Item-properties may influence item–item associations in serial recall

Jeremy B. Caplan, Christopher R. Madan and Darren Bedwell

Department of Psychology, University of Alberta

Author Note

Corresponding author: Jeremy Caplan. jcaplan@ualberta.ca. Psychology Department, Biological Sciences Building, University of Alberta, Edmonton, Alberta T6G 2E9, Canada, Tel:+1.780.492.5265, Fax:+1.780.492.1768.

We thank Alec Solway and Michael Kahana for valuable feedback on the manuscript. Supported by the Natural Sciences and Engineering Research Council of Canada.

Abstract

Attributes of words, such as frequency and imageability, can influence memory for order. In serial-recall, Hulme et al. (2003) found that high-frequency words were recalled worse, and low-frequency words better, when embedded in alternating lists than pure lists. This is predicted by associative chaining, wherein each recalled list-item becomes a recall-cue for the next item. However, Hulme et al. (2003) argued their findings supported positional-coding models, wherein items are linked to a representation of position, with no direct associations between items. They suggested their serial-position effects were due to pre-experimental semantic similarity between pairs of items, which depended on frequency, or a complex tradeoff between item- and order-coding (Morin et al., 2006). We replicated the smooth serial-position effects, but accounts based on pre-existing similarity or item–order tradeoffs were untenable. Alternative accounts based, imageability, phonological/lexical neighbourhood sizes were also ruled out. The standard chaining account predicts that if accuracy is conditionalized on whether the prior item was correct, the word-frequency effect should reappear in alternating lists; however, this prediction was not borne out, challenging this retrieval-based chaining account. We describe a new account, whereby frequency influences the strengths of item–item associations, symmetrically, during study. A manipulation of word-imageability also produced a pattern consistent with item–item cueing at study, but left room for effects of imageability at the final stage of recall. These findings lend further support for the contribution of associative chaining to serial-recall behaviour and show that item-properties may influence serial-recall in multiple ways.

Item-properties may influence item–item associations in serial recall

Introduction

A long-standing, unresolved debate is whether memory for serial lists is supported by item–item associations, as in "associative chaining" models (e.g., Lewandowsky & Murdock, 1989), or by linking list items to a separate representation of order or position, as in "positional-coding" models (e.g., Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1999; Henson, 1998). We consider one argument against chaining based on serial recall of lists alternating in some item-property, like word-frequency. Numerous authors have advanced firm arguments against chaining (e.g., Henson, 1998; Hurlstone, Hitch, & Baddeley, 2014). However, because recent evidence suggests serial-recall could rely on both chaining and positional mechanisms (e.g., Farrell, Hurlstone, & Lewandowsky, 2013; Kahana, Mollison, & Addis, 2010; Kahana, 2012; Serra & Nairne, 2000; Solway, Murdock, & Kahana, 2012), we consider both associative-chaining and positional-coding accounts of the word-frequency effect in serial recall.

In the alternating-list method (Baddeley, 1968), lists are constructed from two pools of items that differ, for example, in high (H) versus low (L) word-frequency. Assuming accuracy differs for lists composed only of high (H-Pure) and only of low (L-Pure) items, the critical comparison is with lists that alternate between pools (starting with either a high, HL-Alt, or low item, LH-Alt). In chaining models, each retrieved item is the cue for the next item. If an item is not retrieved, it cannot cue the next item (but see Lewandowsky & Murdock, 1989). Thus, if H and L items have different recall-probabilities, chaining models should predict better recall of L items from alternating than pure lists, being cued by preceding H versus L items, respectively (e.g., Baddeley, 1968; Henson, Norris, Page, & Baddeley, 1996). That is because the preceding H item is more likely to have been recalled, and thus available to retrieve the next (L) item. Likewise, H items should be disadvantaged in alternating lists, being cued by L versus H items. Chaining models might even predict smooth serial-position curves on alternating lists, if the H→L advantage approximately offsets the L→H disadvantage. In contrast, most published positional-coding models (e.g., Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1998), proceed with recall regardless of whether the previous item was retrieved (but see Farrell, 2012; Howard & Kahana, 1999). Positional-coding models predict the word-frequency effect should be unaffected by list composition, leading to a zig-zag serial-position curve on alternating lists, as they switch from H to L items.

The chaining-like pattern was found for word-frequency; high-frequency words suffer and low-frequency words benefit in alternating lists (Hulme et al., 2003). However, in their introduction, the authors briefly mentioned other researchers' arguments against chaining (which were subsequently disputed; e.g., Kahana, 2012). Because they assumed chaining was ruled out for other reasons, the authors sought to explain their findings without reference to item–item cueing. Pairwise semantic similarity was greatest for pairs of their high-frequency items, intermediate for mixed pairs and lowest for low–low pairs, which paralleled their accuracy effects. They suggested pre-existing similarity enhanced redintegration (deblurring the retrieved, noisy, feature-information to select the response) at the final stage of a positional-coding mechanism (Morin et al., 2006; Stuart & Hulme, 2000), or that smooth serial-position effects were due to tradeoffs between item- and order-memory approximately cancelling (Morin et al., 2006). However, *pairwise* similarity can impair order-memory (Tse, 2009); if similarity is incongruent with serial-order, learning order may be difficult. Also, pairwise similarity may only facilitate serial recall to the same extent as pre-exposure to items individually (Saint-Aubin & Poirier, 2005), which should produce zig-zag effects.

