Retrieval practice decreases processing load of recall: Evidence revealed by pupillometry

Péter Pajkossy[1,2 *], Ágnes Szőllősi[1,2], Mihály Racsmány[1,2]

[1] Institute of Cognitive Neuroscience and Psychology, Hungarian Academy of Sciences,
Budapest, Hungary
(Postal address: 1111 Budapest, Hungary, Egry Jozsef utca 1.)
[2] Department of Cognitive Science, Budapest University of Technology and Economics,
Budapest, Hungary
(Postal address: 1519 Budapest, Hungary, Pf. 286)

[*] All correspondence concerning this article should be addressed to Péter Pajkossy. Postal address: Department of Cognitive Science, Budapest University of Technology and Economics, 1111-Budapest Hungary, Egry József u. 1. E-mail: ppajkossy@cogsci.bme.hu

The authors declare no conflict of interest.

Highlights

Processing load of retrieval was measured by task evoked dilation of the pupil.

Retrieval practice trials systematically decreased processing load of cued recall.

Recall of earlier retested items was less demanding, than recall of restudied items.

Reduced processing load of recall was evident even one week after repeated testing.

Abstract

Retrieval practice is an effective long-term learning strategy. Items practiced through repeated retrieval are resistant to interference, stress, and secondary load, which attributes also characterize automatization in skill learning. In two experiments, we investigated whether retrieval practice is associated with decrease in processing load, which is a further attribute of automatization. Participants first learned paired associates, and then they practiced the items either by repeatedly studying or engaging in retrieval practice. Then their memory was assessed after either five minutes (Experiment 1) or one week (Experiment 2). Processing load was measured by assessing task-evoked pupil dilation during both retrieval practice and later recall. The pattern of results was similar in both experiments. During retrieval practice, processing load decreased during consecutive practice cycles. Moreover, during the final recall test, the retrieval of previously retrieval practiced items required less processing resources, as compared to the retrieval of previously restudied items. Our results suggest that repeated retrieval reduces processing load as well as attentional control involvement during practice and later recall.

Keywords: testing effect, retrieval practice, pupil dilation, automatization, skill learning

1. Introduction

Retrieval practice is considered to be one of the most effective learning strategies (Karpicke & Roediger, 2007; Roediger & Butler, 2011), with findings showing that retrieval practice results in higher memory performance than other forms of learning (such as repeated study). This phenomenon is usually called the testing effect. Most studies found an interaction between the type of practice (retrieval practice vs. repeated study) and the length of delay (minutes vs. days). Whereas repeated study leads to better performance when final recall is preceded by a delay of few minutes, retrieval practice leads to better long-term retention (Thompson, Wenger, & Bartling, 1978; Wheeler, Ewers, & Buonanno, 2003; but see Karpicke, Lehman, & Aue, 2014).

Besides their differential impact on memory retention, memories following retrieval practice have a list of properties, which do not characterize memories following repeated study. Several studies show that repeated retrieval (a) renders information more resistant to interference effects (Racsmány & Keresztes, 2015; Szpunar, McDermott, & Roediger, 2008), (b) remains accessible in multitasking situations where attentional processes are heavily loaded (Mulligan & Picklesimer, 2016), (c) and is resistant against the adverse effect of acute stress (Smith, Floerke, & Thomas, 2016). Note that these attributes of retrieval-practiced memories typically characterize the process of automatization in skill learning (Kuhl, Dudukovic, Kahn, & Wagner, 2007; Newell & Rosenbloom, 1981; Schneider & Chein, 2003; Squire & Zola, 1996).

Accordingly, two recent functional magnetic resonance imaging (fMRI) studies found that items learned through retrieval practice were retrieved faster than repeatedly studied items both after a short delay of 20 minutes (Keresztes, Kaiser, Kovács, & Racsmány, 2014) and a long delay of seven days (Keresztes et al., 2014; van den Broek, Takashima, Segers, Fernández, &

Verhoeven, 2013). Along with the fast performance, the recall of items learned through retrieval practice produced increased activity in the striatal cortex and thalamus and also in the associative cortex during retrieval practice, and a decreased activity in the control network at final recall (for similar results, see Karlsson Wirebring, Wiklund-Hörnqvist, Eriksson, Andersson, Jonsson, & Nyberg, 2015; Wiklund-Hörnqvist, Andersson, Jonsson, & Nyberg, 2017). These findings also resemble brain activation patterns usually observed in skill learning studies (Raichle, Fiez, Videen, MacLeod, Pardo, Fox, & Petersen, 1994). In a recent study investigating reaction time of retrieval, Racsmány, Szőllősi, and Bencze (2018) found that the speed up of cued recall following retrieval practice aligned to a power function which is generally considered to be an important quantitative attribute of automatization in skill learning (Logan, 1988; Schneider & Chein, 2003). Furthermore, the measure of goodness of fit to the individual power functions was associated with long-term retention success.

Based on the similarities and prior findings summarized above, in this study we seek to investigate whether retrieval practice can be characterized by another feature of skill learning: the gradual decrease in processing requirements demanded by the execution of skilled behavior. Most models of controlled attention and working memory (Atkinson & Schiffrin, 1971; Baddeley, 2000; Cowan, 2005; Kahneman, 1973) suggest that there is a capacity limit of resources in information processing: only a limited pool of resources can be allocated to different tasks, and once this capacity is exceeded, performance begins to deteriorate. This effect could be demonstrated by the dual task-procedure, where multiple resource-demanding tasks are used to investigate the characteristics of controlled information processing system (e.g. to investigate whether different tasks tax the same or different resource pools, see e.g. Baddeley & Hitch, 1974). One important characteristic of skilled behavior is that performance is not sensitive to dual-task manipulations, because automatization decreases the resource demand of information processing (Schneider & Chein, 2003).

