Running head: ENCODING, MAINTENANCE, AND RETRIEVAL IN THE LAG EFFECT

Published in *Memory*:
http://www.tandfonline.com/doi/abs/10.1080/09658211.2011.631550#.Ul1aAVD2hak

# Encoding, Maintenance, and Retrieval Processes in the Lag Effect:

# A Multinomial Processing Tree Analysis

Carolina E. Küpper-Tetzel and Edgar Erdfelder

University of Mannheim, Germany

Corresponding author's address:

Carolina E. Küpper-Tetzel

University of Mannheim

L9, 7, room 209

68131 Mannheim, Germany

Phone: +49-621-181 3569

Fax: +49-621-181 3997

Email: kuepper-tetzel@psychologie.uni-mannheim.de

Abstract

Short-term studies on repeated learning of verbatim material have typically revealed an overall benefit of long lags compared to short lags between repetitions. This has been referred to as the lag effect. On educationally relevant time scales, however, an inverted-U-shaped relation between lag and memory performance is often observed. Recently, Cepeda et al. (2009) showed that the optimal lag for relearning heavily depends on the time interval between the last learning session and the final memory test (i.e., the retention interval (RI)). In order to explore the cognitive mechanisms underlying this result in more detail, we independently manipulated both the lag and the RI in a 3 by 2 experimental design and analysed our data using a multinomial processing tree model for free-then-cued-recall data. Our results reveal that the lag effect trends are mainly driven by encoding and maintenance processes rather than by retrieval mechanisms. Our findings have important implications for theories of the lag effect.

Acknowledgements

The authors are grateful to Kerstin Mertz and Jana Scheible for their considerable help with data collection and to Pia Sue Helferich for editing the Multinomial Processing Tree graph. We thank David Riefer and an anonymous reviewer for helpful comments on this manuscript.

# Encoding, Maintenance, and Retrieval Processes in the Lag Effect: A Multinomial Processing Tree Analysis

The lag between initial learning (i.e., when new information is acquired for the first time) and relearning (i.e., when this information is repeatedly studied) has a strong effect on memory performance on a final test. The finding that memory performance benefits from increasing lags between study episodes has been referred to as the *lag effect*. Past studies, however, have profoundly challenged this simple lag effect finding (e.g., Ausubel, 1966; Glenberg & Lehmann, 1980; Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008). They have shown that the retention interval (RI) (i.e., the time between the last study episode and the final test) plays a crucial role in the modulation of the lag effect function. Glenberg and Lehmann (1980), for instance, found that, given a 7-day RI, memory performance increased between the massed practice condition and a 1-day lag condition, but decreased again for a lag of 7 days between learning sessions. In other words, memory performance on the final test administered after one week followed an inverted-U-shaped trend with increasing lag. Recently, Cepeda et al. (2008) and Cepeda et al. (2009) investigated this effect at educationally relevant time scales. Cepeda et al. (2009) conducted two experiments with RIs of 10 and 168 days, respectively, combined with six different lags between initial learning and relearning. They also found that memory performance on the final test followed an inverted-U-shaped trend with increasing lag. The maximum of the lag effect function, however, depended on the length of the RI. More precisely, the optimal lag between study episodes increased as the RI increased. On the basis of an extensive web study, Cepeda et al. (2008) formalised and reinforced this systematic relationship between lag and RI. Taken together, these findings suggest that memory performance follows a nonmonotonic trend with increasing lag, but that the optimal point in time for relearning increases with RI. We use the term “lag effect” in a way that includes these recent findings.

An important aspect that has not been examined yet is the contribution of encoding, maintenance, and retrieval processes to the lag effect trends. Basically, improved memory in the optimal lag condition compared to other study conditions could emerge from any combination of three different influences: (1) enhanced *encoding* during repeated practice leading to a strengthening of the memory trace, (2) improved *maintenance* leading to enduring memory traces and resistance to forgetting until the time of testing, and (3) better *retrieval* during the final test phase. The answer to the question about the exact contributions of these memory processes to the lag effect is – to date – not only unknown, but also has fundamental impact on the evaluation of lag effect explanations. All theories of the lag effect proposed so far aim at explaining the emergence of the lag effect by focussing on different memory processes.

The contextual variability theory (Glenberg, 1976, 1979) states that with increasing lag more different context components are stored along with the to-be-learned information. This boosts the probability of successful retrieval at final test because *more effective retrieval cues* are available due to increased overlap between context components at test and at study. Hence, retrieval should benefit from the increase in context variability associated with longer lags. However, Glenberg (1976) also points out that longer lags must not always translate to increased memory performance. If the RI is short compared to the lag between study sessions inverted-U-shaped memory functions may occur because the retrieval cues at test are biased towards the second learning occurrence, and thus, share less contextual components with the stored memory trace which also contains contextual features from the first learning session.

According to the study-phase retrieval theory (Thios & D’Agostino, 1976), memory performance improves when during the second occurrence of an item its first occurrence is retrieved from memory. The second occurrence serves as a cue initiating automatic study-phase retrieval. Successful study-phase retrieval is assumed to *strengthen the stored memory* *trace*. The benefit of successful retrieval during practice increases with lag because successful study-phase retrieval is more effortful which, in turn, increases performance on the test. However, lags may become too long and lead to a failure in study-phase retrieval decreasing later memory performance.

The Multiscale Context Model (MCM) (Mozer, Pashler, Cepeda, Lindsey, & Vul, 2009) combines the Search of Associative Memory (SAM; Raaijmakers, 2003) and the predictive utility theory (Staddon, Chelaru, & Higa, 2002). SAM encompasses assumptions of the contextual variability and the study-phase retrieval theory. The novel aspect in MCM, however, is the predictive utility assumption. It states that the time that elapses before the reencounter of information (i.e., the lag) determines for how long this information will be *maintained in memory for the future*. More precisely, if the to-be-learned material is relearned after a long lag our memory system will store and, importantly, maintain the material for a longer period of time. By contrast, if the lag is short the material will be available for a short time only.

Taken together, the contextual variability theory, the study-phase retrieval theory, and MCM offer plausible theoretical explanations for the lag effect. They can be distinguished, however, with respect to the underlying memory processes they put forward to explain the lag effect trends. The contextual variability theory emphasises the role of retrieval processes during testing. The study-phase retrieval theory focuses on the importance of encoding processes during repeated practice. MCM advances an adaptive feature of the memory system and introduces the importance of maintenance processes. Thus, in order to evaluate which theory makes the most plausible assumptions in terms of the underlying memory processes it would be crucial to disentangle contributions of encoding, maintenance, and retrieval processes to the lag effect.

A paradigm that has often been used in the past to determine the role of storage (i.e., encoding or maintenance) versus retrieval processes makes use of two memory tests: one test that depends highly on retrieval processes (e.g., free recall) followed by one that depends less on retrieval processes (e.g., cued recall) (see, e.g., Drachman & Leavitt, 1972; Hogan & Kintsch, 1971; Thomson & Tulving, 1970). Consequently, finding a memory effect in free recall, but not in cued recall, suggests that retrieval processes play a major role for the phenomenon. In contrast, if the effect emerges in both memory tests alike, this hints at the conclusion that the phenomenon is rather driven by storage processes (i.e., encoding, maintenance, or both).