Associative-chaining implies serial-order is built upon association-memory. We do not know how Hulme et al.'s manipulation of word-frequency would have affected association-memory outside of a serial-recall task. Madan, Glaholt, and Caplan (2010) found word-frequency enhanced the retrievability of target items in cued-recall (replicated by Criss, Aue, & Smith, 2011). We used Madan et al.'s stimuli so we could extrapolate from association-memory, predicting high-frequency words would be better recalled and thus be more available to cue subsequent items.

Hulme et al. did not match stimuli on imageability, which can enhance serial recall (Miller

& Roodenrys, 2009; Walker & Hulme, 1999), or phonological and orthographic neighbourhood sizes, which might reduce the effectiveness of an item as a cue (Criss et al., 2011; Kahana, 2002; Rehani & Caplan, 2011). Our stimuli were matched on these characteristics.

Madan et al. (2010) also manipulated imageability, with word-frequency controlled. High-imageability increased association-strengths without significantly affecting target-item retrievability. We therefore manipulated imageability, which has not previously been investigated in alternating lists, because we thought both positional-coding and chaining models would predict no imageability effect in alternating lists, so the imageability manipulation might establish our sensitivity to that outcome. As shall be seen, the imageability results suggested a complex effect of imageability on serial-recall.

We also manipulated presentation rate to test whether subject-controlled strategies might be producing smooth serial position effects as was raised (and ruled out) by Morin et al. (2006).

Methods

Participants

University of Alberta students ($N = 151$) in an introductory psychology course, who learned English before the age of six and were comfortable typing, participated for partial course credit. The procedures were approved by a University of Alberta Research Ethics Board. Presentation-rate was manipulated between-subjects (2 s/word: $N = 75$; 4 s/word: $N = 76$).

Materials

Eight-word lists were constructed at random for each participant, from four pools of nouns (Madan et al., 2010): high-frequency, low-frequency, high-imageability, and low-imageability. Each pair of pools of a given type (e.g., high-frequency and low-frequency) was matched on word-length, mean positional bigram frequency, and orthographic neighborhood size (see Madan et al., 2010, for details). For frequency-manipulated pools, imageability was constrained to an intermediate range. For imageability-manipulated pools, frequency was constrainted to an intermediate range. List types were pure-high, pure-low, alternating high–low, or alternating low–high, where “high” and “low” refer to the manipulation of either frequency or imageability. The task was designed using the Python Experiment-Programming Library (Geller, Schleifer, Sederberg, Jacobs, & Kahana, 2007).

Procedure

Participants learned one practice list (not analyzed), and 24 experimental lists, in a study/test procedure. List-type order was random, except every eight consecutive study lists included one list of each condition (H-Pure/L-Pure/HL-Alt/LH-Alt $\times$ imageability/word-frequency manipulation), totalling three lists of each condition. Words were presented centrally, sequentially for study for either two or four seconds each, depending on the between-subjects group, with a 150-ms inter-stimulus interval. At test, participants were asked to type the word list in the order it was presented, pressing “ENTER” after each word. A maximum limit of 15 s was given for each word. A 400-Hz beep (first 80 participants) or silence (remaining participants)[1] was presented for 500 ms to signal that the response was submitted, followed by a 250-ms blank screen. Participants were instructed to type “PASS” to skip words they could not remember. Misspellings or variants of the correct word were scored as incorrect responses.[2] Serial recall was terminated after 45 s.

---

[1] due to sound-driver problems.

[2] We reran the main analyses with a more generous scoring criterion, adapted from a method used by Madan et al. (2010), that used the UNIX spell-checker, `aspell`, to give credit when the correct item was found within the list of suggested “corrections” offered. This set of analyses did not alter the pattern of findings; all significant effects remained significant and all non-significant effects remained non-significant. Thus, spelling errors had a very minimal influence on the results, perhaps in part due to our matching of the word pools on word length, bigram frequency, and orthographic neighborhood.

Results

Accuracy

For strict scoring (Figures 1a,c and 2a,c), a word was correct if it was recalled in its correct position. For lenient scoring (Figures 1b,d and 2b,d), a word was correct if it was on the most recent list, regardless of position (ignoring repetitions).[3] Because there was little trace of cross-over interactions with serial position (Figure 1), we collapsed across serial positions, combining together H items from both HL-Alt and LH-Alt lists, and likewise for L items (Figure 2). The zig-zag prediction now translates into the prediction that H and L differ on pure (left-hand bars) *and* alternating lists (right-hand bars). The smoothness prediction now translates into the prediction that H and L are equivalent on alternating lists (equal accuracy in the right-hand bars). We conducted a mixed, repeated-measures analysis of variance (ANOVA) with design Rate[2,4 s]×List[Pure, Mixed]×Item[High, Low] for each word-property manipulation.[4]

For word-frequency, the alternating lists showed little sign of the telltale zig-zag characteristic predicted by positional-coding models that assume no effect of recalls on subsequent recalls (Figure 1a); rather, alternating lists had fairly smooth serial-position curves with accuracies between accuracy on pure-high and pure-low lists. Supporting this, List×Item was significant, $F(1,149) = 16.0,\ MSe = 0.013,\ p < .0001$, which simple effects explained by a significant main effect of List for both high- and low-frequency words, but in opposite directions, $F(1,149) = 24.1,\ MSe = 0.015,\ p < .00001,\ \eta_p^2 = .14$, with pure > mixed lists; and $F(1,149) = 27.4,\ MSe = 0.014,\ p < .00001,\ \eta_p^2 = .16$, with mixed > pure lists, respectively (Figure 2a). Complementary simple effects found that the main effect of Item was significant for

---

[3] Because lenient-scored data produced the same pattern of significant and non-significant effects, we report analyses only for strict scoring.

