THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
0000, 00 (0), 1–9

Psychology Press
Taylor & Francis Group

Short article

List strength effect without list length effect in recognition memory

Luciano G. Buratto and Koen Lamberts
University of Warwick, Coventry, UK

The study of list length effects (adding items to a list affects memory for the other items) and list strength effects (strengthening some items in a list affects memory for the nonstrengthened items) is important to constrain models of memory. In recognition memory, a list length effect is generally found, whereas a list strength effect is not. Using the switched-plurality procedure in an old–new recognition task (e.g., study *banana*; test *bananas*), we found the opposite pattern. Length manipulations caused no change in memory performance, whereas strength manipulations did. The list strength effect was found when recollection was likely to operate at test (with switched-plurality lures). When recollection was unlikely to operate (with unrelated lures), the strength effect disappeared. The result was observed using both a size judgement task (which has previously produced positive list strength effects) and a pleasantness judgement task (which has not yielded list strength effects before).

The study of interference is an important element of memory research. In particular, two manipulations have been used to assess interference in memory. List length manipulations test how adding items to a list of words affects memory for the other words in the list. A list length effect (LLE) involves better performance on short lists than on long lists. List strength manipulations, on the other hand, test how strengthening items (e.g., by repetition) affects memory for the other, nonstrengthened items in the list. A list strength effect (LSE) occurs when performance on nonstrengthened items is better in pure weak lists (where all items have the same strength) than in mixed lists (where some items have been strengthened).

The LLE has been observed reliably in recall and in recognition. The LSE has also been observed in recall, but most studies have not found an LSE in recognition (see Clark & Gronlund, 1996, for a review). Memory theorists have accepted a positive LLE and a null LSE in old–new recognition as firmly established, and they have designed or modified their models

Correspondence should be addressed to Luciano G. Buratto, Department of Psychology, University of Warwick, Coventry CV4 7AL, UK. E-mail: l.g.buratto@warwick.ac.uk

The authors are grateful to Michael Humphreys, Kenneth Norman, and two anonymous reviewers for their helpful comments on a previous draft of the manuscript.

© 0000 The Experimental Psychology Society

http://www.psypress.com/qjep DOI:10.1080/17470210701566713

accordingly. For example, the classical Search of Associative Memory (SAM) model (Gillund & Shiffrin, 1984), which originally predicted an LSE, has been modified by adding a *differentiation* assumption, whereby strong items become increasingly distinct from weak items. The differentiation assumption, which can account for both the presence of an LLE and the absence of an LSE in recognition, has been incorporated in subsequent memory models (see Criss & McClelland, 2006, for a comparison of two such models).

Recently, however, several studies have cast fresh doubts on the status of strength and length effects in recognition memory. Dennis and Humphreys (2001) found neither an LLE nor an LSE in recognition when several confounding variables (e.g., study–test lag, attention, rehearsal, context reinstatement) were controlled. Norman (2002) found a reliable LSE in item recognition (see Verde & Rotello, 2004, for similar results using the associative recognition paradigm). The findings of a null LLE (Dennis & Humphreys, 2001) and a positive LSE (Norman, 2002; henceforth ∘LLE/ + LSE) highlight the uncertainty over the status of length and strength effects in recognition and call for further investigation of the variables that determine the occurrence of these effects.

There is growing evidence from behavioural, physiological, and pharmacological studies (e.g., Boldini, Russo, & Avons, 2004; Curran, DeBuse, Woroch, & Hirshman, 2006) that two processes may operate during recognition judgements: *familiarity*, a fast, context-insensitive, automatic process, and *recollection*, a slow, context-sensitive, strategic process. The processes have been assumed to operate in a fall-back manner, whereby decisions are based on familiarity if recollection fails. The differential contribution of familiarity and recognition can account for many results in recognition (see Yonelinas, 2002, for a review).

