scientific reports

OPEN

Predicting long-term memory via pupillometry

Oria Pitem & Yaniv Mama✉

Pupillometry research has established that pupil size reflects cognitive processes through autonomic nervous system activity, with high arousal triggering pupil dilation. Studies examining pupil size during encoding have yielded conflicting results regarding its relationship with subsequent memory performance, and few have investigated baseline pupil size. This study examined whether pupil diameter before and during stimulus presentation predicts memory performance. We hypothesized that successfully recalled words would be associated with larger pupils than forgotten words, based on the role of arousal and attention in memory formation. To test these hypotheses, we conducted two experiments in which we tracked ninety-five psychology students' eyes while they performed a long-term memory test. The results depict larger pupil size while studying later successfully retrieved words. Interestingly, this phenomenon also occurs before word presentation (during baseline), which supports the "readiness to remember" (R2R) framework. This implies that pupillary changes while preparing to encode information can indicate later memory performance.

Keywords Pupillometry, Attention, Long-term memory, Arousal, Encoding

The idea of pupil size reflecting cognitive processes was first demonstrated in the pioneering research of Hess and Polt[1]. They noticed a variation in pupil size induced by the emotional valence of the stimuli presented; an increase when watching "positive" images, and a constriction for "negative" images. Beatty and Kahneman[2] observed a variation in peak pupillary diameter depending on the number of digits the participants were asked to recall. They explained that higher cognitive effort or arousal due to task demands caused greater pupil dilation. This claim was supported by Marshall[3], who introduced pupil diameter fluctuation as an instantaneous Index of Cognitive Activity (ICA). The relationship between cognitive capabilities and pupil size stems from the activity of the autonomic nervous system; in high-level arousal, the pupils dilate due to intense secretion of norepinephrine (NE) from the Locus Coeruleus (LC) via sympathetic pathways[4,5]. Dopamine is released in arousing states such as motivation, reward, or novelty. It exerts an excitatory action on the wakefulness-promoting neurons of the LC, resulting in increased alertness[6,7]. Furthermore, the Adaptive Gain Theory[8] suggests that baseline (tonic) pupil size correlates with LC neuron firing rate, as well as with success in attentional tasks such as detection, decision-making, discrimination, and also learning rates in uncertainty conditions[9]. This is based on the inverted U-shaped curve relationship between tonic LC activity (arousal) and performance; at low levels, people are unfocused and perform poorly. At intermediate levels, attention is optimal, and performance peaks. At high levels, people become overstimulated and distracted, causing performance to decline[8,10,11]. Interestingly, Micula et al.[12] revealed a U-shaped relationship between pupil diameter and memory recall performance, wherein pupil dilation initially corresponded with enhanced memory function up to an optimal threshold, beyond which further dilation was associated with diminished recall abilities. Their theoretical framework suggests this pattern emerges from the allocation of cognitive resources, where excessive attentional resources devoted to stimulus identification may reduce the capacity available for memory encoding. Supporting these findings, numerous studies have provided convergent evidence that larger pupil size is associated with successful information retrieval[13–18].

Other researchers examined how pupillary responses during the study phase relate to subsequent memory. Some research has demonstrated that an increase in pupil size while studying is associated with better subsequent memory[12,19–22]. However, other studies have found that decreased pupil size predicts successful memory formation[23–25]. The apparent relationship between pupil size and memory performance exhibits complex dynamics that can be attributed to various experimental conditions, including the distinction between novel and anticipated target detection, as well as different memory retrieval processes such as recollection-based and familiarity-based recognition[6].

The issue regarding long-term memory prediction via pupil size remains open, as the question arises whether it is possible to observe the variation in pupil size for recalled vs. forgotten items even before stimulus onset. Nassar et al.[9] found that expectations of frequent changes in a dynamic environment are related to larger baseline

Department of Psychology, Ariel University, Ariel, Israel. ✉email: Yanivma@ariel.ac.il

pupil diameter, which could be related to expectations of novelty[6]. Cronin et al.[23] found that smaller baseline pupil size during encoding predicted better recognition accuracy, though they did not examine this relationship for recalled versus forgotten items or across different sound conditions. Several other studies examined baseline pupil size and task performance but found no significant effect[24,26].

Cohen et al.[27] investigated the brain activity in regions linked to encoding new information before stimulus onset during the study phase. They found that activity in the cingulo-opercular network seconds before an event begins can predict whether this event will subsequently be remembered. They also recorded pupil size but found no difference between conditions. However, the experiment's environment (inside an MRI scanner) may have had a significant effect. The loud noises and claustrophobic environment of MRI scanners can affect stress levels[28], potentially confounding task-evoked pupillary responses[29] and therefore impacting the reliability of pupillometry studies in MRI scanners.

