

Reading memory formation from the eyes

Anne Bergt^{1,*} | Anne E. Urai^{2,3,*}  | Tobias H. Donner^{2,3}  | Lars Schwabe¹ 

¹Department of Cognitive Psychology, Institute of Psychology, University of Hamburg, Hamburg, