An event-related potential study of the testing effect: Electrophysiological evidence for context-dependent processes changing throughout repeated practice

Dorottya Bencze a,b,1, Ágnes Szöllősi a,b,1, Kornél Németh a, Mihály Racsmány a,b

a Department of Cognitive Science, Budapest University of Technology and Economics, Egry József utca 1, 1111 Budapest, Hungary
b Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Magyar Tudósok Körútja 2, 1117 Budapest, Hungary

ARTICLE INFO

Keywords:
Testing effect
Repeated retrieval practice
Contextual features
Event-related potentials
Difference due to memory effects

ABSTRACT

The testing effect refers to a special form of performance improvement following practice. Specifically, repeated retrieval attempts improve long-term memory. In the present study we examined the underlying mechanisms of the testing effect as a function of time by investigating the electrophysiological correlates of repeated retrieval practice. We additionally investigated the ERP waveforms of the repeated practice phase as a function of the accuracy on the final test in a “difference due to memory” (Dm) analysis. We found a parietally distributed, increased positive amplitude between 500 and 700 ms, and a more positive parietal wave between 700 and 1000 ms in the later relative to the early phases of retrieval practice. We found parietal Dm effects in the same two time windows in the retrieval practice condition with a more positive amplitude predicting retrieval success on the final test. We interpret the earlier waveform as a component associated with episodic recollection and the later ERP as a component related to post-retrieval evaluation processes. Our results demonstrate the important role of these retrieval-related processes in the facilitating effect of retrieval practice on later retrieval, and show that the involvement of these processes changes throughout practice.

1. Introduction

1.1. The beneficial impact of retrieval practice on memory retention: The testing effect

Developing effective learning strategies has long been of particular scientific interest to psychologists as well as to representatives of many other professions. For this reason alone, it seems to be an especially important observation that retrieval in itself boosts memory. As it was described in the classic paper of Tulving (1967), a memory test “serves other functions besides that of measuring the amount or degree of learning” (page 181). That is, retrieval (or testing) can also be considered as an efficient learning/practice strategy. The testing effect refers to a special form of performance improvement following practice; specifically, retrieval practice improves memory retention more than additional study opportunities do (for overviews, see e.g., Putnam, Nestojko, & Roediger, 2016; Roediger & Butler, 2011; Roediger & Karpicke, 2006). Typically, the benefit of testing can only be observed when there is a relatively long delay – at least days or even weeks – between practice and the final memory test (see e.g., Karpicke & Roediger, 2008; but see e.g., Smith, Roediger, & Karpicke, 2013 for the short-term advantage of testing). This finding, therefore, points to the conclusion that retrieval reduces the rate of forgetting and that it plays an active role in long-term memory retention. Although, it seems that one retrieval attempt is beneficial in terms of long-term memory retention, a practice phase with repeated retrieval cycles after initial learning is more beneficial in terms of long-term memory performance than one single retrieval attempt (Racsmány, Szöllősi, & Bencze, 2018; Racsmány & Keresztes, 2015; Szpunar, McDermott, & Roediger, 2008; Wheeler & Roediger, 1992).

Interestingly though, our knowledge regarding the testing effect phenomenon is mostly based on behavioural findings and the underlying neurobiological processes remained relatively understudied (e.g., Karlsson Wirebring et al., 2015; Keresztes, Kaiser, Kovács, & Racsmány, 2014; van den Broek, Takashima, Segers, Fernández, & Verhoeven, 2020).

https://doi.org/10.1016/j.biopsycho.2022.108341
Received 31 August 2021; Received in revised form 14 April 2022; Accepted 18 April 2022
Available online 20 April 2022
0301-0511/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
2013; Wiklund-Hornqvist, Stillejs, Andersson, Jonsson, & Nyberg, 2021; Wing, Marsh, & Cabeza, 2013; for an overview, see van den Broek et al., 2016). In a typical experiment investigating the testing effect phenomenon, after initial learning participants practice the material by either retrieval or restudy. When only behavioural measures (i.e., recall success and/or response latency) are analysed, there is no means, however, to directly investigate those mechanisms that act during restudy practice. Neuroimaging techniques can provide additional important information as they allow the online measurement of such processes. Electrophysiological correlates of memory functions have been extensively studied and described in the literature and event-related potential (ERP) studies reported several distinct effects related to processes that take place during retrieval and effects that predict the success of later retrieval. The ERPs that correspond with different retrieval processes are investigated using mainly recognition memory tasks, analysing old/new effects by contrasting the ERP signals of studied old items and unstudied new items. Meanwhile processes related to later retrieval success manifest in “difference due to memory” (Dm) effects, also known as subsequent memory effects, which are calculated by contrasting the ERPs of items that were successfully retrieved in a later memory task and items that were not retrieved later (for reviews, see e.g., Friedman & Johnson, 2000; Paller & Wagner, 2002; Rugg & Curran, 2007). This comprehensive knowledge on the ERP correlates of retrieval processes and future retrieval success is a useful means in the study of the underlying mechanisms of the testing effect. Specifically, repeated retrieval practice is more beneficial, as compared to restudy practice, however, this difference can only be measured on the basis of memory performance on a final test when only behavioural methods are considered. It is also important what leads to this difference in final recall success, and to answer this question, it is a useful approach to investigate the (neural) processes taking place during various types of practice and their relationship with final retrieval success, as measured by the Dm effects. In addition, retrieval practice assumingly leads to changes in processes related to episodic memory (such as access to contextual details; see e.g., Karpicke, Lehman, & Aue, 2014), and these processes can be captured with specific ERP components.

1.2. The event-related potential correlates of memory retrieval and later retrieval success

One of the most frequently studied retrieval-related ERP components is known most commonly as the late positive component (LPC, also referred to as the left parietal old/new effect). This relatively late positive wave is most pronounced on the left parietal electrode sites for verbal stimuli between 500 and 700 ms after stimulus onset (Friedman & Johnson, 2000). The LPC is thought to reflect recollection, retrieving information with its episodic details, and with the feeling of remembering. Specifically, it is larger for memories when the context is successfully retrieved (Wilding & Rugg, 1996), larger for “Remember” responses (known to reflect recollection) than for “Know” responses (that are associated with familiarity-based memory decisions) (Duzel, Yonelinas, Mangun, Heineze, & Tulving, 1997), and increases following study-test repetitions (Johnson, Kreiter, Russo, & Zhu, 1998). Additionally, the LPC is proposed to reflect high decision accuracy and/or confidence (e.g., Addante, Ranganan, Olichney, & Yonelinas, 2012; Addante, Ranganan, & Yonelinas, 2012; Finnigan, Humphreys, Dennis, & Geffen, 2002; Rubin, Van Petten, Glisky, & Newberg, 1999). This latter finding further corroborates the idea that the LPC is related to recollection, as some current models of recognition memory suppose that high decision confidence is accompanied by recollective remembering (Yonelinas, 2001a, 2001b).

ERP components following the LPC are suggested to correspond with slower, mainly control functions and monitoring processes associated with retrieval. One of the major late components described in recognition memory studies is called the late posterior negativity (LPN). It is observed as a late negative shift over midline posterior areas, starting around 600–800 ms after stimulus onset and lasting further in the epoch (Herron, 2007; Leynes & Kakadia, 2013). The LPN is mainly present in tasks that require source monitoring and the retrieval of contextual information. Also, the LPN reflects extended retrieval processing presumably serving to reconstruct the original study episode when memory features are not readily available or need continued evaluation (Johansson & Mecklinger, 2003, Mecklinger, Rosburg, & Johansson, 2016). The LPN typically appears as an old/new effect with a more negative going amplitude for old compared to new items, but it has been observed in studies that directly compared the ERPs of the retrieval of items that did not differ in novelty. Leynes, Grey, and Crawford (2006) reported a more negative LPN amplitude during the recognition of actions that participants acted out previously without the physical objects necessary for the task, compared to actions that were performed with objects that aided retrieval with additional sensory information. Another study (Hellerstedt & Johansson, 2016) found a difference in the LPN in a task where participants were required to retrieve exemplars of categories from semantic memory by completing word-stems. The LPN was more negative for unsuccessful retrieval attempts compared to successful semantic retrievals, since unsuccessful retrieval presumably engaged additional search and monitoring of possible exemplar candidates. The LPN in these studies seemed to be more negative when participants needed to inspect the retrieved information more thoroughly or search for and monitor more associated information even when comparing previously studied items or items retrieved from semantic memory. Note that these latter findings do not contradict the results presented above (Herron, 2007: Leynes & Kakadia, 2013), considering that here the comparison is not between old and new items.

