Mem Cogn (2015) 43:489–499
DOI 10.3758/s13421-014-0470-6

Order recall in verbal short-term memory: The role of semantic networks

Marie Poirier · Jean Saint-Aubin · Ali Mair · Gerry Tehan · Anne Tolan

Published online: 10 October 2014

Abstract In their recent article, Acheson, MacDonald, and Postle (Journal of Experimental Psychology: Learning, Memory, and Cognition 37:44–59, 2011) made an important but controversial suggestion: They hypothesized that (a) semantic information has an effect on order information in short-term memory (STM) and (b) order recall in STM is based on the level of activation of items within the relevant lexico-semantic long-term memory (LTM) network. However, verbal STM research has typically led to the conclusion that factors such as semantic category have a large effect on the number of correctly recalled items, but little or no impact on order recall (Poirier & Saint-Aubin, Quarterly Journal of Experimental Psychology 48A:384–404, 1995; Saint-Aubin, Ouellette, & Poirier, Psychonomic Bulletin & Review 12:171–177, 2005; Tse, Memory 17:874–891, 2009). Moreover, most formal models of short-term order memory currently suggest a separate mechanism for order coding—that is, one that is separate from item representation and not associated with LTM lexico-semantic networks. Both of the experiments reported here tested the predictions that we derived from Acheson et al. The findings show that, as predicted, manipulations aiming to affect the activation of item representations significantly impacted order memory.

Keywords Short-term memory · Working memory · Order recall · Immediate memory · Activated long-term memory

We are all familiar with the experience of reading an article in our field of expertise. Expressions are recognized, some arguments and ideas are anticipated, and grasping the experimental logic is facilitated by our understanding of the strategies in the area. Thus, our previous knowledge of the constituents of the article significantly supports our understanding of the work. In important ways, this example illustrates one of the most fundamental functions that memory performs: allowing the past to support and guide our present interactions with the world. This is the issue that motivated the present work; in the experiments reported here, we examined the interaction between semantic knowledge and the last few seconds of our most recent past—the content of verbal short-term memory (STM).

Here, we viewed STM as a less general system than working memory. More specifically, STM was defined as the system that carries out the temporary maintenance of information necessary for many mental or cognitive operations and tasks (Baddeley, 1986). Generally, STM is recognized as playing an important role in everyday cognition (Cowan, 1999; Majerus, 2009). Moreover, the role of STM for order has also been highlighted in cognitive development, and in particular in learning new words (Cowan, 1999; Majerus & Boukebza, 2013). One of the roles of STM that is regarded as central is the short-term maintenance of the order of events (Majerus, 2009). As a simple example, consider keying in a new security code, address, or phone number. These can, of course, be written down, but even in order to do so, they must

M. Poirier (✉) · A. Mair
Department of Psychology, City University London, Northampton Square, London, UK EC1V 0HB
e-mail: M.Poirier@city.ac.uk

J. Saint-Aubin
Moncton University, Moncton, New Brunswick, Canada

G. Tehan
University of Southern Queensland, Toowoomba, Queensland, Australia

A. Tolan
Australian Catholic University, Banyo, Queensland, Australia

be maintained in memory long enough for the writing down to take place.

Short- and long-term memory

Until relatively recently, the literature examining how the lexico-semantic properties of verbal items affect performance in STM tasks was sparse. However, the present work bears witness to the growing interest in this area, with recent research having systematically explored the relationship between language organization in long-term memory (LTM) and verbal short-term recall (e.g., Acheson, MacDonald, & Postle, 2011; Hamilton & Martin, 2007; Majerus, 2009; R. C. Martin, 2006; Tehan, Humphreys, Tolan, & Pitcher, 2004; Thorn & Page, 2009). Nevertheless, less work has been done on the factors typically associated with semantic LTM. The studies reported here tested a controversial hypothesis that suggests that semantic LTM plays an important role in verbal STM, and more specifically in short-term order memory.

The role of LTM in short-term recall

The study of LTM's contributions to verbal short-term recall—as well as the study of STM in its own right—has typically relied on a classic task: immediate serial recall. In this task, a small number of items are presented—usually between five and seven—and participants must attempt to recall them, in their order of appearance, immediately after list presentation. It is well established that multiple factors associated with long-term knowledge of the language have a significant impact on the performance of this task. Word frequency and familiarity have positive effects on immediate serial recall (Poirier & Saint-Aubin, 1996), as have concreteness (Walker & Hulme, 1999) and lexicality (Hulme, Maughan, & Brown, 1991; Saint-Aubin & Poirier, 2000; for a review, see Saint-Aubin & Poirier, 1999a). This is also true at a sublexical level (Roodenrys, 2009); for example, when trying to remember nonwords, items containing more-familiar phonemic components are recalled better (Thorn & Frankish, 2005). Currently, it can be argued that there are two general classes of views that address these findings. The first are typically known as *redintegration* accounts, and the second suggest that verbal STM relies more directly on *long-term representations*.

Redintegration From the redintegration perspective, immediate recall is a two-step process. It is assumed that participants first encode verbal materials into phonological forms, as is suggested by the seminal multicomponent model first proposed by Baddeley and Hitch (1974; Baddeley, 1986). In the absence of rehearsal, these representations are thought to rapidly become degraded, through either decay or interference. At the point of recall, a retrieval mechanism produces a phonological representation as a candidate for output. The memory trace may or may not be degraded (but see Roodenrys & Miller, 2008). If the trace is intact, then recall will not be problematic. However, if the trace is degraded, a second step is initiated. Long-term lexical/phonological information is accessed in an attempt to reconstruct the item (e.g., accessing knowledge of words to complete a fragmented trace, somewhat like filling in the gaps in cr_ _odi_e). This reconstruction process is often referred to as *redintegration* (Hulme, Maughan, & Brown, 1991; Schweickert, 1993). It has been used to explain lexicality, word frequency, concreteness, and imageability effects upon serial recall (e.g. Saint-Aubin & Poirier, 2005). However, recent ideas about the contribution of long-term representations to STM have started to move away from dual-process accounts (i.e., degradation of phonological STM followed by redintegration). For example, Thorn, Frankish, and Gathercole (2009), after reviewing their work on phonotactic and lexical frequency, concluded that long-term knowledge impacts immediate recall accuracy in two ways: by strengthening the representations that support performance, and by influencing the reconstruction process. Romani, McAlpine, and Martin (2008) suggested a similar conclusion after a series of studies examining the effects of concreteness on a range of STM tasks.