Although performance profiles in the free-then-cued-recall paradigm provide valuable information, a more fine-grained analysis of the memory processes involved would be preferable. Multinomial processing tree (MPT) models (Batchelder & Riefer, 1999; Erdfelder et al., 2009) offer such an analysis by providing separate estimates of the cognitive processes underlying performance scores. In the past, MPT models have been used successfully to disentangle storage and retrieval contributions to well-known memory phenomena, for example, the bizarreness effect (Riefer & Rouder, 1992), the recognition failure effect (Riefer & Batchelder, 1995), retroactive inhibition (Bäuml, 1991, 1996), and, more recently, the enactment effect (Steffens, Jelenec, Mecklenbräuker, & Thompson, 2006; Steffens, Jelenec, & Mecklenbräuker, 2009).

Thus, in order to evaluate prevailing theories of the lag effect, we examined the role of encoding, maintenance, and retrieval processes by assessing memory performance with both free and cued recall tests and by analysing the data with an extended[1] version of Rouder and Batchelder's (1998) storage-retrieval MPT model for a free-then-cued-recall paradigm.

In our study, we assessed joint effects of different lags and RIs on memory performance. Critically to our approach, memory performance was assessed by a free recall test on weakly associated cue-target word pairs immediately followed by a cued recall test for the target word given the cue word. Based on the performance on the cued recall test at the end of practice combined with the performances on the free and cued recall final tests administered after the RI, 12 observable events can occur for each word pair (Table 1).

(Table 1 about here)

Based on the observed frequencies of these 12 events, MPT modelling allows estimation of seven parameters representing underlying memory processes: one probability of associative encoding ($e$), two probabilities of associative maintenance in memory until the final test ($m_s$ and $m_u$ for maintenance following successful vs. unsuccessful initial cued recall, respectively), two probabilities of successful retrieval in free and cued recall ($r_f$ and $r_c$, respectively), and, finally, two probabilities of single word retrieval in free recall in case of successful vs. unsuccessful associative encoding or maintenance ($s$ and $u$, respectively).

(Figure 1 about here)

To facilitate the understanding of this method, the extended MPT model is presented as a processing tree diagram in Figure 1. It has 32 branches, each terminating in one of the 12 events summarized in Table 1. Each of these branches represents a possible sequence of encoding, maintenance, and retrieval processes underlying performance in free and cued recall. Specifically, a word pair is either encoded as an association with probability $e$ or is not encoded as an association with the complementary probability 1-$e$. In case of successful associative encoding, cued recall at the end of practice may be successful with probability $r_c$ or fail with probability 1-$r_c$. Hence, parameter $r_c$ represents the probability of successful associative retrieval in a cued recall test and is likely to be close to one whenever the cue-target association is stored in memory at the time of testing. Associative maintenance to the time of testing occurs upon encoding and successful initial cued recall with probability $m_s$. If associative maintenance of the word pair is successful ($m_s$), associative retrieval during final cued recall may be successful with probability $r_c$ or unsuccessful with the complementary probability 1-$r_c$. In either case, the intact word pair may be retrieved during free recall with probability $r_f$ resulting in event $E_1$ (in case of successful final cued recall) or $E_4$ (in case of unsuccessful final cued recall). Associative retrieval during free recall, however, may fail with probability 1-$r_f$, so that the intact word pair cannot be retrieved as association. Nevertheless, each word of a pair may be independently retrieved during free recall with probability $s$ or not retrieved with probability 1-$s$, so that both words of a pair ($E_1$ or $E_4$, depending on successful vs. unsuccessful final cued recall, respectively), exactly one word ($E_2$ or $E_5$, depending on successful vs. unsuccessful final cued recall, respectively), or neither word is recalled ($E_3$ or $E_6$, depending on successful vs. unsuccessful final cued recall, respectively). In contrast, if maintenance of the word pair association in memory fails (1-$m_s$), the final cued recall will also fail. However, items may be retrieved ($u$) or not retrieved (1-$u$) as nonassociated single words during free recall, so that both words of a pair ($E_4$), exactly one word ($E_5$), or neither word is recalled ($E_6$)[2]. Moreover, successful associative maintenance may also occur following unsuccessful cued recall at the end of learning, albeit with probability $m_u$ that may, in principle, differ from $m_s$. From there, the MPT tree progresses as described above except that the branches terminate in event categories $E_7$ to $E_{12}$ instead. Finally, associative encoding can fail altogether with probability 1-$e$. This implies a failure of both cued recall at the end of practice and final cued recall. However, items may be retrieved ($u$) or not retrieved (1-$u$) independently as nonassociated single words during free recall, so that both words of a pair ($E_{10}$), exactly one word ($E_{11}$), or neither word is recalled ($E_{12}$). Summing up the branch probabilities that terminate in the same observable event, we obtain the following set of model equations for the 12 possible events:

$p(E_1) = e\, r_c^2\, m_s\, [r_f + (1-r_f)\, s^2]$
$p(E_2) = 2\, e\, r_c^2\, m_s\, (1-r_f)\, s\, (1-s)$
$p(E_3) = e\, r_c^2\, m_s\, (1-r_f)\, (1-s)^2$
$p(E_4) = e\, r_c\, [m_s\, (1-r_c)\, (r_f + (1-r_f)\, s^2) + (1-m_s)\, u^2]$

$p(E_5) = 2\ e\ r_c\ [m_s\ (1\text{-}r_c)\ (1\text{-}r_f)\ s\ (1\text{-}s) + (1\text{-}m_s)\ u\ (1\text{-}u)]$
$p(E_6) = e\ r_c\ [m_s\ (1\text{-}r_c)\ (1\text{-}r_f)\ (1\text{-}s)^2 + (1\text{-}m_s)\ (1\text{-}u)^2]$
$p(E_7) = e\ (1\text{-}r_c)\ r_c\ m_u\ [r_f + (1\text{-}r_f)\ s^2]$
$p(E_8) = 2\ e\ (1\text{-}r_c)\ r_c\ m_u\ (1\text{-}r_f)\ s\ (1\text{-}s)$
$p(E_9) = e\ (1\text{-}r_c)\ m_u\ r_c\ (1\text{-}r_f)\ (1\text{-}s)^2$
$p(E_{10}) = e\ (1\text{-}r_c)\ [m_u\ (1\text{-}r_c)\ (r_f + (1\text{-}r_f)\ s^2) + (1\text{-}m_u)\ u^2] + (1\text{-}e)\ u^2$
$p(E_{11}) = 2\ [e\ (1\text{-}r_c)^2\ m_u\ (1\text{-}r_f)\ s\ (1\text{-}s) + e\ (1\text{-}r_c)\ (1\text{-}m_u)\ u\ (1\text{-}u) + (1\text{-}e)\ u\ (1\text{-}u)]$
$p(E_{12}) = e\ (1\text{-}r_c)\ [m_u\ (1\text{-}r_c)\ (1\text{-}r_f)\ (1\text{-}s)^2 + (1\text{-}m_u)\ (1\text{-}u)^2] + (1\text{-}e)\ (1\text{-}u)^2$

On the basis of the observed event frequencies, the seven model parameters ($e$, $m_s$, $m_u$, $r_c$, $r_f$, $s$, and $u$) are then estimated using standard maximum likelihood techniques (Hu & Batchelder, 1994) as implemented in freely available software for MPT models (e.g., Moshagen, 2010). The parameters of Rouder and Batchelder's original free-then-cued-recall MPT model have previously been validated experimentally by applying specific manipulations that were assumed to influence one parameter (e.g., retrieval) while leaving others (e.g., storage) unaffected (see Rouder & Batchelder, 1998). Thus, it has been established that the model not only fits empirical data but also provides parameters that capture the intended memory processes selectively. Our MPT model keeps all the basic assumptions of Rouder and Batchelder's (1998) model and represents a straightforward extension by (1) differentiating between processes of encoding and processes of maintenance to the time of testing and by (2) providing an estimate of associative retrieval during cued recall in addition. Therefore, it seems appropriate to use our extended MPT model to measure encoding, maintenance, and retrieval contributions to the trends in the lag effect and, more importantly, to evaluate theoretical explanations of the lag effect.