Whereas the dual task method only enables us to indirectly estimate processing requirements, pupillometry provides an online measure of processing load. As was first demonstrated by Hess and Polt (1964), and later observed for several domains of information processing (for a review, see Beatty, 1982), pupil dilates as a function of task difficulty. Based on these findings, Kahneman (1973) proposed that such task evoked pupil responses (TEPRs) can be used as an online measure of the intensive aspect of attention: it indexes the amount of processing resources which is required to perform a task[1]. Accordingly, TEPR is associated with several variables influencing the processing requirements of a task. The magnitude of TEPRs was shown to be related to memory load (Kahneman & Beatty, 1966; Unsworth & Robison, 2015), dual task manipulation (Karatekin, Couperus, & Marcus, 2004), and the complexity of syntactic processing (Just & Carpenter, 1993).

Recent research shed light on the neurobiological underpinnings of TEPRs: changes in pupil size reflect the activation of the brain stem nucleus locus coeruleus (LC) that innervates large parts of the cortex through noradrenergic projections (Aston-Jones & Cohen, 2005; Joshi, Li, Kalwani, & Gold, 2016; Murphy, O'Connell, O'Sullivan, Robertson, & Balsters, 2014). Burst-like firing of the LC is suggested to reset functional networks in the cortex (Bouret & Sara, 2005) or to coordinate cortical networks responsible for information processing by changing the neural gain in specific cortical networks (Aston-Jones & Cohen, 2005; Eldar, Niv, & Cohen, 2016). These phasic LC responses might correspond to TEPRs, thus pupil dilation might signal rapid changes in neural gain accompanying task-relevant processing (Ashton-Jones & Cohen, 2005; Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010). Relatedly, in an fMRI study, Alnæs, Sneve, Espeseth, van de Pavert, and Laeng (2014) showed that pupil size is correlated with the activity of cortical networks responsible for goal-driven, top-down attentional processes.

As summarized above, recent theories and empirical results confirm the conceptualization of pupil dilation, as a correlate of processing load. Therefore, we used pupillometry to test whether repeated testing leads to a decrease in processing load, similarly to skill learning. After an initial learning period, paired associates were practiced either by retrieval or by restudy practice (retested vs. restudied word pairs), and participants' memory was tested after either a short-term delay (five minutes, Experiment 1) or a long-term retention interval (one week, Experiment 2).

First, we predicted that processing load, measured by TEPRs, would decrease after repeated cycles of retrieval practice. Second, because decreased RT during the recall of previously retested items was shown for both short- and long-term delay (Keresztes et al., 2014), we expected that reduced processing load would remain over both short-term and long-term delays. That is, we predicted that the retrieval of previously retested items would be accompanied by lower TEPRs, than the retrieval of previously restudied items.

Importantly, during the practice phase, the comparison of restudy and retest conditions might be misleading, because the timing and specificity of cognitive processes are very different for the two practice strategies. Because of this, restudy practice served as a control condition only in the case of the final test. During the final test phase, the underlying cognitive processes are identical, as the task (cued recall) is the same for all items irrespective of whether they were practiced by retrieval or restudy.

2. Materials and methods

2.1. Participants

We recruited 34 participants in Experiment 1(15 women; M_{age} = 21.8 years, SD_{age} = 2.0), and 46 participants in Experiment 2 (32 women; M_{age} = 21.7 years, SD_{age} = 1.7). The sample size was higher in Experiment 2, because we expected higher exclusion rate in Experiment 2, due to lower recall level associated with long-term, than with short-term recall (see later).

All participants were Hungarian undergraduate students. Subjects received either money or extra course credits. Participants gave written informed consent. The work described has been carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans. The research project was approved by the United Ethical Review Committee for Research in Psychology, Hungary.

Due to data loss and tracking difficulties, we excluded 3-3 participants in both Experiment 1 and Experiment 2. In Experiment 2, two participants did not show up for the second session. Furthermore, we also excluded participants, who had less than three correct retrieval trials (less than a recall rate of 15%) in any cycle during retrieval practice or in any condition of the final test. For this reason, one participant in Experiment 1 and 8 participants in Experiment 2 were excluded. This was necessary, because computing peak dilations for conditions with only a few trials becomes unreliable due to noise inherent in each data series. This latter exclusion criterion increases the reported recall levels in our samples, but does not affect the pattern of results related to memory performance (e.g., testing effect is present both with or without participants with low memory performance).

After the exclusion procedure, the final sample size was n = 30 (14 women; M_{age} = 21.9 years, SD_{age} = 2.1) in Experiment 1, and n = 33 (22 women; M_{age} = 21.6 years, SD_{age} = 1.5) in Experiment 2.

2.2. Memory paradigm

Stimuli were 40 Swahili-Hungarian word pairs translated from Nelson and Dunlosky (1994). The Swahili and the Hungarian words were randomly paired for each participant. That is, for different participants, different Hungarian words were paired to the same Swahili word.

The memory task consisted of three phases: an initial learning phase, a practice phase, and a final test. The only one difference between the two experiments was that the final test was preceded by a five-minute delay in Experiment 1 (short-term delay), whereas there was a one-week retention interval prior to the final test in Experiment 2 (long-term delay).