In this article, we report a recognition-memory experiment in which we manipulated list length and strength, lure relatedness, and encoding task. The first aim of our study was to test the hypothesis that previous ∘LLE/∘LSE (Dennis & Humphreys, 2001) were obtained as a result of low recollection rates at test. Length and strength manipulations may impair recollection more than familiarity (Norman & O'Reilly, 2003). Accordingly, previous null results could be explained by a relatively small contribution of recollection processes at test. If *targets* (studied items) and *lures* (nonstudied items) are highly dissimilar, familiarity alone may be a reliable basis for recognition judgements. However, if targets and lures are similar, familiarity alone may not be diagnostic (e.g., one needs to recollect seeing *banana* to reject lure *bananas*). It has been shown that such a *recall-to-reject* process—the recall of mismatching features in test lures—is involved in recognition tasks with high, but not with low, target–lure similarity (Rotello, Macmillan, & Van Tassel, 2000). Although different definitions of recollection have been proposed, we treat recollection as a recall-to-reject process for the purposes of this paper (for alternative definitions, see Diana, Reder, Arndt, & Park, 2006). In Dennis and Humphreys' (2001) experiments, targets and lures were dissimilar, which raises the possibility that responses were based on familiarity alone. To investigate this possibility, we varied target–lure similarity (to manipulate the likelihood of recall-to-reject) and assessed its impact on length and strength effects.

The second aim of this study was to evaluate the role played by possible confounds in previous studies. We tested whether the LSE previously observed in recognition (Norman, 2002) was not the result of a longer study list in the strong condition and that the null LLE and LSE (Dennis & Humphreys, 2001) were not a consequence of participants' use of covert rehearsal strategies at study. Norman (2002) compared weak short lists with long strong lists. Because repeating study items entails a longer list, length of list presentation and list strength were confounded. Our experiment incorporated several of the design features introduced by Norman (2002, Exp. 2) with the addition of a list length manipulation to test for the possibility that the additional study time may have accounted for some of the observed LSE. The fact that many previous studies have not found an LSE in recognition argues against an effect of study time, as the stronger lists in those studies were also longer, and yet no difference was found. However, because encoding time in most studies has been fixed (and long), rather than self-paced (and short), participants could have used some of the encoding time to rehearse previously presented items. This is particularly true for strength manipulations where study items are repeated; after one repetition, participants may decide that they have already learned the item and use the remaining study time to practise weaker items. Norman (2002) used short encoding times (1.15 s) to prevent this type of rehearsal. Dennis and Humphreys (2001), on the other hand, used longer fixed encoding times (3 s). In an attempt to reduce the possible contribution of rehearsal, we used a self-paced encoding task, in which participants were encouraged to move on to the next item as quickly as possible.

Finally, we aimed to assess the importance of procedural differences between Dennis and Humphreys (2001) and Norman (2002), who used different encoding conditions. Dennis and Humphreys (2001) used a pleasantness rating task at study, whereas Norman used a size judgement task, in which participants had to decide whether a typical exemplar of given word (e.g., *banana*) would fit into a shoebox present in the experimental room. As argued by Norman (2002), the purpose of the size judgement task was to increase the chances of trace overlap, as all words would presumably be encoded with a common referent (the shoebox), thus increasing memory interference. In the pleasantness rating task, by contrast, each word may have been encoded with a different referent, reducing the chances of interference. To test the possibility that the encoding task may have been responsible for the LSE observed by Norman (2002) or the null LSE reported by Dennis and Humphreys (2001), we manipulated encoding task (size vs. pleasantness judgement).

To summarize, our study differed from Norman (2002) in the length manipulation, variable encoding time, and the use of both size and pleasantness judgement tasks at encoding. Our study differed from Dennis and Humphreys (2001) in our use of similar lures and the two encoding tasks.

Method

Participants

A total of 132 University of Warwick undergraduates (53 males; mean age = 21.5 years) participated in the study. Participants were tested individually. Each session took about 45 minutes, and participants were paid £5 (about $10).

Materials

The stimuli were 360 imageable, concrete, familiar and medium-frequency nouns from the MRC Psycholinguistic Database. The words were screened for semantic similarity using latent semantic analysis so that none of the items was strongly related to another (for more details, see Norman, 2002). A total of 30 words were used as fillers. The remaining 330 words were randomly assigned to 11 groups of 30 words, matched for word characteristics. Words were classified as *target* (if presented both at study and test), *interference* (if presented at study but not at test), or *lure* (if presented at test but not at study). The lures were further classified as *SP lures* (switched-plurality; e.g., study *banana*, test *bananas*) or *unrelated lures* (e.g., study *banana*, test *car*). Of the 11 word groups, 3 consisted of targets, 5 contained interference words, and 3 consisted of unrelated lures. SP lures were constructed by switching the plurality of half the targets (from singular to plural or vice versa). All plural forms were generated by adding *s* to their singular form. A distinct word sample was produced for each participant.