Psycholinguistic and LTM network models Over the past two decades, the redintegration hypothesis was the dominant view of LTM effects on short-term recall. Currently, however, another class of models is becoming increasingly influential. Although the models in this group are more heterogeneous, they suggest that the LTM representations and the systems involved in language processing are more closely related to short-term recall than the redintegration hypothesis suggests (see, e.g., Acheson & MacDonald, 2009). In its typical form, the redintegration hypothesis restricts the influence of LTM representations to the retrieval stage of short-term recall. The psycholinguistic and LTM network models that we refer to here propose that there is considerable overlap between STM tasks and language processing; hence, the semantic, lexical, and sublexical networks that are widely thought to underlie language representations are viewed as supporting STM. In essence, these models are mostly moving away from the classic suggestion that verbal STM relies on a separate system. Rather, the premise is that processing linguistic information for recall involves the activation of the relevant long-term networks; in turn, the characteristics of these networks will influence performance.

Burgess and Hitch (2006), for example, offered a computational/network model of verbal STM in which items are represented within lexical and phonological interconnected networks. More recently, in order to explain the effects of a number of lexical and sublexical variables, Roodenrys (2009) proposed that an interactive network model was necessary in which various levels of representation, including letter, phonemic, and lexical levels, are activated and compete with each other. Other recent models explicitly include semantic levels of representation, also. This group includes the computational model proposed by Gupta (2003, 2009), the conceptual models proposed by Cowan (1999; Cowan & Chen, 2009) and Majerus (2009), the psycholinguistic models proposed by N. Martin and Gupta (2004) and R. C. Martin (2006), and from cognitive neuroscience, the proposals of Acheson et al. (2011) and Buchsbaum and D'Esposito (2008).

Making a choice amongst the models described above depends on a number of developments, one of which is a better understanding of how semantic memory influences STM performance. Assuming that these models are appropriate, then semantic LTM should influence STM performance in predictable ways. As of yet, however, there has been little detailed investigation of semantic LTM effects in short-term recall in healthy adults. Exceptions include the work on categorical similarity, the work of Romani et al. (2008) on concreteness, and the recent work of Acheson et al. (2011).

Categorical similarity Poirier and Saint-Aubin (1995; Saint-Aubin & Poirier, 1999a, b; Saint-Aubin, Ouellette, & Poirier, 2005) reexamined the widely held idea that similarity amongst list items in immediate serial recall has an adverse effect upon STM for *order* recall. Although this finding is highly reliable when phonological similarity is manipulated, Poirier and Saint-Aubin (1995) argued that this was not necessarily the case with semantic similarity. In their experiments, they explored semantic similarity effects on both item and order memory; participants studied lists of items that either were from one semantic category or were unrelated to each other. They found that categorical similarity was advantageous to item memory but had little effect upon order memory; in effect, across conditions, order errors were proportional to the number of items recalled (although see Saint-Aubin et al., 2005). Since more items were recalled for categorized lists, there was a proportional increase in order errors. In explaining their results, they suggested that the taxonomic category could be used as an extra retrieval cue supporting recall; this led to better item recall and a stable level of order errors per item.

However, assuming that semantic LTM underpins STM performance suggests another explanation of the semantic category effect and generates further predictions. The latter idea relates to the widespread idea of mutual activation between semantically related items, such as those that belong to the same semantic category (see Atkins & Reuter-Lorenz, 2008, for evidence of spreading activation effects in STM tasks).

For instance, Saint-Aubin et al. (2005) suggested that increased access to same-category items might depend on their long-term associative links (see also Hulme, Stuart, Brown, & Morin, 2003). Items from the same category tend to co-occur more frequently than items taken from different categories, and this is thought to strengthen their associative links in memory (Deese, 1960; Stuart & Hulme, 2000). This is in line with many conceptualizations of lexico-semantic memory in other fields, which often depict semantic/lexical memory in terms of a network of associatively related items; activation in one part of the network can spread and influence recall of other items in the network. It seems plausible that activating multiple items in an associative network might produce higher levels of activation and support recall.

A related idea was put forward by Acheson et al. (2011), although coming from a somewhat different perspective. Importantly, their particular proposal led us to develop novel, specific, and testable predictions. A quote from their article makes their view clearer (emphasis ours): "After initial encoding, lexical activation is determined by repeated interaction with semantic and phonological representations. *Serial ordering errors occur when the relative activation levels of the lexical items change because of this interaction*. . . . If the maintenance of information in verbal WM is achieved by virtue of activation of language-production architecture, this leads to the prediction that disrupting semantic processing should influence the relative activation of lexical-level representations, thus influencing serial ordering" (Acheson et al., 2011, p. 46). Acheson et al. used a dual-task strategy to show that when the interference task involves semantic processing, more order errors are produced than with a spatial task. Interestingly, this effect disappeared with nonwords; that is, we found no differential disruption by the semantic dual task when the primary task involved items with no meaning.