The initial learning phase consisted of five consecutive cycles. In each cycle, participants were presented with the word pairs in random order (five sec/word pair, pre-stimulus interval [PSI]: 500 ms). Word pairs were presented in the center of the computer screen with the Swahili word on the left and the Hungarian word on the right. Before each learning cycle, the instruction to memorize the word pairs was presented on the computer screen. The relatively high number of initial learning cycles was necessary to achieve a high criterion level which is critical for retrieval practice experiments using no feedback during retrieval practice (Karpicke et al., 2014; Smith, Roediger, & Karpicke, 2013).

The initial learning phase was followed by the practice phase with a five-minute delay between them. During this delay, participants were given arithmetic (distractor) tasks.

The practice phase consisted of five consecutive cycles. Each practice cycle consisted of a restudy practice and a retrieval practice block. The order of the restudy and retrieval practice blocks varied randomly across the practice cycles. That is, for half of the participants, two cycles began with restudy practice and three cycles with retrieval practice (the order of these blocks was counterbalanced across the participants). For the remaining participants, two cycles began with retrieval practice and three cycles with restudy practice.

In the restudy practice blocks, participants were presented with half of the material (20 word pairs) in random order. Circumstances of the restudy practice blocks were identical to those in the initial learning phase except for a five-second PSI in the restudy practice blocks. In the retrieval practice blocks, the remaining 20 Swahili words were presented as cues in random order (five sec/word pair, PSI: five sec). Subjects were asked to say out loud the Hungarian equivalents of the Swahili words. The experimenter recorded participants’ responses.

We used a relatively long (five sec) PSI during both restudy and retrieval practice allowing the pupil to return to its baseline level after the previous trial. We refer to this period between two stimuli as a baseline period. To decrease different sources of noise in pupil data, we equated luminance features during baseline periods and stimulus presentation: a sequence of nonsense character was shown on the screen during the baseline period at the same location, where cues and targets were presented during stimulus presentation.

The practice phase was followed by either a delay of five minutes (while participants were given arithmetic tasks) in Experiment 1, or a longer retention interval of one week in Experiment 2. In the final test, each word pair was tested only once. Circumstances of the final test were identical to those in the retrieval practice blocks of the practice phase.

Participants were seated in front of the screen and the remote eye-tracer, no chin-rest was used. Furthermore, to avoid biases in pupil size measurement caused by the gaze point of the participant (pupil foreshortening error, see e.g., Hayes & Petrov, 2016), we instructed the participants to maintain their fixation during the experimental trials on a relatively narrow rectangle presented on the screen.

2.3. Data processing and statistical analysis

2.3.1. Behavioral data

Memory performance was assessed by computing the percentage of words recalled correctly in each retrieval practice cycle and in the two conditions of the final test.

2.3.2. Pupil size data: preprocessing

Pupil size was measured using a SMI RED500 remote eye-tracking system (SensoMotoric Instruments, Teltow, Germany). Data recording frequency was 120 Hz and pupil size was measured in millimeters.

During data processing, we first removed data points with zero value (blinks and missing data). Then we removed each data point which exceeded the mean pupil size value of the specific trial by 3 SD. Then these missing data points were replaced by linear interpolation. As a measure of data quality, we computed the percentage of interpolated data points for each participant. This value exceeded by none of the participants 40%, which was defined as exclusion criterion following others (Kang, Huffer, & Wheatley, 2014; Smallwood et al., 2011). The mean percentage of interpolated data points was 14.7% (*SD* = 3.3) during the retrieval practice trials and 14.6% (*SD* = 2.6) during the final test trials.

2.3.3. Pupil dilation during retrieval practice and final test

We averaged data for each time point during a trial (baseline period and cue presentation), for each participant and for each condition, to get an averaged time-series of the pupil data for each subject and condition. Initial examination of these curves revealed that the presentation of the cue word is accompanied by a sudden increase of the pupil size (similarly to other demonstrations of TEPRs, see e.g. Beatty, 1982). This is also visible in Figure 1, which shows the grand-average curve of participants' pupil data during the practice cycles computed from individual curves.

To compute characteristic measures of pupil dilation, at first, all data points during retrieval were baseline corrected by subtracting from all data points the mean pupil size value of the 500 ms preceding the onset of stimulus presentation. Then, using these curves, for each subject and for each condition, we specified the maximum increase of pupil size in the period of 4500 ms following the onset of the retrieval trial. This was defined as peak dilation. Figure 1A, 2A-B and 3A present baseline corrected grand-average curves, averaged across participants, for each condition separately.

2.3.4. Statistical analysis

When analyzing data of the practice phase, for recall success and also for peak dilation we conducted repeated measures ANOVAs with five levels (Cycle 1-5). These measures were computed by involving only correct retrieval trials. First order polynomial contrasts were used to test whether there is a linear trend in change of recall success and peak dilation.

To analyze differences in the final test phase between the retrieval practice and restudy conditions, paired-samples *t*-tests were used for both recall percentage and peak dilation. For each ANOVA, Greenhouse-Geisser correction was used, when it was necessary.