Despite previous evidence, the underlying mechanisms of preparatory processes that precede long-term memory remain unclear. Madore and Wagner[30] recently presented the readiness to remember (R2R) framework, which aims to uncover what affects a person's ability to retain information in one moment and in another to forget. According to R2R, the initial state of attention partially explains the ability to recollect information, mediated by the strength of encoding. In addition, there are differences between individuals in their ability to be attentionally prepared. Therefore, the physiological state of readiness to remember affects task performance and changes from moment to moment according to the level of attention. The tendency to use the average baseline pupil size is very common in cognitive research, assuming it represents noise or a control condition to which the pupil size should be compared[31]. However, it is known that pupil size, similar to our attention, fluctuates from moment to moment[21,32]. It may be assumed that the momentary arousal level affects attention, which can be reflected in pupil size, even before the study phase begins. Therefore, in this study, we aspired to understand whether the pre-stimulus pupil size during the studying phase can indicate future memory performance. First, based on previous findings, we hypothesized that while studying successfully recalled words, they would be characterized by a larger pupil diameter than forgotten words. This is based on the finding that pupils dilate in response to NE secretion, indicating a heightened state of attention, which might contribute to long-term memory performance. Second, this enhancement in absolute pupil size for later recalled words will also be valid during baseline periods preceding stimulus presentation in the study phase. We assume this is due to a physiological "readiness" for remembering, similar to arousal or paying attention. Since long-term memory performance can be examined in various ways, in the current work, we used two tasks (recognition and free recall) in two experiments to test the hypotheses detailed above. We expect the results to be consistent in both experiments; in Experiment 1, we will observe pupil size while studying for a free recall test, and in Experiment 2, we will do the same in a recognition (old vs. new) test.

Experiment 1 results

Item-by-item analysis

This analysis was performed to verify that the observed variations in pupil size are due to arousal, indicating later memory performance, rather than being confounded by a specific word's effect (stimulus-evoked responses). Therefore, we conducted a repeated measures ANOVA on baseline-corrected pupil size values, revealing no statistically significant difference between all 30 words ($F(4.61, 36.87) = 1.51$, $p = .214$) in the study phase. Therefore, pupil size was not affected by the semantic meaning of a particular word.

Comparison between recalled vs. forgotten trials during the study phase

The mean accuracy rate across participants was 36.82% (SD = 14.12%). The pupil size was larger while studying later recalled trials than forgotten trials (BF10 = 86.58, Cohen's d = 0.61). This is also true during baseline; pupil size in recalled trials was larger than in forgotten trials at 500 ms before word presentation (BF10 = 72.18, Cohen's d = 0.60). This is considered robust evidence favoring the alternative hypothesis against the null hypothesis[33]. See Fig. 1 for a visual representation of the pupil size variation in both conditions during the study phase.

Experiment 1 discussion

According to our findings, the subjects' future recall capacities were significantly differentiated by pupil size during the study phase. Interestingly, the absolute pupil size was larger during the baseline for later recalled words even before the word appeared on the screen. These findings suggest that higher arousal levels before studying words improve attention and lead to better memory abilities, as reflected in the larger pupil size, regardless of the content of the stimulus. In Experiment 2, our objective was to replicate these results in a recognition test.

Experiment 2 results

Item-by-item analysis

A repeated measures ANOVA on baseline-corrected pupil size values revealed no statistically significant difference between all 30 "old" words ($F(8.94, 80.50) = 1.45$, $p = .183$) in the study phase.

Comparison between "hit" vs. "Miss" trials during the study phase

The mean hit rate across participants was 65.66% (SD = 12.47%). The findings from Experiment 1 were replicated; pupil size was larger while studying later recalled trials than forgotten trials (BF10 = 1583.24, Cohen's d = 0.95). This was also true during baseline, as pupil size in recalled trials was larger than in forgotten trials at 500 ms prior to word presentation (BF10 = 49.96, Cohen's d = 0.70). See Fig. 2 for a visual representation of the pupil size variation in both conditions during the study phase.

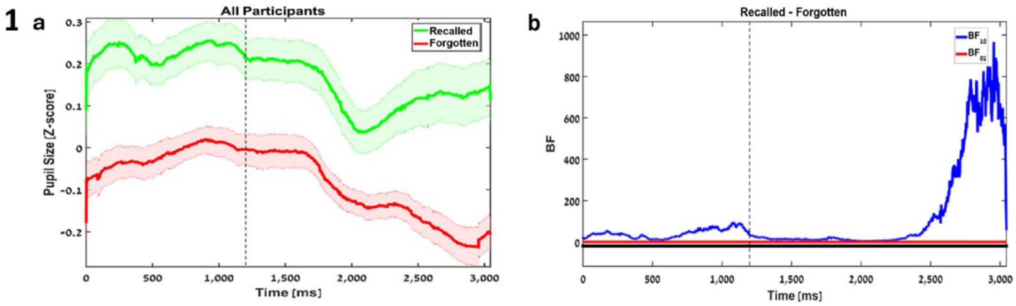

Fig. 1. Pupil size comparison between recalled vs. forgotten trials during the study phase in Experiment 1. (**a**) The horizontal curves represent pupil size in Z-scores in each condition, and the shaded areas represent the standard error from the mean. The vertical dashed line represents the word's appearance (the end of the baseline). (**b**) Bayesian pupil size comparison between recalled and forgotten trials. BF10 (blue curve) and BF01 (red curve) are presented as a function of time. The black horizontal line represents the significant contrasts between conditions.