## Method

### *Participants*

Sixty-two persons participated in this experiment. Two participants were excluded from all analyses because they severely underperformed during the relearning session (cued recall performance < 40%). The remaining 60 participants were current students or alumni of the University of Mannheim. Thirty-seven were female, mean age 22.45 (range, 18-33).

*Materials*

The word material consisted of 30 weakly associated cue-target word pairs from 30 common categories. All words were concrete German nouns taken from German production norms (Hager & Hasselhorn, 1994). Both words of a pair always came from the same category (e.g., *foods*). The cue word was a weakly associated word of the category (e.g., *candy*, production index < 0.02) and the respective target word was one of the four most frequently produced words for that category (e.g., *bread*). Weakly associated word pairs were used to avoid inferential processes during recall. In order to ensure that significant effects would not be due to word list characteristics, we constructed four word lists, each containing 30 target words with high associations to their respective category and to each other (e.g., category: *foods,* target words: *bread, meat, butter, vegetables*). These four target word lists were counterbalanced across participants. Thus, the same cue word was combined with one target word that varied depending on the active word list.

*Design*

The experiment consisted of two learning sessions separated by a lag and one final test session occurring after the RI. The lag between learning sessions was either 0 days, 1 day, or 11 days and the RI was either 7 or 35 days. This resulted in a 3 x 2 between-subjects design. These intervals were chosen on the basis of Cepeda et al. (2008), who suggest an inverted-U-shaped memory trend in the 7-day RI group (i.e., peak at a 1-day lag) and a positive linear trend in the 35-day RI group for lags increasing from 0 to 11 days. Participants were randomly assigned to their experimental condition, subject to the constraint of their availability for attending sessions on different days. Four experimental conditions contained 10 participants each and two conditions contained 9 and 11 participants, respectively.

*Procedure*

Participants attended two learning sessions and one test session. During the first learning session, participants worked on two *study-test trials*. The 1-day lag and 11-day lag group were dismissed after these trials and returned after the respective lag to their second learning session. The 0-day lag group worked for five minutes on an unrelated distractor task and continued with the second learning session on the same day. The second learning session contained one more *study-test trial*. After the second learning session, all participants were dismissed and returned after a RI of either 7 or 35 days to the final test session.

A *study-test trial* consisted of a study phase, a distractor task, and a test phase. During the study phase, each word pair was presented for three seconds separated by a 750 milliseconds interstimulus interval. Word pairs were presented in a different random order in each study phase. After the study phase, participants worked for two minutes on an unrelated arithmetic task. Afterwards, participants were administered a free recall test immediately followed by a cued recall test. Both memory tests were self-paced. For the free recall test, they were instructed to write all word pairs they could remember on a lined recall sheet. It was pointed out to them that if they could only remember one word of a pair, they should nevertheless write it down. For the cued recall test, participants were presented with the cue word and asked to recall and type the correct target word. They were asked to recall all 30 target words given the cue. The cues were presented sequentially in random order. Participants could skip to the next cue word if they could not remember the target word. They were not provided with feedback about their performance during the memory tests.

On the final test session, participants worked on a free recall test immediately followed by a cued recall test. Afterwards, they were compensated for their participation and debriefed. The whole experiment lasted about 1.5 hours divided over three sessions. All participants were required to attend three sessions on three different days. Participants in the 0-day lag condition who completed the experiment within two sessions worked for 15 minutes on an unrelated experiment during their third session. This was done to establish comparable motivational conditions for all participants.

## Results

### *Learning performance*

Participants showed good learning performances in both memory tests at the end of the first learning session after two study-test trails, 52% free-recall accuracy and 83% cued-recall accuracy. Moreover, as expected, free and cued recall performances assessed by the end of the first learning session were not affected by lag, $F(2,57) = 1.31$, $p = .279$, $\eta_p^2 = .04$ and $F(2,57) = 1.24$, $p = .297$, $\eta_p^2 = .04$, respectively.

Memory performance in the relearning session (i.e., after the lag) showed the expected decline in free and cued recall as a consequence of increasing lag. We performed multiple comparisons and adjusted the α-level of .05 using the Holm-Bonferroni correction method (Holm, 1979). All significance tests are reported with *p*-values corresponding to two-tailed tests even when directed hypotheses were imposed. The Welch test was applied whenever Levene’s test indicated unequal variances at $\alpha = .05$. Memory performance on both memory tests was significantly lower during practice after an 11-day lag compared to a 1-day lag (free-recall accuracy: 48% vs. 69%; cued-recall accuracy: 78% vs. 96%), $t(32.76) = 3.51$, $p = .001$, $\eta^2 = 0.27$ and $t(20.80) = 3.74$, $p = .001$, $\eta^2 = 0.40$ for free and cued recall, respectively. Also, the decrease between a 0-day lag and an 11-day lag was significant (free-recall accuracy: 70% vs. 48%; cued-recall accuracy: 92% vs. 78%), $t(38) = 3.52$, $p = .001$, $\eta^2 = 0.25$ and $t(23.98) = 2.81$, $p = .010$, $\eta^2 = 0.25$ for free and cued recall, respectively. There was no significant difference in memory performance between the 0-day and the 1-day lag condition, all *ts* $\leq$ 1.69, *ps* $\geq$ .10.

*Final test performance*

An initial analysis of the final test performance revealed that there was no systematic effect due to the counterbalance factor *word list* neither for free recall nor for cued recall, $F(3,56) = 1.16$, $p = .332$, $\eta_p^2 = .06$ and $F(3,56) = 1.05$, $p = .376$, $\eta_p^2 = .05$, respectively. Thus, for further analyses, data were collapsed across the four word lists. An α-level of .05 was used for all statistical tests. Again, significance tests are reported as two-tailed tests even in case of directed predictions.

Not surprisingly, participants recalled more word pairs on both memory tests after a 7-day RI (free recall: *M* = 41%, *SD* = 20; cued recall: *M* = 74%, *SD* = 23) than after a 35-day RI (free recall: *M* = 19%, *SD* = 17; cued recall: *M* = 43%, *SD* = 23), $t(58) = -4.43$, $p < .001$, $\eta^2 = 0.25$ and $t(58) = -5.12$, $p < .001$, $\eta^2 = 0.31$ for free and cued recall, respectively.