There are a number of reasons why this hypothesis is important:

(1) Knowledge-based effects have typically been considered as affecting item recall rather than order recall; this is especially true of semantic effects. Establishing that semantic factors influence order recall would be significant for extant theories of serial and short-term memory.
(2) Formal models of serial STM typically do not pay attention to LTM contributions, even though many instantiations (connectionist models) require that the hypothetical networks that sustain performance be trained before STM for order can be modeled (Botvinick & Plaut, 2006; Lewandowsky & Farrell, 2008; Page, 2005).
(3) Finally, although some reviews of serial order coding have typically discarded an activation-based account of order representation, this was founded on logical argument rather than empirical verification. The Acheson et al. (2011) data obviously argue in the other direction;

moreover, as we now discuss, here we present further tests of the idea that order coding relies at least in part on sematic networks.

One established way of "disrupting" semantic processing is by using associates that are highly related to a target item. This is the strategy that we adopted in the first experiment reported here. At first glance, the Acheson et al. (2011) quote above could be taken to imply that semantically related lists should generate more order errors than control lists, because the latter have reduced levels of interitem activation. Multiple studies have suggested that this is not the case—but controversy surrounds this point (see Saint-Aubin, et al., 2005, and Tse, 2009). As we mentioned earlier, order errors are proportional to item recall, and semantically related lists produce better item recall.

According to the hypothesis just reviewed (hereafter, *ANet*, for the activated network view), manipulating the semantic activation level of item representations within a list can influence serial ordering in predictable ways. Before we turn to the specifics of Experiment 1, we will outline a basic model that calls upon principles that have broad empirical and theoretical support in the field (Hurlstone, Hitch, & Baddeley, 2014). In no way is this a full-fledged model; our aim was to suggest as simple an architecture as possible, but one that would (a) rely on principles/mechanisms that are broadly agreed upon when it comes to immediate serial memory and (b) make specific, testable predictions in relation to manipulations of semantic activation and order. Our suggestion is that, apart from a semantic network that can support activation, the following elements are required: (1) encoding that produces a primacy gradient and (2) a response selection mechanism that relies on competitive cueing.

Simply put, a primacy gradient means that each successive item presented is encoded with diminishing strength (Grossberg, 1978a, b). Most formal models of short-term serial recall include or imply a primacy gradient, since such a mechanism is necessary to account for the typical form of the typical serial recall curve (Hurlstone et al., 2014). This curve plots correct-in-position recall as a function of presentation order. In the case of immediate serial recall, the curve shows pronounced primacy, as well as a small recency effect for the last item(s). The said recency depends on the materials and testing conditions. How the proposed primacy gradients are conceptualized and justified varies across models. For example, in their primacy model, Page and Norris (1998, 2009) suggest that the primacy gradient could be produced by the association of each incoming item with a start-of-sequence context, with the strength of the association diminishing with distance from the said context (i.e., the fourth word in a sequence would be farther from the start-of-list context than the first). A number of other systems for producing primacy gradients have been suggested; Hurlstone et al. (2014) have provided a review of various implementations of the principle. Here, the most parsimonious view would be that the said primacy gradient is represented within activation levels in a semantic network; however, other architectures could also be envisaged. The important point is that to account for immediate serial recall performance, an encoding that generates a primacy gradient appears to be a reasonable assumption.

Another mechanism that has broad support was also put forward by Grossberg (1978a, b) and is usually known as *competitive queuing* (CQ; Houghton, 1990). Competitive queuing can be thought of as a noisy competition between the activated response candidates; the system is important, since it can transform the parallel activation of the items captured by the primacy gradient into a serial sequence of responses. One way of describing the operation of a generic CQ mechanism is as follows: The activations represented within the primacy gradient are fed forward to the CQ response selection mechanism; there, items compete for selection on the basis of their activation levels, and mutual inhibition and noise make the process error-prone. The most activated item is typically selected, unless activation levels are too low or competition leads to the wrongful selection of another item. For instance, if noise makes it difficult for an item to be selected when appropriate (i.e., there is no winner of the competition within a threshold number of iterations/attempts), then this item as well as all of the remaining ones become less likely to be selected because of the increased pool of candidates and increased mutual inhibition (reducing activation). Importantly, CQ systems typically suppress the activation of any selected response, preventing perseveration.

These relatively simple building blocks provide the needed architecture for the ANet predictions tested in Experiment 1. In essence, the suggestion is as follows. Presenting a list of items for immediate serial recall generates activation in the lexico-semantic network, with activation following a primacy gradient. At the point of recall, the dynamics of the CQ mechanism predicts that the first item is very likely to be output first and will then be suppressed, removing it from the competition for the next response. The second item is then the most likely winner of the competition for response selection, and so forth.

In Experiment 1, we manipulated the level of activation of a target item to test the prediction that this would increase order errors for that item, making it likely that the CQ mechanism would select this item earlier because of its heightened activation; this early selection would mean that activation affected the order in which items were recalled. Lists of six visually presented items were used; the experimental lists contained a target item, presented in Position 5. The three first items of these lists were strong associates of the target, whereas control lists contained the same three associates in Positions 1–3, but the item in Position 5 was unrelated (see Table 1 for list examples).

For the experimental lists, we expected that the first three items would activate the target (fifth item) within LTM networks, making its activation level seem more like that of the earlier list items. On the basis of the ANet view and the summary model described above, the prediction was that the target fifth item would migrate toward earlier positions more often than would a nontarget item studied in the same position.

Basically, the prediction from this version of the ANet hypothesis involves one of the characteristics of order errors in immediate serial recall, known as the *locality constraint* (Henson, Norris, Page, & Baddeley, 1996). It is well established that when a list item is recalled in an incorrect position, it is more likely to migrate to a neighboring list position (Estes, 1972). So, the third item is more likely to be recalled in the second or fourth output position than in the first or seventh. In other words, order errors obey a rule whereby displacements are increasingly unlikely as one moves away from the actual presentation position of the item. The main prediction of Experiment 1 was that the locality constraint would still apply to the target item, but not as strictly as it applied to the comparable control item: The target item would be expected to be recalled in earlier positions more often than was the corresponding item in the control condition.