3. Results

3.1. Experiment 1: short-term delay

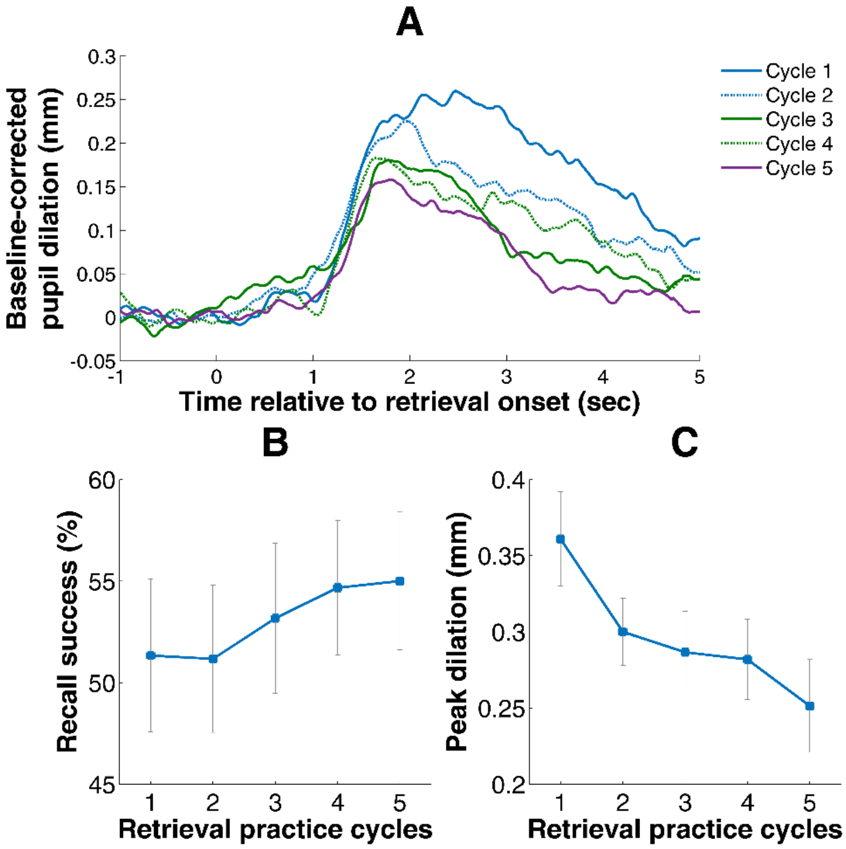
Figure 1A shows grand average pupil dilation curves for the five retrieval practice cycles, whereas Figure 1B-C depicts the change in recall percentage and peak dilation during retrieval practice. We found a gradual increase in recall rate during the retrieval practice phase, main effect of Cycle: $F(4, 116) = 5.11$, $p < .01$ (after Greenhouse-Geisser correction, epsilon = 0.79),

η_p^2 =.15, linear trend: $F(1, 29) = 13.79$, $p < .001$, $\eta_p^2 = .32$. The increase in memory performance was accompanied by a gradual decrease in the size of the pupillary response (see Figure 1A and C). Peak dilation decreased linearly as a function of practice cycles, main effect of cycle: $F(4, 116) = 3.64$, $p < .01$, $\eta_p^2 = .11$, linear trend: $F(1, 29) = 11.79$, $p < .01$, $\eta_p^2 = .29$.

Final test performance after the five-minute delay is depicted in Figure 2. We found superior memory for the restudied items, $t(29) = 9.24$, $p < .001$, $d = 1.79$. Furthermore, the retrieval of previously retested items were associated with smaller TEPRs, than the retrieval of previously restudies items, peak dilation: $t(29) = 2.26$, $p < .05$, $d = 0.39$.

Figure 1.

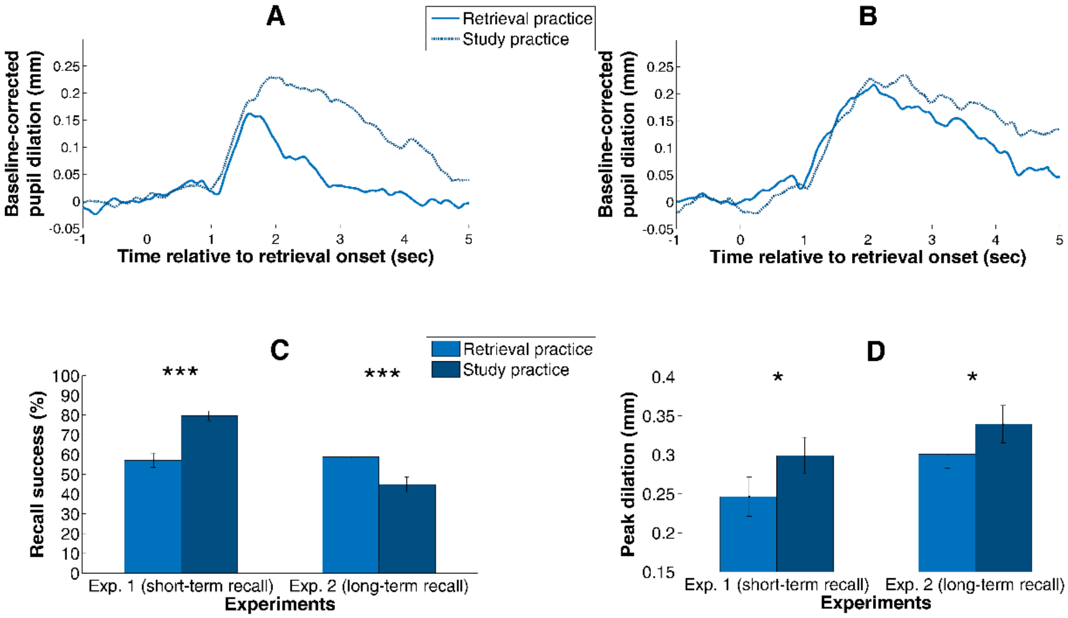
Experiment 1 – Recall Success and Task-Evoked Pupil Responses during Retrieval Practice



Notes. A) Task-evoked pupillary responses in correct retrieval trials of the five practice cycles. Baseline corrected values are presented: the mean of the 500 msec preceding trial onset is subtracted from each data point. B) Mean recall percentage values in the five retrieval practice cycles. C) The maximum increase of pupil diameter during retrieval (peak dilation), as a function of practice cycles. Peak dilation values are computed using only data from correctly retrieved trials. Note that grand average values from Panel A) do not fully correspond to the average peak dilation values in Panel C) because peak dilations occurred at different time points for different trials. (Error bars represent the standard error of the mean.)