Along with the research of retrieval-related neural activity at the time of the EEG recording, electrophysiological signals that predict the success of future retrieval are also widely studied. The Dm effects are calculated by comparing EEG signals of items that were successfully retrieved in a future memory test to items that were not retrieved later (Paller, & Wagner, 2002). The majority of the Dm effect studies focused on neural correlates of later retrieval during encoding, reporting various differences with larger positive amplitudes for later remembered items (Paller & Wagner, 2002; Werkle-Bergner, Müller, Li, & Lindenberger, 2006). Larger Dm effects were reported for later “Remember” vs. “Know” responses (Voss & Paller, 2009) and for the encoding of inter-item associations (Weyerts, Tendolkar, Smid, & Heineze, 1997). Additionally, Dm effects were more consistently described in studies that used a recall task for subsequent retrieval as compared to recognition memory tasks (Paller, McCarthy, & Wood, 1988).

When investigating the underlying mechanisms of the testing effect, the electrophysiological correlates of memory retrieval - including recollective (LPC) and post-retrieval processes (LPN) - can be analysed to achieve a clearer understanding of processes that act during practice. Additionally, examining Dm effects to identify processes that predict future retrieval success helps to characterise the aspects of retrieval-practice that might facilitate future retrieval.

1.3. The event-related potential correlates of the testing effect and retrieval-based learning

Interestingly, ERP studies of the testing effect are relatively few. Recently there have been some findings regarding the oscillatory correlates of the effects of retrieval practice as well (see e.g., Ferreira, Marfil, Staudigl, Bajo, & Hanslmayr, 2014; Hanslmayr, Staudigl, Aslan, & Bäuml, 2010; Pastötter & Bäuml, 2016; Pastötter, Schicker, Niednerhuber, & Bäuml, 2011), however, reviewing this topic in more detail is beyond the scope of the present study. Among the ERP studies investigating the effects of retrieval practice some focused mainly on differences in the final test preceded by some form of practice.

Most studies investigating the testing effect by comparing the final recall success for items practiced with retrieval-practice compared to either non-practiced items (Spitzer, Hanslmayr, Opitz, Mecklinger, &
Bäuml, 2008; Rosburg, Johansson, Weigl, & Mecklinger, 2015) or elaborative retrieval (Liu, Mao, Peng, Lu, & Guo, 2019). These studies found a more positive amplitude in the LPC, a component associated with recollection, for items practiced by retrieval. Further testing effect studies analysed the ERPs from the practice phase by investigating Dm effects based on whether the practiced items were later remembered or forgotten. The results of these studies showed an increased LPC during practice being associated with either current (Bai, Bridger, Zimmer, & Mecklinger, 2015; Bridge & Paller, 2012; Liu, Tan, & Reder, 2018) or later retrieval success (Bai et al., 2015; Liu, Rosburg, Gao, Weber, & Guo, 2017). Regarding later ERP components (compared to the LPC), all studies found that in case of retrieval practice, a more positive amplitude after 700 ms predicted greater retrieval success on a later test (Bai et al., 2015; Bridge & Paller, 2012; Liu et al., 2017, 2018). This effect could be interpreted as the items that were more accessible on a later test requiring less post-retrieval evaluation and showing a more positive-going LPN (Bai et al., 2015; Liu et al., 2017). There has been only one study that specifically investigated processes occurring throughout multiple repetitions of retrieval practice (and not throughout only one or two repetitions) and its relations to the long-term effect of testing (Rafidi, Hulbert, Brooks, & Norman, 2018). Based on the EEG pattern changes throughout retrieval practice, Rafidi and colleagues (2018) found that the competition from related memories decreased across successful retrieval attempts during practice, and the degree of this decline predicted future retrieval, indicating a contribution of this process to the long-term beneficial effects of testing.

In sum, one set of previous studies investigated the electrophysiological correlates of final recall following retrieval practice (Liu et al., 2019; Rosburg et al., 2015; Spitzer et al., 2008). Studies investigating the underlying mechanisms of the events during practice focused on the electrophysiological correlates of retrieval practice as a function of final recall success (Bai et al., 2015; Bridge & Paller, 2012; Liu et al., 2017, 2018). However, the majority of these studies investigated the testing effect phenomenon after only one (Bai et al., 2015; Liu et al., 2017, 2019; Rosburg et al., 2015; Spitzer et al., 2008) or two practice cycles (Bridge & Paller, 2012; Liu et al., 2018). Previous behavioural results showed that repeated retrieval is more beneficial in terms of long-term memory retention than one retrieval attempt (Racsmany et al., 2018; Wheeler & Roediger, 1992; see also Karpicke et al., 2014), indicating differences between the initial and later stages of practice. For this reason, we designed an experiment to examine the electrophysiological correlates of repeated practice.

1.4. Study objectives

We conducted an experiment to investigate the electrophysiological changes during repeated practice. For this reason, following the initial (intentional) learning of paired associates (word pairs), participants practiced the study material in six subsequent cycles and the EEG recording was performed during this phase of the memory task. Our first aim was to analyse the mechanisms that act during practice as a function of time, therefore, we compared the electrophysiological correlates of the first half of practice (first three cycles) to the remaining three practice cycles. Our second aim was to investigate the relationship between the neural activity during practice and subsequent retrieval success. For this purpose we compared EEG recordings of items that were successfully recalled on a final test following a one-week delay to items that were not recalled later.

Half of the word pairs were practiced by retrieval while participants were required to recall the target words in response to the cue words. The other half of the word pairs were practiced by restudy while participants were presented again with the material to (re)study it. To get a clear understanding whether changes in electrophysiological responses are due to retrieval-specific processes or repetition (practice) itself, we used restudy as a control condition. As it was previously outlined, analysing behavioural measures does not provide an opportunity to directly assess study practice, and only indirect conclusions can be made on the basis of the later retrieval of the restudied items. The crucial point here is that memory tests show only which memories are accessible at a given time, and we have no information on the underlying processes that occur at encoding (see Roediger & Marsh, 2012). Taken into account this critical issue, we believe that investigating the electrophysiological correlates of restudy and its comparison to testing reveal important underlying mechanisms that act during different forms of repeated practice.

In sum, the major aim of our study was to investigate episodic retrieval-related processes during repeated tests. We investigated the change of electrophysiological activity throughout repeated retrieval and restudy practice, as well as the ERP differences during practice that were predictive of later retrieval success. We believe that retrieval-related processes are instrumental to the beneficial effects of retrieval practice (as compared to restudy practice) on final recall. In other words, as we assumed that subsequent recall success is reflected in the retrieval-related processes of the retrieval-practice phase, we aimed to analyse the Dm effects during practice. We analysed the ERP changes in three time windows in which previous retrieval-practice studies that similarly used word-pairs as stimuli (Bai et al., 2015; Liu et al., 2017, 2018, 2019) reported changes in retrieval-related ERP components as well as changes related to later memory success.

2. Materials and methods

2.1. Participants

Participants were 25 undergraduate students (native speakers of Hungarian) recruited from different universities in Budapest, Hungary. The data of two participants were excluded from the analysis due to an extremely low performance on the final test of the memory task (lower than 7% recall rate with only one response in either the restudy or the retest practice condition). Therefore, we analysed the data of 23 participants (8 men; age range: 19–29 years, $M = 22.3, SD = 1.9$). Required sample size was based on previous studies investigating the electrophysiological correlates of the testing effect (Bai et al., 2015; Liu et al., 2018; Rosburg et al., 2015) and on previous behavioural studies that found better long-term memory performance after retest practice as compared to restudy practice (e.g., Racsmany, Szollösí, & Marián, 2020; Storm, Friedman, Murayama, & Bjork, 2014).

All participants gave written informed consent at the beginning and at the end of each of the two experimental sessions and received money for participation. Participants had no history of psychiatric and neurological disorders as well as dyslexia, were not prescribed any medication known for influencing cognitive functions, and all had normal or corrected-to-normal vision. The study was approved by the United Ethical Review Committee for Research in Psychology, Hungary. The study was carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

2.2. Stimuli, experimental design, and procedure

Stimuli were 60 Swahili-Hungarian word pairs translated from Nelson and Dunlosky (1994). The stimuli were presented in the middle of the computer screen on grey background, with the word pairs next to each other with the distance between the two words kept at approximately 7.2° visual angles. Participants were seated 57 cm (22.4") from the screen with their head position fixed by keeping their chin in a chin rest. The recording has taken place in a dimly lit sound attenuated room.