Of greatest interest, however, were the different trends in the two RI conditions as a function of increasing lag. To revisit, we expected that with increasing lag (0-day < 1-day < 11-day) memory performance would follow an inverted-U-shaped trend (i.e., negative quadratic trend) in the 7-day RI group and a positive linear (or at least monotonically increasing) trend in the 35-day RI group. Descriptively, the data fit our expectations nicely (see Figure 2). In fact, a significant negative quadratic trend emerged in the 7-day RI condition for cued recall, $t(27) = 2.08$, $p = .048$, $\eta^2 = 0.14$, and a marginally significant quadratic trend occurred for free recall, $t(27) = 1.71$, $p = .099$, $\eta^2 = 0.10$. In contrast, the linear trend was neither significant for free recall, $t(27) = 1.07$, $p = .292$, $\eta^2 = 0.04$, nor for cued recall, $t(27) = -0.25$, $p = .807$, $\eta^2 < 0.01$. In the 35-day RI condition, however, a significant positive linear trend was detected for both free recall, $t(27) = 2.24$, $p = .033$, $\eta^2 =$

0.16, and cued recall, $t(27) = 2.79$, $p = .010$, $\eta^2 = 0.22$. The quadratic trend was not significant in this condition, neither for free recall, $t(27) = 0.71$, $p = .482$, $\eta^2 = 0.02$, nor for cued recall, $t(27) = 0.52$, $p = .606$, $\eta^2 = 0.01$.

(Figure 2 about here)

*Model-based analyses*

Our extended MPT model for free-then-cued-recall was used to disentangle encoding, maintenance, and retrieval contributions to free and cued recall in the final memory test. For model-based analyses, the frequencies of the 12 event categories were calculated for each participant and aggregated separately for each of the 3 x 2 = 6 experimental conditions, resulting in $N = 1,800$ data points in total. The Type I error level was set to $\alpha = .05$ for all model-based analyses. A sensitivity analysis was performed using G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009). This analysis showed that with a sample size of $N = 1,800$ data points, a significance level of $\alpha = .05$, and a desired power of $1-\beta = .95$, the detectable effect size for $G^2$ goodness-of-fit tests based on $df \leq 30$ is $\omega \leq 0.14$ (i.e., a small effect; cf. Cohen, 1988). Thus, all $G^2$ tests reported below allowed detecting already small deviations from the model. The multiTree software (Moshagen, 2010) was used for all MPT model analyses reported here. To revisit, our extended free-then-cued-recall MPT model contains seven parameters ($e$, $m_s$, $m_u$, $r_c$, $r_f$, $s$, and $u$) per condition (see Figure 1), that is, $6 \cdot 7 = 42$ parameters across all six conditions. Hence, the overall goodness-of-fit test has $6 \cdot (12-1) - 42 = 24$ degrees of freedom. This general model fitted the data well ($G^2(24) = 19.65$, $p = .716$). To specify our model as parsimoniously as possible and to increase the precision of parameter estimates, we tested the additional restriction that the maintenance probabilities $m_s$ and $m_u$ can be set equal in each condition. A priori, we hypothesised that $m_s > m_u$, because success in associative cued recall at the end of practice is likely to have positive effects on subsequent maintenance. However, as revealed by a $G^2$ difference test, this effect was not significant in our data, $\Delta G^2(6) = 1.44$, $p = .964$. Consequently, we report results based on the more parsimonious version of our extended model that contains a single maintenance parameter *m* only.[3] The overall goodness-of-fit test for this restricted model has $6 \cdot (12\text{-}1) - 36 = 30$ degrees of freedom and indicates an excellent fit to the data ($G^2(30) = 21.09$, $p = .885$).

*Parameter Estimates*. Of greatest interest for our research question are the probability estimates for associative encoding *e*, associative maintenance *m*, and associative retrieval $r_f$ presented in the upper, middle, and lower chart of Figure 3, respectively.

(Figure 3 about here)

The associative encoding parameter *e* followed an inverted-U-shaped trend with increasing lag in both RI conditions (Figure 3, upper chart). More specifically, the probability of associative encoding increased significantly between the 0-day and the 1-day lag condition, $\Delta G^2(1) = 6.60$, $p = .010$ and $\Delta G^2(1) = 5.90$, $p = .015$, and decreased between the 1-day lag and the 11-day lag condition, $\Delta G^2(1) = 57.12$, $p < .001$ and $\Delta G^2(1) = 27.36$, $p < .001$, for the 7- and 35-day RI group, respectively. In line with our expectations, associative encoding was not affected by the length of the RI, $\Delta G^2(3) = 4.28$, $p = .233$. This result is important because it shows that the additional associative encoding parameter *e* in our extended MPT model can be considered as a valid measure of encoding processes at practice that is not affected by the length of the RI.

The parameter for associative maintenance *m*, however, was affected differently by the length of the RI (Figure 3, middle chart). In the 7-day RI group, associative maintenance increased between the 0-day lag and the 1-day lag as well as the 11-day lag, $\Delta G^2(1) = 20.49$, $p < .001$ and $\Delta G^2(1) = 17.25$, $p < .001$, respectively. There was no difference in associative maintenance between the 1-day and the 11-day lag, $\Delta G^2(1) = 0.06$, $p = .813$. In the 35-day RI condition, associative maintenance increased significantly between the 0-day lag and the 1-

day lag, $\Delta G^2(1) = 15.44$, $p < .001$, and increased further between the 1-day and the 11-day lag, $\Delta G^2(1) = 18.20$, $p < .001$.

In addition, we tested the difference in *m* between the two RI conditions for each lag. As expected, better associative maintenance of the material emerged in the 7-day RI condition than in the 35-day RI condition for all lag conditions, $\Delta G^2(1) = 104.60$, $p < .001$, $\Delta G^2(1) = 112.51$, $p < .001$, and $\Delta G^2(1) = 30.52$, $p < .001$, for 0-, 1-, and 11-day lag, respectively. Again, this validates the current MPT model because the length of the RI should have a strong negative impact on maintenance of the to-be-learned material to the time of testing.

The parameter estimates for associative retrieval $r_f$ during free recall are presented in the lower chart of Figure 3. In the 7-day RI condition, associative retrieval increased significantly between the 0-day lag and the 1-day lag as well as the 11-day lag, $\Delta G^2(1) = 6.79$, $p < .009$ and $\Delta G^2(1) = 7.51$, $p = .006$, respectively. We found the same results in the 35-day RI group, for the comparison between 0-day and 1-day lag, $\Delta G^2(1) = 4.45$, $p = .035$, and for the comparison between 0-day and 11-day lag, $\Delta G^2(1) = 4.12$, $p = .043$. Importantly, the probability of associative retrieval did not differ significantly between the two distributed learning conditions, neither for the 7-day RI, $\Delta G^2(1) = 0.15$, $p = .695$, nor for the 35-day RI, $\Delta G^2(1) = 0.04$, $p = .849$.

In addition, we analysed the difference in associative retrieval between the two RI groups for each lag condition. Not surprisingly, the probability of associative retrieval in free recall was significantly smaller after a 35-day RI than after a 7-day RI for all lag conditions, $\Delta G^2(1) = 5.49$, $p = .019$, $\Delta G^2(1) = 4.91$, $p = .027$, and $\Delta G^2(1) = 7.46$, $p = .006$, for a 0-, 1-, and 11-day lag, respectively.

