Neuropsychologia 146 (2020) 107550

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: http://www.elsevier.com/locate/neuropsychologia

Specifying the neural basis of the spacing effect with multivariate ERP

A.S.N. Kim [a,b,*], M. Wiseheart [a,d], A.M.B. Wong-Kee-You [a], B.T. Le [b], S. Moreno [c,e], R. S. Rosenbaum [a,b,**]

[a] Department of Psychology, York University, Toronto, Canada
[b] Rotman Research Institute, Baycrest, Toronto, Canada
[c] School of Engineering, Simon Fraser University, Burnaby, Canada
[d] LaMarsh Centre for Child and Youth Research, York University, Toronto, Canada
[e] Digital Health Hub, Simon Fraser University, Surrey, Canada

ARTICLE INFO

Keywords:
Spacing effect
Distributed practice
Event-related potentials (ERPs)
Episodic memory
Partial least squares (PLS)
Study-phase retrieval hypothesis

ABSTRACT

The spacing effect refers to the finding that, given a fixed amount of study time, a longer interval between study repetitions improves long-term retention (e.g., Cepeda et al., 2006; Ebbinghaus, 1885/1967; Melton, 1970). Although the spacing effect is a robust and reliable finding in the memory literature, its cognitive and neural mechanisms remain unclear. We used event-related potentials (ERPs) to investigate the neural correlates of the spacing effect in the context of the study-phase retrieval hypothesis, which posits that repeated exposure of an item serves as a reminder of one's previous experience with the item, thereby promoting long-term retention. ERPs were recorded from 30 healthy young adults as they studied pairs of words under three levels of lag, corresponding to 0, 4, or 12 intervening pairs between the first and second occurrences of a target pair. We used two study-phase tasks that differed in the degree of retrieval that was required. During the test phase, participants were tested on paired-associate recall. The results demonstrated a significant effect of spacing on memory performance. However, the effect of encoding task and the interaction between encoding task and spacing were not significant. The results of the partial least squares analyses, which are not constrained by time window or electrode selection, revealed a spacing effect on the ERP data for both study-phase tasks; this effect occurred late in the epoch and was most salient over the centro-parietal scalp region. The results add to the literature on the neural correlates of the spacing effect by providing a more comprehensive account compared to past ERP findings that were focused on testing specific ERP components. They also call for further investigation on the various theoretical accounts of the spacing effect.

Episodic memory is essential to everyday life activities, ranging from remembering where we parked our car to recollecting the details of an important conversation. Attempts to distill episodic memory in the laboratory with well-controlled, word-list learning experiments helped to confirm that episodic memory benefits from distributed practice: given a fixed amount of study time, a longer interval between study repetitions improves long-term retention (e.g., Balota et al., 1989; Madigan, 1969; Melton, 1970). This spacing effect was first noted by Ebbinghaus, 1964/1967), who Tulving credited with providing the groundwork for what he later referred to as episodic memory (Tulving, 1983, 1985). Although the spacing effect has been demonstrated empirically in numerous, diverse populations (Balota et al., 1989; Green et al., 2014; Kim et al., 2018), in both lab-based and real-world settings (Kim, Wong-Kee-You, Wiseheart and Rosenbaum, 2019; Madigan, 1969; Sobel et al., 2011), in different modalities (Hintzman et al., 1973) and memory test formats (Madigan, 1969; Shaughnessy et al., 1972), there is little consensus amongst cognitive theorists on why distributing study events improves long-term retention compared to when study events are massed. Elucidating the neural mechanisms underlying the spacing effect could help provide a better understanding of why spacing benefits memory, and also help predict and optimize the effect in healthy and patient populations. With these aims, in this study we investigated event-related potential (ERP) correlates of the spacing effect, while testing the study-phase retrieval hypothesis, one of the leading accounts of the spacing effect as described further below.

Only a handful of ERP studies have investigated the spacing effect,

* Corresponding author. Department of Psychology, York University, Toronto, Canada.
** Corresponding author. Department of Psychology, York University, Toronto, Canada.
E-mail addresses: a_kim@yorku.ca (A.S.N. Kim), shaynar@yorku.ca (R.S. Rosenbaum).

https://doi.org/10.1016/j.neuropsychologia.2020.107550
Received 9 August 2019; Received in revised form 9 June 2020; Accepted 29 June 2020
Available online 30 June 2020
0028-3932/

and nearly all used a continuous recognition procedure (e.g., Van Strien et al., 2007; Kim et al., 2001). Some of these studies yielded mixed results, and at least one study did not find significant effects of spacing on the recorded potentials (Friedman, 1990). Nevertheless, a core set of ERP components underlying the spacing effect emerged in these studies, including the P300 and N400 (e.g., Van Strien et al., 2007; Kim et al., 2001; Feng et al., 2019).[1] The P300, also known as the "late positive component" (Sutton et al., 1965, 1967; Friedman et al., 1975; Picton, 1992), "P3", or "P3b," because it is the third major peak in the waveform that is positive (Picton, 1992; Luck, 2005; Herron et al., 2003), relates to target probability and is maximally recorded at the midline centro-parietal electrode locations (Picton, 1992; Polich, 2007). Although originally observed in the context of target evaluation and effects of attention (Sutton et al., 1965), later studies showed that a larger encoding-phase P3 is elicited by stimuli that were subsequently retrieved on a memory test compared to stimuli that were not subsequently retrieved (for a review, see Polich, 2007). The N400, on the other hand, is an ERP component with negative polarity that typically peaks around 400 ms after stimulus onset and is observed over the centro-parietal scalp region. This negative ERP deflection has been associated with semantic processing of meaningful information (Kutas and Federmeier, 2011). For example, words that are semantically incongruent within the context of a sentence have been shown to elicit a larger N400 compared to semantically congruent words (Kutas and Hillyard, 1980).

Two studies have implicated the P300 and N400 in the spacing effect in healthy young adults using a continuous recognition paradigm with words, but with different interpretations. Kim et al. (2001) demonstrated a larger P300 with an earlier latency for massed words compared to spaced words. This finding was interpreted in the context of template matching, where the latency and amplitude of the P300 is thought to vary as a function of how well an item matches a given template (i.e., memory of a stimulus event; Hillyard et al., 1971): the closer the match, the earlier and larger the P300. The findings of Kim et al. appeared to be in line with this reasoning, as words in the massed condition would serve as a closer match to the previously presented word compared to words in the spaced condition. Van Strien et al. (2007) also found a larger P300 with a shorter latency for words in the massed condition compared to words in the spaced condition. In contrast to Kim et al. (2001), Van Strien et al. interpreted these findings in the context of repetition priming, reasoning that better and faster recognition of a word results in earlier P300 peaks and larger amplitude repetition effects. However, this is thought to be coupled with reduced encoding and, consequently, reduced recall. Van Strien and colleagues highlighted that this is consistent with general patterns of performance associated with the spacing effect: massed items are greeted with better recognition during a continuous recognition paradigm but worse subsequent recall, whereas the opposite is true for spaced items.

Both Kim et al. (2001) and Van Strien et al. (2007) also found a larger N400 for words in their spaced conditions compared to words in their massed conditions. Kim et al. interpreted this finding in the context of memory search, such that a larger N400 is elicited with the increase in search that takes place to obtain the meaning of an item from long-term memory. Kim et al. suggested that words in the massed condition better matched the preceding context and therefore did not require as much of a memory search as words in the spaced condition. Van Strien et al. (2007), on the other hand, interpreted the effect of spacing on the N400 in relation to deficient semantic processing. The general idea is that under conditions of massed repetition, the semantic representation activated by the first occurrence of a word is still activated at the time of its second occurrence. Consequently, additional semantic processing is not required, resulting in the absence or attenuation of the N400. Under conditions of spaced repetition, however, additional semantic processing will occur, as reflected by the N400.

In a separate study, Zhao et al. (2015) investigated the neural correlates of the spacing effect using a study/test paradigm. Parallel EEG and functional magnetic resonance imaging (fMRI) data were collected while participants encoded words that were repeated in a massed or spaced condition. In contrast to earlier ERP studies, the words in the massed condition were repeated after one to three intervening words, whereas words in the spaced condition were repeated after 25 to 35 words. Each target word was repeated three times. After a retention interval of 24 h, participants were given an unexpected recognition test. Zhao and colleagues focused their ERP investigation on the frontal N400 (FN400) component, as it relates to familiarity-based recognition. Interestingly, they found that upon the third occurrence of a word, the FN400 was more positive for the massed condition compared to the spaced condition. However, the authors did not report whether there was an effect of spacing on the N400 and/or P300.

The respective roles of the N400 and P300 in the spacing effect, as well as the generalizability of these components across various retrieval paradigms that control for retention interval, have yet to be reconciled. Additionally, the question of whether there are other ERP components associated with the spacing effect, and whether this varies based on the retrieval paradigm, also requires further investigation. The present study investigated ERP correlates of spaced versus massed practice using a paired-associate paradigm. In contrast to previous ERP studies on the spacing effect that were constrained to specific electrode clusters and/or time windows, we analyzed our data using partial least squares (PLS), a data-driven, multivariate approach that identifies the strongest effects in a data set without predefining the electrode selection, time window, or patterns in the ERP waveforms (McIntosh et al., 1996). The data-driven approach to our analyses allowed us to assess the generalizability of past studies that relied on recognition paradigms and focused on the P300 and N400. We extended investigations of the spacing effect to a paired-associate paradigm, while also testing if and to what extent other ERP components reflected lag and/or spacing effects (e.g., the amount of variance accounted for by the lag and/or spacing effects across different time windows). Moreover, we controlled the retention interval between the second presentation of target pairs and their recall. Although frequently overlooked, controlling for retention interval is essential, as target items presented towards the end of a study list are more likely to be retrieved on the final test and often have the longest lags between the initial and repeated occurrence (Balota et al., 1989; Cepeda et al., 2006). The present study also incorporated three lag conditions (corresponding to 0, 4, or 12 intervening pairs between the first and second encounter of a target), which allowed us to investigate the impact of lag on memory performance and its underlying neural response.

Although there are implications for theoretical accounts of the spacing effect, none of the studies reviewed incorporated a direct test of these accounts within the design of the experimental tasks used. We adapted the paradigm used by Thios and D'Agostino (1976), as it resulted in significant spacing effects under conditions in which study-phase retrieval was made necessary, but not under conditions in which the need for retrieval was eliminated. To do so, we used two encoding conditions that varied in the degree of retrieval required for the encoding task. As described further below, study-phase retrieval should be necessary in one of our encoding conditions (retrieve), but not the other (read), providing the opportunity to test the study-phase retrieval hypothesis while investigating the ERP correlates of the spacing effect. The study-phase retrieval hypothesis posits that repeated exposure of an item must remind the learner of their previous experience with the item in order for it to be effective for long-term retention (Raaijmakers, 2003; Thios & D'Agostino, 1976). Provided that it is

[1] For the latter component, the 'P' corresponds to the positive polarity of the component, and the '300' corresponds to the latency, in ms, of the component peak as it was first reported (Sutton et al., 1965). However, the latency of the P300 has been shown to vary widely, often exceeding 500 ms, since it is elicited following the completion of stimulus evaluation (Karis, Fabiani & Donchin, 1984; Polich, 2007).

successful, the benefit of study-phase retrieval on subsequent memory is assumed to be greater as the difficulty of study-phase retrieval increases.

Based on the results of relevant past ERP studies (Kim et al., 2001; Van Strien et al., 2007; Zhao et al., 2015), we expected to find a smaller N400 and larger late positive component (LPC) in response to massed compared to spaced conditions. We also sought to explore any differences in the ERP signatures corresponding to the two encoding conditions across the different levels of lag. Although we did not have specific predictions for how the ERPs for the read and retrieve conditions might differ as a function of lag, any difference would be of particular interest if a spacing effect was found in one condition (e.g., retrieve condition) but not the other (e.g., read condition), as the corresponding ERP data would provide further insight into the mechanisms underlying the spacing effect. In line with past studies that support the study-phase retrieval hypothesis (Thios & D'Agostino, 1976; Johnston and Uhl, 1976), we expected to find a larger spacing effect for the retrieve condition compared to the read condition, as long as study-phase retrieval was successful. Together with these findings, we also expected an ERP repetition effect (Johnson, 1995; Rugg, 1995), indicated by larger component amplitudes elicited by the second occurrence of target items compared to the first occurrence.

1. Method

1.1. Participants

Thirty-nine participants were recruited from the participant pool at the Rotman Research Institute at Baycrest and screened to ensure that they were (1) not diagnosed with any medical condition or taking any medications that might alter brain function; (2) between the ages of 18–30 years; and (3) fluent English speakers with (4) normal or corrected-to-normal vision. All participants provided written informed consent, as approved by the Baycrest Research Ethics Board, prior to their participation. Participants who contributed 14 or more artifact-free trials for each of the conditions were included in the analyses. The data from five participants were excluded because there were too few trials in one or more of the conditions to allow for ERP analysis. The data from three participants were excluded from analyses due to corruption of the data files, and the data from one participant were excluded because they performed below chance. Of the 30 remaining participants, 15 were assigned to the read condition (eight females; mean age: 24 years, range: 18–29 years) and 15 were assigned to the retrieve condition (eight females; mean age: 24 years, range: 19–29 years). We chose a sample size that would allow at least 80% power to answer all of our hypotheses, based on an a priori power analysis and an effect size of $f = 0.53$, which was determined based on previously reported effect sizes for significant effect of spacing on ERP data (Kim et al., 2001; Ferrari et al., 2015).

1.2. Design

We used a 2 x 3 mixed factorial design. The two encoding groups (the between-subjects factor) consisted of the read group and retrieve group. In both groups, participants were presented with pairs of English nouns, and target pairs were presented a second time after one of three lags: 0, 4, or 12 intervening pairs (Lag-0, Lag-4, Lag-12, respectively). Lag (or spacing) was the within-subjects factor. ERPs (mean amplitude) recorded during the study phase, corresponding to the presentation of a word pair, served as the dependent variable.

1.3. Materials

The pool of experimental words consisted of 456 English nouns of two syllables, derived from the MRC Psycholinguistic Database (Coltheart, 1981; Wilson, 1988). This pool was divided into six lists that were each composed of 38 pairs of words. In each list, there were 12 target pairs that were presented twice. Four buffer pairs were used at the beginning of the list and seven buffer pairs were used at the end of the list. Each list also included 15 filler pairs that were each presented once. None of the pairs appeared in more than one list for a given participant, and the pairings never changed. The pairs were counterbalanced across participants in the experiment to ensure that presentation of each pair in the read and retrieve groups and across lag conditions was equivalent.

1.4. Procedure

Each participant took part in one experimental session, which consisted of 6 study-test cycles. During the study phase, participants studied 38 unique pairs of words. In each list, there were 12 target pairs that were presented twice to the participant. Thus, each list consisted of 50 word-pair presentations. Participants were instructed to memorize the pairs for a subsequent memory test, with the hope that this explicit memory task would help encourage participants to encode both words of a pair. Each session began with a practice block to familiarize the participants with the experimental task. Each of the participants confirmed that they understood the task upon completion of the practice block. Participants then completed the 6 study-test cycles of the task. After each study-test cycle, participants received a short break (approximately 5 min).

The ERP trial corresponded to the presentation of target word pairs during the study phase. The pairs were presented on a computer monitor, and the presentation of each pair began with a centrally presented fixation cross for 200 ms, followed by the presentation of the word pair for 3 s. The words of a pair were presented simultaneously, one above the other. This was followed by a response screen for 3 s. During the initial presentation of each pair (Fig. 1a), participants were instructed to say the second word of the pair (the word presented at the bottom) aloud when they viewed the response screen. Word-pair trials were separated by a blank screen that appeared for 1–3 s. Upon the second presentation of target pairs, participants performed a different task depending on the encoding group to which they were assigned (Fig. 1b). For the read group, participants were instructed to perform the same task as that performed during the first presentation of the word pair. For the retrieve group, the second presentation of a target pair began with a fixation cross for 200 ms, followed by presentation of the first word of a pair with a question mark ("?") beneath it. The first word of a pair served as the cue. Participants were then presented with a response screen, during which they were to recall aloud the word that was previously paired with the cue. For example, for the pair "JEWEL – TICKET," participants were presented with "JEWEL" with a "?" beneath it during the second presentation of the pair. The participant's task was to recall the word "TICKET" aloud. Participants were not provided with feedback about the accuracy of their responses. Following Thios and D'Agostino (1976), the test phase occurred immediately after the study phase. During the test phase, participants were tested on paired-associate recall for each of the pairs presented during the preceding study phase. The first word of a pair was used as a cue to recall the word that it was paired with. For each pair that was cued for paired-associate recall, participants were presented with a central fixation cross for 200 ms, followed by the cue word and a "?" beneath it for 7s (Fig. 2). Participants were asked to say aloud the word that was presented with the cue, and the experimenter made a button press to indicate participants' responses. The presentation of the next cue word was triggered by the button press to or when the time limit of 7 s was reached. An incorrect response and lack of a response were both registered as an error.

1.5. Behavioral data analysis

The results of the test phase paired-associate recall test were analyzed as the proportion of correct responses. Only those target pairs with correct responses in the study phase (read or recalled correctly in

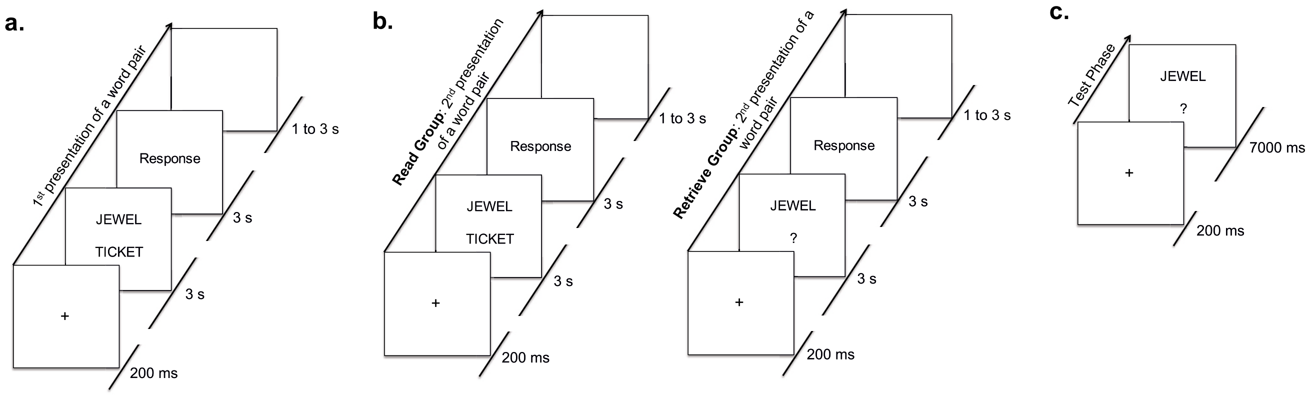

Fig. 1. Study-phase behavioural procedure. (A) First presentation of a pair of words. (B) Second presentation of a pair of words according to encoding group: read vs. retrieve. (C) Test-phase behavioural procedure.