[4] An initial ANOVA with design Rate[2,4 s]×Property[Frequency, Imageability]×List[Pure, Mixed]×Item[High, Low found no significant interactions with Rate ($p > 0.1$), indicating that memory was qualitatively similar for both presentation rates. The main effect of Property was non-significant, suggesting that the word pools were well matched for overall difficulty. We therefore analyze word-frequency and imageability separately and present data collapsed across presentation rates.

pure lists, $F(1,149) = 92.7,\ MSe = 0.017,\ p < .00001,\ \eta_p^2 = .38$, but not alternating lists, $p > .5, \eta_p^2 = .001$ (Figure 2a).

Imageability serial-position curves had the zig-zag characteristic, but also failed to touch the pure-list functions (Figure 1c); perhaps imageability enhances serial-recall in more than one way. Supporting this, List×Item was significant, $F(1,149) = 16.0,\ MSe = 0.013,\ p < .0001$, which simple effects explained as a pure-list advantage for high-imageable words, $F(1,149) = 8.6,\ MSe = 0.014,\ p < .01,\ \eta_p^2 = .054$, but the opposite for low-imageable words, $F(1,149) = 8.8,\ MSe = 0.011,\ p < .01,\ \eta_p^2 = .055$. The high-imageability advantage was found for *both* pure, $F(1,149) = 42.4,\ MSe = 0.021,\ p < .00001,\ \eta_p^2 = .22$, and alternating lists, $F(1,149) = 17.4, MSe = 0.005,\ p < .0001, \eta_p^2 = .10$ (Figure 2c).

Prior-list intrusions

If word-frequency, but not imageability, influences target-item accessibility here as it did for association-memory (Madan et al., 2010), then high-frequency words from prior lists should be intruded more than low-frequency words, whereas imageability of words should not affect intrusion rate. Measuring the number of prior-list intrusions divided by the number of item-errors (prior-list intrusions, extra-list intrusions and omissions), we conducted two ANOVAs with the design Rate×List×Item, excluding participants with no intrusions. No effects involving Rate were significant, $p > .05,\ \eta_p^2 < .06$.

For word-frequency, there were more high- than low-frequency intrusions (Figure 3a), as expected (main effect of Item, $F(1,61) = 23.8,\ MSe = 0.001,\ p < .00001$). However, intrusion rate depended on list type, as List×Item was significant, $F(2,122) = 6.83,\ MSe = 0.002,\ p < .01$. This was explained by simple effects; as can be seen in Figure 3a), this word-frequency effect applied to pure-high, $F(1,61) = 18.8,\ MSe = 0.002,\ p < .00001,\ \eta_p^2 = .24$, and alternating, $F(1,61) = 4.24,\ MSe = 0.001,\ p < .05,\ \eta_p^2 = .07$, but not pure-low lists, $p > .5,\ \eta_p^2 = 0.005$. Thus, high-frequency words are easier to produce than low-frequency words, but frequency may be used to screen responses.

For imageability, as expected, high-imageable words were not easier to produce overall (Figure 3b), as the main effect of Item was non-significant. Intrusion rate did depend on list type, List×Item: $F(2,100) = 3.65,\ MSe = 0.001,\ p < .05$. Simple effects found significantly more low-imageable intrusions only on pure-low lists ($p < .05$). Thus, participants may have used imagery to guide guessing.

Distance functions

According to the item–order hypothesis, processing low-frequency items diverts resources from encoding their order, but this occurs less on alternating lists (Morin et al., 2006). Therefore, low-frequency items should be recalled closer to their correct position in alternating lists than in pure lists. However, distance functions showed no trace of this for either word-frequency or imageability (Figure 4).

Item-cueing at test

In the standard chaining account, low-frequency items are retrieved less often, and are thus less available as cues for subsequent high-frequency items on an alternating list, and vice-versa for high-frequency items facilitating subsequent low-frequency items. It follows that if recall were conditionalized on accuracy of the prior item, the word-frequency effect should re-appear on alternating lists. This was not the case; conditionalizing on correct prior responses (Figure 5a), the pure-list word-frequency effect was still obtained, but with no trace of a word-frequency effect for alternating lists (nor conditionalizing on errors; Figure 5c). Interestingly, the high-imageable advantage was found when conditionalized on both correct and error prior responses (Figure 5b,c). Thus, although word-frequency appears not to act via item-cueing at test, imageability may.

Discussion

The pattern expected by the standard positional-coding account was not found. Instead, item-property effects reduced (imageability) or vanished (word-frequency) in alternating lists.