The experimental design and procedure are illustrated in Fig. 1. Participants performed a memory task consisting of three phases: an initial learning phase (followed by a 30-minute delay), a repeated practice phase, and a final test phase. The first experimental session included the initial learning and repeated practice phases; the final test
was performed in the second experimental session seven days after the first session. EEG recording was performed during the repeated practice phase of the first experimental session.

2.2.1. Initial learning

In the initial learning phase all 60 word pairs were presented to the participants in random order for five consecutive cycles. All word pairs were presented in each learning cycle with the Swahili word on the left and its Hungarian equivalent on the right side of the screen. We decided to present the study material five times in this phase, because previous studies have pointed out that the beneficial effect of testing on long-term memory performance is assured following successful retrieval attempts and multiple pre-practice learning trials improve memory performance even at the initial stages of retrieval practice (see Karpicke et al., 2014; Racsmány et al., 2020). Each word pair was displayed for 5000 ms, with a 500 ms inter-stimulus interval (ISI). Before each learning cycle, participants were instructed to memorise the word pairs as well as they could and they could start the learning cycle by pressing the Space bar. There was no delay between the learning cycles. The initial learning phase was followed by a 30-minute delay, while the EEG cap and the electrodes were placed on the participants’ scalp.

2.2.2. Repeated practice

In the repeated practice phase the participants practiced all word pairs in six cycles. Each practice cycle consisted of a retest and a restudy block (the order of the two block types varied randomly across the practice cycles). The word pairs were randomly assigned into the retest or restudy conditions.

In the retest blocks participants performed a cued recall task. The rationale for using a recall task (instead of a recognition memory test) in the repeated practice phase was that most studies investigating the testing effect also used some form of recall tasks (for overviews, see Karpicke et al., 2014; Putnam et al., 2016; Roediger, 2007). The Swahili word remained on the screen for the entire 8000 ms even if an answer was presented. In each retest block 30 Swahili words were presented together with their Hungarian counterparts in random order (8000 ms/word pair; ISI: 1000 ms, blank screen for 100–500 ms and fixation cross for remaining 500–900 ms). Before each restudy block participants were instructed to memorise the word pairs. At the beginning of each practice block participants could start the task after reading the instructions by pressing the Space bar. There was no delay between the practice cycles.

2.2.3. Final test

Following a 7-day retention interval, participants’ memory for all 60 word pairs was tested in a final test phase. The final test differed from restudy practice in that participants were required to press the Space bar on the keyboard as soon as the corresponding word out loud as soon as it came to mind, to be recorded by the experimenter. The purpose of verbal responses was to reduce the likelihood of excessive head movements that could occur when participants are searching for answer keys on the keypad. The Swahili words were displayed for 8000 ms and the participants had that much time to give their answer before the next item was presented (following a 1000-ms ISI, with the first part showing a blank screen for a jittered period of 100–500 ms, followed by a fixation cross in the remaining 500–900 ms). We kept the ISI constant between trials because previous studies suggested that the time interval between retrieval attempts affect memory performance (e.g., Karpicke & Bauernschmidt, 2011; Karpicke & Roediger, 2007). The Swahili word remained on the screen for the entire 8000 ms even if an answer was presented.

In each retest block 30 Swahili words were presented in random order in the location where they were previously displayed on the screen in the initial learning phase. Participants were instructed to say the corresponding word out loud as soon as it came to mind, to be recorded by the experimenter. The purpose of verbal responses was to reduce the likelihood of excessive head movements that could occur when participants are searching for answer keys on the keypad. The Swahili words were displayed for 8000 ms and the participants had that much time to give their answer before the next item was presented (following a 1000-ms ISI, with the first part showing a blank screen for a jittered period of 100–500 ms, followed by a fixation cross in the remaining 500–900 ms). We kept the ISI constant between trials because previous studies suggested that the time interval between retrieval attempts affect memory performance (e.g., Karpicke & Bauernschmidt, 2011; Karpicke & Roediger, 2007). The Swahili word remained on the screen for the entire 8000 ms even if an answer was presented.

In each retest block 30 Swahili words were presented together with their Hungarian counterparts in random order (8000 ms/word pair; ISI: 1000 ms, blank screen for 100–500 ms and fixation cross for remaining 500–900 ms). Before each restudy block participants were instructed to memorise the word pairs. At the beginning of each practice block participants could start the task after reading the instructions by pressing the Space bar. There was no delay between the practice cycles.

Fig. 1. Experimental design and the procedure of the memory task. Note(s). Participants were presented with Swahili-Hungarian word pairs in the initial learning phase. This phase consisted of five subsequent cycles and all 60 word pairs were presented in each cycle. Initial learning was followed by a repeated practice phase that contained six subsequent practice cycles. Each cycle consisted of a retest and a restudy block. Former refers to a cued recall task where memory for half of the stimulus set was tested, whereas latter refers to the re-presentation of the remaining 30 word pairs. Seven days after practice, memory for all word pairs was tested in the final test phase (again, in the form of a cued recall task). EEG recording was performed during practice. ISI = inter-stimulus interval.
of the testing effect, otherwise the main focus of the study was the electrophysiological changes observed during the repeated practice phase. Stimulus presentation was controlled by Matlab 2008a (MathWorks, Natick, Massachusetts, US) using the Psychtoolbox 3.0.9 (Braunard, 1997).

2.3. Electrophysiological recording

EEG data were recorded using a Brain-Amp (BrainProducts GmbH, Munich, Germany) amplifier from 32 Ag/AgCl scalp electrodes placed according to the international 10/10 electrode system (Chatrian, Lettich, & Nelson, 1985). Eye movements were recorded using two electrodes placed on the outer canthi of the eyes, one electrode placed on the infraorbital ridge of the right eye and one placed on the forehead above the right eye. The averaged earlobes served as reference, and the ground electrode was placed on the forehead. All input impedances were kept below 10 kΩ; the EEG sampling rate was 1000 Hz.

The EEG data were analysed using EEGlab (Delorme & Makeig, 2004) and ERP lab (Lopez-Calderon, & Luck, 2014) toolboxes developed in the Matlab computing environment (MathWorks, Natick, MA, USA). The data were bandpass filtered offline removing frequencies below 0.1 Hz and above 70 Hz with an additional 50 Hz notch filter to remove noise from powerlines. The EEG data were segmented from −500–1800 ms relative to stimulus onset (i.e., beginning of display of the word pair in the restudy condition and beginning of display of the cue word in the restudy condition). The segmented data were re-referenced to a common average and were baseline-corrected using a 500 ms pre-stimulus baseline. We used a common average reference for the reason that we wished to reduce bias toward any component orientation (based on Dien, 2017) when planning to examine ERP changes with diverse topography (see the studies of Bai et al., 2015; Liu et al., 2017). Eye-movement related artifacts were removed using the automatic artifact detection algorithm ADJUST (Mognon, Jovovich, Bruzzone, & Buiatti, 2011) based on independent component analysis (ICA, see Delorme & Makeig, 2004).

2.4. Statistical analysis

We used an alpha level of p < .05 for all statistical tests. We used Greenhouse-Geisser correction to adjust for the lack of sphericity in repeated measures analyses of variance (ANOVAs). We report Cohen d value as a measure of effect size for t-tests and partial eta squared (ηp²) value as a measure of effect size for ANOVAs.

2.4.1. Behavioural data

To analyse the change in memory performance in the restudy condition of the practice cycles the recall rates were compared between the six practice cycles by conducting a repeated-measures ANOVA with six levels. Former studies pointed out that long-term recall success is better and response latency is shorter following restudy practice relative to restudy practice (Keresztes et al., 2014; Kubik, Jönsson, Knopf, & Mack, 2018; Marián et al., 2018; Racsmány et al., 2018; van den Broek et al., 2013). Therefore, to verify the relative long-term efficiency of retrieval-based learning, we analysed these two behavioural measures at the final recall of all word pairs seven days after the repeated practice phase. Memory performance in the final test was analysed by conducting two-tailed paired-samples t-tests to compare recall rates as well as reaction times between the restudy and restudy conditions.

2.4.2. Electrophysiological analysis

Our aim was to investigate the ERP differences observed throughout practice (effects of repeated practice) and the difference between items that were later retrieved or not retrieved (Dm effects). We selected three time windows for the analysis (300–500 ms, 500–700 ms, and 700–1000 ms) based on previous studies that investigated ERP differences in retrieval practice related to subsequent retrieval using cued recall tests with word pairs as stimuli (Bai et al., 2015; Liu et al., 2017; Liu et al., 2018). Mean ERP amplitudes were taken from a set of three frontal (F3, Fz, F4) and three parietal (P3, Pz, P4) electrodes based on previous studies (Bai et al., 2015; Liu et al., 2018; Spitzer et al., 2008).

