p = .333$] affected memory for order.

A 2 × 2 repeated measures ANOVA was performed on the total item errors by condition. This analysis revealed significant main effects of frequency [$F(1,39) = 118.02$, $MS_e = 0.122$, $p < .001$] and concreteness [$F(1,39) = 58.02$, $MS_e = 0.104$, $p < .001$] and a significant frequency × concreteness interaction [$F(1,39) = 12.27$, $MS_e = 0.103$, $p = .001$]. Higher frequency and more concrete words yielded better memory for items, and the change in effect of frequency and concreteness was such that the effect of one variable was smaller at higher levels of the second variable.

Of all the item error categories, omissions formed the largest contribution, and these were analyzed separately to ascertain the sensitivity of items as a function of condition to the likelihood that recall would fail altogether. This analysis yielded significant main effects—namely, frequency [$F(1,39) = 79.84$, $MS_e = 0.123$, $p < .001$] and concreteness [$F(1,39) = 38.50$, $MS_e = 0.084$, $p < .001$]—and the interaction reached borderline significance [$F(1,39) = 4.10$, $MS_e = 0.092$, $p = .050$]. The remaining error categories were not analyzed by category, since the data were considered too sparse and the totals too small to be meaningful. Repetitions, EEI, and ISI did not vary much across conditions, whereas the LFLC rates for both IEI and PA categories are many times those of any other condition. This is likely to be due to the substitution of the LFLC item *truce* with *truth* (an item from the HFLC set) on some occasions.

Discussion

Several results of this experiment are consistent with those in previous research. The main effects were in line with prior findings; high-frequency words were recalled better than low-frequency words (Allen & Hulme, 2006; Hulme et al., 1997; Hulme et al., 2003; Majerus & Van der Linden, 2003; Poirier & Saint-Aubin, 1996; Roodenrys & Quinlan, 2000; Stuart & Hulme, 2000; Tehan & Humphreys, 1988; O. C. Watkins & Watkins, 1977), concrete words were recalled better than abstract ones (Allen & Hulme, 2006; Romani et al., 2008; Walker & Hulme, 1999), and the effect of serial position was broadly consistent with that in experiments in which visual presentation of words was used (Poirier & Saint-Aubin, 1996; Roodenrys & Quinlan, 2000; O. C. Watkins & Watkins, 1977).

Importantly, however, this experiment has also identified a new characteristic of LTM influence on STM recall: the frequency × concreteness interaction. The size of the effect of word frequency will be dependent on the concreteness of the items used, and vice versa. Support for the direction of this interaction is given by R. C. Martin et al. (1999), who reported the results for a group of controls on a list repetition task involving sets of words that varied in terms of imageability and frequency, when investigating the STM performance of an anomic patient M.S. They did not report results for the controls by each condition, collapsing performance across imageability and frequency, but did report a marginally significant interaction, consistent with the direction observed in the present experiment.

The presence of an interaction between lexical-semantic variables is seen to be a natural outcome of the architecture of language-based models (N. Martin, 2008), although it suggests that feedback activation to phonological representations is functionally limited, in that activation from items that are both highly frequent and highly concrete is not substantially greater than the lexical-semantic activation from items either highly frequent or highly concrete. This result can be likewise accommodated within the dual-redintegration framework of Walker and Hulme (1999) if both redintegrative mechanisms operate on item retrieval. In this case, the interaction occurs because of the greater proportion of items successfully redintegrated by both systems when frequency and concreteness are high.

Analysis of the present experiment identified that neither variable affected memory for order. For word frequency, this has been the general finding when recall was immediate (Allen & Hulme, 2006; Hulme et al., 2003; Morin et al., 2006; Poirier & Saint-Aubin, 1996; Saint-Aubin & Poirier, 2005; Stuart & Hulme, 2000), although two experiments have been reported in which a small difference in memory for order was detected (Hulme et al., 2003, Experiment 2; Tse & Altarriba, 2007). Word concreteness has also been reported not to be associated with memory for order (Walker & Hulme, 1999); however, Allen and Hulme (Experiment 1) identified a small advantage for concrete words. It would appear that order effects, when present, are slight and may arise from demand characteristics of individual experiments or variations in participant pools. Despite such anomalies, it is clear that the results for order errors, in combination with those for item recall as reflected in total item errors, identify the impact of lexical-semantic variables as one affecting item memory (Hulme et al., 1997; Hulme et al., 2003; Poirier & Saint-Aubin, 1996; Romani et al., 2008; Walker & Hulme, 1999).

The findings that both variables and their interaction influence the total number of item errors is repeated in the pattern for the dominant item error category of omissions. Both main effects have been reported previously (Allen & Hulme, 2006; Hulme et al., 1997; Hulme et al., 2003; Stuart & Hulme, 2000). In terms of redintegration, these results reflect the relative advantage higher levels of LTM variables can provide in the retrieval of items at recall: a greater proportion of items degraded beyond a recoverable state for phonological or semantic redintegration when word frequency is low and items are abstract; intermediate levels of omission when items are either high frequency or highly concrete but not both; and the least proportion of omissions when word frequency is high and items are concrete. This last circumstance corresponds to the event where a degraded STM trace is most likely to be recovered from either system. Language-based models would posit that higher levels of concreteness and word frequency assist the maintenance of phonological representations at encoding and during retention. Such items are less likely to be unidentifiable by the point of recall, because they provide greater lexical-semantic

feedback activation to phonological traces across the life of the trial.

Both variables were found to interact with serial position; however, there are a number of qualifications to be considered before any conclusions regarding these effects can be drawn. First, the reanalysis of the data for the last five serial positions showed that the frequency × serial position interaction retained significance, whereas the concreteness × serial position interaction failed, implying that the latter had been more influenced by a ceiling effect in the full analysis. The combination of presentation and recall modalities used in the present experiment is comparatively rare; however, Schweikert, Chen, and Poirier (1999) reported a reanalysis of data from a similar investigation by Roodenrys and Hulme testing the frequency effect in five-item lists, where they found a greater effect for the middle serial positions. The frequency × serial position interaction generated in the present instance could be viewed in a similar way. However, other complications in this case include the anomalies present in the third and fourth serial positions for the LFHC condition and their influences on the frequency and concreteness effects across positions, and the possible extension of a ceiling effect operating on the second serial position. Furthermore, the smaller concreteness effect might be constrained by power issues with respect to the detection of patterns across the recall curve. It would appear that in the present case, in which both effects have been observed within the same experiment, the argument that frequency and concreteness serial position interactions are different results from concessions in data interpretation, rather than from the presence of distinctly different patterns. Although it is possible that contextual factors, such as the choice of combination of presentation and recall modalities, may have contributed to a lack of clear difference between serial position interactions in this experiment, as it stands, the evidence for distinct redintegrative systems, as Walker and Hulme (1999) would predict, is not overly convincing.