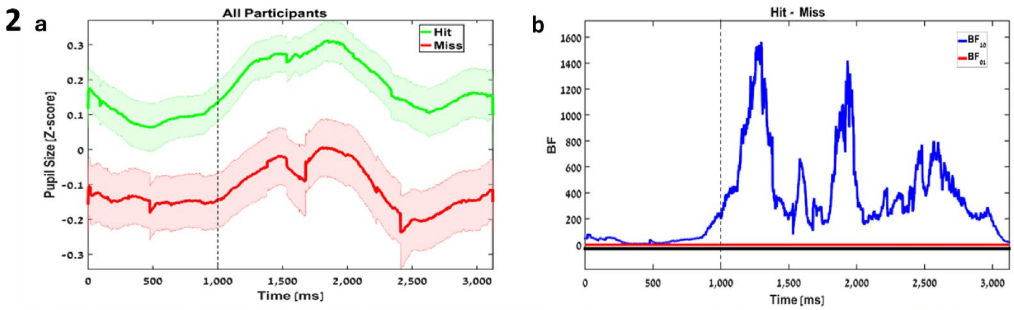

Fig. 2. Pupil size comparison between "Hit" vs. "Miss" trials during the study phase in Experiment 2. (**a**) The horizontal curves represent pupil size in Z-scores in each condition, and the shaded areas represent the standard error from the mean. The vertical dashed line represents the word's appearance (the end of the baseline). (**b**) Bayesian pupil size comparison between "Hit" and "Miss" trials. BF10 (blue curve) and BF01 (red curve) are presented as a function of time. The black horizontal line represents the significant contrasts between conditions.

Experiment 2 discussion

The results obtained from Experiment 2 presented once again larger pupil size during the study phase for recalled compared to forgotten words, regardless of the word's semantic meaning. The variation in pupil size between both conditions was observed 500 ms before word presentation.

Additional analyses

Serial position and pupil size: To examine whether serial position moderates or explains the observed link between baseline pupil size and subsequent memory success, we performed a moderation analysis using PROCESS macro for SPSS (Model 1, version 4.2). **Experiment 1**: The overall model was significant, $F(3, 26) = 15.20$, $p < .001$, $R^2 = 0.637$. Pupil size significantly predicted accuracy ($b = -78.42$, $SE = 17.49$, $t = -4.48$, $p < .001$, 95% CI [-114.36, -42.47]), as did the serial position ($b = -1.06$, $SE = 0.20$, $t = -5.19$, $p < .001$, 95% CI [-1.48, −0.64]). The interaction was significant ($b = 3.35$, $SE = 1.01$, $t = 3.33$, $p = .003$, 95% CI [1.28, 5.42]), indicating that the serial position moderated the effect of pupil size on accuracy. Specifically, the negative effect of baseline pupil size on accuracy in the first trials ($b = -61.79$, $p < .001$) weakened at mid-list positions ($b = -26.45$, $p = .014$) and at final serial positions became positive, though not statistically significant ($b = 8.89$, $p = .57$), as shown in Fig. 3a. The interaction explained an additional 15.5% of the variance, $\Delta R^2 = 0.155$, $F(1, 26) = 11.10$, $p = .003$.

Experiment 2: The overall model was significant, $F(3, 26) = 20.73$, $p < .001$, $R^2 = 0.705$. Pupil size significantly predicted accuracy ($b = -50.90$, $SE = 19.19$, $t = -2.65$, $p = .014$, 95% CI [-90.35, -11.44]), as did the serial position ($b = -0.76$, $SE = 0.25$, $t = -2.98$, $p = .006$, 95% CI [-1.28, −0.24]). The interaction was significant ($b = 3.52$, $SE = 1.04$, $t = 3.37$, $p = .002$, 95% CI [1.38, 5.66]), indicating that the serial position moderated the effect of pupil size on accuracy. Specifically, the same effect as Exp. 1 was observed; at early serial positions, larger baseline pupil size predicted lower accuracy ($b = -33.44$, $p = .040$), became nonsignificant at mid-list ($b = 3.64$, $p = .76$), and was significantly positive by the final positions ($b = 40.73$, $p = .024$), as shown in Fig. 3b. The interaction explained an additional 12.9% of the variance, $\Delta R^2 = 0.129$, $F(1, 26) = 11.39$, $p = .002$. Collectively, these findings reflect a dynamic interaction between baseline pupil-linked arousal and serial position, creating a pattern that allows prediction of subsequent memory success.

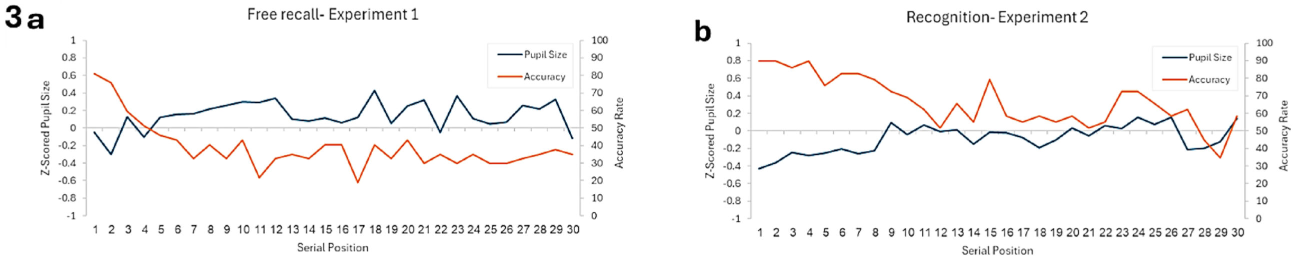

Fig. 3. The left y-axis displays z-scored baseline pupil size, while the right y-axis shows the accuracy rate (percentage). The x-axis represents the serial position throughout the study phase. Two lines are plotted: pupil size (blue line) and accuracy (orange line). The graphs illustrate the moderation effect from Experiment 1 (3a) and Experiment 2 (3b).