Last but not least, estimates of the associative cued recall probability $r_c$ (not shown in Figure 3) differed significantly between conditions, $\Delta G^2(5) = 24.25$, $p < .001$. However, as expected, all six $r_c$ parameter estimates were very close to 1, ranging between .96 and 1.00 with a mean of .99. This result is roughly in line with the original free-then-cued-recall MPT model (Rouder & Batchelder, 1998) which is based on the simplifying approximation $r_c = 1$.

## Discussion

Our data convincingly show that different lags of relearning can affect memory performance either in a linear or in a negatively accelerated quadratic manner depending on the length of the RI. More precisely, in the 7-day lag condition, we revealed an inverted-U-shaped trend with increasing lag. Memory performance in this condition peaked at a 1-day lag and decreased for shorter or longer lags. In contrast, in the 35-day RI condition, memory performance increased with lag, thereby suggesting that memory performance improves from a 0-day to an 11-day lag. Thus, we successfully replicated the lag effect trends detected by Cepeda et al. (2008).

The model-based analyses contribute to a better understanding of the underlying cognitive processes. Our extended MPT model for free-then-cued-recall based on Rouder and Batchelder (1998) fit the empirical data successfully. Not surprisingly, associative retrieval decreased with the length of the RI. The lag effect trends, however, were particularly driven by processes captured by the associative encoding parameter $e$ and the associative maintenance parameter $m$. We found a systematic interplay between encoding and maintenance processes that influenced memory performance in the final test and that was mediated by the length of the RI.

More precisely, associative encoding revealed an inverted-U-shaped trend in both the 7-day and the 35-day RI condition alike. This result represents the drop in learning performance after a long lag (i.e., 11 days) compared to a short lag (i.e., 1 day) – a result that is in line with earliest findings in memory research (Ebbinghaus, 1885/1964). Memory performance after a short RI of 7 days was particularly affected by this inverted-U-shaped trend in associative encoding. In this condition, associative maintenance and associative retrieval (captured by $r_f$) both increased between the massed and the spaced conditions, but there was no further significant difference between the two spaced conditions. In contrast, in the long 35-day RI condition, the detrimental effect of poor encoding after a long lag of 11 days was outweighed by better associative maintenance processes. Thus, the positive effect of increasing lag for memory performance after a RI of 35 days was most certainly due to enhanced associative maintenance processes. Note that retrieval processes cannot explain this effect because the associative retrieval parameter $r_f$ reflected the advantage of spaced practice over massed practice only and was not sensitive to lags of different lengths. However, the latter finding is consistent with previous research suggesting that spaced practice, in general, leads to enhanced retrieval compared to massed practice (e.g., Batchelder & Riefer, 1980).

Taken together, encoding processes dictated the inverted-U-shaped memory performance in the 7-day RI condition, whereas maintenance processes to the time of testing were responsible for the linear increasing memory performance in the 35-day RI condition with increasing lag. Thus, given a long RI, the detrimental effects of decreased encoding as a consequence of a long lag are outweighed by better maintenance processes of the encoded memory traces.

Consequently, a theory that attempts to explain the lag effect must emphasise the crucial roles and interplay of encoding and maintenance processes for this learning effect. Thus, the contextual variability theory, which conceives retrieval processes at test as most important, cannot be considered as a potential candidate for explaining the processes producing the curvilinear lag effect trend. Nevertheless, the MPT results show that the contextual variability theory can account for the spacing effect since the advantage of spaced over massed practice was reflected in the associative retrieval parameter.

Our model findings suggest that both the study-phase retrieval theory and the predictive utility assumption of MCM provide the most plausible explanations for the lag effect trends. In accordance with the study-phase retrieval account, better encoding occurred after a lag of 1 day compared to the massed condition. However, if the lag becomes too long (e.g., 11-day lag), successful study-phase retrieval may fail, which leads to decreased encoding. This, in turn, has negative effects on the later memory performance. The inverted-U-shaped memory performance on the final test after a 7-day RI is consistent with this explanation. MCM endorses the crucial role of forgetting and proposes that more enduring memory traces are stored after a long lag compared to a short lag between study episodes. Using MPT modelling, we were able to find that the linear increasing lag effect in the 35-day RI condition was specifically affected by the model parameter that captures maintenance of the material to the time of testing. Thus, consistent with MCM, the memory traces that were encoded after an 11-day lag were maintained better than memory traces that were encoded after a 1-day lag. This proved to be particularly beneficial if memory performance was assessed after a long 35-day RI.

Taken together, our MPT analyses showed that the study-phase retrieval theory and the predictive utility assumption of MCM offer the most plausible assumptions in regard to the underlying cognitive mechanisms for the lag effect. The current findings suggest that the shift of optimal lag with RI is not due to a single mechanism, but rather to a systematic interplay of encoding and maintenance processes that is moderated by the length of the RI.

## References

Ausubel, D. P. (1966). Early versus delayed review in meaningful learning. *Psychology in the Schools, 3,* 195-198. doi: 10.1002/1520-6807(196607)3:3<195::AID-PITS2310030302>3.0.CO;2-X

Batchelder, W. H., & Riefer, D. M. (1980). Separation of storage and retrieval factors in free recall of clusterable pairs. *Psychological Review,* 87, 375-397. doi: 10.1037/0033-295X.87.4.375

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review, 6*, 57-86. doi: 10.3758/BF03210812

Bäuml, K.-H. (1991). Experimental analysis of storage and retrieval processes involved in retroactive inhibition – The effect of presentation mode. *Acta Psychologica*, *77*, 103-119. doi: 10.1016/0001-6918(91)90026-V

Bäuml, K.-H. (1996). A Markov model for measuring storage loss and retrieval failure in retroactive inhibition. *Acta Psychologica*, *92*, 231-250. doi: 10.1016/0001-6918(95)00018-6

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology, 56*, 236-246. doi: 10.1027/1618-3169.56.4.236

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science, 19*, 1095-1102. doi: 10.1111/j.1467-9280.2008.02209.x

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. ed.). Hillsdale, NJ: Erlbaum.

Drachman, D. A., & Leavitt, J. (1972). Memory impairment in the aged: Storage versus retrieval deficit. *Journal of Experimental Psychology, 93,* 302-308. doi: 10.1037/h0032489

Ebbinghaus, H. (1885/1964). *Memory: A contribution to experimental psychology*. Oxford England: Dover.

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie – Journal of Psychology, 217,* 108-124. doi: 10.1027/0044-3409.217.3.108

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149-1160. doi: 10.3758/BRM.41.4.1149

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior, 15,* 1-16. doi: 10.1016/S0022-5371(76)90002-5

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition, 7*, 95-112. doi: 10.3758/BF03197590

Glenberg, A. M., & Lehmann, T. S. (1980). Spacing repetitions over 1 week. *Memory & Cognition, 8*, 528-538. doi: 10.3758/BF03213772

Hager, W., & Hasselhorn, M. (1994). *Handbuch deutschsprachiger Wortnormen* [Handbook of German word norms]. Göttingen: Hogrefe.