However, the examination of serial position in terms of effects in primacy and recency portions of the curve failed to support predictions from the language-based viewpoint, since, in each case, the effects in the primacy positions of serial recall were not greater than those in the recency positions (Romani et al., 2008). These comparisons are also constrained to some degree by the presence of ceiling effects in the early part of the curve and the anomalies associated with recall in the LFHC condition; however, it does appear unlikely that greater effects would have otherwise resulted in the primacy region. It is also possible that design features in this task—namely, presentation and recall modalities—have contributed to the extent to which lexical-semantic effects have been displayed across the curve by limiting the masking from the phonological record for the latter serial positions (Romani et al., 2008). Since previous opportunities to observe performance in this context have been few, and language-based models have grown from an auditory-verbal perspective (e.g., N. Martin, Saffran, & Dell, 1996), such impacts of task constraints may not have been widely considered.

EXPERIMENT 2

Experiment 2 was a replication of Experiment 1 with a different choice of presentation and recall modalities. The motivation for this experiment was twofold, relating to the possibility of a rival explanation for the existence of the frequency × concreteness interaction in Experiment 1 and, second, a desire to reduce the ambiguity surrounding the serial position interactions observed in the experiment.

Although Experiment 1 had used visual presentation with a view to placing greater emphasis on lexical-semantic representations, short-term traces from visually presented material are known to rapidly degrade, and with them their capacity to support recall (Tolan & Tehan, 1999). Furthermore, one argument regarding the differences between concrete and abstract words involves additional sensory, as well as semantic, encoding available for concrete but not abstract items (Paivio, 1991). It is possible, therefore, that the presentation modality may have encouraged the use of different recall strategies across conditions, offering an alternative explanation for the observed interaction between frequency and concreteness. For example, when items are highly concrete, a visual imagery strategy might attenuate the difference between high- and low-frequency words, since reliance on processes reflecting the contribution of lexical-semantic effects would be reduced. A visual imagery strategy would be much less effective with abstract words, however, leading to a greater dependence on whatever phonological and semantic features of items are retained. Differences in frequency effect between these conditions would be observed in these circumstances.

Hence, it is possible that visual presentation emphasizes a nonsystematic contribution that is less likely to exist in experiments in which auditory presentation is used. In these cases, phonological encoding is direct and has greater duration than does encoding from visually presented stimuli (Penney, 1989; Tolan & Tehan, 1999). We were interested to determine whether the interaction between frequency and concreteness would be replicated under conditions in which phonological encoding and, arguably, phonologically based recall strategies had greater influence.

Given the level of performance observed in Experiment 1, the change in presentation modality raised the possibility that ceiling effects for early serial positions would eventuate. Walker and Hulme (1999) had shown that, for concreteness stimuli with auditory presentation, the first few items in a list are recalled less well when written, instead of spoken, recall is employed. Thus, in order to minimize the likelihood of ceiling effects operating in Experiment 2, written recall was chosen to capture output.

This selection of presentation and output modalities has the additional benefit of testing performance for frequency and concreteness under conditions that have been reported elsewhere, for both frequency (O. C. Watkins & Watkins, 1977) and concreteness (Romani et al., 2008; Walker & Hulme, 1999). This allows for a more direct comparison of the behavior of effects across serial position in tasks

with similar design features and tests the generalizability of the current task manipulations with reference to these.

Method

Participants. Approximately 120 University of Wollongong undergraduate students participated in the experiment as a class exercise. From this initial pool, the data set was reduced for several reasons. First, some data were lost due to answer sheet formatting problems. From the remaining participants, all those who had English as a second language were removed. Any participants who had inadvertently skipped a trial (see the Procedure section below) were omitted. Because this was a class-based exercise, experimental list files were used multiple times (five unique files of each of four counterbalanced orders of conditions). Furthermore, the number of times each experimental list file was used varied according to class size (up to 20 students). In order to determine data for the final analysis, the number of unique files and number of participants who were tested with these files were matched against counterbalancing constraints within the remaining participant pool. List files that had fewer than 3 eligible participants were eliminated; this determined the greatest possible number of list files per counterbalancing condition. In cases in which there were more eligible participants for a particular list file or there were more list files for a particular counterbalancing condition, random selection of participants or files was used to determine inclusion. This process resulted in a final participant pool of 48 students (9 male, 39 female) with a mean age of 22.0 years ($SD = 3.2$ years).

Materials. The stimulus sets were 20 of the list sets used in Experiment 1. However, since auditory presentation of list items was being used in this experiment, some additional considerations were required. Some items were homophones—for example, *sun* and *steak*—and accordingly, it was important to establish that any changes arising from the presence of homophones in the stimuli were unlikely to change the factorial manipulation of the independent variables. This was done by examining the frequency counts of any homophones of list items and altering their values by summing homophone frequencies. In addition, where possible (and in the majority of instances, this was the case), the concreteness ratings of the homophones were extracted, and weighted average concreteness ratings, using the individual frequency counts as the basis for the weighted contributions, were derived.

A MANOVA was performed on the amended word sets (Wilks's lambda, $\Lambda = .023$, $p < .001$), the analysis demonstrating once again that (1) high-frequency sets differed significantly from low-frequency sets [$F(3,44) = 27.925$, $MS_e = 2,784.360$, $p < .001$; Tukey's HSD homogeneous subset analysis identified high- and low-frequency conditions as significantly different], and (2) high-concreteness sets differed significantly from low-concreteness sets [$F(3,44) = 95.29$, $MS_e = 3,537.866$, $p < .001$; Tukey's HSD homogeneous subset analysis identified high- and low-concreteness conditions as significantly different].

The stimuli were digitally recorded in a native Australian female voice, using the ProTools LE software on a G4 Macintosh computer converted to sound files. A response sheet that asked for demographic information (age, sex, first language) and provided spaces for written recall of list items was used for data collection.