Experiment 1

Method

Participants A total of 40 adults took part (14 men and 26 women, age range from 18 to 57, mean = 27); they were offered a small fee (£7) for participating.

Materials The experiment comprised 32 lists, with 16 experimental and 16 control lists. We first generated a set of 16 lists in which the first three items were strong associates of a target word, on the basis of the University of South Florida norms (Nelson, McEvoy, & Schreiber, 2004). These words, when used as cue words in a semantic association/production task, generate the target as a strong associate. More specifically, the cue words had to have a forward association strength with the target that was above 0.2; also, they were excluded if they had a backward association strength above of 0.1. The target was placed in the 5th position of each list; the cue words were placed in Positions 1, 2, and 3; and the remaining positions (4 and 6) were filled with unrelated words. The same words were then used again to create a further set of 16 control lists, so that each word was used twice within the experiment; more specifically, each participant encountered the words; however, the condition in which they encountered the words for the first time (and the second) was counterbalanced across participants. Control lists had the same three associates in the first positions, in the same order. The last three words were a random selection from the filler words and from the targets associated with other lists. The 32 lists thus created were then mixed to create four sets, with a different, quasirandom order of lists. This was done such that a given trio of related words was presented once in the first block of 16 lists and once in the second block of 16 lists. Also, each block of 16 lists contained eight experimental and eight control lists. Each participant was only presented with one set of 32 lists, with the sets counterbalanced across participants. To be clear, each participant studied each word twice, once in the first block of 16 lists and once in the second; however, the order of the condition encountered first was counterbalanced across participants. A bespoke computer program controlled stimulus presentation and response collection.

Table 1 Sample experimental and control lists

Experimental List Examples					
officer	badge	siren	fence	police	tractor
band	record	concert	yellow	music	tourist
Control List Examples					
officer	badge	siren	music	tourist	yellow
band	record	concert	tractor	fence	police

Procedure Participants were tested individually, in soundproofed cubicles, within a session lasting approximately 20 min. Following the instructions, they completed two practice trials. A fixation cross appeared in the center of the screen for 2 s, indicating that the first word was about to be presented. Words appeared sequentially on the screen for one and a half seconds each, and were separated by a 500-ms blank screen. Right after the six words from a list had been presented, participants were to type them into response boxes, in the order in which they had appeared in the list, starting with the word presented first. If they did not remember a word, they were asked to type the letter “b” and to proceed to the following position. The program prevented participants from typing a response if the previous one was not entered or if the enter key had not been pressed. They were not allowed to backtrack to correct a previous response.

Results and discussion

The hypothesis examined here relates to the recall of the critical word and its control both appearing in the fifth position of their respective lists. The ANet view predicts that there would be more movement toward earlier positions for the fifth item when the first three words presented were strong associates of said target.

Table 2 presents correct-in-position scores (i.e., to be scored correct, the item must be recalled in its presentation position), as well as the item recall score (i.e., an item scored correct if it is recalled, irrespective of position). The table also presents means for the critical fifth item. As a perusal of the table shows, the overall correct-in-position performance is very similar in both conditions. Table 2 also presents item recall scores; as can be seen, Item 5 was recalled better in the experimental condition. Item scores are usually higher than correct-in-position scores because if an item is recalled in the wrong position it will be given a correct-in-position score of zero; however, it will be considered correct with item scoring. It can be seen, for example, that for the control lists, the correct-in-position score for Item 5 is .58, whereas it is .68 when item scoring is used. This difference is larger for Item 5 in the experimental condition, for which the correct-in-position performance is .58 and item scoring leads to a .75 mean. The implication is that Item 5 was recalled out of position more often in the experimental than in the control condition.

As would be expected in light of the content of Table 2, no statistically reliable effect emerged for the correct-in-position scores. With respect to the item scores, paired-sample *t* tests showed no reliable difference for the overall means, but there was a significant difference for Position 5 [$t(39) = 2.5$, $p = .017$].

Figure 1 shows the percentages of trials on which the item studied in Position 5 was actually recalled in another position—essentially the error frequency per position—for the target fifth item. As can be seen, the rate with which the fifth word was recalled in an incorrect position appears to be higher for the experimental condition than for the control condition, particularly for Positions 2 and 3. A 2 (condition) × 5 (error position) repeated measures analysis of variance (ANOVA) revealed that errors were significantly more frequent for the experimental condition [$F(1, 39) = 12.63$, $MSE = 0.56$]. We also found a significant effect of position [$F(4, 156) = 16.76$, $MSE = 0.85$] and a significant interaction [$F(4, 156) = 2.75$, $MSE = .52$]. Simple main effect tests showed that Item 5 migrated more often to Positions 2 and 3 in the experimental condition—whereas there was no evidence of more migrations to Position 6.