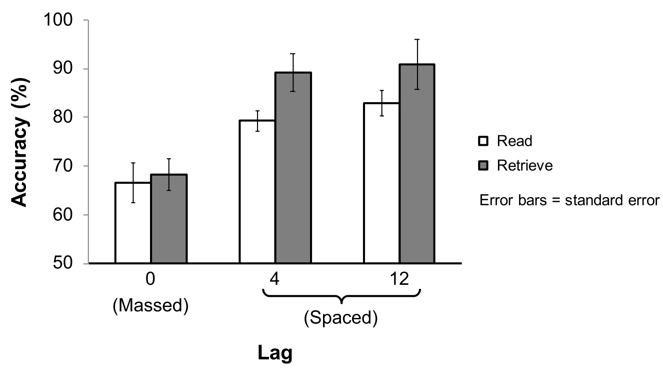

Fig. 2. Mean paired-associate recall performance as a function of encoding group and spacing condition. Error bars = standard error.

the read and retrieve groups, respectively) were considered in the analysis. A mixed-factor analysis of variance (ANOVA) was performed on these data to assess whether both encoding groups (read and retrieve) demonstrated a spacing effect.

1.6. Electrophysiological recording

Stimuli were presented visually on an LCD computer screen (Viewsonic VG932m) 60 cm from the participant during EEG recording. Continuous EEG activity was recorded using Ag/AgCl active electrodes (BioSemi). The electroencephalogram was recorded from 64 scalp electrodes based on the 10/20 system in a BioSemi electrode cap, with a Common Mode Sense (CMS) active electrode and Driven Right Leg (DRL) passive electrode serving as ground. An additional electrode was placed on each of both mastoids, and lateral to and below each eye to monitor eye movements and to cover the whole scalp evenly (a total of six additional electrodes). A bandpass of 0.05–50 Hz (Kane et al., 2000; Kim et al., 2009; Kim et al., 2012) and a sampling rate of 2048 Hz were used during recording. An average reference was used for ERP analyses. All off-line analyses were performed using Brain Electrical Source Analysis software (BESA; version 5.3.7). Ocular artifacts were detected using BESA's automatic artifact correction, which uses a principal component analysis that models eye movements. A predefined source model was applied to the data, accounting for HEOG, VEOG, and blink activities. The EEG data for each participant was visually scanned with the artifact correction applied to check for any noticeable ocular artifacts that had not been removed from the data. In these cases, any remaining ocular artifacts were corrected using BESA's manual selection and definition of artifacts for correction.

ERPs extracted from the EEG were epoched into 1200 ms segments and time-locked to the presentation of target pairs during the study-phase. The epoch began 200 ms before the presentation onset of a pair and lasted for another 1 s after the presentation of a pair. The epoch was baseline corrected to the 200 ms segment preceding the presentation onset of a pair. The resulting ERPs were submitted to two sets of PLS analyses (described below) to explore any effects of repeated presentation and lag on the ERP data.

The first PLS analysis assessed whether there were any effects of repetition in the study-phase ERP data. Thus, we investigated ERP differences as a function of word-pair occurrence (first vs. second occurrence of a pair) across the three lag conditions (Lag-0, Lag-4, Lag-12). The second PLS analysis explored whether the ERP patterns that accounted for most of the variance in the data reflected differences in memory performance as a function of increasing lags between the first and second occurrence of a pair during study.

1.7. PLS analysis

PLS is a multivariate analysis approach used to reduce the data and identify the strongest experimental effects, including when during the epoch and where over the scalp these effects occurred, in a data-driven manner. PLS describes the relation between a set of independent variables and a large set of dependent measures (McIntosh et al., 1996; http://www.rotman-baycrest.on.ca/pls). This is accomplished by computing the optimal least-squares fit to the portion of a cross-block covariance matrix that is attributable to experimental manipulations or behaviour (Wold, 1982). In this matrix, the sequence of time-points for each electrode is ordered as columns, and participants within each experimental condition are ordered as rows. The PLS output includes latent variables (LVs) that describe patterns across experimental conditions (referred to as design scores) expressed at each time point in terms of ERP amplitude. The number of LVs extracted is equal to the number of experimental conditions, with the first LV accounting for the most variance. A salience is then calculated for the LV at each electrode at each time point. The polarity and magnitude of the electrode salience denote the direction and strength, respectively, of the identified differences among the experimental conditions shown in the design scores. Scalp scores for each LV are obtained to measure how strongly each individual participant contributes to the patterns depicted by the LV. These scores are the dot product of the original data matrix and the saliences, resulting in a single value for each participant.

Different versions of PLS have been used in the literature, and this technique has been applied previously to analyze ERP data (Kim et al., 2009; Vallesi et al., 2009; Hay et al., 2002). Mean-centered PLS identifies the maximal experimental effects in the data set. This version of PLS is referred to as mean-centered, because the means for each column in the ERP amplitude data matrix described above are subtracted from the grand mean to produce a mean-centered deviation matrix. The LVs are then derived from singular value decomposition (SVD) of this mean-centered deviation matrix. In the present study, a mean-centered PLS analysis was used to identify where on the scalp the strongest experimental effects were expressed. Two mean-centered PLS analyses were conducted, one to examine the effects of repeated presentation and on the other to examine the spacing effect.

Statistical significance of the whole spatiotemporal pattern expressed by each LV was evaluated by a permutation test using 500 permutations across the different experimental conditions (Edgington, 1980; McIntosh et al., 1996). Permutations consist of sampling without replacement to reassign the order of conditions for each participant. Mean-centered PLS was recalculated for each new permuted sample, and the number of times the permuted singular values exceeded the observed singular values in each LV was calculated and expressed as a probability. An LV was considered significant at $p < .05$. A bootstrap test of 100 bootstrap samples was used to assess the stability of the saliences identified for each LV, at each electrode and time point. This was done to detect those portions of the ERP waveforms that express robust experimental effects across participants and to circumvent the effects of possible outliers (Efron and Tibshirani, 1986). Bootstrap samples were produced by sampling with replacement and keeping fixed the assignment of experimental conditions to each participant. Mean-centered PLS was computed for each bootstrap sample. The ratio of the salience to its standard error, estimated through the bootstrap procedure, approximately corresponds to a z-score. Bootstrap ratios equal to or greater than 2.81 (roughly corresponding to a p-level of .005) were chosen as the cut-off for stable non-zero saliences. Thus, whereas permutation tests were used to determine the significance of each LV, the stability of the saliences identified for each LV was determined using bootstrap estimates of the salience standard errors.

Since both the bootstrap estimates and permutation tests involve resampling, it is important to note that when an SVD is performed on a resampled matrix, reflection (a sign change in saliences) and axis rotation (a change in the order of extracted LVs) can occur. These reflections and rotations, however, are arbitrary and can be corrected using a Procrustes rotation of the resampled SVD outcome to the original SVD outcome (for details on the Procrustes rotation, please refer to McIntosh and Lobaugh, 2004). Whereas PLS using SVD identifies the maximal experimental effects in the data set, an alternate version of PLS can be used to constrain the analysis so the specified contrasts represent 100% of the cross-block covariance. This allows the experimenter to test specific hypotheses about the data, and is made possible by the relation between the singular values from the SVD to the total sums-of-squares from the data matrix, where the singular values are derived from the sums-of squares from the data matrix and an SVD is not performed. Consequently, a Procrustes rotation is not needed, which is why this particular version of PLS is referred to as "non-rotated task PLS." Similar to mean-centered PLS, in non-rotated PLS, the significance of the specified contrast and the corresponding electrode saliences are assessed through the permutation tests. The bootstrap ratios are used to assess the time-points at which there is a stable difference between the contrasted conditions. A non-rotated analysis was conducted to investigate *similarities in* the effect of lag (or spacing) on the ERP recordings across the two encoding groups (read and retrieve), as described further below.

1.7.1. Repetition effects

The first analysis of the ERP data investigated the effects of repeated presentation (first vs. second occurrence of a pair) and level of spacing (Lag-0, Lag-4, Lag-12). The encoding-related ERPs recorded at each electrode site were sorted separately for each presentation of a target pair as a function of lag, yielding the following 6 categories of ERPs: (a) first occurrence of target pairs in the Lag-0 condition (Pres1_Lag-0); (b) first occurrence of target pairs in the Lag-4 condition (Pres1_Lag-4); (c) first occurrence of target pairs in the Lag-12 condition (Pres1_Lag-12); (d) second occurrence of target pairs in the Lag-0 condition (Pres2_Lag-0); (e) second occurrence of target pairs in the Lag-4 condition (Pres2_Lag-4); and (f) second occurrence of target pairs in the Lag-12 condition (Pres2_Lag-12). Although the ERPs to the first occurrence of all target pairs were not expected to differ across the lag conditions, we organized these ERPs by lag so that they could be compared to ERPs elicited by the second occurrence of the same targets as a function of lag. Individual participant waveforms were calculated by averaging the ERPs corresponding to these 6 categories. The mean (SD; range) trial counts contributing to the grand-mean waveforms were as follows: 21 (2.65; 14–25) for Pres1_Lag-0; 21 (2.28; 16–24) for Pres1_Lag-4; 21 (2.65; 15–24) for Pres1_Lag-12; 20 (2.42; 14–24) for Pres2_Lag-0; 20 (2.90; 14–25) for Pres2_Lag-4 and 21 (3.35; 14–24) for Pres2_Lag-12. The ERPs were then submitted to a mean-centered PLS analysis.