Figure 2

Recall Success and Task-Evoked Pupil Responses During the Final Test in Experiment 1-2.



Notes. A-B) Task-evoked pupillary responses during retrieving previously retested and previously restudied words, on the final test, either after short-term delay in Experiment 1 (A) or after long-term delay, Experiment 2 (B). Baseline corrected values are presented: the mean of the 500 msec preceding trial onset is subtracted from each data point. C) Recall rates for correctly recalled words, in Experiment 1 (short-term delay) and Experiment 2 (long-term delay). D) The maximum increase of pupil diameter during retrieval (peak dilation), for the previously retrieved and restudied words, in Experiment 1 (short-term delay) and Experiment 2 (long-term delay). Note that the grand average values of Panel A-B) do not fully correspond to the average peak dilation values of Panel C-D), because peak dilations occurred at different time points for different trials. (* $p < .05$; *** $p < .001$. Error bars represent the standard error of the mean.)

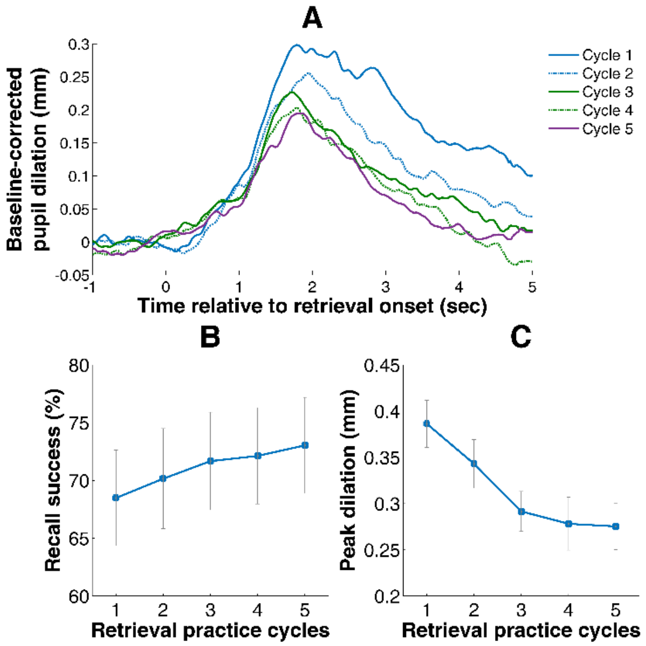
3.2. Experiment 2: long-term delay

In Experiment 2, we replicated the pattern of results found in the practice phase of Experiment 1 (see Figure 3A-C). During repeated retrieval cycles, we found an increase in mean recall percentage, main effect of cycle: $F(4, 128) = 6.59$, $p < .001$ (after Greenhouse-Geisser correction, epsilon = 0.64), $\eta_p^2 = .17$, linear trend: $F(1, 32) = 12.86$, $p < .01$, $\eta_p^2 = .29$, and a decrease in peak pupil dilation, main effect of cycle: $F(4, 128) = 8.31$, $p < .001$, $\eta_p^2 = .21$, linear trend: $F(1, 32) = 32.12$, $p < .001$, $\eta_p^2 = .50$.

Contrary to Experiment 1, on the final test (after a one week delay), we found superior memory for the previously retested items, $t(32) = 4.86$, $p < .001$, $d = 0.77$. Importantly, the pattern of TEPRs, however, was similar to Experiment 1: peak dilation was smaller for the previously retested items than for the restudied items, $t(32) = 2.10$, $p < .05$, $d = 0.38$ (see also Figure 2).

Figure 3.

Experiment 2: Recall Success and Task-Evoked Pupil Responses during Retrieval Practice



Notes. A) Task-evoked pupillary responses, averaged for correct retrieval trials, separately for the five cycles. Baseline corrected values are presented: the mean of the 500 msec preceding trial onset is subtracted from each data point. B) Mean recall percentage values in the five retrieval practice cycles. C) The maximum increase of pupil diameter during retrieval (peak dilation), as a function of practice cycles. Peak dilation values are computed using only data from correctly retrieved trials. Note that grand average values from Panel A) do not fully correspond to the average peak dilation values in Panel C) because peak dilations occurred at different time points for different trials. (Error bars represent the standard error of the mean.)

4. Discussion

In two studies, we aimed to test how retrieval practice affects processing load of retrieval, as measured by pupil dilation. First, we found that repeated retrieval cycles were accompanied by a decrease in pupil dilation. Second, we showed that the retrieval of previously retested materials evokes smaller pupil dilation than the retrieval of previously restudied information. This pattern of results suggests that repeated retrieval leads to a decrease in processing load, as pupil dilation can be regarded as an online measure of processing load (Kahneman, 1973) or controlled attention (Alnæs et al., 2014; Unsworth & Robison, 2015).