Our stimuli were matched for neighborhood sizes, and the word-frequency manipulation controlled for imageability and vice-versa, ruling out several possible alternative interpretations of our findings. The lack of interaction with presentation-rate rules out participant-controlled strategies, including covert rehearsal, producing smooth serial-position effects (Morin et al., 2006). Further, item–order tradeoff accounts are difficult to reconcile with the stability of the distance functions across list types (Figure 4) and the similarity between lenient and strict scoring (Figure 1).

The semantic-similarity account

Following Hulme et al. (2003), we tested the semantic-similarity account of smooth serial-position effects, estimating similarity with $cos(\theta)$, where θ was the angle between vector-representations of items obtained via Latent Semantic Analysis (Landauer & Dumais, 1997). Larger $cos(\theta)$ indicates greater similarity. Reported by Madan et al. (2010) with our stimuli, for word-frequency, $cos(\theta)$ was greatest for HH pairs ($M \pm SD = .17 \pm .16$), but did not differ substantially between HL and LL pairs ($.077 \pm .080$ and $.05 \pm .11$, respectively). For imageability, the values were $.10 \pm .14$, $.078 \pm .080$ and $.12 \pm .15$, respectively. Unlike Hulme et al. (2003), our similarity pattern is different than the pattern of accuracy across conditions for both the word-frequency and imageabililty manipulations, challenging semantic-similarity accounts.

The item–item cueing account at recall

Although it was an appealing idea that serial-recall behaviour might follow from memory for associations, and consistent with greater rates of high-frequency words as prior-list intrusions, the conditional analyses (Figure 5) are incompatible with this view, and suggest a dissociation between memory in paired-associate (Madan et al., 2010) and serial-recall procedures. Serial lists and associations may be remembered quite differently (Murdock & Franklin, 1984), but it might be possible, in the future, to reconcile them (Caplan, 2005), explaining differences by considering that in serial lists, each item is both a cue and a target.

An account based on associative-chaining effects at study

Our findings may be compatible with item-properties acting on item–item associations during study. This may follow from current formulations of association-memory models: if one item is processed more effectively at study, the benefit distributes (multiplies through), such that memory for the association is strengthened regardless of how it is tested. Consider a convolution-based model (like Murdock, 1982). To store a pair:

$$\mathbf{w} = \mathbf{f} * \mathbf{g}, \tag{1}$$

where vectors are set in boldface, * denotes the convolution operation, **w** stores the memory and **f** and **g** are the item-vectors to be associated, with an encoding strength of 1. Now, if the representation of **f** were stronger during study, implemented by multiplying the item by a scalar, $\alpha > 1$:

$$\mathbf{w} = \alpha\mathbf{f} * \mathbf{g} \equiv \alpha(\mathbf{f} * \mathbf{g}). \tag{2}$$

Because scalars multiply through, the encoding strength of the pair is effectively strengthened by a factor α. The same result holds if **g** were strengthened, and in matrix models (e.g., Humphreys, Bain, & Pike, 1989). If word-frequency strengthens items like this during encoding, then because in alternating lists, each nearest-neighbor association includes exactly one H and one L word, HL and LH associations will be equivalent in strength, stronger than LL associations and weaker than HH associations on average. Thus, our findings are consistent with an account of word-frequency effects that acts during the study phase of an associative-chaining model. This account is consistent with word-frequency strengthening associations in cued-recall (Madan et al., 2010), but do not explain why the large effect of frequency on target-retrievability in cued-recall Madan et al. seem virtually absent in serial-recall.

A redintegration account of imageability

A redintegration account might partly apply to imageability, which affected serial-recall on alternating lists. High-imageable items may receive a boost in retrievability while not affecting the item's effectiveness as a cue for the next item. For example, consider a three-word list, CHILD–PONY–HAY. The participant may represent this list with two images, a child riding a pony and a pony eating hay. Suppose the participant remembers the first image, but erroneously produces HORSE instead of PONY. HORSE would be an error, but might be just as effective as a cue for HAY as PONY.

Positional-coding models

Our pattern of results does not rule out positional-coding models, but does suggest some constraints on positional-coding model design. With important exceptions, positional-coding models typically assume recall proceeds from one position to the next without being influenced by the outcome of the prior retrieval (e.g., Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1998). Item-properties are thus thought to affect a late stage of the model such as redintegration or response selection, after positional cueing is complete. This allowed positional-coding models to explain so-called "immunity" effects in lists alternating in phonological confusability, wherein non-confusable items were seemingly unaffected by neighboring list items. However, Farrell and Lewandowsky (2003) showed that immunity was a coincidental, approximate cancelling of facilitating and impairing effects on non-confusable items in alternating lists, challenging those model-accounts. Smooth serial-position effects in alternating lists further challenge models that assume recall proceeds regardless of the outcome of each recall.

One solution could be to assume an item is retrieved with some contextual (Howard & Kahana, 1999), or positional information (Caplan, 2005; Rehani & Caplan, 2011). Alternatively, Lewandowsky and Farrell (2008) added to their positional-coding model, a closed-loop encoding rule to explain phonological confusability effects in alternating lists: encoding strength is inversely related to the match of an item to the current contents of memory (Lewandowsky &

Murdock, 1989).[5] Finally, Farrell's (2012) positional-coding model assumed items are stored in groups of items. If a group can be facilitated as a whole, this model might produce smooth serial-position effects. Alternatively, item-properties might affect group boundaries (Farrell, 2012); if participants form groups of size two in alternating lists, that might produce smooth serial-position effects.