Stimulus-evoked responses: To further explore whether word presentation increased the difference already present at baseline, we conducted additional analyses including baseline subtraction[31]. Specifically, we computed the within-subject change in pupil size by subtracting each bin (20 ms) of the post-stimulus data from the 500 ms baseline average pupil size for each trial. Trials with fewer than 25 baseline bins or more than 20% missing data were excluded. Missing values ($\leq 20\%$) were interpolated using cubic splines. Participants with less than five valid trials in each condition (recalled vs. forgotten) were excluded from the analysis. Finally, we performed two paired samples t-tests to compare the means of recalled and forgotten trials (or hit and miss) for each experiment. In the free recall task (Experiment 1), no significant difference was found between conditions, $t(36) = -0.96$, $p = .34$ (two-tailed), suggesting that while a baseline difference in pupil size exists between the conditions, it does not appear to increase following stimulus presentation. This implies that the observed post-stimulus effect likely reflects a continuation of pre-existing baseline differences rather than stimulus-evoked modulation. In contrast, the recognition task (Experiment 2) revealed a significant difference between conditions, $t(28) = 2.06$, $p = .049$ (two-tailed). This suggests that the change in pupil size after the stimulus cannot be explained just by pre-existing baseline differences, but instead indicates an effect caused by the stimulus itself in this specific experiment.

Experiment 1 methods
Participants
Forty-three native Hebrew-speaking undergraduate students (36 females, mean age 22.63 years old, SD = 2.13) from Ariel University participated in the memory task in exchange for course credit. The inclusion criterion was normal or corrected vision by soft contact lenses. Exclusion criteria were: (a) reported history of attention deficit disorder/learning disabilities/brain injuries/decrease in memory/visual impairments, (b) consumption of drugs or coffee on the day of the experiment. They were encouraged to avoid using makeup such as mascara and eyeliner, as the eye tracker can mistakenly detect it as part of the pupil[37]. All participants signed an informed consent form before their participation. They were asked to report any medications they consume regularly in case this might affect their pupil size. The experiment was conducted in a dimly illuminated room without windows in the Neuro-cognition Lab at Ariel University. The experiment was approved by the ethical committee of the Department of Psychology at Ariel University (ID: AU-SOC-YM-20211121). This research was performed in accordance with relevant guidelines and regulations based on the Declaration of Helsinki.

Tools
We used the EyeLink Portable DUO device (SR Research Ltd., Ontario, Canada) to examine pupil variability during the study phase. The experiment was programmed using the EyeLink Experiment Builder software package (SR Research Ltd., version 2.4.77). The EyeLink recorded both eyes in binocular mode at a sample rate of 1000 Hz. The eye tracker was connected to a laptop while recording and transmitting data to the stationary computer screen (1920×1080). This screen presented the stimuli while located 78 cm (± 1) from the subject during the experiment. To ensure maximum accuracy, we fixated the subject's head with a chin rest and guided them to look only at the screen.

Stimuli
The pool of items consisted of 30 common and neutral Hebrew words (see Appendix- Table 1). The words were all bi-syllabic nouns, three to five letters long (M = 3.83, SD = 0.69), with language frequencies greater than 12 per million[38].

Procedure
The experiment began with general instructions regarding the calibration of the eye tracker, in which the participants were asked to gaze toward five drifting points on the screen. Next, additional instructions for the memory task appeared on the screen: "Words will appear sequentially at the center of the screen. Please remember as many as possible (silent reading) for a later recall test." During the study phase, thirty words were displayed in black Arial font (45 pt.) on a white background in random order. Each word appeared in the center of the screen for 1800 milliseconds, following a 1200-millisecond baseline blank screen. These timing

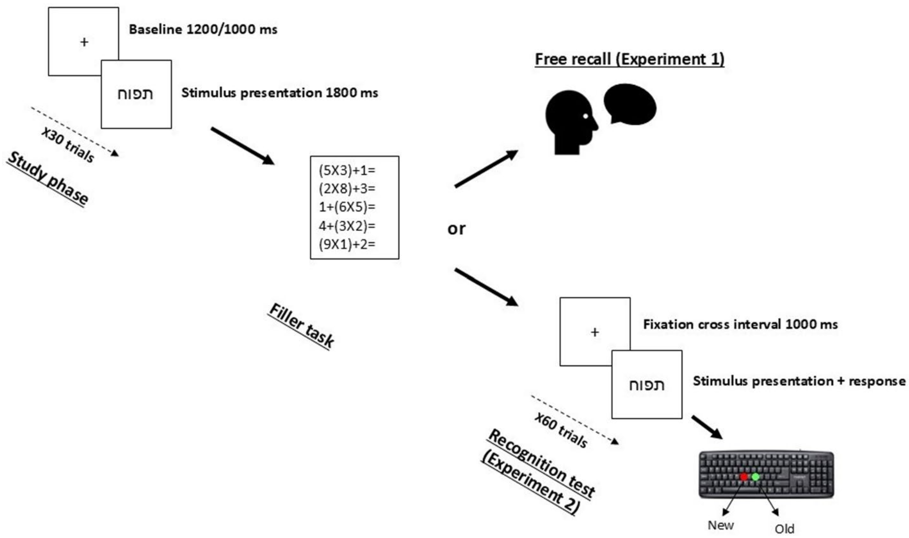

Fig. 4. The experimental procedure for Experiments 1 and 2. In the study phase, the baseline was a blank screen presented for 1200 ms (Exp. 1) or contained a fixation cross (Exp. 2) presented for 1000 ms. The words appeared after the baseline at every trial. After 30 repetitions, the subjects were asked to complete a filler task. After completion, the subjects freely recalled the words studied (Exp. 1) or pressed the "new" or "old" button on the keyboard while observing 60 words appear on the screen sequentially (Exp. 2).