Table 2 Mean recall across positions and for Position 5

	All Positions	Position 5
Correct-in-Position Scores		
Control lists	.71	.58
Experimental lists	.71	.58
Item Recall Scores		
Control lists	.78	.68
Experimental lists	.79	.75

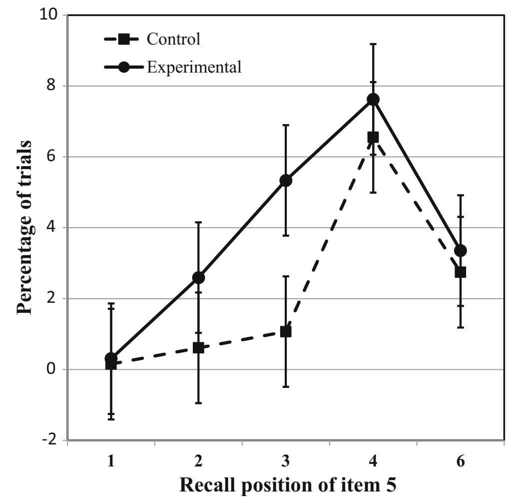

Fig. 1 Percentages of trials showing an error for Item 5 as a function of presentation position; only the erroneous recall positions are plotted on the *x*-axis. Error bars represent 95 % confidence intervals computed according to the method of Loftus and Masson (1994) for within-subjects factors. When the difference between two means is significant, those confidence intervals do not overlap by more than half the distance of one side of an interval (Masson & Loftus, 2003)

These findings support the predictions derived from the ANet account: When the first three items in a list are strong associates of the fifth item, the latter tends to migrate more than does a control item appearing in the same position; as expected, the target item migrated toward typically better recalled positions rather than toward the posterior position (6).

These predictions were derived on the basis of the idea that item order is coded as an activation primacy gradient within the lexico-semantic network that supports language representation, and the results lend support to this view. However, an alternative interpretation of this pattern of data is less interesting. This competing interpretation suggests that the fifth item is more frequently recalled with the first three related items because of a grouping strategy. Although the task instructions emphasized ordered recall, participants might have subjectively grouped the related items and this could have generated order errors. Essentially, the alternative hypothesis suggests that the results are an artifact of a study/recall strategy rather than an indication that semantic activation plays a role in order encoding and maintenance. This being said, it is important to note that the said strategy could well originate from the fact that recall relies on activated semantic networks, and that this makes maintaining clustered and related items easier. For the next experiment, we used lists that eliminated any advantage that grouping could involve, making the use of such a strategy useless, and hence very unlikely.

Experiment 2

Experiments 2a and 2b were based on a reanalysis of the previously published findings of Saint-Aubin et al. (2005). In their study, the experimental lists contained items that were all from the same semantic category (vegetables, sports, clothing, etc.). They can hence be expected to be reasonably close neighbors within the proposed semantic network. On the basis of the ANet view, we would expect heightened co-activation for these lists, relative to control lists containing unrelated items. Importantly, one would not expect any special grouping strategy for the categorized lists as all the items are from the same category. The control lists were constructed by re-organizing the items from the semantically related condition so that each word within a list was from a different semantic category. Each condition involved the same items overall. In Experiment 2a the lists were studied in silence, whereas in Experiment 2b, participants engaged in articulatory suppression. Both semantic category and suppression were manipulated between participants. There were $N = 70$ in each group for the silent conditions, and $N = 56$ in the two suppression conditions (categorized or control lists). All lists were seven items long, and 14 lists were presented in each condition. The details of the methodology are otherwise similar to the study reported above and can be found in Saint-Aubin et al. (2005).

Because the lists used in these experiments were seven items long, we examined the recall of Items 5 and 6. These seemed like the best candidates, since a reasonable number of errors would need to be made for reliable migration analyses to be possible. In an immediate serial recall task, the highest performance is typically observed for the first few items; the last item (7) is of less interest, since it can only migrate in one direction.

What are the predictions for this experiment? When the differences in correct recall between categorized and noncategorized items are examined, what is typically found is that the whole curve moves upward for the related items; that is, a categorization advantage is present and does not interact with serial position (provided ceiling and floor effects are avoided; see Poirier & Saint-Aubin, 1995, and Saint-Aubin & Poirier, 1999b). The number of items recalled is higher for categorized lists, and there is a proportional increase in order errors. As before, because of the heightened activation presumed to accompany the presentation of a categorized list, we predicted that the items studied in Position 5 and 6 would tend to migrate forward (up the positions) more than controls.

However, as the reviewers pointed out, with the items all being taken from the same category, the straightforward expectation would be a similar level of coactivation across items, with the result that the entire series would be in a higher state of activation than a control list. Why would Items 5 and 6 migrate upward more than other items in the list?

To clarify this prediction, we need to consider the operation of the basic model described previously in a bit more detail. This is necessary in order to account for a feature of the data obtained in Experiment 1 and to justify the migration prediction made above.

An examination of Fig. 1 shows that for the *control condition*, Item 5 more often moved upward, toward Position 4, than downward, toward Position 6. Hence, even for control items—at least for the less well recalled positions—movement forward, toward earlier positions, was more likely. How could the primacy-gradient-plus-CQ mechanism produce this behavior? In order to answer this question, one must consider how the described system can produce blank responses (i.e., no item recalled in position *X*) and how the system can lead to an item not being recalled at all (item errors).[1]

In order to illustrate the proposed functioning of the CQ system, let us consider the recall of Item 5. Assume that noise and inhibition from the remaining items make the level of activation of Item 5 drop, and its successful retrieval in brought into question. On the basis of empirical errors rates, overall, the most likely outcome of this situation is a blank response. The second most likely possibility is the retrieval of the strongest competitor—based on activation within the primacy gradient, Item 6 (assuming that previous items were recalled and suppressed). If no item is recalled, none of the remaining response choices would be suppressed, reducing the probability of recall of the nonretrieved Item 5 as well as of the other available candidates, because of increased competition/mutual inhibition. If Item 6 is recalled, Item 5 remains in the competition, and again, the general probability of recall reduces through competition, although to a lesser degree, since Item 6 is now suppressed.