Importantly, the behavioral measures of memory performance were partly independent from pupillometric data. The pattern of behavioral results was that one could expect inspecting the vast literature of the testing effect (see Karpicke & Roediger, 2007). Specifically, after a short-term delay, a superior recall of restudied items was shown, whereas after a long-term delay, a superior recall of retested items was evident. Importantly, after both short- and long-term delays, the retrieval of previously retested items was accompanied by lower TEPRs in comparison with the retrieval of previously restudied items. This pattern of data suggest that the retrieval of retested items requires less processing resources, but this is not related to retrieval success. This decoupling can be explained if we acknowledge that the amount of processing resources required for a task is not necessarily related to performance in a task (e.g. dual-task manipulations do not necessarily affect accuracy measures heavily, see e.g. Baddeley & Hitch, 1974; Baddeley, Lewis, Eldridge, & Thomson, 1984; Karatekin et al., 2004). If no distraction or secondary task is present and the capacity is enough to perform the task, then performance can be independent of processing load. Specifically, retrieval practice makes the search for target items more automatic, but this does not guarantee that the targets will be found. The decoupling of behavioral measures and TEPRs is similar to the results of Keresztes et al. (2014) who found that retested items were retrieved quicker after both short-term and long-term delay, however, response speed was independent of average recall performance.

Our findings support the assumption that retrieval promotes the retention of practiced memories through automatized reactivation. Retrieval-based learning has a range of attributes that also characterizes automatization process during skill learning. Memories learned through retrieval practice show low-level of forgetting (Karpicke & Roediger, 2007; Roediger & Butler, 2011) and are resistant to the disturbing effects of interference and acute stress (Racsmány & Keresztes, 2015; Smith et al., 2016), just like skills (Schneider & Chein, 2003). Theories of skill learning assume that a kind of automatization process characterizes skill acquisition (Logan, 1988; Newell & Rosenbloom, 1981; Schneider & Shiffrin, 1977). In a recent study, Racsmány et al. (2018) showed that reaction time of retrieval during retrieval practice followed a power function speed up, typically observed in skill learning processes. The present study reveals another similarity between retrieval-based learning and skill acquisition: both are characterized by a reduction in processing load related to retrieval. In other words, the results of both studies suggest that retrieval practice decreases the involvement of attentional control and increases the level of automatization of cued recall.

Interestingly, our results are in line with a recent study by van Rijn, Dalenberg, Borst, and Sprenger (2012). The authors manipulated the load of memory in a short-term memory task and similarly to our results, they also found decreased TEPRs during repeated practice of paired-associates. It was concluded that repeated retrieval led to higher memory strength and this was reflected in decreased pupil dilation. In our interpretation, decreased TEPRs do not reflect changes in memory strength, but changes in the process of retrieval (i.e. automatization).

It is important to highlight that the pattern of our results is best explained in terms of changes in processing load. The pupil size responses we measured were tightly locked to stimulus presentation, which excludes the possibility, that the demonstrated effects are influenced by more general emotional or arousal driven changes (e.g., luminance, arousability or emotional significance of stimuli, see e.g. Beatty & Lucero-Wagoner, 2000; Bradley, Miccoli, Escrig, & Lang, 2008). For TEPR decrease during the practice phase, additional confounding factors should be considered. During retrieval practice, stimulus familiarity and general arousal levels changed simultaneously with the level of practice, thus these factors might also explain the decrease in TEPRs during repeated retrieval practice. Note however, that these factors are not specific to retested items (i.e. participants repeatedly encounter the restudied word pairs during the practice), and so do not explain the differences in TEPRs during the final recall. In particular, if stimulus familiarity or participant fatigue would have caused the decrease of TEPRs during retrieval practice in Experiment 1, then this should have exerted the same effect on the recall of both previously retested and restudied items. Therefore, we suggest that the pattern of results in the two experiments can be best conceptualized as changes in processing load.

5. Conclusion

In sum, our results suggest that repeated retrieval practice decreases retrieval-related processing load, as assessed by pupil dilation. It was also found that pupil dilation during final recall was significantly higher for items that were practiced through repeated study both after a five-minute and a one-week delay. These results support the assumption that retrieval practice promotes long-term retention of practiced information through diminishing involvement of attentional control in retrieval and preserves long-term learning through automatized processing of specific cue-target associations.

Acknowledgements

This work was supported by the 2017-1.2.1-NKP-2017-00002 Research Grant (National Brain Research Program, Hungary) and the NKFI K124094 Research Grant. Péter Pajkossy was supported by the Bolyai János Research Scholarship of the Hungarian Academy of Sciences.

References

Alnæs, D., Sneve, M.H., Espeseth, T., Endestad, T., van de Pavert, S.H.P., Laeng, B., 2014. Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. J. Vision 14, 1-20. http://dx.doi.org/10.1167/14.4.1

Aston-Jones, G., Cohen, J.D., 2005. An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. Annu. Rev. Neurosci. 28, 403-450. http://dx.doi.org/10.1146/annurev.neuro.28.061604.135709

Atkinson, R.C., Shiffrin, R.M., 1971. The control of short-term memory. Sci. Am. 225, 82-90. http://dx.doi.org/10.1038/scientificamerican0871-82

Baddeley, A.D., 2000. The episodic buffer: A new component of working memory? Trends Cogn. Sci. 4, 417-423. http://dx.doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A.D., Hitch, G., 1974. Working memory. Psychol. Learn. Motiv. 8, 47-89. http://dx.doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A.D., Lewis, V., Eldridge, M., Thomson, N., 1984. Attention and retrieval from long-term memory. J. Exp. Psychol. General 113, 518-540. http://dx.doi.org/10.1037/0096-3445.113.4.518

Beatty, J., 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychol. Bull. 91, 276-292. http://dx.doi.org/10.1037/0033-2909.91.2.276

Beatty, J., Lucero-Wagoner, B., 2000. The pupillary system. In: Cacioppo, J.T., Tassinary, L.G., Berntson, G.G. (Eds.), Handbook of psychophysiology, 2nd Edition, Cambridge University Press, Cambridge, pp. 142-162.