Design

Each participant attended one session. The session consisted of three experimental blocks, each containing one of three different list types: *weak interference short* (WIS; 30 target items presented once and 30 interference items presented once), *weak interference long* (WIL; 30 target items presented once and 90 interference items presented once), and *strong interference* (SI; 30 target items presented once and 30 interference items presented three times). Participants were tested on all three list types, with list order balanced across participants. A retroactive design was used: All target words were presented before any of the interference items were repeated. This prevented participants from distinguishing targets from interference items (thus preventing differential rehearsal of targets during study). Encoding condition (size vs. pleasantness judgement) was manipulated between participants (with 66 participants in each condition), and list type (WIS, WIL, and SI) was manipulated within participants.

Procedure

Stimuli were presented on a 43-cm cathode-ray tube (CRT) monitor. Each session consisted of four blocks: a practice block and three experimental blocks. Each block consisted of three phases: study, video game, and test. In the study phase, participants were presented with 60 (WIS), 120 (WIL), or 120 (SI) items. A total of 10 extra items were used as fillers (5 at the start and 5 at the end) of each study list to control for primacy and recency effects. Participants were warned that some items might appear several times. They were also informed that their memory would be tested. Participants were instructed to decide either whether or not a typical instance of the object denoted by the word would fit into a shoebox (size judgement) or whether the instance would be considered pleasant (pleasantness judgement). For practical reasons, participants in the size judgement condition were all tested before participants in the pleasantness condition. All participants were instructed to pay attention to the plurality of the words, imagining a single instance for singular words and two instances for plural words. Participants provided size or pleasantness ratings on a scale from 1 to 6 (*definitely yes* to *definitely no*). The task was self-paced, with an upper display time limit of 3,000 ms, with 500 ms of interstimulus interval. A video game task was used to balance study–test lag across list types (210 s for WIS, no game for WIL and SI). The test list consisted of 60 words (15 old items, 15 SP lures, and 30 unrelated lures). Words appeared one at a time on the screen. The participants were instructed to rate their confidence in their judgement on a scale from 1 to 6 (*definitely old* to *definitely new*). They were also told to try and recall studied words when possible (e.g., when presented with SP lures) to help them increase memory accuracy.

Results

Testing the hypotheses outlined in the Introduction required three comparisons: (a) studied items versus unrelated lures (SU); (b) studied items versus related lures (SSP); and (c) related lures versus unrelated lures (SPU). SU and SSP discriminations measure how much more likely participants were to say "old" to targets than to unrelated and related lures, respectively. If Dennis and Humphreys' (2001) null results were due to relatively low recollection rates, we should observe a ∘LLE/∘LSE in SU and a +LLE/+LSE in SSP. The only difference between the two comparisons is the degree of target–lure similarity, which directly affects the likelihood that recollection plays a role at test (Rotello et al., 2000). If length and strength manipulations act by reducing recollection, we should observe a decrease in performance in SSP, where recollection is presumably important, but not in SU, where familiarity alone may be sufficient for correct old–new discrimination.

Moreover, we should find negative effects (−LLE/−LSE) in SPU. A negative LLE occurs when performance on short lists is worse than that on long lists. Similarly, a negative LSE occurs when performance on nonstrengthened items is worse on pure weak lists than on strong lists. In the SPU comparison, related lures are analysed as targets. This provides a measure of *pseudodiscrimination* (how much more likely are participants to say "old" to related lures than to unrelated lures). High recollection rates should produce weak pseudodiscrimination, because related lures would be confidently rejected. Low recollection rates should increase pseudodiscrimination, because related lures would be mistaken for targets.