Conclusion

Smooth serial-position effects in serial-recall of alternating lists challenge models that assume recall proceeds without regard to the outcome at the prior position, and accounts that look only to the redintegration phase. Rather, our findings point to effects of item-properties at study, either on the strength of item–item associations, or in some relational manner in positional-coding models.

---

[5] At face-value, this account does not seem applicable to word-frequency and imageability manipulations. Furthermore, our high-frequency pairs and Hulme et al.'s (2003) had the greatest similarity, so a closed-loop rule may wrongly predict an inverted word-frequency effect for pure lists.

References

Baayen, R., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database (Release 2) [CD-ROM].* Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].

Baddeley, A. D. (1968). How does acoustic similarity influence short-term memory? *Quarterly Journal of Experimental Psychology*, *20*(3), 249-264.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillatory-based memory for serial order. *Psychological Review*, *107*(1), 127-181.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*, 551-581.

Caplan, J. B. (2005). Associative isolation: unifying associative and order paradigms. *Journal of Mathematical Psychology*, *49*(5), 383-402.

Criss, A. H., Aue, W. R., & Smith, L. (2011). The effects of word frequency and context variability in cued recall. *Journal of Memory and Language*, *64*(2), 119-132.

Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review*, *119*(2), 223-271.

Farrell, S., Hurlstone, M. J., & Lewandowsky, S. (2013). Sequential dependencies in recall of sequences: filling in the blanks. *Memory & Cognition*, *41*(6), 938-952.

Farrell, S., & Lewandowsky, S. (2003). Dissimilar items benefit from phonological similarity in serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 838-849.

Geller, A. S., Schleifer, I. K., Sederberg, P. B., Jacobs, J., & Kahana, M. J. (2007). PyEPL: a cross-platform experiment-programming library. *Behavior Research Methods*, *39*(4), 950-958.

Henson, R. N. A. (1998). Short-term memory for serial order: the Start-End Model. *Cognitive Psychology*, *36*(2), 73-137.

Henson, R. N. A., 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*(1), 80-115.

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 923-941.

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*(4), 500-518.

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, *96*(2), 208-233.

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*(2), 339-373.

Kahana, M. J. (2002). Associative symmetry and memory theory. *Memory & Cognition*, *30*(6), 823-840.

Kahana, M. J. (2012). *Foundations of human memory*. New York: Oxford University Press.

Kahana, M. J., Mollison, M. V., & Addis, K. M. (2010). Positional cues in serial learning: the spin-list technique. *Memory & Cognition*, *38*(1), 92-101.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211-240.

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: new data and a model. *Psychology of Learning and Motivation: Advances in Research and Theory*, *49*(1), 1-48.

Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review*, *96*(1), 25-57.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs.

Psychonomic Bulletin & Review, *1*(4), 476-490.

Madan, C. R., Glaholt, M. G., & Caplan, J. B. (2010). The influence of item properties on association-memory. *Journal of Memory and Language*, *63*(1), 46-63.

Miller, L. M., & Roodenrys, S. (2009). The interaction of word frequency and concreteness in immediate serial recall. *Memory & Cognition*, *37*(6), 850-865.

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*(4), 724-729.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*(6), 609-626.

Murdock, B. B., & Franklin, P. E. (1984). Associative and serial-order information: different modes of operation? *Memory & Cognition*, *12*(3), 243-249.

Rehani, M., & Caplan, J. B. (2011). Interference and the representation of order within associations. *Quarterly Journal of Experimental Psychology*, *64*(7), 1409-1429.

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*(3/4), 325-332.

Serra, M., & Nairne, J. S. (2000). Part-set cueing of order information: implications for associative theories of serial order memory. *Memory & Cognition*, *28*(5), 847-855.

Solway, A., Murdock, B. B., & Kahana, M. J. (2012). Positional and temporal clustering in serial order memory. *Memory & Cognition*, *40*(2), 177-190.

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*(3), 796-802.

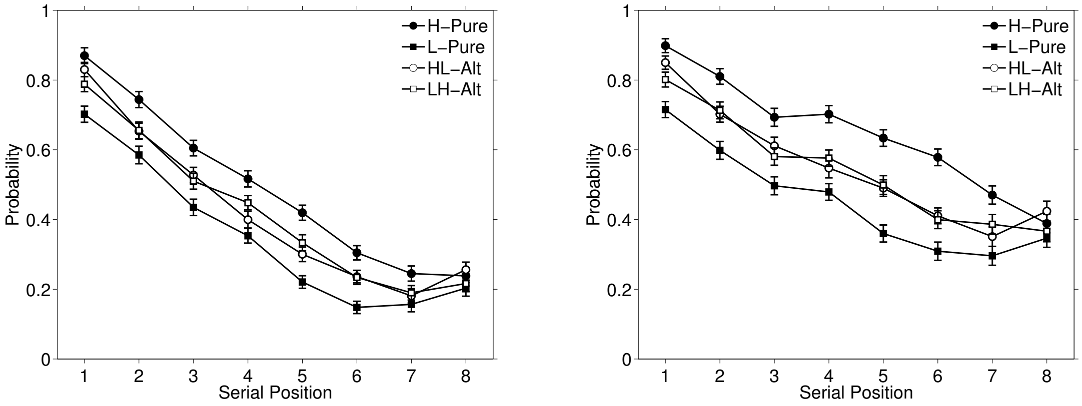
Tse, C.-S. (2009). The role of associative strength in the semantic relatedness effect on immediate serial recall. *Memory*, *17*(8), 874-891.