In the statistical analysis of the effect of repeated practice, all epochs of both the restudy and restudy conditions were analysed. Our aim was to investigate the role of repeated practice separately from retrieval success (immediate or later), therefore we included the restudy trials with unsuccessful retrieval as well. Mean amplitudes were averaged for the first (1–3 practice cycles) and second (4–6 practice cycles) half of practice in the restudy and restudy conditions separately, with 90 epochs in all conditions. Consequently, the categories were as follows: (1) 1st-3rd restudy practice cycles, (2) 4th-6th restudy practice cycles, (3) 1st-3rd restudy practice cycles, and (4) 4th-6th restudy practice cycles. We analysed the Dm effects for all cycles of the repeated practice phase together, to obtain sufficient trial numbers for the ERP analysis. Mean amplitudes were averaged for trials with items that were successfully retrieved in the final test, and for trials that were not retrieved in the final test for both the restudy and restudy conditions. Accordingly, we analysed the following categories: (1) restudied items that were not retrieved later, (2) restudied items that were successfully retrieved later, (3) restudied items that were not retrieved later, (4) restudied items that were successfully retrieved later. The data were collapsed across all practice rounds. The mean number and range of trials within each cycle were as follows: 54 (12-162) restudied, later not retrieved; 126 (18-168) restudied, later not retrieved; 86 (30-168) restudied, later retrieved; 94 (12-150) restudied, later not retrieved trials.

Differences in ERP amplitudes were analysed by conducting separate ANOVAs for the three time windows on the frontal and parietal electrode sites. To capture the effects of practice, we conducted 2 × 2 × 3 repeated measures ANOVAs with Practice Strategy (Retest vs. Restudy), Practice Time (First half vs. Second half of the repeated practice phase) and Laterality (Left [F3/P3] vs. Middle [Fz/Pz] vs. Right [F4/P4]) as within-subject factors. For the Dm analysis, we conducted 2 × 2 × 3 repeated measures ANOVAs with the factors of Practice Strategy (Retest vs. Restudy), Final Retrieval (Later retrieved vs Later not retrieved) and Laterality (Left [F3/P3] vs. Middle [Fz/Pz] vs. Right [F4/P4]). Following significant three-way interactions, follow-up ANOVAs were conducted for the restudy and restudy practice conditions separately. Significant Practice Strategy × Practice Time interactions were followed by further contrast analyses to compare the first and the second half of the repeated practice phase in the two practice strategies separately (with Bonferroni corrections applied to account for multiple comparisons). Significant Practice Strategy × Final Retrieval interactions were similarly followed by separate contrast analyses for the two practice strategies.

3. Results

3.1. Behavioural results: Recall success and reaction times of correct responses

Recall rates (in %) during restudy practice were as follows: Mcycle1 = 57.0 (SE = 5.4), Mcycle2 = 55.4 (SE = 5.1), Mcycle3 = 57.1 (SE = 5.2), Mcycle4 = 58.0 (SE = 5.3), Mcycle5 = 57.4 (SE = 5.3), and Mcycle6 = 57.8 (SE = 5.2). For recall rate in this phase, the ANOVA showed a significant effect, F(3,52, 77.54) = 2.821, p < .036, ηp² = .114. Since the focus of our study was to examine the electrophysiological differences between the first and second half of practice, in a follow-up analysis we compared the cumulative recall rates of practice cycles 1–3 to cycles 4–6 (Mcycle1-3 = 56.5%, SE = 5.2; Mcycle4-6 = 57.7%, SE = 5.2). The paired-samples t-test showed no significant difference, t(22) = 1.936, p = .066, d = 0.404.

In the final test phase (see Fig. 2), participants showed higher recall rate, t(22) = 6.153, p < .001, d = 0.787, and faster reaction times, t (22) = 4.401, p < .001, d = 1.129, for the word pairs practiced by
retesting compared to the restudied items. These results showed the long-term facilitating effect of testing, resulting in better long-term memory performance and faster reaction times. Importantly, all the word pairs recalled in the final test phase were items that were previously recalled in the retest practice phase (except for one participant who recalled one extra item on the final test, relative to practice).

3.2. Electrophysiological results

3.2.1. Effects of repeated practice

The ANOVA results are reported in Table 1, the results of the follow-up post hoc analyses are described below. The ERPs of the selected electrodes from frontal and parietal sites are presented in Fig. 3. Topographic maps with scalp distributions of the voltage differences between the second (cycles 4–6) minus first (cycles 1–3) half of practice are seen in Fig. 4.

On the frontal electrode sites, the contrast analyses in the 500–700 ms time window did not show any significant difference between the first and second half of practice in either the retest, F(1.90, 41.80) = 2.292, p = .275, η² = .052, or the restudy condition, F(1.90, 41.80) = 4.205, p = .093, η² = .092.

On the parietal electrode sites in the 500–700 ms time window we found a significantly more positive amplitude in the second half of practice compared to the first half of the repeated practice phase in the retest condition, F(2, 44) = 20.504, p < .001, η² = .318, but no significant difference in the restudy condition, F(2, 44) = 0.027, p = 1.000, η² < .001. We observed a similar pattern in the 700–1000 ms time window: the contrast analyses showed a more positive amplitude for the second half of practice compared to the first half in the retest condition, F(1.96, 43.10) = 16.826, p < .001, η² = .280, and no difference throughout the repeated practice phase in the restudy condition, F(1.96, 43.10) = 1.144, p = .582, η² = .026.

In sum, our analysis of the effects of repeated practice showed a more positive amplitude with parietal distribution for the second half of practice compared to the first half of practice in the 500–700 ms (see Fig. 4 top middle topographic map) and 700–1000 ms time windows (see Fig. 4 top right topographic map) in the retest condition, but no such difference in the restudy condition (see Fig. 4 bottom row topographic maps). Note, that we analysed all trials; therefore, these ERP changes reflect retrieval attempts and not only successful retrieval.

Table 1

| Time window | Main effects and interactions | Electrode sites | Frontal | Parietal |
|-------------|-------------------------------|----------------|--------|---------|
| A) 300–500 ms | Practice Strategy | F(1, 22) = 1.561, p = .225, η² = .066 | F(1, 22) = 52.960, p < .001, η² = .707 |
|             | Practice Time | p = .329, η² = .015 | p = .327, η² = .129 |
|             | Practice Strategy × Time | p = .067, η² = .144 | p = .112, η² = .111 |
|             | Practice Strategy × Laterality | F(1.74, 38.15) = 10.635, p = .001, η² = .326 | F(1.74, 38.15) = 2.410, p = .124, η² = .099 |
|             | Practice Time × Laterality | F(1.44, 31.73) = 1.272, p = .099, η² = .010 | F(1.44, 31.73) = 0.379, p = .889, η² = .005 |

| B) 500–700 ms | Main effects and interactions | Electrode sites | Frontal | Parietal |
|-------------|-------------------------------|----------------|--------|---------|
|             | Practice Strategy | F(1, 22) = 16.772, p < .001, η² = .433 | F(1, 22) = 47.316, η² = .683 |
|             | Practice Time | p = .187, η² = .009 | p = .950, η² = .302 |
|             | Practice Strategy × Time | p = .033, η² = .190 | p = .003, η² = .344 |
|             | Practice Strategy × Laterality | F(1.43, 30.05) = 2.947, η² = .071 | F(1.43, 30.05) = 2.947, η² = .071 |
|             | Practice Time × Laterality | F(1.50, 33.05) = 0.381, η² = .017 | F(1.50, 33.05) = 0.381, η² = .017 |
|             | Practice Strategy × Time | p = .785, η² = .008 | p = .556, η² = .026 |

| C) 700–1000 ms | Main effects and interactions | Electrode sites | Frontal | Parietal |
|-------------|-------------------------------|----------------|--------|---------|
|             | Practice Strategy | F(1, 22) = 5.483, η² = .178 | F(1, 22) = 2.171, η² = .090 |
|             | Practice Time | p = .029, η² = .200 | p = .155, η² = .090 |
|             | Practice Strategy × Time | p = .645, η² = .029 | p = .111, η² = .347 |
|             | Practice Strategy × Laterality | F(1, 22) = 2.115, η² = .090 | F(1, 22) = 2.115, η² = .090 |
|             | Practice Time × Laterality | p = .160, η² = .088 | p = .030, η² = .196 |
|             | Practice Strategy × Laterality | F(1.38, 30.44) = 3.501, η² = .071 | F(1.38, 30.44) = 3.501, η² = .071 |
|             | Practice Time × Laterality | p = .058, η² = .137 | p = .004, η² = .288 |
|             | Practice Strategy × Laterality | F(1.46, 32.60) = 0.003, η² = .001 | F(1.46, 32.60) = 0.003, η² = .001 |
|             | Time × Laterality | p = .379, η² = .039 | p = .200, η² = .071 |

Note(s). ANOVAs conducted on frontal and parietal electrode sites in the (A’) 300–500 ms, (B’) 500–700 ms, and (C’) 700–1000 ms time windows on ERP data of the repeated practice phase, separated into the first and second half of the practice rounds. Shading indicates significant results.