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior, 10,* 562-567. doi: 10.1016/S0022-5371(71)80029-4

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6,* 65-70.

Hu, X. & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika, 59,* 21-47. doi: 10.1007/BF02294263

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods, 42*, 42-54. doi: 10.3758/BRM.42.1.42

Mozer, M. C., Pashler, H., Cepeda, N. J., Lindsey, R., & Vul, E. (2009). Predicting the optimal spacing of study: A multiscale context model of memory. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems* (p. 1321-1329). La Jolla, CA: NIPS Foundation.

Raaijmakers, J. G. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science: A Multidisciplinary Journal, 27*, 431-452. doi:10.1016/S0364-0213(03)00007-7

Riefer, D. M., & Batchelder, W. H. (1995). A multinomial modeling analysis of the recognition-failure paradigm. *Memory & Cognition, 23*, 611-630. doi: 10.3758/BF03197263

Riefer, D. M., & Rouder, J. N. (1992). A multinomial modeling analysis of the mnemonic benefits of bizarre imagery. *Memory & Cognition, 20*, 601-611. doi: 10.3758/BF03202710

Rouder, J. N., & Batchelder, W. H. (1998). Multinomial models for measuring storage and retrieval processes in paired associate learning. In C. E. Dowling, F. S. Roberts, & P. Theuns (Eds.), *Recent progress in mathematical psychology: Psychophysics, knowledge, representation, cognition, and measurement.* (pp. 195-225). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Staddon, J. E. R., Chelaru, I. M., & Higa, J. J. (2002). Habituation, memory and the brain: The dynamics of interval timing. *Behavioural Processes, 57,* 71-88. doi: 10.1016/S0376-6357(02)00006-2

Steffens, M. C., Jelenec, P., & Mecklenbräuker, S. (2009). Decomposing the memory processes contributing to enactment effects by multinomial modelling. *European Journal of Cognitive Psychology*, *21*, 61-83. doi: 10.1080/09541440701868668

Steffens, M. C., Jelenec, P., Mecklenbräuker, S., & Thompson, E. M. (2006). Decomposing retrieval and integration in memory for actions: A multinomial modelling approach. *Quarterly Journal of Experimental Psychology*, *59*, 557-576. doi: 10.1080/02724980443000764

Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning & Verbal Behavior, 15*, 529-536. doi:10.1016/0022-5371(76)90047-5

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology, 86,* 255-262. doi: 10.1037/h0029997

Footnotes

[1] Rouder and Batchelder's (1998) original MPT model for free-then-cued-recall differs from our extended model mainly by using a single storage parameter that combines encoding and maintenance processes. For our research question, however, it was crucial to differentiate between encoding and maintenance processes. Inclusion of cued recall performances at the end of practice (i.e., in the relearning session after the lag) enabled us to introduce an additional encoding parameter and to decompose Rouder and Batchelder's associative storage parameter $a$ into associative encoding and maintenance parameters, $e$ und $m$, respectively. Keeping the basic assumptions of Rouder and Batchelder's MPT model unchanged, our extension provides separate estimates for associative encoding, maintenance, retrieval in cued recall, and retrieval in the final free recall test.

[2] One reviewer pointed out correctly that the MPT model does not distinguish between the probability to recall the cue and the probability to recall the target. Rather, a single parameter $s$ or $u$ is assigned to the independent retrieval of one word of the pair, irrespective of whether it is a cue or target. This model assumption, most certainly, represents an approximation that allows us to keep the model as simple and parsimonious as possible. Nevertheless, we checked whether cues and targets were recalled equally often. This was indeed the case for all but one experimental condition: Only in the 11-day lag group that was tested after a 35-day RI, participants recalled slightly more targets ($M$ = 11, $SD$ = 5.9) than cues ($M$ = 9, $SD$ = 5.7) in the final free recall test, $t(9) = 3.35$, $p = .008$. In all other five conditions differences were even smaller and not significant (all $t$s ≤ 2.06, all $p$s ≥ .067). Thus, for the sake of simplicity and model parsimony, since the difference in cue and target recall in our experiment was small and the $s$ and $u$ parameters were not of major interest for the current research question, we decided not to distinguish between cue and target retrieval.

[3] We would like to point out that the restriction $m_s = m_u = m$ has negligible effects on estimates of the other parameters. Estimates for $m$ reported in the present paper resemble those for $m_s$ in the unrestricted version of the extended model. In other words, substantive conclusions are not affected by the version of the extended model used for data analyses.

Figure Captions

*Figure 1.* Extended MPT model for a free-then-cued-recall paradigm to disentangle encoding, maintenance, and retrieval processes based on Rouder and Batchelder (1998). The processing tree presents the latent cognitive processes leading to 12 observable event categories: *Successful cued recall at the end of practice, successful final cued recall* and *free recall of the complete word pair* ($E_1$), *exactly one word of the pair* ($E_2$), *or neither word of the pair* ($E_3$) or *successful cued recall at the end of practice, unsuccessful final cued recall* and *free recall of the complete word pair* ($E_4$), *exactly one word of the pair* ($E_5$), *or neither word of the pair* ($E_6$) or *unsuccessful cued recall at the end of practice, successful final cued recall* and *free recall of the complete word pair* ($E_7$), *exactly one word of the pair* ($E_8$), *or neither word of the pair* ($E_9$) or *unsuccessful cued recall at the end of practice, unsuccessful final cued recall* and *free recall of the complete word pair* ($E_{10}$), *exactly one word of the pair* ($E_{11}$), *or neither word of the pair* ($E_{12}$). The transition probabilities between cognitive states (rounded rectangles) are represented by the model parameters ($e$ = probability of associative encoding during study, $m_s$, $m_u$ = probability of associative maintenance to the time of testing upon successful encoding (following successful or unsuccessful cued recall at the end of practice, respectively), $r_c$ = probability of associative retrieval during cued recall, $r_f$ = probability of associative retrieval during free recall, $s$ = probability of associated single word retrieval during free recall, $u$ = probability of nonassociated single word retrieval during free recall).

*Figure 2.* Mean and standard errors of correctly recalled word pairs on the final free recall (upper chart) and final cued recall test (lower chart) as a function of lag and retention interval.

*Figure 3.* Parameter estimates and standard errors for the probability of associative encoding $e$ (upper chart), for the probability of associative maintenance $m$ (middle chart), and for the probability of associative retrieval $r_f$ (lower chart) as a function of lag and retention interval.