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*(5), 1256-1271.

Word-Frequency Manipulation

a Strict Scoring **b** Lenient Scoring



Imageability Manipulation

c Strict Scoring **d** Lenient Scoring

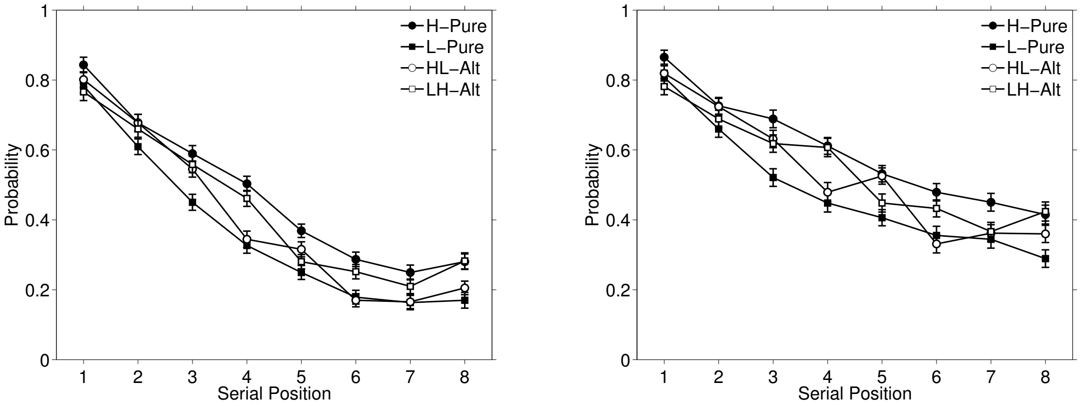
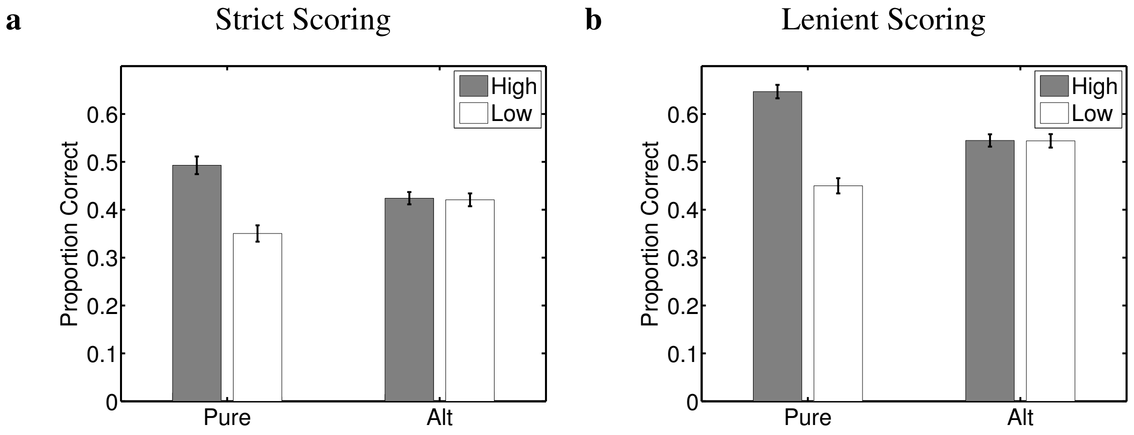


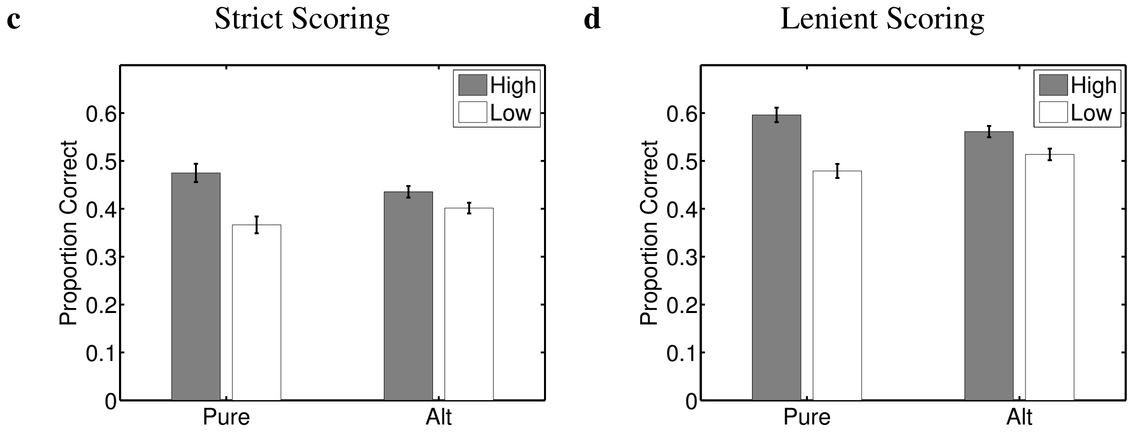
Figure 1. Probability of correct recall as a function of serial position, for high- and low-valued (frequency or imageability), collapsed across 2-s and 4-s presentation rates. a,b, Word-frequency manipulation; c,d, Imageability manipulation. a,c, strict scoring; b,d, lenient scoring. H-Pure: Pure lists of high (frequency/imageability) words. L-Pure: Pure lists of low (frequency/imageability) words. HL-Alt: Alternating lists starting on a high item. LH-Alt: Alternating lists starting on a low item. Error bars are 95% confidence intervals, corrected for inter-individual differences (Loftus & Masson, 1994).