1.7.2. Lag and encoding task effects

This analysis of the ERP data was used to investigate the effect of spacing and only considered ERPs to the second occurrence of target pairs that were subsequently recalled. More specifically, ERPs to the second presentation of a pair were only included for pairs that were read or retrieved correctly during the study phase (corresponding to the read and retrieve groups, respectively) and also recalled correctly during the subsequent test phase. The encoding-related ERPs were sorted and averaged separately for each level of spacing (Lag-0, Lag-4, Lag-12) and electrode site as a function of experimental group (read vs. retrieve), yielding the following three categories of ERPs for each encoding group: (a) second presentation of target pairs in the Lag-0 condition (Lag-0); (b) second presentation of target pairs in the Lag-4 condition (Lag-4); and (c) second presentation of target pairs in the Lag-12 condition (Lag-12). Individual participant waveforms were derived by averaging the ERPs corresponding to these three categories. The mean (SD; range) trial counts going into the grand-mean waveforms for the read condition were 14 (1.05; 14–18) for read_Lag-0, 16 (2.85; 12–22) for read_Lag-4, and 17 (2.57; 14–22) for read_Lag-12. The mean (SD; range) trial counts going into the grand-mean waveforms for the retrieve condition were 15 (2.20; 14–21) for retrieve_Lag-0, 15 (2.28; 14–22) for retrieve_Lag-4, and 15 (2.32; 14–21) for retrieve_Lag-12. The ERPs were then submitted to a mean-centered PLS analysis to investigate the strongest experimental effects in the data, as well as when during the epoch and where over the scalp these effects occurred. A non-rotated PLS analysis was then conducted to examine further similarities in the effect of lag across the two encoding groups (read and retrieve).

2. Results

2.1. Behavioural data

Participants' performance across lag and encoding conditions are listed in Table 1. The results of the ANOVA revealed that a significant main effect of encoding group was not found ($F(1, 28) = 2.09$, $p = .16$, $\eta^2 = .07$), nor was a significant interaction between encoding group and lag found ($F(1.60, 44.74) = 1.94$, $p = .16$, $\eta^2 = .07$). For the effect of Lag, Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated ($\chi^2 = 7.80$, $p = .02$). Therefore, a Greenhouse-Geisser correction was used. There was a significant effect of lag ($F(1.60, 44.74)$

Table 1
Memory performance (%) for each encoding group as a function of lag.

Condition	Lag	Mean	SD
Read	0	66.54	15.71
	4	79.32	8.27
	12	82.95	10.42
Retrieve	0	68.27	12.71
	4	89.20	14.85
	12	90.94	19.69

Note. SD = standard deviation.

= 45.41, $p < .001$, $\eta^2 = .62$) on test phase paired-associate recall performance. A significant linear and quadratic trend was also found for the lag variable ($F(1, 28) = 58.29$, $p < .001$, $\eta^2 = .68$ and $F(1, 28) = 19.90$, $p < .001$, $\eta^2 = .42$, respectively). These trend patterns revealed an increase in memory performance from Lag-0 to Lag-4, followed by a leveling in performance between Lag-4 and Lag-12.

When the data were collapsed across encoding conditions, a repeated-measures ANOVA with a Greenhouse-Geisser correction showed that mean accuracy on the final memory test (for targets that were read/retrieved correctly during the study phase) differed significantly between levels of lag ($F(1.57, 45.58) = 43.99$, $p < .001$, $\eta^2 = .6$). Post hoc tests using a Bonferroni correction revealed that the proportion of correctly recalled pairs were significantly lower in Lag-0 (M = .68, SD = .14) compared to Lag-4 (M = .84, SD = .13, $p < .001$, Cohen's d = 1.18) and Lag-12 (M = .87, SD = .16, $p < .001$, Cohen's d = 1.26). However, the proportion of correctly recalled pairs in the Lag-4 and Lag-12 conditions did not significantly differ (p = .27).

2.2. ERP data

We first describe the main ERP components and then report the PLS results. The stimuli elicited sensory evoked potentials that were maximally recorded over the posterior scalp regions. The main peaks were a positive wave (120 ms), followed by a negative wave (195 ms) that were

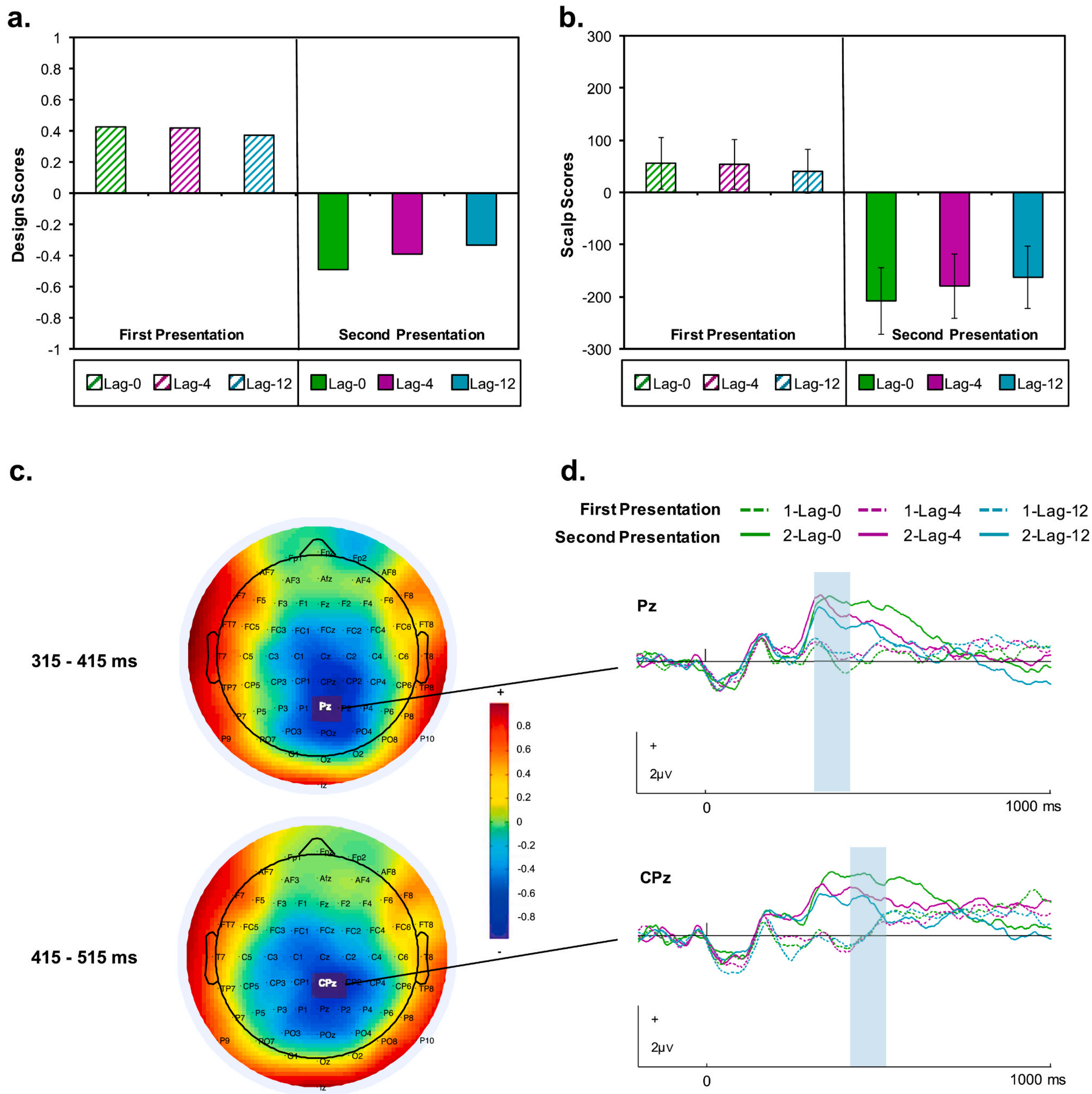

Fig. 3. Mean-centered partial least squares results for the analysis on repeated presentation as a function of lag: LV1. (A) Design scores (arbitrary units) according to presentation sequence (first presentation vs. second presentation) and spacing condition (Lag-0, Lag-4, Lag-12). (B) Scalp scores (arbitrary units) according to presentation sequence and spacing condition. Error bars plot 95% confidence intervals across participants. (C) Topographical distribution of saliences across the scalp at the designated time windows when bootstrap resampling of the electrode saliences indicated that the effect picked up by the LV was stable. (D) ERP waveforms recorded at electrodes with the peak saliences (Pz and CPz). Highlighted time windows correspond to the respective salience maps.

maximally recorded over the posterior scalp regions and of opposite polarity over anterior scalp regions. These visual evoked potentials were followed by a positive wave around 400 ms over the fronto-central scalp region and around 500 ms over the posterior scalp region (LPC).

2.2.1. Repetition effects

To help validate the PLS approach, repetition effects were first analyzed, as these have been consistently demonstrated in ERP studies that rely on univariate approaches to analysis. The first two LVs were significant by permutation tests ($ps < .001$). Design scores corresponding to the first LV (Fig. 6a) indicated that it reflected waveform differences between the first and the second occurrence of pairs across lags (accounting for 46.6% of the cross-block variance; observed singular value = 287.9). Design scores for the second LV (Fig. 4a; accounting for 21.1% of the cross-block variance; observed singular value = 193.5) differentiated the second presentation of pairs in the Lag-0 condition from the Lag-4 and Lag-12 conditions.