parameters were selected based on related previous pupillometric studies (e.g[18,20,24]). Immediately after the study phase, the subjects performed a filler task containing five exercises in the form of A + B*C=, but digits from 1 to 9 were randomly placed instead of A, B, and C (inspired by Kucewicz et al.[20]). Then, they were asked to freely recall as many words as possible while the experimenter documented their report (see Fig. 4 for a visual representation of the experimental procedure). All EDF's (EyeLink Data Files) were analyzed using MATLAB 2019 (The MathWorks Inc., Natick, Massachusetts) with a designated code.

Statistical analysis

Pupil data was processed using "CHAP"[39], an open-source software in MATLAB (The MathWorks Inc., Natick, Massachusetts). First, pupil data from the EyeLink files were converted from arbitrary units to Z-score values based on the mean and standard deviation calculated for each trial separately. Then, we removed outlier samples with Z-scores larger than 2.5 in each trial. Next, we excluded trials with more than 20% missing values from the analysis for each participant (1.82%). We defined a minimum of 5 trials in each condition as a criterion for inclusion in the analysis, which excluded one participant (0.02%). After pre-processing and eliminating trials by the last two standards mentioned above, 42 subjects remained in the analysis (35 females, mean age 22.69 years old, SD = 2.11). Finally, we detected eye blinks using Hershman, Henik, and Cohen's[40] algorithm and filled in missing values using linear interpolation. We aimed to examine pupil size variation throughout all trials and compare each sample of recalled and forgotten words. The conventional approach to analyzing pupillometry data relies on comparing means across conditions within predefined time windows, which collapses multiple samples into single values. This method can conceal valuable temporal information about cognitive processes that evolve over time[41]. Regarding our study, Bayesian statistical methods allow a more comprehensive understanding of arousal dynamics by tracking temporal changes in pupil size within each trial and comparing recalled and forgotten trials. Specifically, Bayes factors represent the probability of discovering a significant difference between both conditions and ruling out the null hypothesis[42]. This measure provides a consistent parameter estimation that improves with sample size, allows researchers to actively accept the null hypothesis when appropriate, and eliminates concerns about alpha inflation from multiple comparisons[41]. These advantages make Bayesian analysis especially suited for pupillometry research, where researchers analyze continuous streams of data points while maintaining statistical validity. To confirm our hypothesis of larger pupil size while studying later recalled compared to forgotten trials, we evaluated the average pupil size in each sample of both conditions across the entire time course (3 s per trial). For our second hypothesis, we performed the same analysis mentioned before, but only for the baseline, which we defined as a 500 ms pre-word presentation.

Experiment 2 methods

Participants

We recruited fifty-two new native Hebrew-speaking undergraduate students (45 females, mean age 22.53 years old, SD = 2.42) from Ariel University to participate in the recognition task in exchange for course credit. The inclusion and exclusion criteria were the same as in Experiment 1. The experiment was conducted in the same room in the Neuro-cognition Lab at Ariel University. The Department of Psychology ethics committee at Ariel University approved the experiment (ID: AU-SOC-YM-20211121), and all participants signed an informed consent form. This research was performed in accordance with relevant guidelines and regulations based on the Declaration of Helsinki.

Stimuli

The pool of items consisted of 60 common and neutral Hebrew words: 30 “old” words, which appeared in the study phase, and 30 “new” words, which appeared only in the recognition phase (see Appendix- Table 1). These words possess the same properties as those presented in Experiment 1.

Procedure

The procedure was similar to Experiment 1 except for a few minor changes. During the 1000-millisecond baseline, a black fixation cross appeared in the center of the white background. Immediately after the study phase and performance of five distraction exercises as in Experiment 1, the participants were instructed to click the red button if the word presented on screen was “new” (did not appear at the study phase) or the green button if the word presented is “old” (appeared at the study phase).” Each word in the recognition test was presented, proceeding with a 1000-millisecond baseline fixation cross screen (see Fig. 4 for visual representation of the experimental procedure).

Statistical analysis

Following the pre-processing procedure, 30 subjects remained in the analysis (26 females, mean age 22.2 years old, SD = 1.83). Twenty-two individuals were excluded from the analysis due to having less than five trials in the “Miss” condition (41%). This criterion was implemented to ensure a reliable comparison between both conditions, meaning that individuals with superior memory performance had to be excluded. Finally, we excluded trials with more than 20% missing values from the analysis for each participant (1.66%).