Bjork, E.L., Bjork, R.A., 2011. Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In: Gernsbacher, M.A., Pew, R.W., Hough, L.M., Pomerantz, J.R. (Eds.), Psychology and the real world: Essays illustrating fundamental contributions to society, Worth Publishers, New York, pp. 56-64.

Bjork, R.A., 1994a. Memory and metamemory considerations in the training of human beings. In: Metcalfe, J., Shimamura, A. (Eds.), Metacognition: Knowing about knowing, MIT Press, Cambridge, pp. 185-205.

Bjork, R.A., 1994b. Institutional impediments to effective training. In: Druckman, D., Bjork, R.A (Eds.), Learning, remembering, believing: Enhancing human performance, National Academy Press, Washington, pp. 295-306.

Bouret, S., Sara, S.J., 2005. Network reset: A simplified overarching theory of locus coeruleus noradrenaline function. Trends Neurosci. 28, 574-582. http://dx.doi.org/10.1016/j.tins.2005.09.002

Bradley, M.M., Miccoli, L., Escrig, M.A., Lang, P.J., 2008. The pupil as a measure of emotional arousal and autonomic activation. Psychophysiology 45, 602-607. http://dx.doi.org/10.1111/j.1469-8986.2008.00654.x

Cowan, N., 2005. Working memory capacity, Psychology Press, Hove.

Eldar, E., Niv, Y., Cohen, J.D., 2016. Do you see the forest or the tree? Neural gain and breadth versus focus in perceptual processing. Psychol. Sci. 27, 1632-1643. http://dx.doi.org/10.1177/0956797616665578

Gilzenrat, M.S., Nieuwenhuis, S., Jepma, M., Cohen, J.D., 2010. Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. Cogn. Affect. Behav. Neurosci. 10, 252-269. http://dx.doi.org/10.3758/CABN.10.2.252

Hayes, T.R., Petrov, A.A., 2016. Mapping and correcting the influence of gaze position on pupil size measurements. Behav. Res. Met. 48, 510-527. http://dx.doi.org/10.3758/s13428-015-0588-x

Hess, E.H., Polt, J.M., 1964. Pupil size in relation to mental activity during simple problem solving. Science 143, 1190-1192. http://dx.doi.org/10.1126/science.143.3611.1190

Joshi, S., Li, Y., Kalwani, R.M., Gold, J.I., 2016. Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. Neuron 89, 221-234. http://dx.doi.org/10.1016/j.neuron.2015.11.028

Just, M.A., Carpenter, P.A., 1993. The intensity dimension of thought: Pupillometric indices of sentence processing. Can. J. Exp. Psychol. 47, 310-339. http://dx.doi.org/10.1016/10.1037/h0078820

Kahneman, D., 1973. Attention and effort, Prentice Hall, Engelwood Cliffs.

Kahneman, D., Beatty, J., 1966. Pupil diameter and load on memory. Science 154, 1583-1585. http://dx.doi.org/10.1126/science.154.3756.1583

Kang, O.E., Huffer, K.E., Wheatley, T.P., 2014. Pupil dilation dynamics track attention to high-level information. PLoS One 9, e102463. http://dx.doi.org/10.1371/journal.pone.0102463

Karatekin, C., Couperus, J.W., Marcus, D.J., 2004. Attention allocation in the dual-task paradigm as measured through behavioral and psychophysiological responses. Psychophysiology 41, 175-185. http://dx.doi.org/10.1111/j.1469-8986.2004.00147.x

Karlsson Wirebring, L., Wiklund-Hörnqvist, C., Eriksson, J., Andersson, M., Jonsson, B., Nyberg, L., 2015. Lesser neural pattern similarity across repeated tests is associated with better long-term memory retention. J. Neurosci. 35, 9595-9602. http://dx.doi.org/10.1523/JNEUROSCI.3550-14.2015

Karpicke, J.D., Lehman, M., Aue, W.R., 2014. Retrieval-based learning: An episodic context account. Psychol. Learn. Motiv. 61, 237-284. http://dx.doi.org/10.1016/B978-0-12-800283-4.00007

Karpicke, J.D., Roediger, H.L., 2007. Repeated retrieval during learning is the key to long-term retention. J. Mem. Lang. 57, 151-162. http://dx.doi.org/10.1016/j.jml.2006.09.004

Keresztes, A., Kaiser, D., Kovács, G., Racsmány, M., 2014. Testing promotes long-term learning via stabilizing activation patterns in a large network of brain areas. Cereb. Cortex 24, 3025-3035. http://dx.doi.org/10.1093/cercor/bht158

Kuhl, B.A., Dudukovic, N.M., Kahn, I., Wagner, A. D., 2007. Decreased demands on cognitive control reveal the neural processing benefits of forgetting. Nat. Neurosci. 10, 908-914. http://dx.doi.org/10.1038/nn1918

Logan, G.D., 1988. Toward an instance theory of automatization. Psychol. Rev. 95, 492-527. http://dx.doi.org/10.1037/0033-295X.95.4.492

Mulligan, N.W., Picklesimer, M., 2016. Attention and the testing effect. J. Exp. Psychol. Learn. Mem. Cog. 42, 938-950. http://dx.doi.org/10.1037/xlm0000227

Murphy, P.R., O'Connell, R.G., O'Sullivan, M., Robertson, I.H., Balsters, J.H., 2014. Pupil diameter covaries with BOLD activity in human locus coeruleus. Hum. Brain Mapp. 35, 4140-4154. http://dx.doi.org/10.1002/hbm.22466

Nelson, T.O., Dunlosky, J., 1994. Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. Memory 2, 325-335. http://dx.doi.org/10.1080/09658219408258951

Newell, A., Rosenbloom, P.S., 1981. Mechanisms of skill acquisition and the law of practice. In: Anderson, J.R. (Ed.), Cognitive skills and their acquisition, Lawrence Erlbaum Associates, Hillsdale, pp. 1-55.