The first LV reflected a larger LPC elicited by the second occurrence of the word pairs compared to the first occurrence. The bootstrap resampling of the electrode saliences indicate that the effect picked up by this LV was stable during the latency-window between about 315 ms and 515 ms over the posterior scalp region. Saliences at all electrodes are shown in Fig. 3c for the following time windows: 315–415 ms and 415–515 ms. ERP waveforms recorded at electrodes with peak saliences

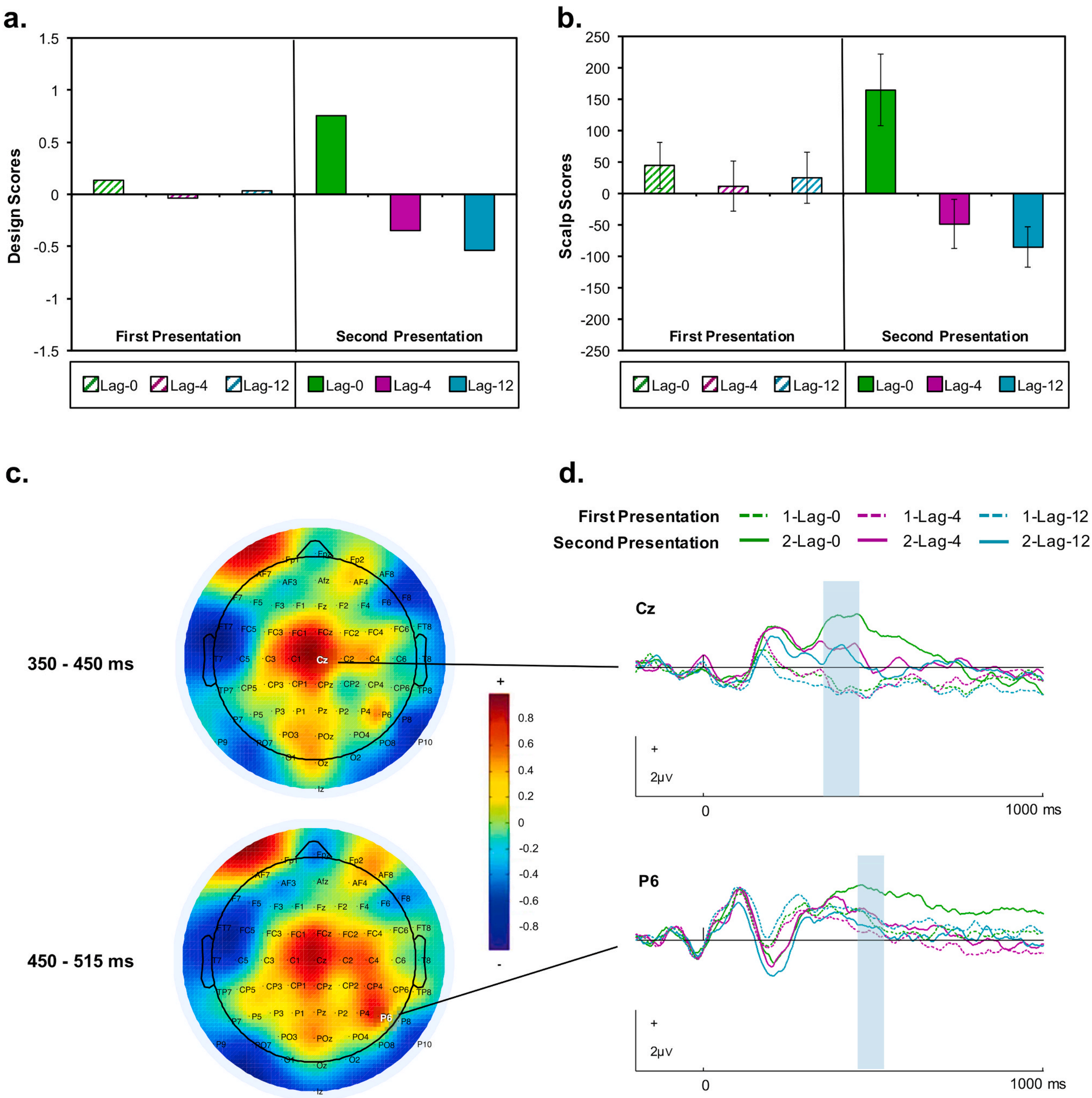

Fig. 4. Mean-centered partial least squares results for the analysis on repeated presentation as a function of lag: LV2. (A) Design scores (arbitrary units) according to presentation sequence (first presentation vs. second presentation) and spacing condition (Lag-0, Lag-4, Lag-12). (B) Scalp scores (arbitrary units) according to presentation sequence and spacing condition. Error bars plot 95% confidence intervals across participants. (C) Topographical distribution of saliences across the scalp at the designated time windows when bootstrap resampling of the electrode saliences indicated that the effect picked up by the LV was stable. (D) ERP waveforms recorded at electrodes with the peak saliences (Cz and P6). Highlighted time windows correspond to the respective salience maps.

(Pz and CPz) are shown in Fig. 3d, with highlighted time windows corresponding to the time windows of the respective salience maps. In the salience maps, negative saliences (shown in blue) indicate where over the scalp waveforms were more negative (or less positive) for the first occurrence of word pairs compared to the second occurrence of words pairs. Conversely, positive saliences (shown in red) indicate where waveforms were more positive (or less negative) for the first occurrence of word pairs compared to the second occurrence of words pairs.

The mean scalp scores, which reflect the degree to which each participant expressed the task–ERP relation indicated by the data, are shown for each category of pairs in Fig. 3b. The scalp scores for each category of pairs are considered statistically reliable if the error bars do not cross 0. Additionally, whereas PLS identifies omnibus patterns that indicate relative differences between a set of conditions, confidence intervals (CIs) provide a sense of how much the conditions differ. Thus, whether CIs overlap or are completely separate can provide insight into which conditions are distinguishable from one another. As seen in Fig. 3b, the error bars (CIs) for the lag conditions overlap for both the first and second occurrence of word pairs, indicating that this first LV reflects an effect of word-pair presentation (first vs. second) but not lag.

The second LV reflected larger positivity over the centro-parietal scalp to the second occurrence of words pairs in the Lag-0 condition vs. the Lag-4 and Lag-12 conditions. In contrast, the mean scalp scores

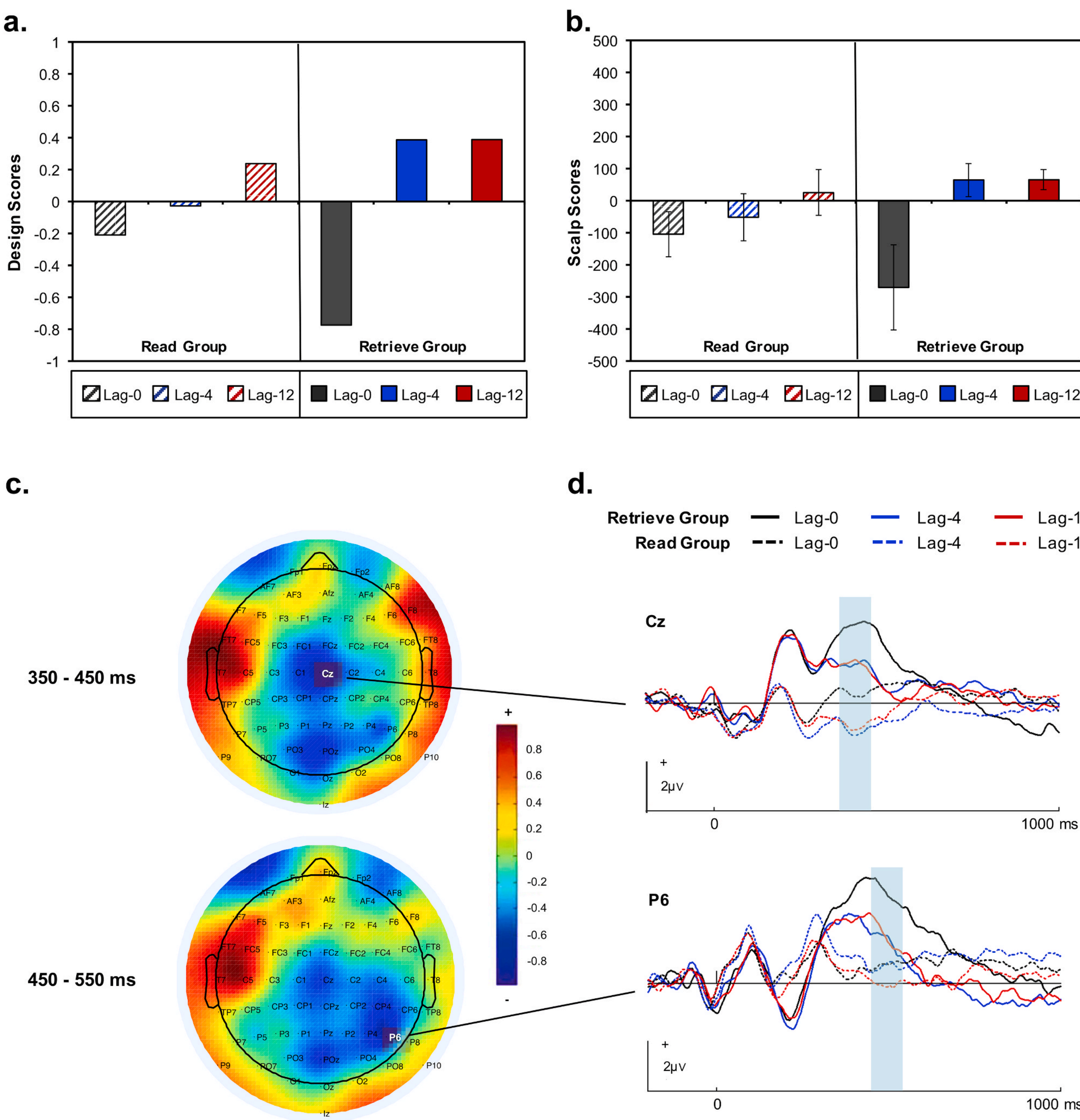

Fig. 5. Between-groups mean-centered partial least squares results for the spacing effect analysis. (A) Design scores (arbitrary units) according to spacing condition (Lag-0, Lag-4, Lag-12) and encoding group (read and retrieve). (B) Scalp scores (arbitrary units) according to spacing condition and encoding group. Error bars plot 95% confidence intervals across participants. (C) Topographical distribution of saliences across the scalp at the designated time windows when bootstrap resampling of the electrode saliences indicated that the effect picked up by LV was stable. (D) ERP waveforms recorded at electrodes with the peak saliences (Cz and P6). Highlighted time windows correspond to the respective salience maps.

(Fig. 4b) indicated that the levels of lag were not distinguishable from one another for the first occurrence of pairs, as the error bars overlapped across the lag conditions. The bootstrap resampling of the electrode saliences indicated that the effect picked up by this LV was stable during the latency-window between about 350 and 450 ms over the fronto-central scalp region, and the LPC was prevalent over the fronto-central and right posterior scalp region between about 450 and 515 ms. The topographical distribution of the electrode saliences is displayed in Fig. 4c for the 350–450 m and 450–515 ms time windows. Regions of the scalp showing positive saliences (shown in red) indicate electrode locations where the ERP amplitudes were more positive (or less negative) for the Lag-0 condition compared to the Lag-4 and Lag-12 conditions. In contrast, negative saliences (shown in blue) indicate scalp regions where the ERP amplitudes were more positive for the Lag-4 and Lag-12 conditions compared to the Lag-0 condition. ERP waveforms recorded at electrodes with peak saliences (Cz and P6) are shown in Fig. 4d. ERPs recorded at both of these electrode sites demonstrated sustained larger positivity for the second occurrence of pairs in the Lag-0 vs. the Lag-4 and Lag-12 conditions. Overall, the second occurrence of pairs elicited larger positive ERPs compared to the first occurrence of pairs within the time window highlighted in Fig. 4d.

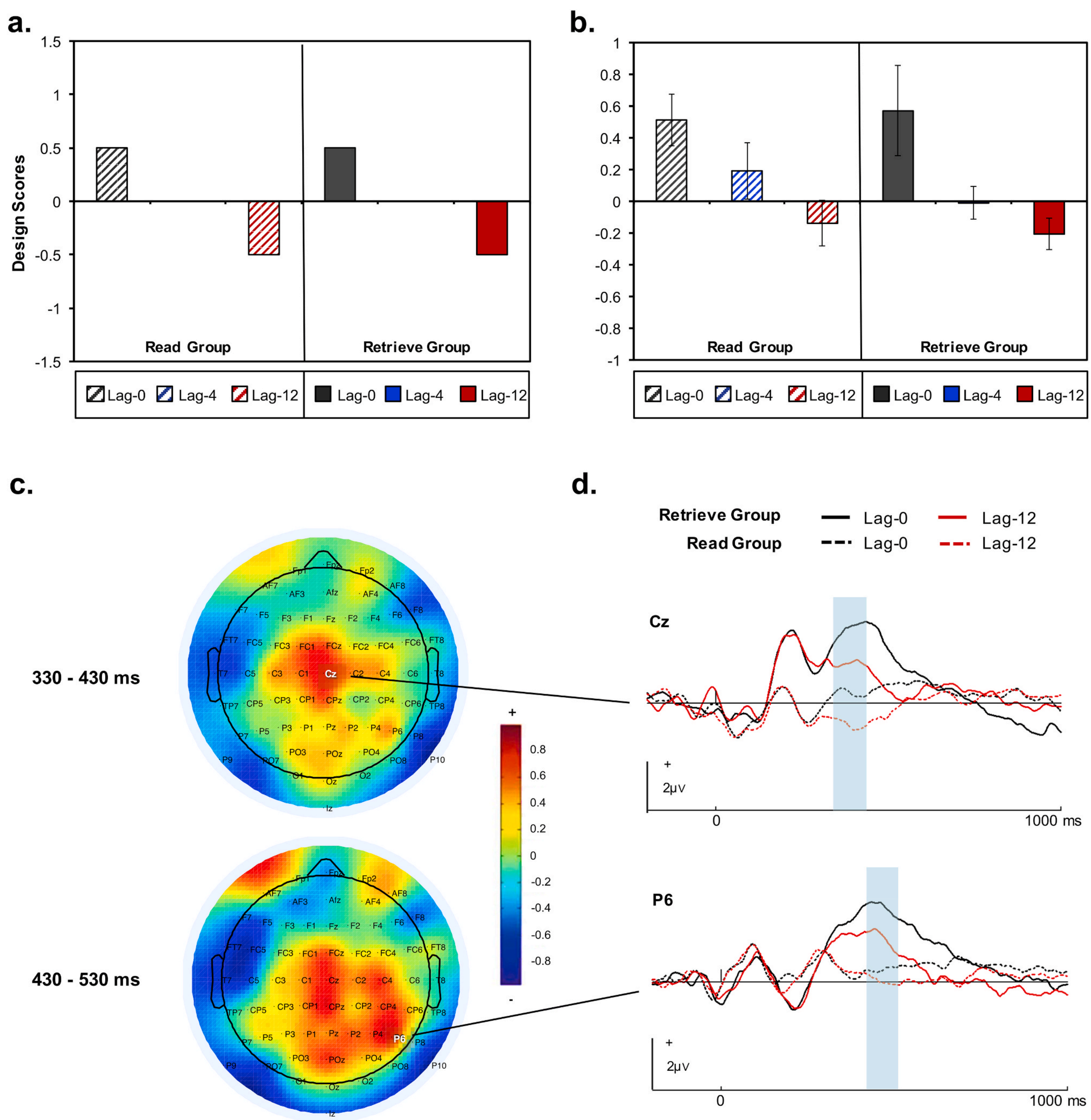

Fig. 6. Between-groups non-rotated partial least squares results for the spacing effect analysis. (A) Design scores (arbitrary units) according to spacing condition (Lag-0, Lag-4, Lag-12) and encoding group (read and retrieve). (B) Scalp scores (arbitrary units) according to spacing condition and encoding group. Error bars plot 95% confidence intervals across participants. (C) Topographical distribution of saliences across the scalp at the designated time windows when bootstrap resampling of the electrode saliences indicated that the effect picked up by LV was stable. (D) ERP waveforms recorded at electrodes with the peak saliences (Cz and P6). Highlighted time windows correspond to the respective salience maps.

2.2.2. Lag and encoding task effects

Mean-centered PLS analyses. The first LV was significant (observed singular value = 288.9, explained cross-block covariance = 38.3%, $p < .001$). The corresponding design scores (Fig. 5a) indicated that this LV reflected waveform differences between Lag-0 vs. Lag-4 and Lag-12 for the retrieve group. The mean scalp scores (Fig. 5b) indicated that the levels of lag were not distinguishable from one another in the read group, as the error bars overlapped across the lag conditions. The topographical distribution of the electrode saliences are shown in Fig. 5c: regions of the scalp showing negative saliences (shown in blue) indicate electrode locations where the ERP amplitudes were more positive for the Lag-0 condition of the retrieve group compared to the Lag-4 and Lag-12 conditions. Conversely, positive saliences (shown in red) indicate scalp regions where the ERP amplitudes were more positive for the Lag-4 and Lag-12 conditions of the retrieve group compared to the Lag-0 condition. The bootstrap resampling of the electrode saliences indicate that the effect picked up by this LV was stable over the central and posterior scalp regions from about 350 to 450 ms and over the right and mid-posterior scalp regions from about 450 to 550 ms after the presentation of a pair. Fig. 5c shows the mean electrode saliences for the 350–450 ms and 450–550 ms time windows. Fig. 5d shows the grand average waveforms recorded from two representative electrodes (Cz and P6), with highlighted time windows corresponding to the respective salience maps. To investigate further the spacing effect on the ERP data from both encoding groups, a non-rotated analysis was conducted to clarify the effect of spacing within each encoding group.

Non-rotated (set contrast) PLS analysis. A between-group analysis was conducted to contrast ERPs elicited by subsequently recalled target pairs in the Lag-0 condition against ERPs elicited by subsequently recalled target pairs in the Lag-12 condition. This analysis allowed the specified contrast to represent 100% of the cross-block covariance, yielding a significant latent variable (observed singular value = 267.2, $p < .001$). Fig. 6a shows the corresponding design scores, and Fig. 6b shows the mean scalp scores for each of the categories of pairs. The bootstrap resampling of the electrode saliences indicated that the effect picked up by this latent variable was stable between about 330 and 430 ms over the central scalp region and between 430 and 530 ms over the central and posterior scalp regions. The topographical distribution of the electrode saliences is shown in Fig. 6c for the 330–430 ms and 430–530 ms time windows. Regions of the scalp showing positive saliences (shown in red) indicate electrode locations where the ERP amplitudes were more positive (or less negative) for the Lag-0 condition compared to the Lag-12 conditions. In contrast, negative saliences (shown in blue) indicate scalp regions where the ERP amplitudes were more positive (or less negative) for the Lag-12 condition compared to the Lag-0 condition. ERP waveforms recorded at electrodes with peak saliences (Cz and P6) are shown in Fig. 6d. The time windows corresponding to the respective salience maps are highlighted in these figures. The mean scalp scores for each category of pairs are shown in Fig. 6b. For both encoding groups, the error bars for the Lag-0 and Lag-12 conditions did not overlap, demonstrating that the conditions were distinguishable for both encoding groups.

3. Discussion

To investigate the ERP correlates of the spacing effect, we used a multivariate approach that did not constrain our analyses to specific clusters of electrodes or time windows, setting the present study apart from previous ERP investigations on the spacing effect that tested specific ERP components. Additionally, we incorporated multiple spacing conditions (Lag-4, Lag-12) with a massed condition (Lag-0) into our study design, which allowed us to investigate the effect of different levels of spacing on the ERP data and memory performance. We did so while controlling for retention interval, which is important for investigations of spacing effects yet often overlooked, and used multiple encoding tasks that potentially differ in the degree of study-phase retrieval to help clarify the neural mechanisms underlying the spacing effect. Our behavioural results demonstrated a significant effect of spacing. However, neither the effect of encoding task nor the interaction between encoding task and spacing was significant. Consistent with these findings, the results of our multivariate ERP analyses revealed a spacing effect for both encoding conditions (read and retrieve) late over the centro-parietal scalp. The results of the present study shed light on the neural basis of the spacing effect, but also call for further investigation on the various accounts of this robust effect.