Table 1. Twelve event categories for a memory paradigm that combines cued recall performance at the end of practice with final free and cued recall performances.

| Cued recall at end of practice | Final cued recall | Final free recall | | |
|---|---|---|---|---|
| | | Both words | Exactly one word | Neither word |
| Correct | Correct | $E_1$ | $E_2$ | $E_3$ |
| | Incorrect | $E_4$ | $E_5$ | $E_6$ |
| Incorrect | Correct | $E_7$ | $E_8$ | $E_9$ |
| | Incorrect | $E_{10}$ | $E_{11}$ | $E_{12}$ |

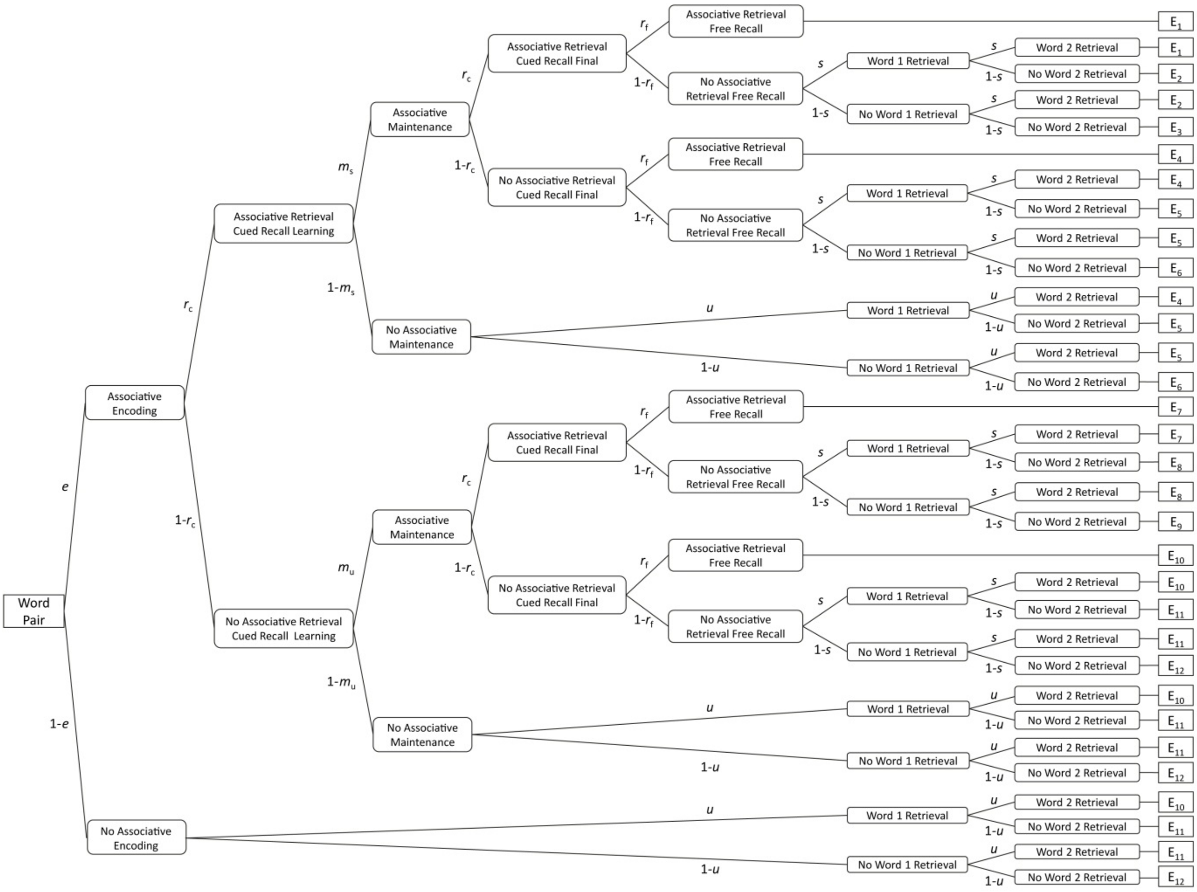
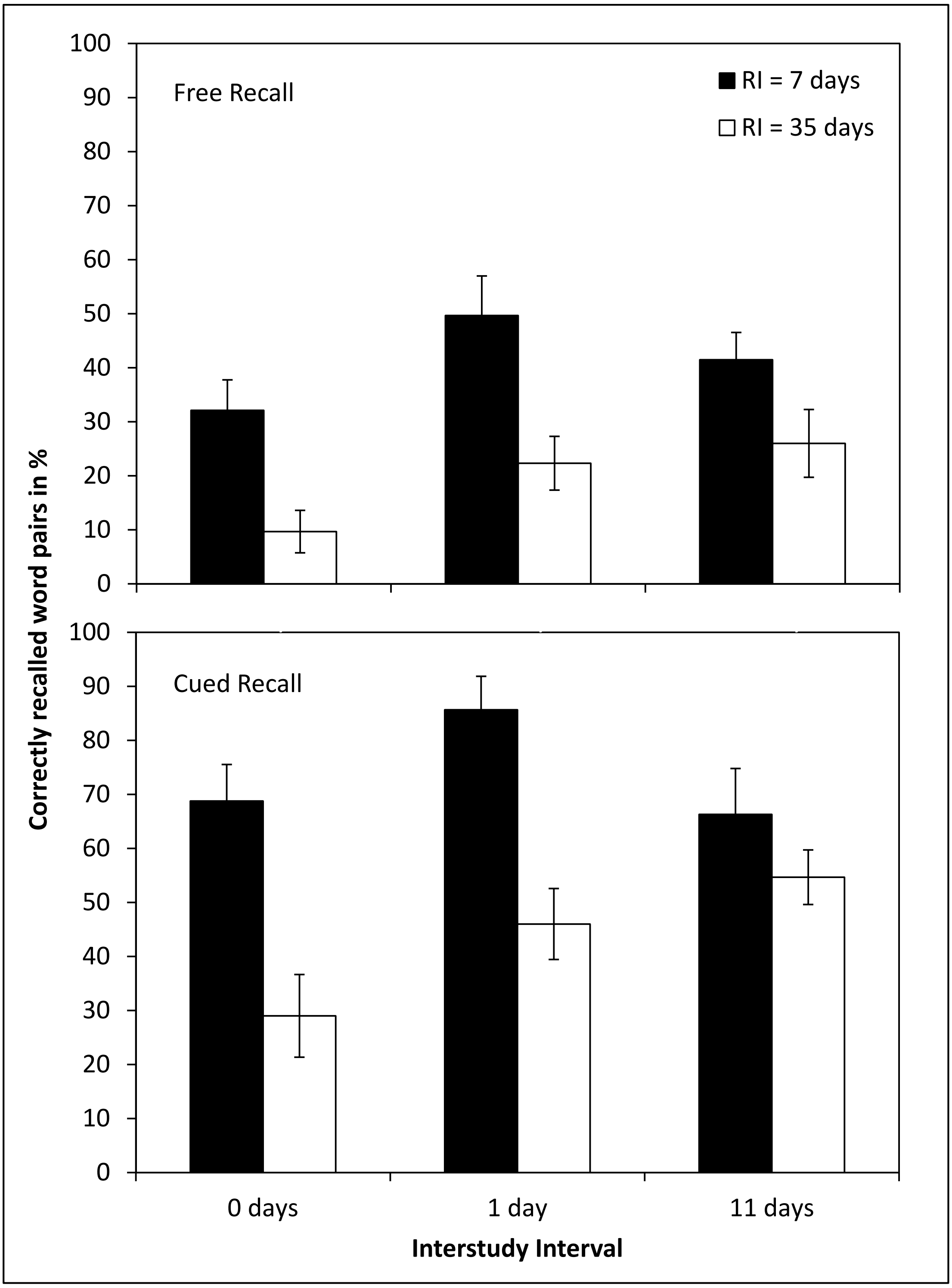