3.2.2. Difference due to memory (Dm) effects

The ANOVA results are reported in Table 2, the results of the follow-up post hoc analyses are described below. The ERPs of the selected electrodes from frontal and parietal sites are presented in Fig. 5. Topographic maps with scalp distributions of the voltage differences between items that were successfully retrieved in the final test and those that were not retrieved later are seen in Fig. 6.

On the frontal electrode sites in the 500–700 ms time window the contrast analyses did not show any significant difference between items that were successfully recalled in the later final test and those that were not recalled later in either the retest, F(1.66, 39.60) = 4.311, p = .089, η² = .098, or the restudy condition, F(1.80, 39.60) = 0.982, p = .355, η² = .024. Similarly, in the 700–1000 ms time window the contrast analysis did not reveal any significant difference in either the retest, F(1.75, 38.50) = 3.269, p = .157, η² = .078, or the restudy condition, F(1.75,
On the parietal electrode sites in the 500–700 ms time window the follow-up ANOVAs revealed a significant Final Retrieval × Laterality interaction in case of the retest condition, \( F(1.72, 37.78) = 9.192, p < .001, \eta^2_p = .295 \). Further contrast analyses showed a more positive amplitude for items that were successfully retrieved in the final test phase compared to items that were not retrieved later on the left parietal (P3) electrode site, \( F(1.81, 39.70) = 26.289, p < .001, \eta^2_p = .398 \), but no such difference on the central (Pz), \( F(1.81, 39.70) = 2.719, p = .321, \eta^2_p = .064 \), or right (P4) electrodes, \( F(1.81, 39.70) = 3.117, p = .256, \eta^2_p = .073 \). In the restudy condition there was no significant interaction, \( F(1.74, 38.24) = 1.372, p = .264, \eta^2_p = .059 \), and no main effect of Final Retrieval, \( F(1, 22) = 2.648, p = .118, \eta^2_p = .107 \). The same pattern was also present in the 700–1000 ms time window with the follow-up ANOVA showing a significant Final Retrieval × Laterality interaction in the retest condition, \( F(1.54, 33.83) = 8.875, p = .002, \eta^2_p = .288 \). Similarly to the previous time window, contrasts analyses showed a significantly more positive amplitude for later retrieved items compared to later forgotten items only on the left parietal site, \( F(1.65, 36.2) = 20.638, p < .001, \eta^2_p = .363 \), and no difference on the central, \( F(1.65, 36.2) = 2.737, p = .320, \eta^2_p = .070 \), and right sites, \( F(1.65, 36.2) = 1.634, p = .628, \eta^2_p = .043 \). In the restudy condition we found no interaction, \( F(1.90, 41.74) = 0.374, p = .679, \eta^2_p = .017 \), and no main effect of Final Retrieval, \( F(1, 22) = 0.679, p = .419, \eta^2_p = .030 \).

To summarise, we found a Dm effect in the retest condition with a more positive amplitude for items that were remembered in the final test.
D. Bencze et al.

Table 2
Summary of the results of the repeated measures ANOVAs conducted to capture the Dm effects.

| A) 300–500 ms | Main effects and interactions | Electrode sites | Frontal | Parietal |
|---------------|-----------------------------|----------------|---------|---------|
| Practice Strategy | F(1, 22) = 1.889, p = 183, η² = .079 | p < .001, η² = .687 |
| Final Retrieval | F(1, 22) = .736, p = 400, η² = .032 | p = .581, η² = .14 |
| Practice Strategy × Final Retrieval | F(1, 22) = 2.735, p = .111 | p = .150, η² = .092 |
| Practice Strategy × Latency | F(1, 78, 30, 25) = 9.952, p < .001, η² = .312 | p = .143, η² = .090 |
| Final Retrieval × Latency | F(1, 72, 27, 68) = 1.023, p = .359, η² = .004 | p = .163, η² = .081 |

B) 500–700 ms
| Main effects and interactions | Electrode sites | Frontal | Parietal |
|-------------------------------|----------------|---------|---------|
| Practice Strategy | F(1, 22) = 16.987, p < .001, η² = .436 | p < .001, η² = .579 |
| Final Retrieval | F(1, 22) = 0.442, p = .513, η² = .020 | p < .001, η² = .446 |
| Practice Strategy × Final Retrieval | F(1, 22) = 7.040, p = .015, η² = .242 | p = .543, η² = .017 |
| Practice Strategy × Latency | F(1, 91, 22, 30) = 2.544, p = .110, η² = .097 | p = .119, η² = .098 |
| Final Retrieval × Latency | F(1, 85, 40, 34) = 3.452, p = .045, η² = .136 | p = .192, η² = .073 |
| Practice Strategy × Final Retrieval | F(1, 44, 10, 60) = 0.024, p = .941, η² = .001 | p = .002, η² = .251 |

C) 700–1000 ms
| Main effects and interactions | Electrode sites | Frontal | Parietal |
|-------------------------------|----------------|---------|---------|
| Practice Strategy | F(1, 22) = 7.426, p = .012, η² = .252 | p = .241, η² = .062 |
| Final Retrieval | F(1, 22) = 0.174, p = .681, η² = .008 | p = .020, η² = .222 |
| Practice Strategy × Final Retrieval | F(1, 22) = 6.887, p = .016, η² = .238 | p = .060, η² = .151 |
| Practice Strategy × Latency | F(1, 36, 20, 18) = 3.612, p = .055, η² = .141 | p = .002, η² = .305 |
| Final Retrieval × Latency | F(1, 41, 31, 11) = 0.405, p = .600, η² = .018 | p = .008, η² = .205 |
| Practice Strategy × Final Retrieval | F(1, 12, 24, 81) = 1.559, p = .226, η² = .066 | p = .011, η² = .188 |

Note(s). ANOVAs conducted on frontal and parietal electrode sites in the (A) 300–500 ms, (B) 500–700 ms and (C) 700–1000 ms time windows on the ERP data of the repeated practice phase, collapsed across all practice rounds. Shading indicates significant results.

compared to items that were not retrieved later, which was present on the left parietal site throughout the 500–700 ms (see Fig. 6 top middle topographic map) and 700–1000 ms time (see Fig. 6 top right topographic map). We found no significant Dm effects in the restudy condition (see Fig. 6 bottom row topographic maps).

4. Discussion

The aim of the present study was to investigate the underlying mechanisms of repeated retrieval practice. Our behavioural results confirmed the presence of the testing effect, as we found better memory performance as well as faster reaction times following rest study practice relative to restudy practice in the final test phase of the memory task. Although these behavioural findings are especially important as they verified the long-term efficiency of subsequent multiple memory tests, the main purpose of the present experiment was to examine the electrophysiological correlates of repeated practice. Specifically, we focused on the ERP correlates of repeated retrieval practice by analysing the ERP changes taking place throughout the course of repeated practice as well as changes during the repeated practice phase that predicted successful later retrieval.

It should be highlighted that we used multiple pre-practice learning trials to reach a relatively high and constant recall success rate in different stages of rest test practice. Consequently, participants recalled a comparable amount of word pairs in the early and later phases of retrieval practice (Mcycle-3 = 57% and Mcycle-6 = 58%). Therefore, we can conclude that the electrophysiological changes reported in the analysis of the effects of repeated practice did not depend on additional material being recalled in the later stages of practice. In other words, changes in ERPs do not specifically and exclusively reflect successful retrieval, instead, it is associated with all retrieval attempts, and the electrophysiological changes reflect different underlying mechanisms of the retrieval of the same material as a function of time. In this line of thought, the results of previous studies such as changes in pupil dilation (e.g., Pajkossy, Szöllősi, & Racsmány, 2019) or in reaction times (e.g., Racsmany et al., 2018) suggest that together with no increase in recall success, there are differences in processes associated with the initial and later stages of repeated retrieval.