Our ERP results indicated a larger positive LPC for pairs in the massed condition (Lag-0) compared to pairs in the spaced conditions (Lag-4 and Lag-12). The results of the data-driven (mean-centered PLS) analysis revealed this effect for the retrieve group but not the read group, raising the possibility that participants were indeed engaged in more demanding retrieval in the 'retrieve' condition compared to the 'read' condition. It is also possible that explicit instructions to engage in study-phase retrieval may have recruited additional neural resources compared to when participants were instructed to simply read a target pair. This ERP pattern may reflect the ease with which pairs were retrieved during the study-phase, with massed pairs being easier to retrieve and eliciting a larger positive LPC compared to the spaced pairs. This finding is in line with past studies that demonstrate a larger positive LPC for massed, compared to spaced, words using continuous recognition paradigms (Van Strien et al., 2007; Kim et al., 2001). Similar to the results of the data-driven analysis described above, the results of the set contrast (non-rotated) analysis showed that, across both encoding groups, the amplitude of the LPC was larger for pairs that were studied within a massed (Lag-0), compared to a spaced (Lag-12), schedule.

In contrast to past ERP investigations on the spacing effect, we did not find an effect of spacing on the N400 or FN400. This may be due to differences in the materials and encoding tasks used across studies. For example, Van Strien et al. (2007) and Kim et al., (2001) both demonstrated an effect of spacing on the N400, and both groups presented their participants with single words in the context of a continuous recognition paradigm. Feng et al., (2019) also found an effect of spacing on the N400 using faces as stimuli and an encoding task that required participants to make age judgments about the faces that they viewed. In all of these studies, participants were tested on recognition during the final memory test. In contrast, in the present study, we used pairs of words with two different encoding tasks that differed in terms of whether participants had to engage in study-phase retrieval, and participants were tested on paired-associate recall during the final memory test. Nevertheless, past studies and the present study all demonstrated an effect of spacing on the LPC, suggesting that this latter component may reflect neurocognitive processes that generally underlie the spacing effect, independent of stimulus type and memory test format.

Our ERP findings included an expected repetition effect – the second occurrence of target pairs elicited larger positive amplitudes over the posterior scalp region compared to the first occurrence of these items. We also found an effect of lag in the ERP data for the second presentation of target pairs, whereas no such effect was found for the first presentation of target pairs. This suggests that the manipulation of lag only impacted the ERPs to the second presentation of pairs, which makes sense since lag was operationalized as the number of intervening word pairs between the repeated presentation of a target pair. The main distinction here was a contrast between target pairs that were massed (Lag-0) and target pairs that were spaced (Lag-4, Lag-12); the spaced conditions resulted in ERP patterns that were more similar to ERPs elicited by the original presentation of a pair compared to its immediate repetition in the massed condition, which may reflect neural processes related to repetition suppression and neural fatigue. In relation to the effect of spaced study events on later memory performance, past fMRI work has shown reduced repetition suppression within the hippocampus/MTL when repetitions are spaced compared to massed (Brozinsky et al., 2005), and the involvement of the hippocampus in better memory performance following study events that are distributed vs. massed (Li and Yang, 2020). It may be that after a certain level of spacing is implemented between repetitions, further spacing will not contribute to any additional benefit to subsequent memory performance. Another possibility is that there are different thresholds of spacing that need to be met to result in better memory performance. Other factors, including the task that participants perform during lags, may also influence the impact that a certain lag has on subsequent memory performance. For example, Thios and D'Agostino (1976) used the same encoding task, but different lags and a different task for participants to perform during the lags, across two experiments, and they found significant differences in participants' memory across all lag conditions in one experiment but not the other.

The finding that the manipulation of encoding task used in the present study did not result in significant differences in subsequent memory performance could be due to one or more factors. First, the encoding task used in the retrieve condition of the present study aligns with the generation effect (Slamecka and Graf, 1978), whereas the encoding task in the read condition aligns with the production effect (MacLeod et al., 2010), also known as the pronunciation effect (Hopkins and Edwards, 1972). The generation effect refers to the benefit of retrieving target information from memory. The production effect, in contrast, refers to the benefit of having studied information aloud, as opposed to silently, in memory. Although generation appears to require greater cognitive effort than production, both the generation and production of targets have been shown to similarly enhance subsequent recognition by approximately 10% (Bertsch et al., 2007; Ozubko and Macleod, 2010). This may help account for why we did not find a main effect of encoding task in the present study. It is also possible that participants in our production (or read) condition still had the cognitive capacity and motivation to also engage in at least some study-phase retrieval, which, in turn, could help explain why we did not find a main effect of encoding task. This could also relate to the use of an intentional memory task: because participants knew that their memory would be tested, they may have attempted to retrieve the pairs when the pairs were repeated. Future research should investigate these potential issues further.

Although the results of the present study do not favour the study-phase retrieval hypothesis, it is also difficult to explain our findings based on the encoding variability hypothesis, another leading account that posits that increasing the spacing between the repetitions of an item at study also increases the likelihood of the item being encoded differently due to changes in the environment (Balota et al., 1989; Bjork and Allen, 1970; Gartman and Johnson, 1972). Under conditions of spaced study, given that the encoding of an item presumably occurs in more than one context, it is more likely that one or more of these contexts will be partly reinstated at later retrieval, thus improving the overall retention of the item (Logan and Balota, 2008). Accordingly, changes in the encoding task across repetitions should also increase encoding variability. Thus, in the present study, encoding variability at the second occurrence of a target pair should have been greater in the retrieve vs. read condition, as the former includes a task change across repeated occurrences of a target (from read to retrieve), whereas the latter does not, such that any task-related encoding variability effects would not be present. If the processing requirements of our encoding tasks contributed to encoding variability, we would have expected to see differences in memory performance between the two encoding groups, but this was not the case. Our results support findings of past studies that have investigated the effects of lag on memory using an encoding task manipulation so that either the same or a different encoding task was performed across repetitions of a target (Bird et al., 1978; Maskarinec and Thompson, 1976; Shaughnessy, 1976). Future research should investigate whether these findings can be reconciled with the encoding variability hypothesis. Finally, the design of the present study does not allow for a direct test of the deficient processing hypothesis, another leading account of the spacing effect that claims that massed items are processed in a shallower manner compared to spaced items, which is thought to result in less efficient encoding and poorer memory (Gerbier and Toppino, 2015 for review). Nevertheless, our findings do not conflict with this account.

4. Conclusion

In this study, we supplement behavioural evidence of the spacing effect with ERP data based on a novel multivariate, data-driven approach to understanding the neural architecture of the spacing effect. A spacing effect was found within each encoding condition, even though they differed from one another in terms of instructions at study, with no significant effect of encoding task or interaction between encoding task and spacing on memory. The results of the ERP analyses, which were not constrained by time window or electrode selection, add to the extant literature on the neural correlates of the spacing effect by providing a more comprehensive account compared to past ERP studies that were focused on testing specific ERP components. The results of the present study cannot readily be explained based solely on the study-phase retrieval or encoding variability hypotheses, leaving open the possibility of the deficient processing hypothesis or a hybrid account as best capturing the spacing effect (Delaney et al., 2010; Lohnas and Kahana, 2014; Lohnas et al., 2011; Raaijmakers, 2003). The use of other encoding and retrieval manipulations together with multivariate, data-driven analysis of neuroimaging data should bring us even closer to understanding the neurocognitive architecture of the spacing effect.

CRediT authorship contribution statement

A.S.N. Kim: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration. **M. Wiseheart:** Conceptualization, Writing - review & editing, Supervision. **A.M.B. Wong-Kee-You:** Validation, Formal analysis, Investigation, Data curation, Writing - review & editing, Visualization. **B.T. Le:** Validation, Investigation, Data curation. **S. Moreno:** Conceptualization, Methodology, Resources, Writing - review & editing, Supervision. **R.S. Rosenbaum:** Conceptualization, Methodology, Resources, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgements

The research reported in this manuscript was supported by a Natural Sciences and Engineering Research Council (NSERC) grant and York Research Chair in Cognitive Neuroscience of Memory to R.S.R. This research was also supported by an NSERC Engage grant (EGP #451345-13) to R.S.R. and A.S.N.K., and a fellowship from NSERC to A.S.N.K. M. W. acknowledges support from NSERC. The authors declare no conflicts of interest.

References

Balota, D.A., Duchek, J.M., Paullin, R., 1989. Age-related differences in the impact of spacing, lag, and retention interval. Psychol. Aging 4 (1), 3.

Bertsch, S., Pesta, B.J., Wiscott, R., McDaniel, M., 2007. The generation effect: a meta-analytic review. Mem. Cognit. 35, 201–210.

Bird, C.P., Nicholson, A.J., Ringer, S., 1978. Resistance of the spacing effect to variations in encoding. Am. J. Psychol. 91, 113–121.

Bjork, R.A., Allen, T.W., 1970. The spacing effect: Consolidation or differential encoding. J. Verbal Learn. Verbal Behav. 9(50, 567–572. https://doi.org/10.1016/S0022-5371(70)80103-7.

Brozinsky, C.J., Yonelinas, A.P., Kroll, N.E.A., Ranganath, C., 2005. Lag-sensitive repetition suppression effects in anterior parahippocampal gyrus. Hippocampus 15, 557–561.

Cepeda, N.J., Pashler, H., Vul, E., Wixted, J.T., Rohrer, D., 2006. Distributed practice in verbal recall tasks: a review and quantitative synthesis. Psychol. Bull. 132 (3), 354.

Coltheart, M., 1981. The MRC psycholinguistic database. Q. J. Exp. Psychol.: Human Experimental Psychology 33A (4), 497–505.

Delaney, P.F., Verkoeijen, P.P., Spirgel, A., 2010. Spacing and testing effects: a deeply critical, lengthy, and at times discursive review of the literature. In: Psychology Of Learning and Motivation, vol. 53. Academic Press, pp. 63–147.