Procedure. The participants were tested in groups of up to 20, and testing took approximately half an hour to complete. Each group was given oral instructions, supported by overhead material, as to how to correctly participate in the experiment. The experiment was conducted on individual IBM-compatible computers connected to headphones, controlled by purpose-written software. To reinforce adherence to correct participation, an instructions screen appeared prior to the participants' commencing the experiment. No practice trials were given. Initiation of each trial occurred when the participants pressed the left mouse button. The program would then present each word in the trial at a rate of one word per second. After the sixth word, an auditory prompt (a *beep*) was played to indicate the commencement of the recall phase. Written recall was according

to the strict serial recall criteria outlined in Experiment 1 and was self-paced. Response sheets were collected at the conclusion of each experimental session.

Results

Recall was calculated in the same manner as that in Experiment 1. The mean number of correctly recalled items by serial position and condition is shown in Figure 3. The ranking of condition by recall performance collapsed across serial positions replicated that in Experiment 1—namely, recall being greatest for the HFHC condition ($M = .760$, $SD = .139$), followed by the HFLC condition ($M = .711$, $SD = .149$), the LFHC condition ($M = .690$, $SD = .127$), and, lastly, the LFLC condition ($M = .595$, $SD = .139$).

Serial recall. An alpha level of .05 was again the criterion for significance. A $2 \times 2 \times 6$ (frequency × concreteness × serial position) repeated measures ANOVA was conducted on the serial recall data. Once again, all the main effects were significant [frequency, $F(1,47) = 61.47$, $MS_e = 0.041$, $p < .001$; concreteness, $F(1,47) = 46.67$, $MS_e = 0.032$, $p < .001$; and serial position, $F(5,235) = 130.85$, $MS_e = 0.089$, $p < .001$]. The serial position curves for this experiment were consistent with those in other experiments that had auditory presentation of material and written recall in that, when compared with those for Experiment 1, there was a marked recency effect for each condition. The frequency × concreteness interaction was found to be significant [$F(1,47) = 5.50$, $MS_e = 0.028$, $p = .023$], and this interaction manifested in the same way as in Experiment 1. The frequency × serial position interaction was significant once more [$F(5,235) = 7.14$, $MS_e = 0.011$, $p < .001$]. Comparing recall for high- and low-frequency conditions, the frequency effect increased for the first three positions and stayed constant until the last position, where it closed again. The concreteness × serial position interaction was significant [$F(5,235) = 4.71$, $MS_e = 0.016$, $p = .001$]. The difference in recall be-
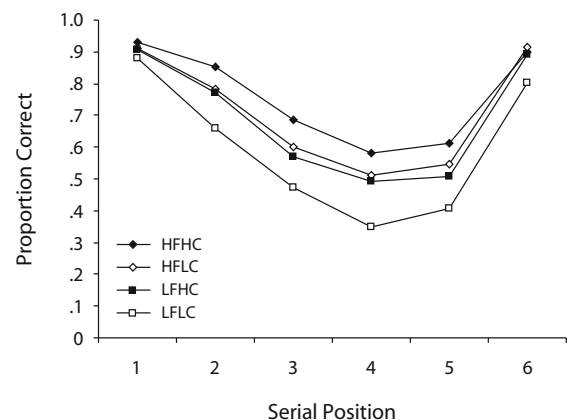

Figure 3. Serial recall of words as a function of frequency and concreteness with auditory presentation and written recall. HFHC, high frequency, high concreteness; HFLC, high frequency, low concreteness; LFHC, low frequency, high concreteness; LFLC, low frequency, low concreteness.

tween high- and low-concreteness words increased to the fourth serial position and then decreased for the final two positions. The three-way interaction was nonsignificant [$F(5,235) = 1.69$, $MS_e = 0.013$, $p = .156$].

Serial position interactions. The level of performance in the HFHC condition for the first serial position indicated that there could be a ceiling effect operating on the distribution of the data. As an additional check, to discount the interaction of each variable with serial position as a consequence of range-restricted data, a further repeated measures ANOVA on the last five serial positions was run. Interactions of each variable with serial position persisted, despite the removal of the first position data; the frequency × serial position interaction was significant [$F(4,188) = 3.16$, $MS_e = 0.011$, $p = .017$], as was the concreteness × serial position interaction [$F(4,188) = 2.74$, $MS_e = 0.015$, $p = .038$], and the three-way interaction remained nonsignificant [$F(4,188) = 1.49$, $MS_e = 0.013$, $p = .216$]. All other results were consistent with those determined from the full data. Despite the persistence of these interactions with serial position, a strong case can be made that these are a direct reflection of the changes in effects in the last serial position. It is apparent that both frequency and concreteness effects throughout the medial positions change very little.

The effects observed for primacy and recency portions of the curve were compared for both frequency and concreteness. The average frequency effect for the primacy positions ($M = .084$, $SD = .104$) was not different from the average effect for recency positions ($M = .103$, $SD = .088$) [$t(47) = -1.39$, $p = .172$]. The comparison for the concreteness effect also yielded a nonsignificant result [$t(47) = -0.34$, $p = .734$], since the average effects for the primacy ($M = .070$, $SD = .014$) and recency ($M = .075$, $SD = .084$) portions of the curve were the same.

Item error analysis. The same classification system as that outlined for Experiment 1 data was employed in this experiment. A summary of the categories for item errors is presented in Table 2.

Conditionalizing order errors yielded rates of .117 for HFHC, .150 for HFLC, .129 for LFHC, and .173 for LFLC. A 2 × 2 repeated measures ANOVA showed significant main effects for frequency [$F(1,47) = 4.60$, $MS_e = 0.003$, $p = .037$] and concreteness [$F(1,47) = 20.54$, $MS_e = 0.003$, $p < .001$]. Thus, significantly more order errors oc-

curred in the recall of lists with low-frequency words and abstract words, respectively. The frequency × concreteness interaction, however, was nonsignificant [$F(1,47) = 0.47$, $MS_e = 0.003$, $p = .495$].

The total item errors were analyzed using a 2 × 2 repeated measures ANOVA. This analysis indicated that the main effects were significant [frequency, $F(1,47) = 78.71$, $MS_e = 0.188$, $p < .001$; concreteness, $F(1,47) = 43.58$, $MS_e = 0.095$, $p < .001$]. A significant frequency × concreteness interaction [$F(1,47) = 8.10$, $MS_e = 0.007$, $p = .007$] was also observed. These results replicated those for Experiment 1—namely, that low-frequency and abstract words generated more item errors than did high-frequency and concrete words, respectively, and that the difference in errors between low- and high-frequency words was smaller for concrete words than for abstract words.

As in Experiment 1, omissions made up the bulk of item errors and were analyzed separately. Patterns for omissions exhibited significant main effects for both frequency [$F(1,47) = 40.68$, $MS_e = 0.148$, $p < .001$] and concreteness [$F(1,47) = 28.64$, $MS_e = 0.078$, $p < .001$]. More omissions occurred with low-frequency than with high-frequency items, and more omissions were evident for abstract items than for concrete items. The interaction, however, was nonsignificant [$F(1,47) = 0.92$, $MS_e = 0.083$, $p = .342$]. The failure to find an interaction in this case appears to have been due to a floor effect on word sets with at least one high-end attribute.

The effects of other item errors were small, in comparison with omissions, and were not subjected to inferential tests. However, it should be noted that more EEI errors were committed in low-frequency conditions, and an order magnitude difference was present in IEI errors, driven predominantly by the substitution of *truth* for *truce*.

Discussion

This experiment confirmed the expected main effects of frequency and concreteness and yielded serial position curves typical for a task employing auditory presentation and written recall (Romani et al., 2008; Walker & Hulme, 1999; O. C. Watkins & Watkins, 1977). In addition, a frequency × concreteness interaction was found in Experiment 2 and followed the pattern observed in Experiment 1. The effect size for each variable was modulated by the level of the second, and effects diminished when

Table 2
Proportions of Items Correctly Recalled and Proportions of
Different Error Categories by Condition in Experiment 2

			Errors						
				Item Errors					
Condition	Correct	Order	Total	Rep	Omis	EEI	ISI	IEI	PA
HFHC	.760	.097	.143	.010	.092	.004	.025	.001	.011
HFLC	.711	.120	.169	.010	.121	.003	.023	.002	.013
LFHC	.690	.098	.212	.016	.144	.011	.025	.003	.010
LFLC	.595	.119	.286	.011	.187	.008	.035	.029	.011

Note—Rep, repetitions; Omis, omissions; EEI, extraexperimental intrusions; ISI, intraset intrusions; IEI, intraexperimental intrusions; PA, phonological approximations; HFHC, high frequency, high concreteness; HFLC, high frequency, low concreteness; LFHC, low frequency, high concreteness; LFLC, low frequency, low concreteness.

items were frequently used or concrete. Replication across two tasks varying in input and output requirements suggests that this interaction is a stable feature of STM recall. In addition, its presence in a context in which a recall strategy reliant on the visual features of words would be less productive—namely, when phonological traces were stronger and the additional burden of written recall would encourage efficient output—reinforces the likelihood that the interaction between these lexical-semantic variables was an outcome of orthodox STM activity.

In this experiment, and unlike in Experiment 1, variations in memory for order across conditions were observed. Memory for order was influenced independently by both variables; however, the magnitudes of these variations were small, between 1% and 2% for word frequency and between 3% and 4% for word concreteness. Small order effects for frequency and concreteness have been reported in other experiments (Allen & Hulme, 2006; Roodenrys et al., 2002; Tse & Altarriba, 2007; Walker & Hulme, 1999). More generally, order effects have been interpreted as additional evidence that the locus of effect for lexical-semantic variables extends well beyond the scope outlined by late-stage redintegration theories (Monnier & Syssau, 2008; Saint-Aubin, Ouellette, & Poirier, 2005).

The overall results for item memory replicated those for Experiment 1. Total item errors indicated that the frequency × concreteness interaction was predominantly a product of better item memory for high-frequency and concrete words that was limited in effect when items were both concrete and frequent in language use. The effects of frequency and concreteness were evident with the pattern of omissions, presumably reflecting the relative advantages higher levels of each variable provide in either restoring or retaining STM traces. The failure to find an interaction in this error category appears to have been due to the low levels of omissions for high-frequency and concrete items and is a testament to the durability of short-term traces that arise from auditory presentation.

In this experiment, both the frequency × position and the concreteness × position interactions were significant and were different from those in Experiment 1; the size of either effect varied little in the medial positions. It is possible that this difference may have been due to the tendency to omit early medial items, so that the latter items in the lists could be outputted before they were lost entirely. The increase in recall performance in the final items was similar for most conditions, except perhaps for the HFHC condition, where performance at recency may have suffered, relative to other conditions, because prerecency items were sufficiently intact to be outputted, thus slowing the output process and increasing the degree of output interference encountered at the recency position (Hulme, Newton, Cowan, Stuart, & Brown, 1999). Alternatively, given the high level of performance at the recency position, the HFHC condition may have been restricted by a ceiling effect.

Therefore, in the context of this experiment, both frequency and concreteness behaved similarly when performance across serial positions is considered. Each was also found to exhibit the same size of effect in primacy and recency portions of the curve. The frequency × serial position interaction can be compared with that found by O. C. Watkins and Watkins (1977) with auditory presentation. These authors determined that the frequency effect was greater in the primacy positions than in the recency positions and, specifically, was absent for the last two positions. This contrasts with the present findings, where the effects of frequency can be seen for all positions but the first. With respect to the concreteness × serial position interaction, two additional experiments are relevant: Walker and Hulme's (1999) Experiment 2 and Romani et al.'s (2008) Experiment 1B. The present results are closer to those of the former, in that the effect was present in the penultimate but not the last position. In the experimenet of Romani et al., this pattern appears to have been reversed. Furthermore, Romani et al.'s finding that the concreteness effect was greater in the primacy than in the recency region was not replicated in Experiment 2. It would appear that the claim regarding the reduction of lexical-semantic effects prior to the last serial position is not as generalized as Romani et al. implied, although the results in this instance do support their contention that serial position interactions for frequency and concreteness should be similar.

GENERAL DISCUSSION

Experiments 1 and 2 were in agreement regarding the pattern of results for correct recall; there is evidence that the effects of word frequency and word concreteness combine in the short-term recall of verbal material. Specifically, the nature of this combination is such that the more concrete the stimuli, the smaller the effect of word frequency. The structure of the language-based models of STM (R. C. Martin et al., 1999; Romani et al., 2008), and the inherent properties that derive from them, are compatible with the finding of an interaction in the data, if it is assumed that a limit exists in the level of activation a lexical node can realize.

Using the structure of these models as a framework, it is proposed that word frequency affects the strength of activation at the lexical layer in the long-term knowledge store (R. C. Martin et al., 1999). When list items vary in word frequency, an advantage to high-frequency items will occur because they activate the lexical nodes for these items more strongly than those for low-frequency words, thus providing stronger feedback activation to the linked nodes at the phonological level. Concreteness, reflecting the richness of semantic representation, is a marker of the quantity of semantic nodes that connect with a lexical node (R. C. Martin & Lesch, 1996). When a semantic feature is activated, it provides feedback activation to lexical entries that are connected to that semantic feature. Although feedback will return to the lexical node that activated it in the first place, it will also activate semantic competitors at the lexical level. Therefore, those lexical items with more semantic features will be better reinforced by interactive activation, and this in turn will assist the preservation of activated phonological representations in LTM. Since concrete words provide greater feedback activation from

the semantic level to the initiating lexical nodes than do abstract words, concreteness effects should be evident between conditions that vary on this attribute. Furthermore, if the level of activation of nodes in the lexical layer has some upper bound, a frequency × concreteness interaction of the type observed could result. That is, the activation levels experienced by lexical nodes of words that are both highly frequent and concrete would be less than the sum of activation levels achieved by items that are highly frequent or concrete alone.

The existence of a frequency × concreteness interaction can also be accommodated within a redintegration account that assumes separate redintegrative capacities for phonological and semantic information (Walker & Hulme, 1999). Word frequency affects the redintegration of short-term phonological traces through the accessibility and availability of phonological long-term representations, in either an item-specific (Hulme et al., 1997; Saint-Aubin & LeBlanc, 2005; Saint-Aubin & Poirier, 2005) or an associative (Hulme et al., 2003; Stuart & Hulme, 2000) manner. Word concreteness influences the strength of semantic representation and the uniqueness of semantic features (Walker & Hulme, 1999) and, therefore, impacts on the likelihood that short-term semantic codes are correctly redintegrated. It is assumed that as items increase in either frequency or concreteness, the likelihood of successful redintegration from phonological or semantic mechanisms, respectively, increases. Therefore, items that are both frequent in use and highly concrete will benefit less in relative terms, because an increasing proportion of these will be reconstructed from both processes.

However, Walker and Hulme's (1999) assertion that separate redintegrative mechanisms are evidenced by the distinct signatures of frequency and concreteness effects across serial position has not been supported in these experiments. Whereas Experiment 1 was inconclusive with regard to genuine differences between serial position interactions for frequency and concreteness, Experiment 2 produced interactions with serial position that were similar for both variables and invariant across the medial positions of the recall curve, as proponents of language-based models have suggested they should be (Romani et al., 2008).

It is therefore important to address the apparent conflict between observations made here and those made by Walker and Hulme (1999) and to offer an explanation as to their origin. A survey of the serial recall literature involving pure lists of word frequency reveals that earlier experiments (e.g., Hulme et al., 1997) in which a definite increase in effect across serial positions until the last item was observed used very small stimulus sets and high numbers of trials. For example, the serial recall experiments of Hulme et al. (1997) employed sets of eight items per condition, and each condition constituted 25 seven-item trials. Thus, each item was presented throughout the course of the experiment, on average, 22 times. Typically, the number of presentations for each item in the majority of other reported experiments has been less than 10, and in the case of open sets of words, this reduces to 1 or 2 presentations. Furthermore, examination of the serial position interactions in experiments in which the number

of item presentations has not been large does not reveal a consistent pattern of performance. Ceiling effects in the first and, sometimes, second serial positions have often been evident, leading to a possible masking of actual effect size in the primacy positions and the artificial creation of statistically significant interactions. The regularity of the increase in effect across serial positions has varied and, in some instances, has diminished further into the list. Therefore, apparent qualitative differences in frequency × serial position interactions have existed in reported experiments, and these differences have been loosely correlated with how often items were used within an experiment.

This is perhaps most clearly demonstrated in a comparison between the frequency × serial position interactions for Hulme et al.'s (2003) Experiment 1 and Experiment 2. The first of these experiments used eight-item closed sets from Hulme et al. (1997) in a comparison of recall performance between six-item lists of pure and alternating frequency. The recall curves for pure lists replicated the pattern identified by Hulme et al. (1997) and displayed a monotonically increasing effect from positions one to five. In contrast, in the second experiment, which was a replication of the first using an open set of stimuli, in which items were presented twice in the life of the task, revealed curves for the recall of pure lists that did not possess this feature. These data were arguably affected by a ceiling effect on the first serial position, masking the true level of performance for high-frequency lists, and the effect sizes were constant for the medial serial positions. Furthermore, these experiments did not block conditions, and so presentation effects of set size on serial position interactions would appear to have survived the intermittent presentation of items belonging to specific conditions.

It can be suggested that this difference between serial position interactions resulting from the use of closed and open item pools resulted from an improvement in recall in the early serial positions for low-frequency words. Recently, research involving preexposure to pairs of items or items individually (Saint-Aubin & Poirier, 2005; Stuart & Hulme, 2000) has shown that recall of familiarized low-frequency items is markedly greater than recall of unfamilarized low-frequency items. It seems likely that intraexperimental familiarity effects occur in circumstances that incorporate the repeated presentation of items. The improved recall for early serial positions might well reflect the incompleteness of familiarization in these circumstances and an inability to sustain processing benefits throughout the list. Thus, the oft-cited serial position interaction for word frequency (e.g., Hulme et al., 1997) may, in reality, be an artifact of experimental method.

The stimuli in the serial recall experiments of Walker and Hulme (1999) were presented seven times within their respective conditions. The present experiments used each stimulus item six times. Therefore, if a corresponding presentation effect were to be present in sets manipulating levels of concreteness, these experiments are adequately matched. Although Romani et al. (2008) used open sets of words for their serial recall experiment, the pattern observed is similar. Furthermore, a difference in this case

relates to the lack of concreteness effect at the penultimate position with open sets and is, therefore, one unlikely to be related to the multiple presentations of items.

One final distinction made by Walker and Hulme (1999) regarding differences in short-term processing between items varying in frequency and concreteness was in relation to backward recall. Hulme et al. (1997) had observed no effect of frequency in this task, whereas Walker and Hulme found a facilitative effect of concreteness. These results were presumed to reflect the use of a semantic retrieval strategy in task execution. Although no direct disconfirmation of this position is available at present, given recent research revealing conditions under which the effects of lexical-semantic variables could go undetected, and given variations in the size of the stimulus sets used by Hulme et al. (1997) and Walker and Hulme, this contrast too should perhaps be treated with suspicion. Hulme et al. (1997) had again used closed sets of eight items for blocks of 25 trials per condition in a design that tested forward and backward recall in lists of differing frequency. Therefore, depending on the counterbalancing of conditions, some participants would have been presented with each item 44 times by the end of the backward recall trials. In contrast, each stimulus in the backward recall experiment of Walker and Hulme was presented only 7 times across the condition to which it belonged. Here, the absence of effect for word frequency may reflect the influence of increased activation of low-frequency words due to repeated presentation, rather than the adoption of a semantic retrieval strategy per se. Furthermore, if a frequency effect was shown to be present under conditions similar to those for Walker and Hulme's word concreteness task, a more general lexical-semantic influence over backward recall performance might be responsible and relate to the degree to which these variables assist in maintaining the intactness of phonological traces (Romani et al., 2008).

Experiments 1 and 2 differed with respect to the results for memory for order. Experiment 1 indicated no involvement of lexical-semantic variables in the sequencing of list items, whereas Experiment 2 showed small effects for frequency and concreteness, with the reordering of items more evident for low-frequency and abstract words, respectively. A review of results across multiple studies suggests that the appearance of order effects is not attributable to particular features of experimental designs, apparently being influenced more by the incidental characteristics of the experiments performed, or perhaps arising from data influenced by individuals whose capacity for maintaining order was poor. However, in these cases, the direction of the effects has been consistent; facilitative effects of order memory have been formerly identified for concreteness (Allen & Hulme, 2006), word pleasantness (Monnier & Syssau, 2008), word frequency (Tse & Altarriba, 2007), and semantic associativity (Tse & Altarriba, 2007), whereas a detrimental effect for semantic similarity (Saint-Aubin et al., 2005) has been observed. It is presumed that, in the latter case, semantic similarity served to confuse the recall of items.

The differences in order retention, when observed, tend to be small, suggesting a minor role for lexical-semantic variables. This picture is consistent with the multitude of experimental research identifying the contributions of lexical-semantic variables in serial recall as ones relating to item information (Majerus, 2008). Despite this, word frequency does affect the perception of presented items, in that high-frequency words are advantaged to a small degree (Allen & Hulme, 2006). It is possible that speech perception processes could, in part, influence the quality of phonological encoding and, therefore, impact serial recall performance from task onset, particularly in the context of a short-term phonological buffer acting as a serial order mechanism. However, Allen and Hulme found no effect of concreteness in an auditory perception task, so a corresponding argument for concreteness is not supported.

In the context of redintegration theory, the presence of an order effect associated with semantic similarity has been previously explained by the admission of semantic features to the short-term trace (Saint-Aubin et al., 2005), and Walker and Hulme's (1999) extension of redintegration includes short-term semantic encoding. Whereas semantic similarity is claimed to increase the confusability of items and compromise memory for order, concreteness is argued to determine the uniqueness of the semantic representation (Walker & Hulme, 1999). If this is the case, degraded short-term traces of abstract items might be more similar than degraded short-term traces of concrete words, thereby facilitating better order memory for concrete words.

The language-based models presented here admit the influence of lexical-semantic variables from the point of item presentation onward, but their role is primarily the maintenance of item identity through interactive activation in LTM. In the model described by Romani et al. (2008), the management of serial order occurs through the maintenance of buffered phonological representations in the placeholder that interact with LTM at the phonological level. These, in turn, are supported by connected lexical-semantic activation. Small order effects might reflect perturbations arising from the interaction of the contents of the placeholder with the phonological representations in LTM as a result of supportive lexical-semantic activation; however, this explanation would benefit from greater specification of the serial order mechanism overseeing the preservation of order memory.

In summary, two experiments have confirmed the presence of a frequency $\times$ concreteness interaction in serial recall. The results of these experiments have been examined in relation to two explanatory frameworks in which frequency and concreteness are considered separately: the dual-redintegration framework of Walker and Hulme (1999) and the language-based model of Romani et al. (2008). Each position is capable of explaining the presence of an interaction between these lexical-semantic variables; however, the redintegration account has greater difficulty accommodating the observations of similar interactions with serial position for both variables. Indeed, the distinction between serial position interactions is a cornerstone of the dual-redintegration theory (Walker & Hulme, 1999).

Closer inspection of the previous experimentation on which this distinction was drawn indicates that confounds with set size may be responsible for differences in observed patterns, supporting the findings of the present work and the predictions drawn from models with a language-based architecture (Romani et al., 2008)—namely, that serial position interactions of lexical-semantic variables should take on a similar form. However, although effects for medial positions were found to be constant, the suggestion that lexical-semantic effects are eliminated in the final two positions of serial recall and that effects are greater in the primacy than in the recency portions of the curve did not generalize to the present experiments. It is likely that additional experiment-specific factors influence the extent to which lexical-semantic effects are reduced in the recency portion of the curve.

Although the present experiments have yielded results that imply that lexical-semantic effects exist at output, they do not directly address specific processes that are responsible for the encoding and maintenance of representations. Evidence from ERP studies have suggested that lexical-semantic activation may operate in the serial recall process during encoding and persist throughout the retention period (Cameron, Haarmann, Grafman, & Ruchkin, 2005; Ruchkin et al., 1999; Ruchkin, Grafman, Cameron, & Berndt, 2003), as language-processing models of STM assume (N. Martin, 2008; N. Martin & Saffran, 1997; R. C. Martin et al., 1999; Romani et al., 2008). However, other researchers (Thorn, Frankish, & Gathercole, 2008; Thorn, Gathercole, & Frankish, 2005) have suggested, on the basis of close examination of item recall errors, that variables such as language familiarity, word frequency, and nonword phonotactic frequency have an influence over the retention of items, as well as their production in serial recall, whereas the effect of lexicality—that is, differential performance for words versus nonwords—is present in serial recall, but not earlier, as indexed by the numbers of completely incorrect recall attempts. These data conflict with the prediction from language-based models, that influences of lexical-semantic variables should be present throughout task execution. In the case of lexicality, the absence of lexical-semantic representations for nonwords should result in immediate differences in the short-term traces that maintain information prior to recall. Accordingly, Thorn et al. (2008; Thorn et al., 2005) were less convinced that lexical-semantic variables play an integrated and ongoing role in the retention of items before recall and proposed that short-term memory activity is punctuated by multiple mechanisms at differing stages of the task, involving different LTM variables. Specifically, they argued that phonotactic frequency and word frequency are involved in the retention of the short-term trace but also are influential in a late-stage redintegration process. In contrast, lexicality effects are the sole product from involvement with the late-stage reconstruction of partially degraded items. This account therefore has similarities with the redintegration model of Walker and Hulme (1999), in that it identifies specific roles and processes for LTM variables, although it relies on alternative data to argue for the separate treatment of these and, so far, has been confined to variables with a phonological base. Continued investigation along these lines might serve to reinstate an interpretation involving redintegration mechanisms with respect to the set of results presented here.