The confidence ratings collected at test were used to construct receiver operating characteristic (ROC) curves for each participant, list type, and encoding condition. Hit rates and false-alarm rates for both encoding conditions are presented in the Appendix. Sensitivity (A_z) was estimated for each individual by fitting an unequal-variance Gaussian model to each participant's confidence data using the RscorePlus maximum-likelihood algorithm (Harvey, 2001). A_z is an estimate of the area under the ROC and exhibits better statistical properties than single-point discrimination measures such as d' or A' (Verde, Macmillan, & Rotello, 2006). The analysis of a derived measure (A_z) was essential, because list strength manipulations are known to affect criterion placement as well as sensitivity measures (e.g., Hirshman, 1995). A_z is a multipoint discrimination measure and thus does not depend on criterion placement.

Sensitivity was computed as follows. First, nine ROC curves were generated for each participant (3 list types: WIS, WIL and SI $\times 3$ discrimination types: SU, SSP, and SPU). Second, a Gaussian model was fitted to each ROC curve. Finally, the model parameters were used to calculate A_z. We filtered out participants whose data provided poor fits in at least one condition (chi-squared p-value <.05; note that this is a very liberal exclusion criterion); 13 participants were excluded from analysis. Alpha was set to .05 (two-tailed) for all analyses, unless otherwise stated.

Repeated-measures analyses of variance (ANOVAs) on the discrimination measure (A_z) with list type (WIS, WIL, and SI) and encoding condition (size and pleasantness) as the independent variables were carried out separately for each discrimination type (SU, SSP, and SPU). There was no main effect of encoding condition and no interaction between list type and encoding condition for any of the discrimination types. Therefore, we analysed the data collapsed across encoding conditions.

The analysis of the aggregated A_z data showed no main effect of list type (∘LLE/∘LSE) in the SU comparison, $F(2, 234) = 0.82$, $MSE = 0.01$. In contrast, a main effect of list type was found in the SSP comparison, $F(2, 234) = 7.08$, $MSE = 0.01$, $p < .01$. LSD post hoc comparisons revealed a +LSE (significant differences between WIS and SI lists and between WIL and SI lists) but no LLE (nonsignificant difference between WIS and WIL lists; ∘LLE/+LSE). An effect of list type was also observed in the SPU comparison, $F(2, 234) = 17.12$, $MSE = 0.01$, $p < .001$. Post hoc comparisons showed a −LSE in this condition (significant differences between WIS and SI lists and between WIL and SI lists) but no LLE (nonsignificant differences between WIS and WIL lists; ∘LLE/−LSE). Table 1 presents A_z measures broken down by encoding condition for different list and comparison types.

We carried out a 3 (list type: WIS, WIL, and SI) $\times$ 2 (discrimination type: SU, SSP) repeated measures ANOVA with A_z as the dependent measure to investigate whether the impairment in discrimination was specific to SI lists. A significant interaction was found, $F(2, 234) = 10.56$, $MSE = 0.01$, $p < .001$, confirming that there was no difference across list types for the SU comparison but that discrimination was lower only for the strong list in the SSP comparison. Figure 1 illustrates these trends. The ROC curve for SI lies below the curves for WIS and WIL in the SSP comparison and above them in the SPU comparison. In the SU condition, the curves largely coincide.

The use of a self-paced encoding task to compare length and strength effects naturally introduces a confound: Study time for SI lists was shorter than study time for WIL lists, because participants responded faster to repeated words in SI lists. Encoding times were indeed shorter for SI lists (M = 1,735 ms) than for both WIS lists (M = 1,826 ms) and WIL lists (M = 1,807 ms), $F(2, 234) = 15.16$, $MSE = 18{,}023$, $p < .001$. More importantly, however, no reliable difference was found between encoding times for target items (WIS = 1,799 ms, WIL = 1,817 ms, SI = 1,824 ms; $F < 1$). The total average difference in encoding time between SI and WIL lists, 8.5 s, corresponds to just a small fraction of the total encoding time (less than 4%). Moreover, the shorter study−test lag for the SI condition should decrease the likelihood

Table 1. *Derived sensitivity (A_z) across encoding conditions and discrimination types*