Racsmány, M., Keresztes, A., 2015. Initial retrieval shields against retrieval-induced forgetting. Front. Psychol. 6, 657. http://dx.doi.org/10.3389/fpsyg.2015.00657

Racsmány, M., Szőllősi, Á., Bencze, D., 2018. Retrieval practice makes procedure from remembering: An automatization account of the testing effect. J. Exp. Psychol. Learn. Mem. Cog. 44, 157-166. http://dx.doi.org/10.1037/xlm0000423

Raichle, M.E., Fiez, J.A., Videen, T.O., MacLeod, A.M., Pardo, J.V., Fox, P.T., Petersen, S.E., 1994. Practice-related changes in human brain functional anatomy during nonmotor learning. Cereb. Cortex 4, 8-26. http://dx.doi.org/0.1093/cercor/4.1.8

Roediger, H., Butler, A.C., 2011. The critical role of retrieval practice in long-term retention. Trends Cogn. Sci. 15, 20-27. http://dx.doi.org/10.1016/j.tics.2010.09.003

Schneider, W., Chein, J.M., 2003. Controlled and automatic processing: Behavior, theory, and biological mechanisms. Cogn. Sci. 27, 525-559. http://dx.doi.org/10.1016/S0364-0213(03)00011-9

Schneider, W., Shiffrin, R.M., 1977. Controlled and automatic human information processing. II. Perceptual learning, automatic attending and a general theory. Psychol. Rev. 84, 1-66. http://dx.doi.org/10.1037/0033-295X.84.2.127

Smallwood, J., Brown, K.S., Tipper, C., Giesbrecht, B., Franklin, M.S., Mrazek, M.D., ... Schooler, J.W., 2011. Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. PloS One 6, e18298. http://dx.doi.org/10.1371/journal.pone.0018298

Smith, A.M., Floerke, V.A., Thomas, A.K., 2016. Retrieval practice protects memory against acute stress. Science 354, 1046-1048. http://dx.doi.org/10.1126/science.aah5067

Smith, M.A., Roediger, H.L., Karpicke, J.D., 2013. Covert retrieval practice benefits retention as much as overt retrieval practice. J. Exp. Psychol. Learn. Mem. Cog. 39, 1712-1725. http://dx.doi.org/10.1037/a0033569

Squire, L.R., Zola, S.M., 1996. Structure and function of declarative and nondeclarative memory systems. Proc. Nat. Acad. Sci. 93, 13515-13522. http://dx.doi.org/10.1073/pnas.93.24.13515

Szpunar, K.K., McDermott, K.B., Roediger, H.L., 2008. Testing during study insulates against the buildup of proactive interference. J. Exp. Psychol. Learn. Mem. Cog. 34, 1392-1399. http://dx.doi.org/10.1037/a0013082

Thompson, C.P., Wenger, S.K., Bartling, C.A., 1978. How recall facilitates subsequent recall: A reappraisal. J. Exp. Psychol. Hum. Learn. Mem. 4, 210 -221. http://dx.doi.org/10.1037/0278-7393.4.3.210

Tyler, S.W., Hertel, P.T., McCallum, M.C., Ellis, H.C., 1979. Cognitive effort and memory. J. Exp. Psychol. Learn. Mem. Cog. 5, 607-617. http://dx.doi.org/10.1037/0278-7393.5.6.607

Unsworth, N., Robison, M.K., 2015. Individual differences in the allocation of attention to items in working memory: Evidence from pupillometry. Psychon. Bull. Rev. 22, 757-765. http://dx.doi.org/10.3758/s13423-014-0747-6.

van den Broek, G.S., Takashima, A., Segers, E., Fernández, G., Verhoeven, L., 2013. Neural correlates of testing effects in vocabulary learning. NeuroImage 78, 94-102. http://dx.doi.org/10.1016/j.neuroimage.2013.03.071

Van Rijn, H., Dalenberg, J., Borst, J., Sprenger, S.A., 2012. Pupil dilation co-varies with memory strength of individual traces in a delayed response paired-associate task. PLoS One 7, e51134. http://dx.doi.org/10.1371/journal.pone.0051134

Wiklund-Hörnqvist, C., Andersson, M., Jonsson, B., Nyberg, L., 2017. Neural activations associated with feedback and retrieval success. Sci. Learn. 2, e12. http://dx.doi.org/10.1038/s41539-017-0013-6

Wheeler, M.A., Ewers, M., Buonanno, J.F., 2003. Different rates of forgetting following study versus test trials. Memory 11, 571-580. http://dx.doi.org/10.1080/09658210244000414

Footnote

[1] Kahneman (1973) referred to this construct as mental effort, but defined it as the amount of processing resources required for a task (see e.g. p. 15). In this manuscript, we use the term processing resources, as the concept of effort has other conceptualizations in memory research (e.g. motivational aspects or desirable difficulties, see e.g. Bjork, 1994a, 1994b; Bjork & Bjork, 2011; Tyler, Hertel, McCallum, & Ellis, 1979), and by this terminological choice we aim to emphasize that we investigate how processing load changes during retrieval practice.