4.1. Effects of repeated practice

Comparing the ERP recordings of the first and second half of practice revealed an increased positive amplitude for the second half of practice in the 500–700 ms time window that was only present in the restest condition. This effect had a parietal distribution, it was present in all analysed parietal electrode sites (P3, Pz, and P4). The time window and the parietal distribution of this effect is comparable to the LPC (Friedman & Johnson, 2000; Rugg & Curran, 2007). As this component reflects the retrieval of the contextual features of an episode (see e.g., Wilding & Rugg, 1996), our findings indicate an increase in the number of accessible contextual/episodic details as a result of repeated retrieval attempts. A couple of previous ERP studies have pointed out that the LPC is associated with recollection (Düzel et al., 1997) and high decision confidence (Finnigan et al., 2002; Rubin et al., 1999). In fact, these psychological constructs and mental processes are interdependent, as retrieval is accompanied by the feeling of remembering (recollection) when episodic/contextual details of memories are accessed during retrieval (Tulving, 1985), and recollective remembering is shown to be associated with high level of confidence (Yonelinas 2001a, 2001b).

The change in the LPC mirrors previous findings regarding the component’s association to retrieval practice. Previous ERP studies investigating the effect of testing on the final test (Liu et al., 2019; Spitzer et al., 2008) found an increased LPC associated with the beneficial effects of retrieval practice. Rosburg and colleagues (2015) also found a more pronounced LPC for the tested compared to the non-tested items. Additionally, they found an increase in the LPC for the tested items between the first and the second retrieval session. Crucially, Rosburg and colleagues applied only two retrieval sessions, whereas there were six subsequent retrieval (test) practice cycles in our experiment. We found a similar pattern of results as Rosburg and colleagues did, suggesting that changes between two retrieval cycles are still present throughout multiple subsequent memory tests.

This pattern of findings on the change in the amplitude of the LPC might fit with certain aspects of the episodic context account of the testing effect (Karpicke et al., 2014; Lehman, Smith, & Karpicke, 2014; Whiten & Karpicke, 2017). This theory suggests that the beneficial effect of testing is in the reconstruction of a study episode together with its contextual details. Accordingly, the LPC is thought to reflect the retrieval of episodic information, specifically, it is larger for memories when contextual details are accessible (e.g., Düzel et al., 1997; Leynes & Phillips, 2008; Wilding & Rugg, 1996; Woroch & Gonsalves, 2010). Importantly, we found no ERP change in the 500–700 ms time window
in the restudy condition suggesting that this change in the LPC specifically characterises retrieval practice and that repeated encounter with the stimuli in itself cannot be the crucial explanatory factor.

Our analysis of the ERP changes throughout repeated practice showed a second, retrieval practice-related difference between the first and second half of practice later in the epoch as well. In the 700–1000 ms time window we found a parietally distributed (P3, Pz and P4 electrode sites) ERP with a more positive amplitude in the second half of retrieval practice compared to the first half. This effect was only present in the practice condition that required retrieval, and was observed in a similarly late time window and the same parietal distribution as the LPN. (Note that we selected and interpreted the analysis of the 700–1000 ms time window based on studies such as Bai and colleagues (2015), but some previous studies (e.g., Leynes et al., 2006) reported that the LPN continues later in the epoch.) The LPN is suggested to reflect retrieval-related evaluative processing of memory features (Johansson & Mecklinger, 2003; Mecklinger et al., 2016), and it is typically observed either as a negative component (e.g., Leynes et al., 2006; Leynes & Kakadia, 2013) or as a negative going curve with a positive polarity (e.g., Groh-Bordin & Frings, 2009; Liu, Wu, Wang,

Fig. 5. ERP waveforms illustrating the Dm effects. Note(s). Grand average ERPs of retested items that were retrieved on the final test (solid blue lines), retested items that were not retrieved on the final test (dotted blue lines), restudied items that were retrieved on the final test (solid yellow lines) and restudied items that were not retrieved on the final test (dotted yellow lines), collapsed across all practice rounds. The ERPs are plotted from 500 ms before stimulus onset to 1800 ms after stimulus onset at left, midline and right electrodes form frontal and parietal sites (F3, Fz and F4 from the frontal, and P3, Pz and P4 from the parietal sites). Selected time windows (300–500, 500–700 and 700–1000) are highlighted in grey.

Fig. 6. Topographical maps illustrating the Dm effects. Note(s). Scalp distributions of the voltage differences between items that were successfully retrieved on the final test and items that were not retrieved on the final test, in the retest and restudy conditions (n.s.: non-significant).

in the restudy condition suggesting that this change in the LPC specifically characterises retrieval practice and that repeated encounter with the stimuli in itself cannot be the crucial explanatory factor.

Our analysis of the ERP changes throughout repeated practice showed a second, retrieval practice-related difference between the first and second half of practice later in the epoch as well. In the 700–1000 ms time window we found a parietally distributed (P3, Pz and P4 electrode sites) ERP with a more positive amplitude in the second half of retrieval practice compared to the first half. This effect was only present in the practice condition that required retrieval, and was observed in a similarly late time window and the same parietal distribution as the LPN. (Note that we selected and interpreted the analysis of the 700–1000 ms time window based on studies such as Bai and colleagues (2015), but some previous studies (e.g., Leynes et al., 2006) reported that the LPN continues later in the epoch.) The LPN is suggested to reflect retrieval-related evaluative processing of memory features (Johansson & Mecklinger, 2003; Mecklinger et al., 2016), and it is typically observed either as a negative component (e.g., Leynes et al., 2006; Leynes & Kakadia, 2013) or as a negative going curve with a positive polarity (e.g., Groh-Bordin & Frings, 2009; Liu, Wu, Wang,
This component shows a more positive amplitude in cases where there is less need for extended evaluation of the retrieved information, for example when the contextual details of the memory are more readily available (e.g., Leynes et al., 2006; Leynes & Kakadiia, 2013). The current results, therefore, demonstrate that following repeated retrieval attempts subsequent recall requires a decreased level of extended search processes and evaluation of the contextual information bound to the target memory. This result, again, is in line with the episodic context account of the testing effect (Karpike et al., 2014; Lehman et al., 2014; Whiffen & Karpicke, 2017), because this theory proposes that following repeated retrieval the searching set during remembering is restricted to the contextual features of the study episode and previous retrieval attempts.

The finding on the change in the amplitude of a component related to post-retrieval evaluation processes is consistent with the automatization account of the testing effect as well (Racsmany et al., 2018). This theory emphasizes that with each retrieval attempt the relation between a cue and a target item strengthens (see also Mulligan & Peterson, 2015) and that, as a consequence of repeated retrieval trials, finding a target item in response to a cue finally becomes fast and automatic. Crucially, automatic behaviors do not require attentional capacities and effort (Hasher & Zacks, 1979; Logan, 1988a, 1988b; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977); consequently, automatic responses are triggered in the presence of the (appropriate) cues without extended monitoring processes. The LPN has been shown to decrease with the repetition of the retrieval task when the task itself remained the same across repetitions even though the to-be-retrieved material varied (Herron, 2007). The author of this study described the LPN as a heterogenous component, the stimulus-locked wave reflecting both the maintenance of retrieved contextual information and retrieval fluency, the ease with which source related information was searched for or retrieved. In this line of thought our results showing a decreasing LPN with repeated retrieval-practice sessions could be interpreted as a result of increasing retrieval fluency across task repetitions.

In one of the studies that investigated the changes in post-retrieval processes during a retrieval-practice phase and a subsequent final test, Rosburg and colleagues (2015) found no change in the LPN when comparing the effect either between retested and unpracticed items, or between the retrieval practice session and the immediate final test. Based on our results that the decrease in the LPN was present when we compared the first half and the second half of a practice period consisting of six repetitions, our findings suggest that the change in the LPN may require prolonged practice to manifest across practice rounds.

### 4.2. Effects of later retrieval success

In the 500–700 ms time window we found an increased positive amplitude for retested items that were successfully recalled following a one-week delay compared to items that were not recalled later. The effect had a similar parietal, but more left-lateralized distribution (P3 electrode site) compared to the change observed during practice in this time window, and was only present in the retest condition. This result combined with the effect of repeated practice reported earlier shows that this parietal, retrieval-related ERP (bearing a functional similarity to the LPC) not only increases with repeated retrieval attempts, but also predicts the success of later memory retrieval.