		Size judgement					*Pleasantness judgement*				
Discrimation type	*List type*	*M*				*SEM*	*M*				*SEM*
Studied vs. unrelated	Weak short	0.90		⊤ n.s.	⊤	0.01	0.91		⊤ n.s.	⊤	0.01
	Weak long	0.92	⊤ n.s.	⊥	n.s.	0.01	0.91	⊤ n.s.	⊥	n.s.	0.01
	Strong	0.91	⊥		⊥	0.01	0.92	⊥		⊥	0.01
Studied vs. switched plurality	Weak short	0.75		⊤ n.s.	⊤	0.02	0.74		⊤ n.s.	⊤	0.02
	Weak long	0.75	⊤ *	⊥	*	0.02	0.74	⊤ *	⊥	*	0.02
	Strong	0.70	⊥		⊥	0.02	0.70	⊥		⊥	0.01
Switched plurality vs. unrelated	Weak short	0.69		⊤ n.s.	⊤	0.02	0.74		⊤ n.s.	⊤	0.01
	Weak long	0.72	⊤ **	⊥	**	0.02	0.70	⊤ **	⊥	*	0.02
	Strong	0.80	⊥		⊥	0.02	0.79	⊥		⊥	0.01

Note: A_z = estimate of the area under the receiver operating characteristic (ROC) curves. n.s = not significant., $^{*}p < .05$; $^{**}p < .01$.

of observing an LSE. Therefore, an explanation of our results in terms of study time differences does not seem feasible.

Discussion

We aimed to test the hypotheses (a) that previous ◦LLE/◦LSE (Dennis & Humphreys, 2001) resulted from weak recollection at test, (b) that previous results (+LSE or ◦LLE/◦LSE) were not influenced by confounds in the experimental design, and (c) that the contrasting results found by Norman (2002) and Dennis and Humphreys (2001) were not caused by differences in encoding task. The results support the three hypotheses. First, no LLE/LSE was found in the SU discrimination (which is not assumed to involve recollection), replicating Dennis and Humphreys' (2001) null results; however, reliable LSEs were found in the recollection-dependent SSP discrimination and SPU pseudodiscrimination, indicating that Dennis and Humphreys' (2001) null LSE may have been caused by a limited (or no) contribution of recollection at test. Second, the LSE found here was reliable in comparison to both weak short lists and weak long lists, suggesting that the longer study lists were not critical to the LSE found by Norman (2002; see also Verde & Rotello, 2004). Moreover, the LSE in the current experiment was found using short encoding times (~1.8 s). Dennis and Humphreys (2001) used longer encoding times (3 s), leaving open the possibility that covert rehearsal may have contributed to their null LSE. Third, encoding task did not have a significant effect on any of the sensitivity measures. Although this null result does not rule out the possibility that the encoding task systematically affects interference, our data provide no indication that such effects would be sufficiently strong to explain the discrepancies between previous studies.

The most common result in the literature has been a failure to find an LSE in recognition. Why, then, did our experiments yield an LSE? Highly confusable lures, such as the SP lures used here, seem to be crucial for eliciting an LSE: No discrimination LSEs have been reported in the literature when only unrelated lures were used at test, even when interference items were repeated 11 times (Diana & Reder, 2005, Exp. 2). Although, Shiffrin, Huber, and Marinelli (1995) did use similar lures and yet found no LSE, the lures they used (nonstudied exemplars from studied categories) did not have a very