Generally speaking, suppression and competition mean that if the correct item is not retrieved and an item is recalled (i.e., the response is not blank), then the most likely item would be the following item (Item 5, in our example), creating the upward migrations observed in the data. If the previous item was not recalled (i.e., Item 4, in our example), it might win the competition and create an item transposition in which Items 4 and 5 were recalled as Items 5 and 4, respectively. However, in combination with this trend, a reduction in retrieval probability would be associated with moving through the primacy gradient, and a further reduction in the probability of retrieval with every error in recall. To summarize, retrieval difficulties open the window for upward movement, and reduced probability of retrieval would mean that this upward movement was not matched by errors in the other direction.

[1] In our data (from three different laboratories) across multiple experiments, item errors were the most high-frequency errors by far when different items were presented on each trial, as was the case here.

These processes operate for the control lists and would also be at play for the categorized lists. However, increased activation would mean more items retrieved, more forward movement, and proportional difficulties retrieving as the CQ mechanism works its way through the primacy gradient. We hence expected an increase in migrations toward earlier positions for semantically categorized lists in Experiment 2a, in which lists were studied in silence. Experiment 2b, in which lists were studied under suppression, was thought of as a replication that could help establish the robustness of the findings in Experiment 2a.

Experiment 2a: Results and discussion

Figure 2a and b summarize the main findings for this data set. As can be seen, more migrations were made for the categorized items than for the control lists. The results for each position were analyzed with two mixed ANOVAs; the between-subjects factor was List Type (categorized or not), and the within-subjects factor was Error Position. For Position 5, we found main effects of list type [$F(1, 138) = 10.05$, $MSE = 0.516$] and position [$F(5, 290) = 82.0$, $MSE = 0.514$], as well as a significant interaction [$F(5, 690) = 4.45$, $MSE = 2.29$]. The same effects were obtained for Position 6, with list type [$F(1, 138) = 24.69$, $MSE = 0.626$], error position [$F(5, 290) = 86.81$, $MSE = 0.718$], and the interaction [$F(5, 690) = 14.10$, $MSE = 0.718$] producing reliable effects. Simple main effect tests revealed the following: For the words studied in the fifth position, the difference between conditions was only significant for recall errors in Position 4. For the items studied in the sixth position, this difference was significant for the errors observed in Positions 4 and 5.

These findings fit nicely with those of Experiment 1; in both experiments, an increase in order errors/migrations for semantically related lists was observed, relative to control lists, as is predicted by the ANet account. Again, the increase in migration toward earlier and *not* toward later positions was found in this experiment.

Experiment 2b: Results and Discussion

Figure 3a and b summarize the main findings for the immediate serial recall data under suppression. As in Experiment 2a, it is clear that more migrations were made for the categorized lists than for the lists containing unrelated items. Here also, the results for each position were analyzed with two mixed ANOVAs with one between-subjects factor, List Type (categorized or not), and the within-subjects factor Error Position. For Position 5, we observed main effects of list type [$F(1, 110) = 15.77$, $MSE = 6.1$] and position [$F(5, 550) = 45.3$, $MSE = 14.8$], as well as a significant interaction [$F(5, 550) = 3.57$, $MSE = 1.16$]. The same effects were obtained for Position 6, with list type [$F(1, 110) = 28.7$, $MSE = 14.0$], error position [$F(5, 550) = 28.72$, $MSE = 11.56$], and the interaction [$F(5, 550) = 1.97$, $MSE = 0.79$] producing reliable effects. Simple main effect tests revealed that migrations were more pervasive for this experiment; for the words studied in the fifth position, the difference between conditions was significant for recall errors in Positions 3 and 4. For the items studied in the sixth position, this difference was significant for the errors observed in Positions 3, 4, and 5.

Again, the overall pattern of effects, as well as the details of the findings, conform to what would be expected in light of the ANet view. In all experiments, there were substantial changes in the error patterns for related items; as predicted, these migrations were toward earlier positions. Why would suppression lead to migrations across more positions than would silent conditions? What would be the effect of suppression on the primacy-gradient-plus-CQ system outlined here?

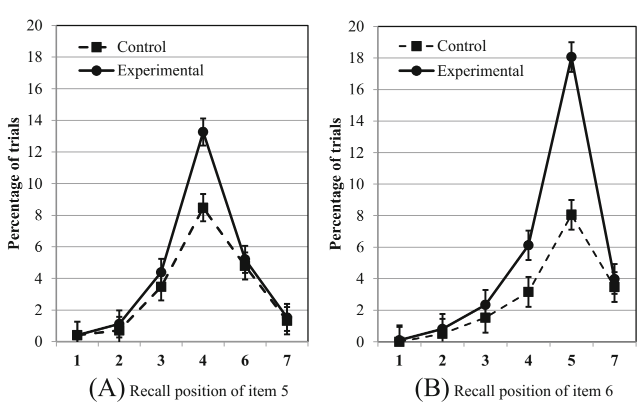

Fig. 2 Error frequencies for (**a**) Item 5 and (**b**) Item 6 as a function of recall position. Error bars represent 95 % confidence intervals computed according to the method of Loftus and Masson (1994) for the between-subjects factor of Similarity. When the difference between two means is significant, those confidence intervals do not overlap by more than half the distance of one side of an interval (Masson & Loftus, 2003).