Lastly, on a practical note, the interaction between frequency and concreteness suggests that the domain over which stimuli are controlled for in one variable will determine the size of the manipulated variable's effect. This moderation may explain, in part, variations in effect sizes observed between experiments. Furthermore, experimenters should be aware that effects can be smaller than anticipated if their control of the other variable is at the higher end of the respective measurement scale.

REFERENCES

ALLEN, R., & HULME, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory & Language*, **55**, 64-88.

BAAYEN, R. H., PIEPENBROCK, R., & VAN RIJN, H. (1993). *The Celex lexical database*. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

BOURASSA, D. C., & BESNER, D. (1994). Beyond the articulatory loop: A semantic contribution of serial order recall of subspan lists. *Psychonomic Bulletin & Review*, **1**, 122-125.

BUCHNER, A., & ERDFELDER, E. (2005). Word frequency of irrelevant speech distractors affects serial recall. *Memory & Cognition*, **33**, 86-97.

CAMERON, K. A., HAARMANN, H. J., GRAFMAN, J., & RUCHKIN, D. S. (2005). Long-term memory is the representational basis for semantic verbal short-term memory. *Psychophysiology*, **42**, 643-653.

CAZA, N., & BELLEVILLE, S. (1999). Semantic contribution to immediate serial recall using an unlimited set of items: Evidence for a multicapacity view of short-term memory. *International Journal of Psychology*, **34**, 334-338.

COLTHEART, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, **33A**, 497-505. Retrieved March 2004 from www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm.

COWAN, N. (1992). Memory span and the timing of spoken recall. *Journal of Memory & Language*, **31**, 668-684.

FORSTER, K. I., & FORSTER, J. C. (1999). DMDX [Computer software]. Tucson: University of Arizona.

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, & Cognition*, **25**, 447-463.

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, & Cognition*, **23**, 1217-1232.

HULME, C., STUART, G., BROWN, G. D. A., & MORIN, C. (2003). High- and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory & Language*, **49**, 500-518.

JEFFERIES, E., FRANKISH, C., & LAMBON RALPH, M. A. (2006). Lexical and semantic influences on item and order memory in immediate serial recognition: Evidence from a novel task. *Quarterly Journal of Experimental Psychology*, **59**, 949-964.

JONES, G. V. (1985). Deep dyslexia, imageability, and ease of predication. *Brain & Language*, **24**, 1-19.

KINTSCH, W., & BUSCHKE, H. (1969). Homophones and synonyms in short-term memory. *Journal of Experimental Psychology*, **80**, 403-407.

LI, S.-C., & LEWANDOWSKY, S. (1995). Forward and backward recall: Different retrieval processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 837-847.

MAJERUS, S. (2008). 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). Hove, U.K.: Psychology Press.

MAJERUS, S., LEKEU, F., VAN DER LINDEN, M., & SALMON, E. (2001). Deep dysphasia: Further evidence on the relationship between phonological short-term memory and language processing impairments. *Cognitive Neuropsychology*, **18**, 385-410.

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.

MARTIN, N. (2008). The influence of long-term knowledge on short-term memory: Evidence for multiple mechanisms. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 220-243). Hove, U.K.: Psychology Press.

MARTIN, N., & SAFFRAN, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology*, **14**, 641-682.

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 & Language*, **52**, 83-113.

MARTIN, R. C., & LESCH, M. F. (1996). Associations and dissociations between language impairment and list recall: Implications for models of STM. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 149-178). Hove, U.K.: Psychology Press.

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.

MONNIER, C., & SYSSAU, A. (2008). Semantic contribution to verbal short-term memory: Are pleasant words easier to remember than neutral words in serial recall and serial recognition? *Memory & Cognition*, **36**, 35-42.

MORIN, C., POIRIER, M., FORTIN, C., & HULME, C. (2006). Word frequency and the mixed-list paradox in immediate and delayed serial recall. *Psychonomic Bulletin & Review*, **13**, 724-729.

MUELLER, S. T., SEYMOUR, T. L., KIERAS, D. E., & MEYER, D. E. (2003). Theoretical implications of articulatory duration, phonological similarity, and phonological complexity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 1353-1380.

MURDOCK, B. B. (1976). Item and order information in short-term serial memory. *Journal of Experimental Psychology: General*, **105**, 191-216.

NEATH, I. (1997). Modality, concreteness, and set-size effects in a free reconstruction of order task. *Memory & Cognition*, **25**, 256-263.

PAIVIO, I. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, **45**, 255-287.

PENNEY, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, **17**, 398-422.

POIRIER, M., & SAINT-AUBIN, J. (1996). Immediate serial recall, word frequency, item identity and item position. *Canadian Journal of Experimental Psychology*, **50**, 408-412.

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.

ROODENRYS, S., HULME, C., LETHBRIDGE, A., HINTON, M., & NIMMO, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 1019-1034.

ROODENRYS, S., & QUINLAN, P. T. (2000). The effects of stimulus set size and word frequency on verbal serial recall. *Memory*, **8**, 71-78.

RUCHKIN, D. S., BERNDT, R. S., JOHNSON, R., JR., GRAFMAN, J., RITTER, W., & CANOUNE, H. L. (1999). Lexical contributions to retention of verbal information in working memory: Event-related brain potential evidence. *Journal of Memory & Language*, **41**, 345-364.

RUCHKIN, D. S., GRAFMAN, J., CAMERON, K., & BERNDT, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral & Brain Sciences*, **26**, 709-777.

SAINT-AUBIN, J., & LEBLANC, J. (2005). Word frequency effects in immediate serial recall of pure and mixed lists: Tests of the associative link hypothesis. *Canadian Journal of Experimental Psychology*, **59**, 219-227.

SAINT-AUBIN, J., OUELLETTE, D., & POIRIER, M. (2005). Semantic similarity and immediate serial recall: Is there an effect on all trials? *Psychonomic Bulletin & Review*, **12**, 171-177.