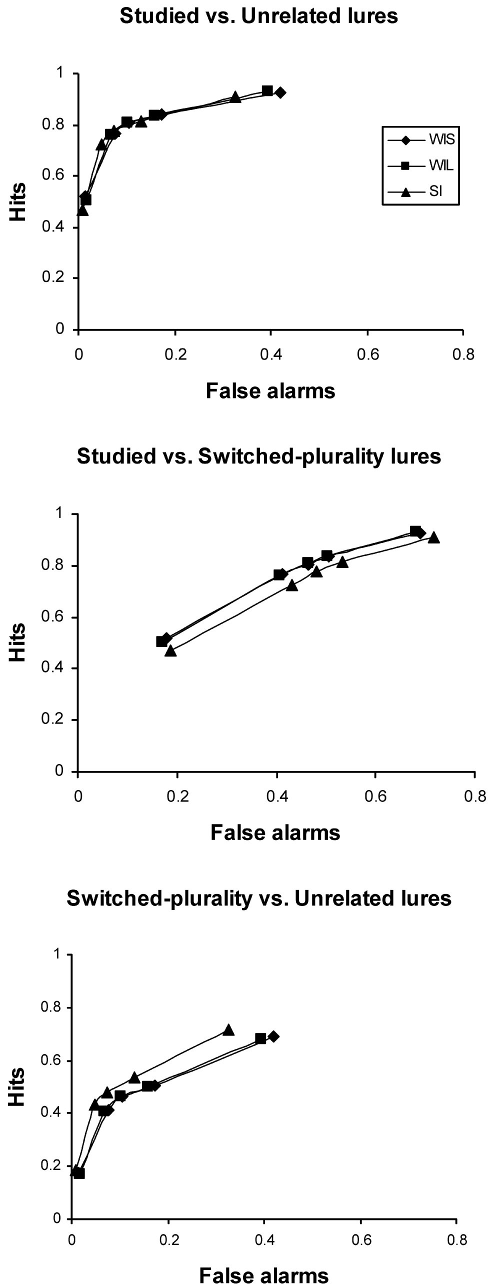

Figure 1. *Receiver operating characteristic (ROC) curves of studied versus unrelated lure discrimination, studied versus switched plurality discrimination, and switched plurality versus unrelated lure discrimination as a function of interference strength collapsed across size and pleasantness encoding conditions. Pooled data;* N *= 119. WIS = weak interference short list; WIL = weak interference long list; SI = strong interference.*

similar counterpart in the study set in the way that the SP lures did. It is thus possible that the participants relied on familiarity to discriminate between studied and nonstudied items. It is also possible that participants in Shiffrin et al. (1995) did not use recollection as often as they could have done, because, unlike the participants in the present experiment, they were not encouraged to use recall at test (see Rotello et al., 2000, Exp. 2, for evidence that recall-to-reject is modulated by strategic control). Finally, encoding time in Shiffrin et al. (1995) was fixed at 3 s, suggesting that their null LSE could have been caused by rehearsal borrowing from stronger to weaker items. Consistent with this possibility, strong items in several of their mixed lists (containing weak and strong items) were worse recognized than strong items in their pure strong lists.

The strength manipulation in the current study renders our procedure similar, in some respects, to that used in the retrieval-practice paradigm (Anderson, Bjork, & Bjork, 1994). By studying interference items repeatedly in the study phase, the participants effectively carried out retrieval practice. As a consequence, it could be argued that the LSE found here is the result of retrieval-induced forgetting. Although we cannot entirely rule out this possibility, it is unlikely that retrieval-induced forgetting played a major role in our results. Anderson et al. (1994) showed that the degree of retrieval-induced forgetting is strongly dependent on the strength of competitors activated at retrieval. In the retrieval-practice task, strong competition has usually been achieved through the use of lists of highly related words (for instance, multiple instances from the same semantic category). However, the task we used involved lists of unrelated words. Because our items were unrelated, competition at retrieval was probably weak, and hence retrieval-induced forgetting is unlikely to have occurred.

Perhaps the most surprising outcome from our experiment was the absence of an LLE in all the conditions. To our knowledge, this is the first time that a recognition LSE is reported without a concurrent LLE, raising questions about the empirical and theoretical status of these effects.

The absence of an LLE in the current study remains puzzling. One possible explanation is that our length manipulation was not powerful enough to elicit an LLE. A recent study using the same controls as those of Dennis and Humphreys (2001) and ourselves found a reliable LLE (Cary & Reder, 2003, Exp. 3). However, the most important difference between Cary and Reder's study and the study we report in this article lies in the long-to-short list length ratios: They used a 4 : 1 length ratio, whereas we followed Dennis and Humphreys (2001) in using a 2:1 ratio. Length ratio per se is not sufficient to account for the conflicting results, because recognition LLEs have been reported with list ratios as small as ours (e.g., Ratcliff, Clark, & Shiffrin, 1990, Exp. 6). However, it is unclear whether those small-ratio LLEs would still occur if all the confounding variables (such as study–test lag or attention drift) highlighted by Dennis and Humphreys (2001) were controlled. What is clear from the results reported here is that strength manipulations are more sensitive to lure similarity than are length manipulations. Whether this relationship changes with increasing length ratios remains an open question.