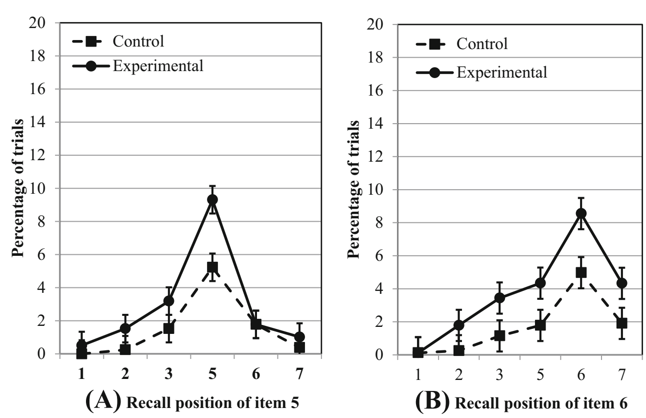

Fig. 3 Error frequencies for (**a**) Item 5 and (**b**) Item 6 as a function of recall position. Error bars represent 95 % confidence intervals computed according to the method of Loftus and Masson (1994) for the between-subjects factor of Similarity. When the difference between two means is significant, those confidence intervals do not overlap by more than half the distance of one side of an interval (Masson & Loftus, 2003)

Perhaps the most parsimonious suggestion is that suppression reduces the resources available for encoding; this would lead to lower activation levels, or a flattened primacy gradient. This account correctly predicts more omissions, and a greater number of omissions could also open the door to migrations across more positions, especially for the somewhat more activated items from the experimental condition.

General discussion

In the introduction to this article, we briefly reviewed a group of models that have been increasingly influential. These views insist on the importance of long-term knowledge in producing the behavior that is typically analyzed when studying short-term memory. Within this category of models, the proposal put forward recently by Acheson et al. (2011) makes a controversial suggestion: order recall in STM should be considered as the results of activation perturbations within existing semantic networks. On the basis of this view, dubbed the *ANet model* in the present article, a series of specific predictions were derived. More precisely, we tested the prediction that words for which the semantic activation was heightened by items within the same list would be more likely to migrate toward earlier positions within the list.

The findings of all three experiments plainly support the ANet perspective and predictions. In the first experiment, we manipulated the content of the first part of the list, so that in 50 % of the trials, assuming the operation of a semantic network, the fifth item's activation was increased. This was predicted to lead to a specific increase in migrations of this item toward earlier positions at the point of recall. The results of Experiment 1 showed precisely that pattern. Experiments 2a and 2b examined the same predictions, while eliminating a more trivial alternative interpretation of the first set of findings (i.e., that the migration of the target fifth item toward earlier positions was due to a grouping strategy). In both of these cases, the hypotheses derived from the ANet model were unequivocally supported.

Taken together, the results presented here are in line with the models that suggest that short-term serial recall relies on the activation of the LTM networks that are associated with language processing (e.g., Acheson & MacDonald, 2009; Cowan, 1999; Cowan & Chen, 2009; Gupta, 2003, 2009; Majerus, 2009; N. Martin & Gupta, 2004; R. C. Martin, 2006; Roodenrys, 2009).

The findings may also prove important for more formal models of serial order. More specifically, one way of looking at the present work is that it provides an empirical test for one of the most frequently proposed mechanisms within these models: primacy gradients (Hurlstone et al., 2014). In effect, many recent models of serial STM successfully account for the serial position curve that is typical of STM recall; the said curve is typified by strong primacy and a diminutive recency effect (typically only involving the last item or so). In order to account for the better recall of the first items, these proposals almost invariably include what is referred to as a *primacy gradient*; that is, they assume a decreasing strength of the encoding of successive items (although other mechanisms are also brought to bear in some instances; see, e.g., Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012). The predictions that were tested here assumed that order recall is guided by a primacy gradient, such that the most activated item is recalled first, followed by the second most activated item, and so on.

A related point relates to the suggestion, included in the Acheson et al. (2011) model, that the order-coding mechanism is integrated into the network that allows for item-level representation (i.e., activation within the lexico-semantic network). Recent quantitative models have typically involved separate mechanisms for coding order and item information. Consider, for example, the interference-based model of Oberauer et al. (2012); they offered a model of complex span that represented order in the same fashion as two previous models of immediate recall (Farrell, 2006; Farrell & Lewandowsky, 2002). In all three models, the authors called upon a distributed neural network that has a two-layer structure, with one layer representing serial positions and the other representing items. Items are encoded through Hebbian associations between item and position representations: The first list item is associated with the first-position representation (viz. a position marker), the second item is associated with the second-position marker, and so on (see also Henson, 1998). Memory for order is maintained by the patterns of association in the weight matrix that connects position markers to item representations. The links from position markers to items are unidirectional (going from the position markers to the items); at the point of recall, the position marker is used as the cue, and it leads to the retrieval of a blurry representation of the target. If one focuses on these aspects of the model, the results presented here can seem problematic. This is because it is not clear that a change in the activation of item representations could lead to perturbation of the associations between position markers and items: That is, the activation runs from the position markers to the items, and not the other way around. This being said, it is of course likely that this could be addressed by some reasonably slight tweak of the model's architecture. Importantly, in their review of the formal models of serial STM, Hurlstone et al.

(2014) also noted that perturbing the activations in one or both layers (i.e., item and order layers) predicts transposition errors akin to those observed in serial recall.

Conclusion

Previous interpretations have insisted that categorized lists have almost all of their effect in terms of increasing item recall (irrespective of position; Saint-Aubin & Poirier, 1999a, b). This increase is accompanied by a proportional increase in order errors. So, if order error proportions are the measure chosen, category would typically have no effect on order. However, Saint-Aubin et al. (2005) did report a statistically reliable effect of categorized lists on the *proportion* of order errors.

The ANet framework discussed here offers a straightforward and parsimonious interpretation of this typical pattern of findings: The representation of the words in an immediate serial recall task relies on available language-processing systems, including activation within and between phonological, sublexical, lexical, and semantic networks. In that sense, our view is well aligned with those suggesting that STM can be conceptualized as activated LTM rather than as a separate system (e.g., Acheson & MacDonald, 2009; Acheson et al., 2011; Cowan, 1999). Categorized lists lead to heightened network activation, which produces better item retrieval as well as perturbation of the representation of item order.