General discussion

Our study demonstrates that baseline pupil size during the study phase is a reliable predictor of successful memory performance, even before the subject has read and comprehended the word’s semantic meaning. Notably, we found that the difference observed at baseline persisted and even grew after word presentation (in Experiment 2). These findings highlight the utility of pupil size as a non-invasive, cost-effective measure of physiological arousal, closely linked to locus coeruleus (LC) activity and rapid cognitive state fluctuations[8,43,44]. Our results are consistent with the Adaptive Gain Theory[8], which suggests that the LC shifts between phasic and tonic modes to regulate task engagement. Phasic LC activity supports focused performance (exploitation), while tonic LC activity encourages exploration. Optimal performance occurs with high phasic and moderate tonic LC activity. Importantly, tonic LC activity is linked to larger baseline pupil size and weaker phasic pupil responses, providing a physiological marker of this balance. In our study, increased baseline pupil size indicated a physiological state of moderate tonic LC activity that primes the brain for heightened cognitive readiness, thereby enhancing learning outcomes.

Previous well-documented studies posit that arousal during encoding enhances the consolidation of memory traces, thereby improving long-term retention[30,45]. Supporting this, recent studies have shown that light exercise enhances memory through activation of neural pathways within the ascending arousal system, with greater pupil dilation during exercise associated with improved memory performance[46]. Similarly, changes in pupil dilation during very-light-intensity exercise predicted improvements in Stroop performance[47]. These findings, along with our results, suggest that baseline pupil size can serve as an arousal indicator that predicts memory performance.

Considering longstanding evidence that recall probabilities are influenced by serial position[34], we considered the possibility that this factor could confound our findings. Although the literature on baseline pupil size across serial positions in memory tasks is limited, two recent studies by Koo et al.[35] and Książek et al.[36] observed gradual increases in baseline pupil size throughout the study phase, which they attributed to rising memory load. Given that pupil size is also sensitive to time-related fluctuations, it was crucial to determine whether baseline pupil size independently predicted memory performance beyond the effects of serial position. Our results from both experiments revealed independent effects of baseline pupil size and serial position on accuracy rates and a significant interaction between them. Specifically, the relationship between pupil size and memory varied across serial positions: it was strongly negative for early-list items, weakened mid-list, and became positive toward the end. These findings suggest that they point to a dynamic interplay in which baseline arousal modulates memory encoding differently as the task unfolds. This supports the interpretation that “readiness to remember”, reflected in pupil size, fluctuates over time but cannot be fully accounted for by serial position effects alone.

A notable limitation of our study was the exclusion of 22 out of 52 subjects in Experiment 2 due to insufficient trials in the “Miss” condition (5 or fewer). We established this principle to ensure a valid comparison between the two conditions in the experiment. It is well-known that “Hit” rates in recognition memory tests tend to be higher because they provide a cue, the word itself, compared to free recall tests[48]. Despite this disadvantage, it was essential for us to create two relatively similar experiments to ensure comparability between experiments rather than increasing task difficulty.

The current study established the relationship between pupil size, physiological arousal, and long-term memory, demonstrating the possibility of predicting memory, a “readiness to remember” state, even before the stimulus appeared. This research has many possible applications in artificial intelligence, particularly in predicting memory capacities by quantifying arousal levels. Future research should investigate whether the pupil size-memory relationship remains stable when scaling up to more extensive learning scenarios. While our study demonstrated the predictive power of pupil size for memory formation with a limited item set, real-world learning often involves longer study sessions with more material. We suggest examining this relationship using larger sets of items presented across longer study sessions. This extended format would better simulate authentic learning conditions, such as studying for an exam. This approach could assist in identifying the optimal study duration and potential break-points by monitoring changes in baseline pupil patterns, which reflect arousal levels over time. Understanding these dynamics could have significant implications for optimizing study schedules and learning strategies in educational settings.

Data availability

All data obtained in this study are available on the Open Science Framework (https://osf.io/gdsnw/). Data analy sis was performed using a dedicated pupillometry MATLAB code by Hershman, Henik & Cohen (2018), which can be found at the following link: https://osf.io/jyz43/.