Ebbinghaus, H., 1964. Memory: A Contribution to Experimental Psychology. Dover, New York, NY (Original work published 1885).

Edgington, E.S., 1980. Randomization Tests. Marcel Dekker, New York.
Efron, B., Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. Stat. Sci. 1, 54–77.
Feng, K., Zhao, X., Liu, J., Cai, Y., Ye, Z., Chen, C., Xue, G., 2019. Spaced learning enhances episodic memory by increasing neural pattern similarity across repetitions. J. Neurosci. 39 (27), 5351–5360.
Ferrari, V., Bradley, M.M., Codispoti, M., Lang, P.J., 2015. Massed and distributed repetition of natural scenes: brain potentials and oscillatory activity. Psychophysiology 52 (7), 865–872.
Friedman, D., 1990. Cognitive event-related potential components during continuous recognition. Psychophysiology 27, 136–148.
Friedman, D., Simson, R., Ritter, W., Rapin, I., 1975. The late positive component (P300) and information processing in sentences. Electroencephalogr. Clin. Neurophysiol. 38 (3), 255–262. https://doi.org/10.1016/0013-4694(75)90246-1.
Gartman, L.M., Johnson, N.F., 1972. Massed versus distributed repetition of homographs: A test of the differential-encoding hypothesis. J. Verbal Learn. Verbal Behav. 11, 801–808.
Gerbier, E., Toppino, T.C., 2015. The effect of distributed practice: Neuroscience, cognition, and education. Trends in Neuroscience and Education 4 (3), 49–59.
Green, J.L., Weston, T., Wiseheart, M., Rosenbaum, R.S., 2014. Long-term spacing effect Benefits in developmental amnesia: case experiments in rehabilitation. Neuropsychology 28 (5), 685.
Hay, J.F., Kane, K.A., West, R., Alain, C., 2002. Event-related Neural Activity Associated with Habit and recollection. Neuropsychologia 40, 260–270.
Herron, J.E., Quayle, A.H., Rugg, M.D., 2003. Probability effects on event-related potential correlates of recognition memory. Cognit. Brain Res. 6, 66–73.
Hillyard, S.A., Squires, K.C., Bauer, J.W., Lindsay, P.H., 1971. Evoked potential correlates of auditory signal detection. Science 172, 1357–1360.
Hintzman, D.L., Block, R.A., Summers, J.J., 1973. Modality tags and memory for repetitions: locus of the spacing effect. J. Verb. Learn. Verb. Behav. 12, 229–238.
Hopkins, R.H., Edwards, R.E., 1972. Pronunciation effects in recognition memory. J. Verb. Learn. Verb. Behav. 11, 534–537.
Johnson, R., 1995. Event-related potential insights into the neurobiology of memory systems. In: Boller, J.C., Grafman, J. (Eds.), Handbook of Neuropsychology, vol. 10. Elsevier, Amsterdam, pp. 135–164.
Johnston, W.A., Uhl, C.N., 1976. The contributions of encoding effort and variabilityto the spacing effect on free recall. J. Exp. Psychol. Hum. Learn. Mem. 2, 153–160.
Kane, K., Picton, T.W., Moscovitch, M., Winocur, G., 2000. Event-related potentials during conscious and automatic memory retrieval. Cognit. Brain Res. 10, 19–35.
Kim, M.S., Kim, J.J., Kwon, J.S., 2001. The effect of immediate and delayed word repetition on event-related potential in a continuous recognition task. Cognit. Brain Res. 11 (3), 387–396.
Kim, A.S.N., Vallesi, A., Picton, T.W., Tulving, E., 2009. Cognitive association formation in episodic memory: evidence from event-related potentials. Neuropsychologia 47, 3162–3173.
Karis, D., Fabiani, M., Donchin, E., 1984. "P300" and memory: Individual differences in the von Restorff effect. Cognit. Psychol. 16 (2), 177–216.
Kim, A.S.N., Binns, M.A., Alain, C., 2012. Neuroelectric evidence for cognitive association formation: an event-related potential investigation. PloS One 7, e34856.
Kim, A.S.N., Saberi, F.M., Wiseheart, M., Rosenbaum, R.S., 2018. Ameliorating episodic memory deficits in a young adult with developmental (congenital) amnesia. J. Int. Neuropsychol. Soc. 24, 1003–1012.
Kim, A.S.N., Wong-Kee-You, A.M.B., Wiseheart, M., Rosenbaum, R.S., 2019. The spacing effect stands up to big data. Behav. Res. Methods. https://doi.org/10.3758/s13428-019-01235-z.
Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu. Rev. Psychol. 62, 621–647.
Kutas, M., Hillyard, S.A., 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207 (4427), 203–205.
Li, C., Yang, J., 2020. Role of the hippocampus in the spacing effect during memory retrieval. Hippocampus 1–12. https://doi-org.ezproxy.library.yorku.ca/10.1002/hipo.23193.
Logan, J.M., Balota, D., 2008. Expanded vs equal spaced retrieval practice in healthy young and older adults. Aging Cognit. Neuropsychol. 15, 257–280.
Lohnas, L.J., Kahana, M.J., 2014. A retrieved context account of spacing and repetition effects in free recall. J. Exp. Psychol. Learn. Mem. Cognit. 40, 755–764.
Lohnas, L.J., Polyn, S.M., Kahana, M.J., 2011. Contextual variability in free recall. J. Mem. Lang. 64 (3), 249–255.
Luck, S.J., 2005. An Introduction to the Event-Related Potential Technique. MIT Press, Cambridge, MA.
MacLeod, C.M., Gopie, N., Hourihan, K.L., Neary, K.R., Ozubko, J.D., 2010. The production effect: delineation of a phenomenon. J. Exp. Psychol. Learn. Mem. Cognit. 36, 671–685.
Madigan, S.A., 1969. Intraserial repetition and coding processes in free recall. J. Verb. Learn. Verb. Behav. 8 (6), 828–835.
Maskarinec, A.S., Thompson, C.P., 1976. The within-effect distributed practice effect: tests of the varied context and varied encoding hypotheses. Mem. Cognit. 4, 741–746.
McIntosh, A.R., Lobaugh, N.J., 2004. Partial least squares analysis of neuroimaging data: applications and advances. Neuroimage 23, S250–S263.
McIntosh, A.R., Bookstein, F.L., Haxby, J.V., Grady, C.L., 1996. Spatial pattern analysis of functional brain images using partial least squares. Neuroimage 3 (3), 143–157.
Melton, A.W., 1970. The situation with respect to the spacing of repetitions and memory. J. Verb. Learn. Verb. Behav. 9 (5), 596–606.
Ozubko, J.D., Macleod, C.M., 2010. The production effect in memory: evidence that distinctiveness underlies the benefit. J. Exp. Psychol. Learn. Mem. Cognit. 36 (6), 1543–1547.
Picton, T.W., 1992. The P300 wave of the human event-related potential. J. Clin. Neurophysiol. 9, 456–479.
Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118 (10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019.
Raaijmakers, J.G., 2003. Spacing and repetition effects in human memory: application of the SAM model. Cognit. Sci. 27 (3), 431–452.
Rugg, M.D., 1995. ERP studies of memory. In: Rugg, M.D., Coles, M.G.H. (Eds.), Electrophysiology of Mind: Event-Related Brain Potentials and Cognition. Oxford University Press, Oxford, UK, pp. 132–170.
Shaughnessy, J.J., 1976. Persistence of the spacing effect incidental learning conditions. Mem. Cognit. 4, 369–377.
Shaughnessy, J.J., Zimmerman, J., Underwood, B.J., 1972. Further evidence on the MP-DP effect in free-recall learning. J. Verb. Learn. Verb. Behav. 11 (1), 1–12.
Slamecka, N.J., Graf, P., 1978. The generation effect: delineation of a phenomenon. J. Exp. Psychol. Hum. Learn. Mem. 4, 592–604.
Sobel, H.S., Cepeda, N.J., Kapler, I.V., 2011. Spacing effects in real-world classroom vocabulary learning. Appl. Cognit. Psychol. 25 (5), 763–767.
Sutton, S., Braren, M., Zubin, J., John, E.R., 1965. Evoked potential correlates of stimulus uncertainty. Science 150, 1187–1188.
Sutton, S., Tueting, P., Zubin, J., John, E.R., 1967. Information delivery and the sensory evoked potential. Science 155, 1436–1439.
Thios, S.J., D'Agostino, P.R., 1976. Effects of repetition as a function of study-phase retrieval. J. Verb. Learn. Verb. Behav. 15 (5), 529–536.
Tulving, E., 1983. Elements of Episodic Memory. Oxford University Press, Oxford.
Tulving, E., 1985. Memory and consciousness. Can. Psychol. 26 (1), 1–12.
Vallesi, A., Stuss, D.T., McIntosh, A.R., Picton, T.W., 2009. Age-related differences in processing irrelevant information: evidence from event-related potentials. Neuropsychologia 47, 577–586.
Van Strien, J.W., Verkoeijen, P.P., Van der Meer, N., Franken, I.H., 2007. Electrophysiological correlates of word repetition spacing: ERP and induced band power old/new effects with massed and spaced repetitions. Int. J. Psychophysiol. 66 (3), 205–214.
Wilson, M.D., 1988. The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. Behav. Res. Methods Instruments Comput. 20, 6–11.
Wold, H., 1982. Soft modelling: the basic design and some extensions. In: Wold, H. (Ed.), Systems under Indirect Observation: Causality-Structure-Prediction. Part II. North-Holland Publishing Company, Amsterdam, pp. 1–54.
Zhao, X., Wang, C., Liu, Q., Xiao, X., Jiang, T., Chen, C., Xue, G., 2015. Neural mechanisms of the spacing effect in episodic memory: a parallel EEG and fMRI study. Cortex 69, 76–92.