Word-frequency Manipulation



a Strict Scoring **b** Lenient Scoring

Imageability Manipulation



c Strict Scoring **d** Lenient Scoring

Figure 2. Probability of correct recall as a function of high/low item, manipulation, list composition (pure versus mixed lists), and scoring method (strict versus lenient), collapsed across presentation rates. Error bars are 95% confidence intervals, corrected for inter-individual differences (Loftus & Masson, 1994).

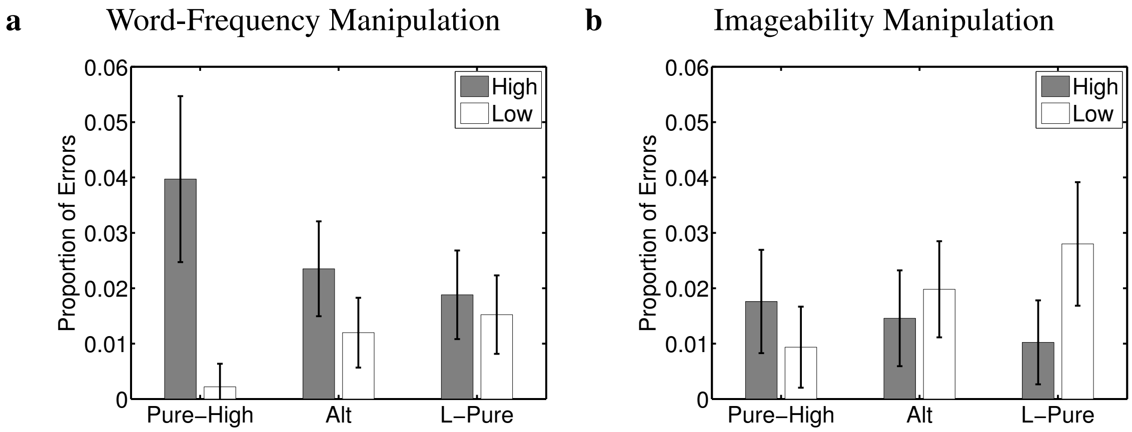


Figure 3. Prior-list intrusion rate (proportion of total errors) for each manipulation, as a function of the list type during which the intrusion occurred, and whether the item was high or low in value. Error bars are 95% confidence intervals, corrected for inter-individual differences (Loftus & Masson, 1994).

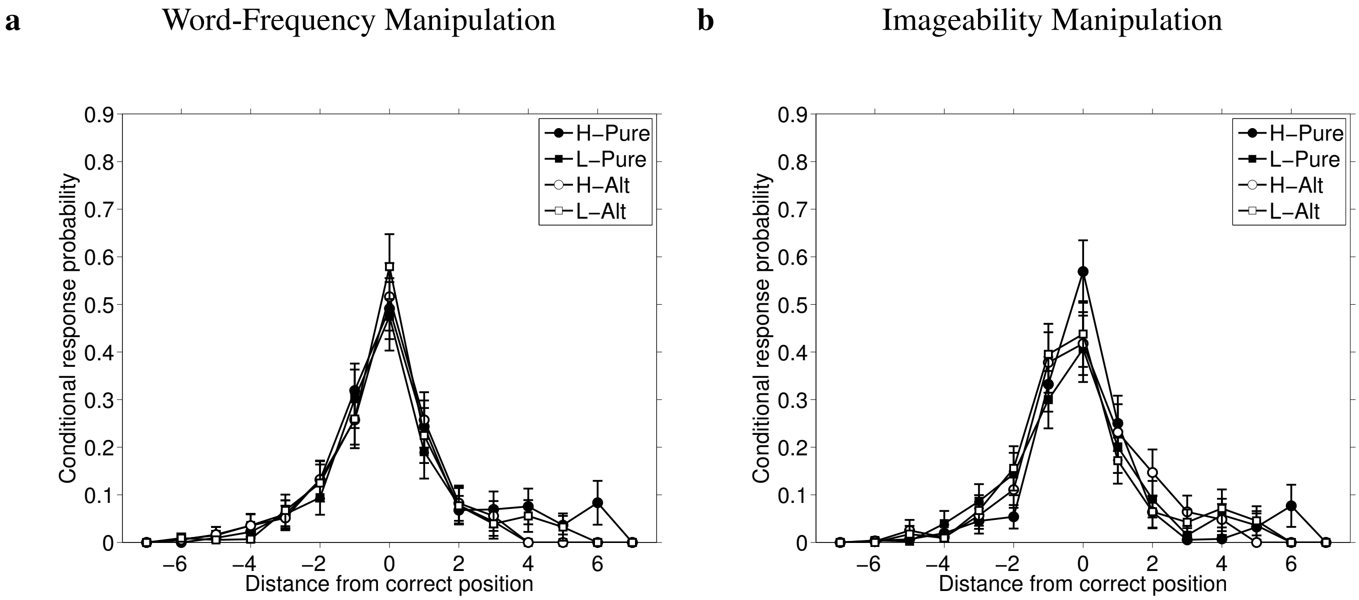


Figure 4. Distance functions for responses following the first order-error (as recommended by Solway et al., 2012). Conditional response probability is plotted as a function of lag from the item's correct position, for the word-frequency (a) and imageability (b) manipulations. Distances of zero are responses of items in their correct position, and the sharpness of the peak can be interpreted as a measure of positional-certainty (order-memory). Error-bars are 95% confidence intervals based on standard error of the mean.