Received: 14 October 2024; Accepted: 30 June 2025
Published online: 01 July 2025

References

1. Hess, E. H. & Polt, J. M. Pupil size as related to interest value of visual stimuli. *Science* **132** (3423), 349–350. https://doi.org/10.1126/science.132.3423.349 (1960).
2. Beatty, J. & Kahneman, D. Pupillary changes in two memory tasks. *Psychonomic Sci.* **5** (10), 371–372. https://doi.org/10.3758/BF03328444 (1966).
3. P Marshall, S. The index of cognitive activity: measuring cognitive workload. *IEEE Conf. Hum. Factors Power Plants*. 75–79. https://doi.org/10.1109/hfpp.2002.1042860 (2002).
4. Liu, Y., Rodenkirch, C., Moskowitz, N., Schriver, B. & Wang, Q. Dynamic lateralization of pupil dilation evoked by LC activation results from sympathetic not parasympathetic contributions. *Cell. Rep.* **20** (13), 3099–3112. https://doi.org/10.1016/j.celrep.2017.08.094 (2017).
5. Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H. & Balsters, J. H. Pupil diameter covaries with BOLD activity in human locus coeruleus. *Hum. Brain. Mapp.* **35** (8), 4140–4154. https://doi.org/10.1002/hbm.22466 (2014).
6. Kafkas, A. Eyes on memory: pupillometry in encoding and retrieval. *Vision* **8** (2), 1–13. https://doi.org/10.3390/vision8020037 (2024).
7. Samuels, E. & Szabadi, E. Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part I: principles of functional organisation. *Curr. Neuropharmacol.* **6** (3), 235–253. https://doi.org/10.2174/157015908785777229 (2008).
8. Aston-Jones, G. & Cohen, J. D. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* **28**, 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709 (2005).
9. Nassar, M. R. et al. Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat. Neurosci.* **15** (7), 1040–1046. https://doi.org/10.1038/nn.3130 (2012).
10. Unsworth, N. & Robison, M. K. Tracking arousal state and Mind wandering with pupillometry. *Cogn. Affect. Behav. Neurosci.* **18** (4), 638–664. https://doi.org/10.3758/s13415-018-0594-4 (2018).
11. Yerkes, R. M. & Dodson, J. D. The relation of strength of stimulus to rapidity of habit-formation. *J. Comp. Neurol. Psychol.* 459–482. https://doi.org/10.1037/h0073415 (1908).
12. Micula, A. et al. A glimpse of memory through the eyes: pupillary responses measured during encoding reflect the likelihood of subsequent memory recall in an auditory free recall test. *Trends Hear.* **26** https://doi.org/10.1177/23312165221130581 (2022).
13. Goldinger, S. D. & Papesh, M. H. Pupil dilation reflects the creation and retrieval of memories. *Curr. Dir. Psychol. Sci.* **21** (2), 90–95. https://doi.org/10.1177/0963721412436811 (2012).
14. Jagiello, R., Pomper, U., Yoneya, M., Zhao, S. & Chait, M. Rapid brain responses to familiar vs. Unfamiliar Music – an EEG and pupillometry study. *Sci. Rep.* **9** (1), 1–13. https://doi.org/10.1038/s41598-019-51759-9 (2019).
15. Kahneman, D. & Beatty, J. Pupil diameter and load on memory. *Science* **154** (3756), 1583–1586. https://doi.org/10.1126/science.154.3756.1583 (1966).
16. Otero, S. C., Weekes, B. S. & Hutton, S. B. Pupil size changes during recognition memory. *Psychophysiology* **48** (10), 1346–1353. https://doi.org/10.1111/j.1469-8986.2011.01217.x (2011).
17. Pajkossy, P., Szőllősi, Á. & Racsmány, M. Pupil size changes signal hippocampus-related memory functions. *Sci. Rep.* **10** (1), 1–15. https://doi.org/10.1038/s41598-020-73374-9 (2020).
18. Võ, M. L. H. et al. The coupling of emotion and cognition in the eye: introducing the pupil old/new effect. *Psychophysiology* **45** (1), 130–140. https://doi.org/10.1111/j.1469-8986.2007.00606.x (2008).
19. Kafkas, A. & Montaldi, D. Striatal and midbrain connectivity with the hippocampus selectively boosts memory for contextual novelty. *Hippocampus* **25** (11), 1262–1273. https://doi.org/10.1002/hipo.22434 (2015).
20. Kucewicz, M. T. et al. Pupil size reflects successful encoding and recall of memory in humans. *Sci. Rep.* **8** (1), 1–8. https://doi.org/10.1038/s41598-018-23197-6 (2018).
21. Miller, A. L. & Unsworth, N. Variation in attention at encoding: insights from pupillometry and eye gaze fixations. *J. Experimental Psychology: Learn. Memory Cognition*. https://doi.org/10.1037/xlm0000797 (2019).
22. Papesh, M. H., Goldinger, S. D. & Hout, M. C. Memory strength and specificity revealed by pupillometry. *Int. J. Psychophysiol.* **83** (1), 56–64. https://doi.org/10.1016/j.ijpsycho.2011.10.002 (2012).
23. Cronin, S. L., Lipp, O. V. & Marinovic, W. Pupil dilation during encoding, but not type of auditory stimulation, predicts recognition success in face memory. *Biol. Psychol.* **178** (November 2022), 108547. https://doi.org/10.1016/j.biopsycho.2023.108547 (2023).
24. Kafkas, A. & Montaldi, D. Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *Q. J. Experimental Psychol.* **64** (10), 1971–1989. https://doi.org/10.1080/17470218.2011.588335 (2011).
25. Naber, M., Frässle, S., Rutishauser, U. & Einhäuser, W. Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. *J. Vis.* **13** (2), 1–20. https://doi.org/10.1167/13.2.11 (2013).
26. Hopstaken, J. F., van der Linden, D., Bakker, A. B. & Kompier, M. A. J. The window of my eyes: task disengagement and mental fatigue covary with pupil dynamics. *Biol. Psychol.* **110**, 100–106. https://doi.org/10.1016/j.biopsycho.2015.06.013 (2015).
27. Cohen, N. et al. Prestimulus activity in the Cingulo-Opercular network predicts memory for naturalistic episodic experience. *Cereb. Cortex*. **30** (3), 1902–1913. https://doi.org/10.1093/cercor/bhz212 (2020).
28. Lawal, O., Regelous, P. & Omiyi, D. Supporting claustrophobic patients during magnetic resonance imaging examination– the patient perspective. *Radiography* **29** (6), 1108–1114. https://doi.org/10.1016/j.radi.2023.09.008 (2023).
29. Rojas-Thomas, F. et al. Impact of acute psychosocial stress on attentional control in humans. A study of evoked potentials and pupillary response. *Neurobiol. Stress*. **25** (December 2022), 100551. https://doi.org/10.1016/j.ynstr.2023.100551 (2023).
30. Madore, K. P. & Wagner, A. D. Readiness to remember: predicting variability in episodic memory. *Trends Cogn. Sci.* **26** (8), 707–723. https://doi.org/10.1016/j.tics.2022.05.006 (2022).