Strength and length effects have been investigated in memory research because they can be used to test the assumptions of computational models (for reviews, see Clark & Gronlund, 1996; Diana et al., 2006). In particular, the *bind cue decide model of episodic memory* (BCDMEM, Dennis & Humphreys, 2001) may be difficult to reconcile with the present findings. According to BCDMEM, only context noise (the number of contexts in which a word has been seen before) causes interference in recognition memory tasks, whereas item noise (determined by the number of words seen in the same context) should not matter. Because the present task is an item noise task, a null LSE is predicted. The fact that we found a positive effect challenges the context-noise assumption. Furthermore, because words are represented as individual nodes in BCDMEM, regardless of plurality, it is unclear how the model would be able to distinguish between targets (e.g., *banana*) and new similar lures (e.g., *bananas*).

Our findings (∘LLE/ + LSE), together with Dennis and Humphreys' (2001, ∘LLE/∘LSE) and Norman's (2002, + LSE), suggest that further research on the conditions under which strength and length effects occur is needed, and that further model development may be required.

Original manuscript received 16 May 2007
Accepted revision received 27 June 2007
First published online day month year

REFERENCES

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1063–1087.

Boldini, A., Russo, R., & Avons, S. E. (2004). One process is not enough! A speed–accuracy tradeoff study of recognition memory. *Psychonomic Bulletin and Review*, *11*, 353–361.

Cary, M., & Reder, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory and Language*, *49*, 231–248.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin and Review*, *3*, 37–60.

Criss, A. H., & McClelland, J. L. (2006). Differentiating the differentiation models: A comparison of the retrieving effectively from memory model (REM) and the subjective likelihood model (SLiM). *Journal of Memory and Language*, *55*, 447–460.

Curran, T., DeBuse, C., Woroch, B., & Hirshman, E. (2006). Combined pharmacological and electrophysiological dissociation of familiarity and recollection. *Journal of Neuroscience*, *26*, 1979–2009.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*, 452–477.

Diana, R. A., & Reder, L. M. (2005). The list strength effect: A contextual competition account. *Memory and Cognition*, *33*, 1289–1302.

Diana, R. A., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin and Review*, *13*, 1–21.

Gillund, G., & Shiffrin, R. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1–67.

Harvey, L. O., Jr. (2001). *Parameter estimation of signal detection models: RscorePlus user's manual*. Boulder, CO: Author.

Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 302–313.

Norman, K. A. (2002). Differential effects of list strength on recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 1083–1094.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, *110*, 611–646.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect. 1: Data and discussion. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *16*, 163–178.

Rotello, C. M., Macmillan, N. A., & Van Tassel, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory and Language*, *43*, 67–88.

Shiffrin, R. M., Huber, D. E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 267–287.

Verde, M. F., Macmillan, N. A., & Rotello, C. M. (2006). Measures of sensitivity based on a single hit rate and false alarm rate: The accuracy, precision, and robustness of d', A_z, and A'. *Perception and Psychophysics*, *68*, 643–654.

Verde, M. F., & Rotello, C. M. (2004). Strong memories obscure weak memories in associative recognition. *Psychonomic Bulletin and Review*, *11*, 1062–1066.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441–517

APPENDIX

Hits and false alarms across encoding conditions and discrimination types

		HR Targets			*FAR Unrelated lures*			*FAR Switched plurality*		
Encoding task	*List type*	*M*		*SEM*	*M*		*SEM*	*M*		*SEM*
Size ($N = 61$)	Weak short	0.79	n.s. *	0.02	0.10	n.s. *	0.01	0.42	n.s. n.s.	0.03
	Weak long	0.83	*	0.02	0.09	*	0.01	0.45	n.s.	0.03
	Strong	0.75		0.02	0.07		0.01	0.44		0.02
Pleasantness ($N = 58$)	Weak short	0.82	n.s. n.s.	0.02	0.11	n.s. *	0.01	0.48	n.s. n.s.	0.03
	Weak long	0.78	n.s.	0.02	0.10	*	0.01	0.46	n.s.	0.03
	Strong	0.79		0.02	0.07		0.01	0.49		0.03

Note: HR = hits; FAR = false-alarms; *M* = mean; *SEM* = standard error of the mean; n.s. not significant, $^{*}p < .05$.