Our aim in this article was to test specific predictions derived from the Acheson et al. (2011) proposal; the latter suggests that STM relies on the LTM networks available for language processing. Our findings produced a pattern that was very much in line with the derived predictions. The results support models in which STM relies on activated LTM representations and networks.

References

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin, 135,* 50–68. doi:10.1037/a0014411

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

Atkins, A. S., & Reuter-Lorenz, P. A. (2008). False working memories? Semantic distortion in a mere 4 seconds. *Memory & Cognition, 36,* 74–81. doi:10.3758/MC.36.1.74

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. VIII, pp. 47–89). New York: Academic Press.

Botvinick, M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review, 113,* 201–233.

Buchsbaum, B. R., & D'Esposito, M. (2008). The search for the phonological store: From loop to convolution. *Journal of Cognitive Neuroscience, 20,* 762–778. doi:10.1162/jocn.2008.20501

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55,* 627–652.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge: Cambridge University Press.

Cowan, N., & Chen, Z. (2009). How chunks form in long-term memory and affect short-term memory limits. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 86–101). Hove: Psychology Press.

Deese, J. (1960). Frequency of usage and number of words in free-recall: The role of association. *Psychological Reports, 7,* 337–344.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton, E. Martin, & Committee on Basic Research in Education (Eds.), *Coding processes in human memory* (pp. 161–190). Washington, DC: Winston.

Farrell, S. (2006). Mixed-list phonological similarity effects in delayed serial recall. *Journal of Memory and Language, 55,* 587–600.

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review, 9,* 59–79. doi:10.3758/BF03196257

Grossberg, S. (1978a). Behavioral contrast in short-term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology, 17,* 199–219.

Grossberg, S. (1978b). A theory of human memory: Self-organization and performance of sensory–motor codes, maps, and plans. In R. Rosen & F. M. Snell (Eds.), *Progress in theoretical biology* (Vol. 5, pp. 233–374). New York: Academic Press.

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *Quarterly Journal of Experimental Psychology, 56,* 1213–1236. doi:10.1080/02724980343000071

Gupta, P. (2009). A computational model of nonword repetition, immediate serial recall, and nonword learning. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 108–135). Hove: Psychology Press.

Hamilton, A. C., & Martin, R. C. (2007). Proactive interference in a semantic short-term memory deficit: Role of semantic and phonological relatedness. *Cortex, 43,* 112–123.

Henson, R. N. (1998). Short-term memory for serial order: The start–end model. *Cognitive Psychology, 36,* 73–137.

Henson, R. N., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology, 49A,* 80–115.

Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish, & M. Zock (Eds.), *Current research in natural language generation* (pp. 287–319). London: Academic Press.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30,* 685–701.

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 and Language, 49,* 500–518.

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin, 140,* 339–373.

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 49, pp. 1–48). San Diego: Academic Press.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1,* 476–490. doi:10.3758/BF03210951

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

Majerus, S., & Boukebza, C. (2013). Short-term memory for serial order supports vocabulary development: New evidence from a novel word learning paradigm. *Journal of Experimental Child Psychology, 116,* 811–828.

Martin, R. C. (2006). Semantic short-term memory, language processing, and inhibition. In A. S. Meyer, L. R. Wheeldon, & A. Krott (Eds.), *Automaticity and control in language processing* (pp. 162–191). Hoboken: Taylor & Francis.

Martin, N., & Gupta, P. (2004). Exploring the relation-ship between word processing and verbal STM: Evidence from associations and dissociations. *Cognitive Neuropsychology, 21,* 213–228.

Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57,* 203–220. doi:10.1037/h0087426

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36,* 402–407. doi:10.3758/BF03195588

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

Page, M. (2005). Connectionist models of short-term memory for serial order. In G. Houghton (Ed.), *Connectionist models in cognitive psychology* (pp. 215–219). Hoboken: Taylor & Francis.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105,* 761–781. doi:10.1037/0033-295X.105.4.761-781

Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society B, 364,* 3737–3753. doi:10.1098/rstb.2009.0173

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *Quarterly Journal of Experimental Psychology, 48A,* 384–404.

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. J. (2009). Explaining phonological neighbourhood effects in short-term memory. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 177–197). Hove: Psychology Press.

Roodenrys, S., & Miller, L. M. (2008). A constrained Rasch model of trace redintegration in serial recall. *Memory & Cognition, 36,* 578–587. doi:10.3758/MC.36.3.578

Saint-Aubin, J., & Poirier, M. (1999a). The influence of long-term memory factors on immediate serial recall: An item and order analysis. *International Journal of Psychology, 34,* 347–352.

Saint-Aubin, J., & Poirier, M. (1999b). 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. (2000). Immediate serial recall of words and nonwords: Tests of the retrieval-based hypothesis. *Psychonomic Bulletin & Review, 7,* 332–340.

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.

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. doi:10.3758/BF03196364

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

Stuart, G., & Hulme, C. (2000). The effects of word co-occurrence on short-term memory: Associative links in long-term memory affect short-term memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 796–802. doi:10.1037/0278-7393.26.3.796

Tehan, G., Humphreys, M. S., Tolan, G. A., & Pitcher, C. (2004). The role of context in producing item interactions and false memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 107–119. doi:10.1037/0278-7393.30.1.107

Thorn, A. S. C., & Frankish, C. R. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 729–735. doi:10.1037/0278-7393.31.4.729

Thorn, A., & Page, M. (2009). *Interactions between short-term and long-term memory in the verbal domain*. Hove: Psychological Press.

Thorn, A., Frankish, C. R., & Gathercole, S. E. (2009). 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: Psychology Press.

Tse, C. S. (2009). The role of associative strength in the semantic relatedness effect on immediate serial recall. *Memory, 17,* 874–891.

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1256–1271. doi:10.1037/0278-7393.25.5.1256