31. Mathôt, S., Fabius, J., Van Heusden, E. & Van der Stigchel, S. Safe and sensible preprocessing and baseline correction of pupil-size data. *Behav. Res. Methods*. **50** (1), 94–106. https://doi.org/10.3758/s13428-017-1007-2 (2018).
32. Robison, M. K. & Unsworth, N. Pupillometry tracks fluctuations in working memory performance. *Atten. Percept. Psychophysics*. **81** (2), 407–419. https://doi.org/10.3758/s13414-018-1618-4 (2019).
33. Kelter, R. Bayesian alternatives to null hypothesis significance testing in biomedical research: A non-technical introduction to bayesian inference with JASP. *BMC Med. Res. Methodol.* **20** (1), 1–13. https://doi.org/10.1186/s12874-020-00980-6 (2020).
34. Murdock, B. B. The serial position effect of free recall. *J. Exp. Psychol.* **64** (5), 482–488. https://doi.org/10.1037/h0045106 (1962).
35. Koo, M. et al. Recall of reverberant speech in quiet and four-talker babble noise. *Brain Sci.* **11** (7). https://doi.org/10.3390/brainsci11070891 (2021).
36. Książek, P., Zekveld, A. A., Fiedler, L., Kramer, S. E. & Wendt, D. Time-specific components of pupil responses reveal alternations in effort allocation caused by memory task demands during speech identification in noise. *Trends Hear.* **27**, 1–19. https://doi.org/10.1177/23312165231153280 (2023).
37. Winn, M. B., Wendt, D., Koelewijn, T. & Kuchinsky, S. E. Best practices and advice for using pupillometry to measure listening effort: an introduction for those who want to get started. *Trends Hear.* **22**, 1–32. https://doi.org/10.1177/2331216518800869 (2018).
38. Icht, M., Mama, Y. & Algom, D. The production effect in memory: multiple species of distinctiveness. *Front. Psychol.* **5**, 74928. https://doi.org/10.3389/fpsyg.2014.00886 (2014).
39. Hershman, R., Cohen, N. & Henik, A. CHAP: an open source software for processing and analyzing pupillometry data. *J. Vis.* **17** (10), 543. https://doi.org/10.1167/17.10.543 (2017).
40. Hershman, R., Henik, A. & Cohen, N. A novel Blink detection method based on pupillometry noise. *Behav. Res. Methods*. **50**, 107–114. https://doi.org/10.3758/s13428-017-1008-1 (2018).
41. Hershman, R., Milshtein, D. & Henik, A. The contribution of Temporal analysis of pupillometry measurements to cognitive research. *Psychol. Res.* **87** (1), 28–42. https://doi.org/10.1007/s00426-022-01656-0 (2023).
42. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **90** (430), 773–795. https://doi.org/10.1080/01621459.1995.10476572 (1995).
43. Foote, S. L., Bloom, F. E. & Aston-Jones, G. Nucleus locus ceruleus: new evidence of anatomical and physiological specificity. *Physiol. Rev.* **63** (3), 844–914. https://doi.org/10.1152/physrev.1983.63.3.844 (1983).
44. Yamazaki, Y. et al. A possible contribution of the locus coeruleus to arousal enhancement with mild exercise: evidence from pupillometry and neuromelanin imaging. *Cereb. Cortex Commun.* **4** (2), 1–8. https://doi.org/10.1093/texcom/tgad010 (2023).
45. Kleinsmith, L. J., Kaplan, S. & Trate, R. D. The relationship of arousal to short- and long-term verbal recall. *Can. J. Psychol. / Revue Canadienne De Psychologie*. **17** (4), 393–397. https://doi.org/10.1037/h0083278 (1963).
46. Suwabe, K. et al. Improvement of mnemonic discrimination with acute light exercise is mediated by pupil-linked arousal in healthy older adults. *Neurobiol. Aging*. **133** (May 2023), 107–114. https://doi.org/10.1016/j.neurobiolaging.2023.09.006 (2024).
47. Kuwamizu, R. et al. Pupil dynamics during very light exercise predict benefits to prefrontal cognition. *NeuroImage* **277(June)** https://doi.org/10.1016/j.neuroimage.2023.120244 (2023).
48. Haist, F., Shimamura, A. P. & Squire, L. R. On the relationship between recall and recognition memory. *J. Experimental Psychology: Learn. Memory Cognition*. **18** (4), 691. https://doi.org/10.1037/0278-7393.18.4.691 (1992).

Acknowledgements
The authors thank Gal Atun for valuable contributions to programming Experiment 1 and assisting with the eye-tracker operation, Harai A. Bester for meaningful assistance in data analysis, and Ronen Hershman for insightful discussions and guidance during data analysis using CHAP.

Author contributions
All authors contributed equally to this current paper.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-09703-7.

Correspondence and requests for materials should be addressed to Y.M.