SAINT-AUBIN, J., & POIRIER, M. (1999). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *Quarterly Journal of Experimental Psychology*, **52A**, 367-394.

SAINT-AUBIN, J., & POIRIER, M. (2005). Word frequency effects in immediate serial recall: Item familiarity and item co-occurrence have the same effect. *Memory*, **13**, 325-332.

SCHWEIKERT, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, **21**, 168-175.

SCHWEIKERT, R., CHEN, S., & POIRIER, M. (1999). Redintegration and the useful lifetime of the verbal memory representation. *International Journal of Psychology*, **34**, 447-453.

STUART, G., & HULME, C. (2000). The effects of word co-occurrence on short-term memory: Associative links in long-term memory affect short-term memory performance. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 796-802.

TEHAN, G., & HUMPHREYS, M. S. (1988). Articulatory loop explanations of memory span and pronunciation rate correspondences: A cautionary note. *Bulletin of the Psychonomic Society*, **26**, 293-296.

THORN, A. S. C., FRANKISH, C. R., & GATHERCOLE, S. E. (2008). The influence of long-term knowledge on short-term memory: Evidence for multiple mechanisms. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 198-219). Hove, U.K.: Psychology Press.

THORN, A. S. C., GATHERCOLE, S. E., & FRANKISH, C. R. (2005). Redintegration and the benefits of long-term knowledge in verbal short-term memory: An evaluation of Schweikert's (1993) multinomial processing tree model. *Cognitive Psychology*, **50**, 133-158.

TOLAN, G. A., & TEHAN, G. (1999). Determinants of short-term forgetting: Decay, retroactive interference, or proactive interference? *International Journal of Psychology*, **34**, 285-292.

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**, 675-690.

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, & Cognition*, **25**, 1256-1271.

WATKINS, M. J. (1977). The intricacy of memory span. *Memory & Cognition*, **5**, 529-534.

WATKINS, O. C., & WATKINS, M. J. (1977). Serial recall and the modality effect: Effects of word frequency. *Journal of Experimental Psychology: Human Learning & Memory*, **3**, 712-718.

APPENDIX
Stimulus Sets and Stimulus Attributes for Experiments 1 and 2

Condition	Item	Conc.	Freq.	Phonemes	Letters	PNS	Phonological Similarity		
							Onset	Nucleus	Coda
LFLC	verb	337	3	3	4	6	0.28	0.27	0.20
	truce	335	3	4	5	9	0.30	0.35	0.20
	grief	303	15	4	5	12	0.34	0.23	0.20
	hint	312	26	4	4	12	0.29	0.25	0.25
	guess	247	31	3	5	11	0.35	0.27	0.20
	myth	334	28	3	4	16	0.34	0.25	0.20
	blame	293	23	4	5	5	0.29	0.23	0.26
	pause	306	36	3	5	59	0.24	0.39	0.33
	fate	255	35	3	4	36	0.27	0.23	0.20
	theme	336	37	3	5	11	0.28	0.23	0.26
	proof	328	33	4	5	5	0.27	0.35	0.20
	harm	244	33	3	4	23	0.29	0.41	0.26
	M	302.50	25.25	3.42	4.58	17.08	0.30	0.29	0.23
	SD	35.63	12.10	0.51	0.51	15.83	0.03	0.07	0.04
HFLC	hope	261	163	3	4	25	0.36	0.28	0.24
	cause	287	174	3	5	59	0.32	0.31	0.21
	truth	261	134	4	5	6	0.30	0.34	0.18
	risk	290	85	4	5	6	0.32	0.31	0.27
	deal	342	193	3	4	30	0.35	0.34	0.22
	chance	254	178	4	6	6	0.37	0.38	0.26
	rate	308	211	3	4	46	0.32	0.29	0.20
	rule	286	128	3	4	30	0.32	0.34	0.22
	cost	348	204	4	4	18	0.32	0.38	0.24
	style	344	107	4	5	10	0.34	0.31	0.22
	force	331	242	3	5	27	0.35	0.31	0.20
	claim	331	100	4	5	7	0.31	0.29	0.30
	M	303.58	159.92	3.50	4.67	22.50	0.33	0.32	0.23
	SD	34.98	49.19	0.52	0.65	17.21	0.02	0.03	0.03
LFHC	tail	613	36	3	4	35	0.24	0.29	0.27
	barn	614	12	3	4	33	0.25	0.35	0.21
	steak	646	12	4	5	20	0.28	0.29	0.36
	sheep	622	20	3	5	23	0.37	0.36	0.24
	rope	608	44	3	4	38	0.40	0.29	0.24
	chalk	634	9	3	5	22	0.37	0.30	0.26
	thumb	638	27	3	5	15	0.26	0.32	0.22
	fox	605	16	4	3	26	0.28	0.37	0.31
	broom	613	7	4	5	10	0.31	0.39	0.22
	pond	623	19	4	4	16	0.23	0.37	0.26
	crane	606	5	4	5	15	0.36	0.29	0.21
	bell	620	42	3	4	29	0.25	0.31	0.27
	M	620.17	20.75	3.42	4.42	23.50	0.30	0.33	0.26
	SD	13.17	13.55	0.51	0.67	8.88	0.06	0.04	0.04
HFHC	heart	605	164	3	5	28	0.27	0.32	0.20
	bird	602	103	3	4	46	0.23	0.29	0.17
	head	603	310	3	4	38	0.27	0.28	0.17
	land	604	272	4	4	18	0.35	0.29	0.24
	sun	617	152	3	3	37	0.35	0.27	0.21
	meal	602	91	3	4	37	0.31	0.36	0.23
	glass	635	144	4	5	3	0.41	0.32	0.25
	ball	615	112	3	4	40	0.23	0.35	0.23
	dog	610	116	3	3	19	0.26	0.36	0.38
	skin	614	102	4	4	12	0.33	0.33	0.21
	film	604	122	4	4	6	0.28	0.33	0.26
	horse	613	133	3	5	23	0.27	0.34	0.25
	M	610.33	151.75	3.33	4.08	25.58	0.30	0.32	0.23
	SD	9.52	68.99	0.49	0.67	14.24	0.05	0.03	0.05

Note—Conc., concreteness; Freq., Celex-based frequency; PNS, phonological neighborhood size; LFLC, low frequency, low concreteness; HFLC, high frequency, low concreteness; LFHC, low frequency, high concreteness; HFHC, high frequency, high concreteness.