Word Pair
e
Associative Encoding
$r_c$
Associative Retrieval Cued Recall Learning
$m_s$
Associative Maintenance
$r_c$
Associative Retrieval Cued Recall Final
$r_f$
Associative Retrieval Free Recall
E1
$1\text{-}r_f$
No Associative Retrieval Free Recall
s
Word 1 Retrieval
s
Word 2 Retrieval
E1
1-s
No Word 2 Retrieval
E2
1-s
No Word 1 Retrieval
s
Word 2 Retrieval
E2
1-s
No Word 2 Retrieval
E3
$1\text{-}r_c$
No Associative Retrieval Cued Recall Final
$r_f$
Associative Retrieval Free Recall
E4
$1\text{-}r_f$
No Associative Retrieval Free Recall
s
Word 1 Retrieval
s
Word 2 Retrieval
E4
1-s
No Word 2 Retrieval
E5
1-s
No Word 1 Retrieval
s
Word 2 Retrieval
E5
1-s
No Word 2 Retrieval
E6
$1\text{-}m_s$
No Associative Maintenance
u
Word 1 Retrieval
u
Word 2 Retrieval
E4
1-u
No Word 2 Retrieval
E5
1-u
No Word 1 Retrieval
u
Word 2 Retrieval
E5
1-u
No Word 2 Retrieval
E6
$1\text{-}r_c$
No Associative Retrieval Cued Recall Learning
$m_u$
Associative Maintenance
$r_c$
Associative Retrieval Cued Recall Final
$r_f$
Associative Retrieval Free Recall
E7
$1\text{-}r_f$
No Associative Retrieval Free Recall
s
Word 1 Retrieval
s
Word 2 Retrieval
E7
1-s
No Word 2 Retrieval
E8
1-s
No Word 1 Retrieval
s
Word 2 Retrieval
E8
1-s
No Word 2 Retrieval
E9
$1\text{-}r_c$
No Associative Retrieval Cued Recall Final
$r_f$
Associative Retrieval Free Recall
E10
$1\text{-}r_f$
No Associative Retrieval Free Recall
s
Word 1 Retrieval
s
Word 2 Retrieval
E10
1-s
No Word 2 Retrieval
E11
1-s
No Word 1 Retrieval
s
Word 2 Retrieval
E11
1-s
No Word 2 Retrieval
E12
$1\text{-}m_u$
No Associative Maintenance
u
Word 1 Retrieval
u
Word 2 Retrieval
E10
1-u
No Word 2 Retrieval
E11
1-u
No Word 1 Retrieval
u
Word 2 Retrieval
E11
1-u
No Word 2 Retrieval
E12
1-e
No Associative Encoding
u
Word 1 Retrieval
u
Word 2 Retrieval
E10
1-u
No Word 2 Retrieval
E11
1-u
No Word 1 Retrieval
u
Word 2 Retrieval
E11
1-u
No Word 2 Retrieval
E12

Figure 1



Free Recall
RI = 7 days
RI = 35 days
Cued Recall
Correctly recalled word pairs in %
0
10
20
30
40
50
60
70
80
90
100
0 days
1 day
11 days
Interstudy Interval


Figure 2

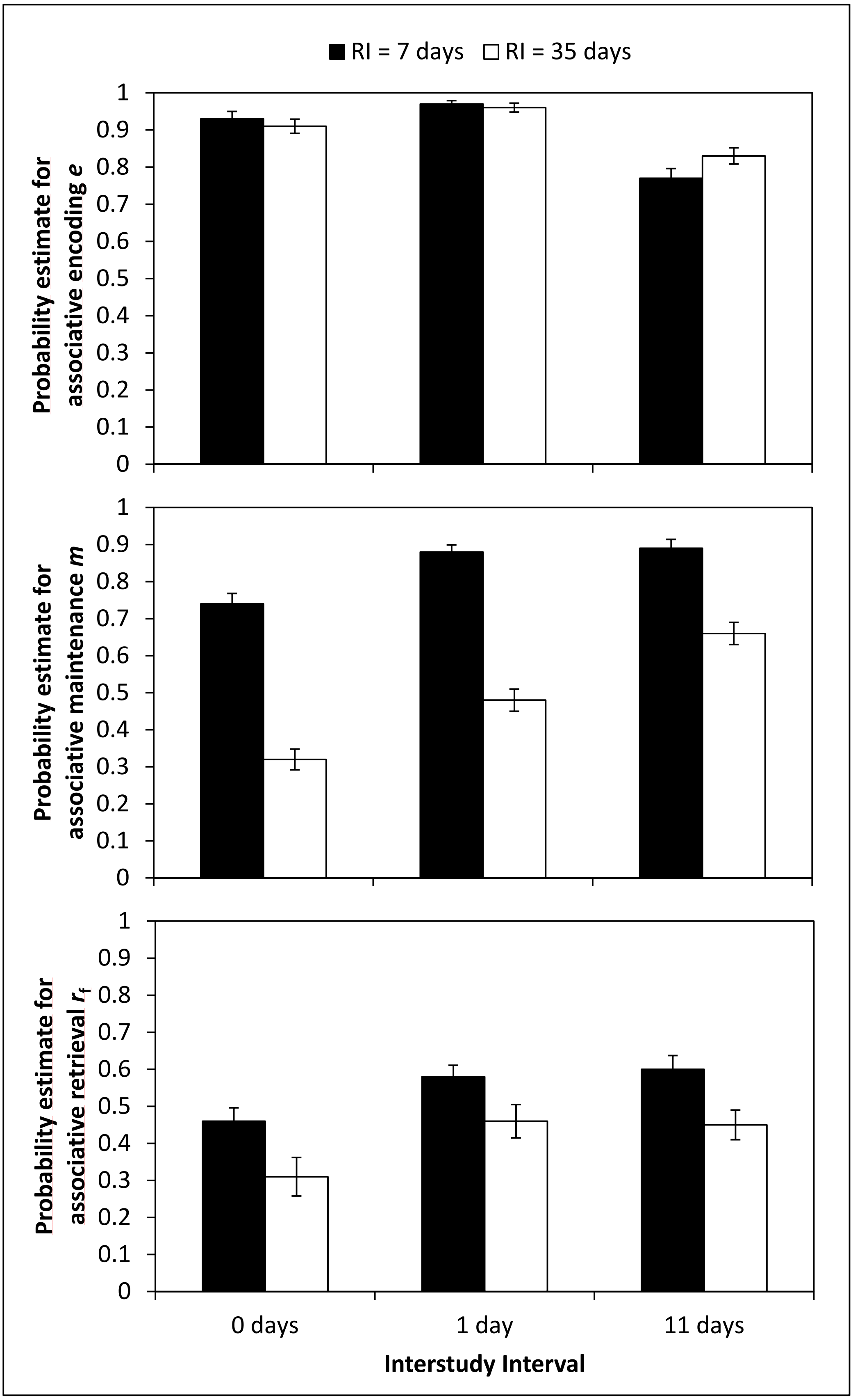

RI = 7 days
RI = 35 days
Probability estimate for associative encoding e
Probability estimate for associative maintenance m
Probability estimate for associative retrieval $r_f$
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 days
1 day
11 days
Interstudy Interval

Figure 3