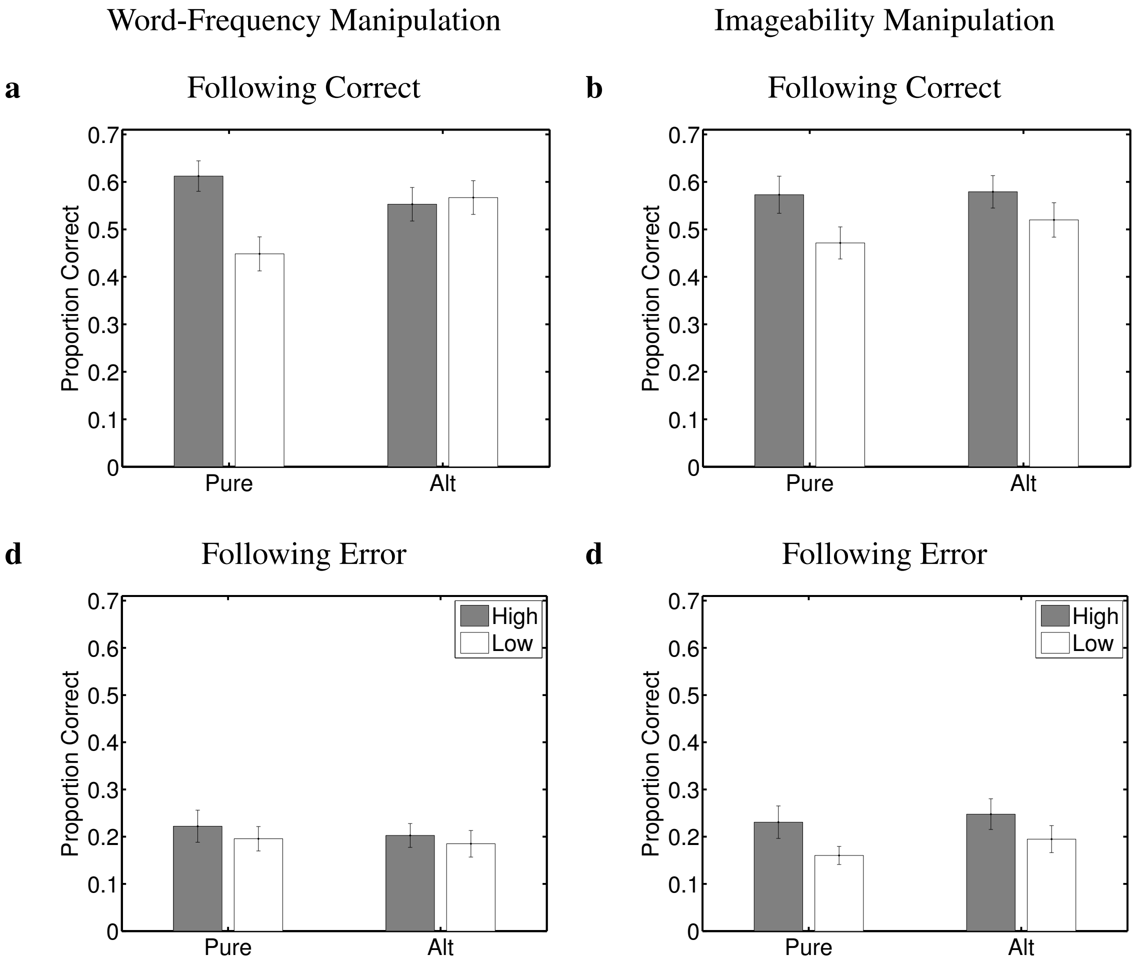


Figure 5. Serial-position curves conditionalized on the prior item being correct (a,b; strict scoring criterion) or incorrect (c,d), for the word-frequency manipulation (a,c) and the imageability manipulation (b,d); compare with Figure 2. Error bars are 95% confidence intervals, uncorrected. Due to many missing values, t tests are unpaired, as follows: Word-frequency effect following correct responses for pure lists, $t(297) = 6.69, p < 0.0001$; for alternating lists: $t(297) = -0.55, n.s.$. Following error responses for pure lists: $t(299) = 1.22, n.s.$; for alternating lists: $t(299) = 0.92, n.s.$. Imageability effect following correct responses for pure lists: $t(298) = 3.85, p < 0.001$; for alternating lists: $t(298) = 2.33, p < 0.05$. Following error responses for pure lists: $t(296) = 3.51, p < 0.001$; for alternating lists: $t(300) = 2.40, p < 0.05$.