Previous retrieval practice studies that investigated ERPs during practice as a function of later retrieval also reported a change in a similar component. Specifically, Bai and colleagues (2015) found a more positive amplitude between 500 and 700 ms for items that were successfully retrieved on a later test. The authors reported that this Dm effect resembled the widespread topographical distribution of retrieval success effect during the practice phase of their experiment, indicating that it reflects recollection-related processes comparable to the LPC. Another study (Liu et al., 2018) found the same widely distributed Dm effect with a slightly earlier beginning (400–700 ms). Liu and colleagues (2017) reported a similar, but parietally distributed Dm effect for retrieval-practice, which was presumed to reflect recollective processes as well. These studies, in line with our results, demonstrate the significant role of recollective processes in promoting the facilitating effect of retrieval-practice on later retrieval. Our results extend these findings by showing evidence that recollection-related processes are predictive to later retrieval success even in case of multiple cycles of repeated retrieval practice, and importantly, that the involvement of these processes seems to increase throughout practice.

In the later, 700–1000 ms time window we found a Dm effect with a more positive amplitude for later successfully retrieved items compared to items that were not retrieved later. The effect had a parietal distribution similar to the practice-related effect observed in this later time window, but had a more left lateralized scalp distribution (P3 electrode site). Based on the late time window, parietal distribution and presence only in the retrieval-practice condition, it is likely that this effect reflects the same processes as the practice-related changes reported in the 700–1000 ms time window and is functionally comparable to the LPN. Even though the LPN generally has a relatively symmetrical posterior topography (see Johansson & Mecklinger, 2003; Mecklinger et al., 2016), but there have been studies reporting an LPN with a more left lateralized distribution (e.g., Hellerstedt & Johansson, 2016; Johansson, Stenberg, Lindgren, & Rosén, 2002; Rugg, Schloerscheidt, & Mark, 1998).

Similarly to our results, a previous study (Bai et al., 2015) reported a more positive-going LPN for retested items that were successfully retrieved on a later test compared to items that were later forgotten, the latter said to be less accessible and requiring more extensive evaluation. In the study of Liu and colleagues (2017) a comparable, widely distributed late Dm effect was observed in the retrieval practice condition, extending even later in the epoch. The authors concluded that the effect reflects contextual information retrieval and manipulation of the retrieved information. An alternative interpretation of this late Dm effect for retrieval practice is that it represents re-encoding processes engaging after the information was retrieved (Bridge & Paller, 2012; Liu et al., 2018), connected to post-retrieval information evaluation processes (Liu et al., 2018). In case of the late Dm effect observed in our study, our results are consistent with these previous findings showing that later retrieval success is predicted by more positive amplitude. However, contrary to Liu and colleagues (2018) who reported the disappearance of this effect after one successful retrieval attempt, we observed this difference in the combined data of six retrieval practice cycles, and with a distinctly parietal distribution. Additionally, along with the finding that the decreased need for the engagement of these post-retrieval processes predicted successful future retrieval, our analysis of the repeated practice effects showed that these processes also decreased over the time course of repeated retrieval practice cycles.

### 4.3. Repeated retrieval facilitates access to episodic details

Our findings related to final recall success (Dm effects) showed a similar spatiotemporal distribution to the effects related to repeated retrieval practice. This pattern, along with the similarity to the findings of previous retrieval practice studies (Bai et al., 2015; Bridge & Paller, 2012; Liu et al., 2017, 2018) indicates that processes observed throughout repeated retrieval practice and subsequent recall success are interrelated. In other words, retrieval-related processes changing throughout practice (including access to episodic/contextual details) - and not only practice or repetition in itself - are instrumental to the beneficial effects of retrieval practice on final recall.

It is also important to interpret the findings on the change in the amplitudes of the LPC and the LPN together. We found a decrease in the LPN indicating a decrease in search processes for episodic details. Although our data on the LPC indicate that there is an increase in access to contextual details as a function of repeated retrieval practice, this...
finding is in line with our data on the LPN. Specifically, previous studies suggest that there is less need for additional search processes in cases when the contextual details are more readily available resulting in a decrease in the LPN (Johansson & Mecklinger, 2003, Mecklinger et al., 2016). In sum, it seems that repeated retrieval promotes recollective remembering (as suggested by our findings on the LPC) and that accessing the contextual/episodic details of a memory does not require extra monitoring/extended search processes (as indicated by our findings on the LPN).

Traditionally, recollection is suggested to be slow and effortful (Yonelinas, 2001a, 2001b, 2002; see also Atkinson & Juola, 1973). However, several studies of episodic autobiographical memory make a distinction between the indirect (or generative, strategic) and direct (or associative) ways of retrieval (Conway, 2005; Conway & Pleydell-Pearce, 2000; Moscovitch, 1995). The indirect way of episodic memory retrieval requires iterative search processes, including the elaboration of a cue (often followed by further elaborations when needed) and extended monitoring processes. In contrast, direct retrieval is typically fast and automatic. Importantly, theories of episodic memory stress the central role of appropriate cues in direct retrieval (Conway, 2005; Ehlers et al., 2002). That is to say, when there is a strong association between a cue and a target memory, there is no need for iterative search processes, because a highly specific cue can automatically trigger the retrieval of the target memory together with its contextual details. In brief, recollection can be experienced without extended monitoring/search processes and it seems that repeated retrieval supports direct/automatic access to the target information and its episodic/contextual details.

Relatively, retrieval practice has been shown to promote greater processing of cue-target associations while weakening the processing of associations across cue-target pairs (such as common categorical or semantic information; see Mulligan & Peterson, 2015). On a similar note, a recent EEG study reported that repeated, successful retrieval practice across multiple practice cycles resulted in reduced competition from related memories (other studied items or related knowledge from outside of the study context). The reduction of competition between retrieval attempts also predicted retrieval success of the item following a one-week delay (Raﬁdi et al., 2018). Our interpretation that repeated retrieval promotes a strong association between cues and target memories is in line with these results, as a direct access to the target information would result in a reduced competition from other items.

5. Summary and conclusions

Summarising the most important results of the present experiment, we found a change in the electrophysiological correlates of retrieval and post-retrieval processes across multiple consecutive memory tests. We suggest that with each retrieval attempt, new relations are created between the target memories and a variety of contextual cues. Also, as a result of repeated retrieval, it seems that there is a reduced need for extended monitoring/search processes, attentional control, and further evaluation. Our findings suggest that repeated retrieval supports automatic access to memories together with their contextual features.

In brief, based on a long line of previous studies, it seems that there are several factors that ensure the long-term efficiency of retrieval practice. Some theories of the testing effect phenomenon emphasise the importance of the contextual cues and the automatisation of retrieval (Karpicke et al., 2014; Racsmány et al., 2019). Our results support these theories and suggest that these factors both contribute to the beneficial impact of test-enhanced learning. Moreover, our findings also show that multiple retrieval trials involve different processes relative to the early stages of retrieval practice indicating the power of repeated retrieval practice.

Open Practices Statement

The experiment reported in this article was not formally preregistered.

Declaration of interest

The authors declare no conflict of interest.

Data Availability

Datasets related to this article are available at an open source data repository (Open Science Framework, OSF, https://osf.io/heqyn/).

Acknowledgement

This work was supported by the 2017-1.2.1-NKP2017-00002 Research Grant (Hungarian Brain Research Program, Hungary) and by the NKFI (National Research, Development and Innovation Office, Hungary) K124098 Research Grant. Ágnes Szőlősí was supported by the ÚNKP-20-4-II-BME-13 New National Excellence Program of the Ministry for Innovation and Technology, Hungary, from the source of the National Research, Development and Innovation Fund, Hungary. Kornél Németh was supported by the National Research, Development and Innovation Office, Hungary (PD18_128361) and ÚNKP-18-4 New National Excellence Program of the Ministry of Human Capacities, Hungary, and János Bolyai Research Scholarship of the Hungarian Academy of Sciences, Hungary (ÚNKP-Bolyai-i). We thank Anita Lencsés for her help in data collection. We thank Karina Cozma for her help in data analysis. We thank Balázs Knakker for the technical comments.

References

Addante, R. J., Ranganath, C., Olichney, J., & Yonelinas, A. P. (2012). Neurophysiological evidence for a recollection impairment in amnesia patients that leaves familiarity intact. *Neuropsychologia, 50*, 3004–3014. https://doi.org/10.1016/j.neuropsychologia.2012.07.038

Addante, R. J., Ranganath, C., & Yonelinas, A. P. (2012). Examining ERP correlates of recognition memory: Evidence of accurate source recognition without recollection. *Neuroimage, 62*, 439–450. https://doi.org/10.1016/j.neuroimage.2012.04.031

Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. *In S. Kornblum (Ed.), Attention and performance IV*, pp. 1–58. New York, NY: Academic Press.

Bai, C. H., Bridger, E. K., Zimmer, H. D., & Mecklinger, A. (2015). The beneficial effect of testing: An event-related potential study. *Frontiers in Behavioral Neuroscience, 9*, 248. https://doi.org/10.3389/fnbeh.2015.00248

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436. https://doi.org/10.1163/156856897X14007

Bridge, D. J., & Paller, K. A. (2012). Neural correlates of reactivation and retrieval-induced distortion. *Journal of Neuroscience, 32*, 12144–12151. https://doi.org/10.1523/JNEUROSCI.1378-12.2012

van den Broek, G., Takashima, A., Wiklund-Hansson, C., Wirebring, L. K., Segers, E., Verhoeven, L., & Nyberg, L. (2016). Neurocognitive mechanisms of the “testing effect”: A review. *Trends in Neuroscience and Education, 5*, 52–66. https://doi.org/10.1016/j.tine.2016.05.001

van den Broek, G. S. E., Takashima, A., Segers, E., Fernandez, G., & Verhoeven, L. (2013). Neural correlates of testing effects in vocabulary learning. *Neuroimage, 78*, 94–102. https://doi.org/10.1016/j.neuroimage.2013.03.071

Chartian, G. E., Lettich, E., & Nelson, P. L. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked EEG activities. *The American Journal of EEG Technology, 25*, 83–92. https://doi.org/10.1080/00029228.1985.11080163

Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language, 53*, 594–628. https://doi.org/10.1016/j.jml.2005.08.005

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review, 107*, 261–288. https://doi.org/10.1037/0033-295X.107.2.261

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dien, J. (2017). Best practices for repeated measures ANOVAs of ERP data: Reference, regional channels, and robust ANOVAs. *International Journal of Psychophysiology, 111*, 42–56. https://doi.org/10.1016/j.ijpsycho.2016.09.006

Düzel, E., Yonelinas, A. P., Mangun, G. R., Heizner, H. J., & Tulving, E. (1997). Event-related brain potential correlates of two states of conscious awareness in memory.
Liu, X. L., Tan, D. H., & Reder, L. M. (2018). The two processes underlying the testing effect—Evidence from Event-Related Potentials (ERPs). Neuropsychologia, 112, 77–85.

Liu, Y., Yao, M., Peng, Y., Lu, B., & Guo, C. (2019). Dissociating neural correlates of retrieval practice and elaborative practice in associative recognition memory. Neuronetwork, 30, 985–992. https://doi.org/10.1186/s13408-000000013-0

Liu, Y., Ji, Z., Li, B., Wang, C., Gao, C., and Guo, C. (2017). Differentiation of subsequent memory effects between retrieval practice and elaborative study. Biological psychology, 127, 134–147. https://doi.org/10.1016/j.biopsycho.2017.05.010

Logan, G. D. (1988a). Toward an instance theory of automation. Psychological Review, 95, 492–527. https://doi.org/10.1037/0033-295X.95.2.492

Logan, G. D. (1988b). Automaticity, resources, and memory: Theoretical controversies and practical implications. Human Factors, 30, 583–598. https://doi.org/10.1518/0018720887850504

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. Frontiers in Human Neuroscience, 8, 213. https://doi.org/10.3389/fnhum.2014.00211

Mahan, L., Sottili, A., & Racsmay, M. (2018). Anodal transcranial direct current stimulation of the right dorsolateral prefrontal cortex improves long-term retention of reencountered memories. Cortex, 108, 80–91. https://doi.org/10.1016/j.cortex.2018.07.012

Manel, M. A., Anderson, J. L., Derbis, M. H., & Morissette, N. (2007). Testing the testing effect in the classroom. European Journal of Cognitive Psychology, 19, 494–513. https://doi.org/10.1080/09541440701526154

Mecklinger, A., Rosburg, T., & Johansson, M. (2016). Reconstructing the past: The late posterior negativity (LPN) in episodic memory. Neuroscience and Biobehavioral Reviews, 68, 621–638. https://doi.org/10.1016/j.neubiorev.2016.06.024

Mogoson, A., Jovicich, J., Bruzzone, L., & Busoli, M. ORJ ADSJST: An automatic EEG artifact detector based on the use of spatial and temporal features. Psychophysiology, 48, 229–240. https://doi.org/10.1111/jep.12016

Moschov, M. (1995). Recovered consciousness: A hypothesis concerning modularity and episodic memory. Journal of Clinical and Experimental Neuropsychology, 17, 593–610. https://doi.org/10.1073/pnas.92.4.1963

Mulligan, N. W., & Peterson, D. J. (2015). Negative and positive testing effects in terms of item-specific and relational information. Journal of Experimental Psychology: Learning, Memory, and Cognition, 41, 859–871. https://doi.org/10.1037/xlm0000056

Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multirial learning of Swahili-English translation equivalents. Memory, 2, 325–335. https://doi.org/10.1080/096582104100017405

Palksay, P., Sottili, A., & Racsmay, M. (2019). Retrieval practice decreases processing load of recall: Evidence from pupillometry. International Journal of Psychophysiology, 143, 88–95. https://doi.org/10.1016/j.ijpsycho.2019.07.002

Paller, K. A., McCarthy, G., & Wood, C. C. (1988). ERPs predictive of subsequent recall and recognition performance. Biological Psychology, 26, 269–276. https://doi.org/10.1016/0301-0511(88)90023-8

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory: Toward an instance theory of automatization. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28, 966–968. https://doi.org/10.1037/0278-7393.28.5.966

Pastor, D., & Baud, K. H. T. (2016). Reversing the testing effect by feedback: Behavioral and electrophysiological evidence. Cognitive, Affective, and Behavioral Neuroscience, 16, 473–488. https://doi.org/10.3758/s13415-016-0407-6

Pastor, D., Schöck, S., Niederbrunner, J., & Baud, K. H. T. (2014). Retrieval during learning facilitates subsequent memory encoding. Journal of Experimental Psychology: Learning, Memory, and Cognition, 37, 287–297. https://doi.org/10.1037/xlm0000210

Putnam, A. L., Nestojko, J. F., & Roediger, H. L. (2016). Improving student learning: Two strategies that make it stick. In J. C. Halvord, J. M. Lodge, & J. Hattie (Eds.), From the laboratory to the classroom: Translating the science of learning for teachers (pp. 106–133). New York, NY: Routledge.

Racsmay, M., & Kerezets, A. (2015). Initial retrieval shields against retrieval-induced forgetting. Frontiers in Psychology, 6, 657–668. https://doi.org/10.3389/fpsyg.2015.00657

Racsmay, M., Sottili, A., & Bencze, D. (2018). Retrieval practice makes procedure from remembering: An automatization account of the testing effect. Journal of Experimental Psychology: Learning, Memory, and Cognition, 44, 157–166. https://doi.org/10.1037/xlm0000043

Racsmay, M., Sottili, A., & Marián, M. (2020). Reversing the testing effect by feedback is a matter of performance criterion at practice. Memory and Cognition, 48, 1161–1170. https://doi.org/10.3758/s13421-020-01041-5

Rafidi, N. S., Hulbert, J. C., Brooks, P. P., & Norman, K. A. (2018). Reductions in retrieval-induced forgetting: A search for a unifying property. Current Biology, 28, 992–999. https://doi.org/10.1016/j.cub.2018.02.047

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. Trends in Cognitive Sciences, 15, 20–27. https://doi.org/10.1016/j.tics.2010.09.007

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. Perspectives on Psychological Science, 1, 181–210. https://doi.org/10.1111/j.1745-6916.2006.00002.x

Roediger, H. L., & Marsh, E. J. (2012). Episodic and autobiographical memory. In J. M. Healy, R. W. Proctor, & I. B. Weiner (Eds.), A Handbook of Psychology, 4, Experimental Psychology (pp. 475–497). Hoboken: John Wiley & Sons Inc.

Rosburg, T., Johansson, M., Weigl, M., & Mecklinger, A. (2015). How does testing affect retrieval-related processes? An event-related potential (ERP) study on the short-term effects of repeated retrieval. Cognitive, Affective, and Behavioral Neuroscience, 15, 195–218. https://doi.org/10.3758/s13423-014-0130-y
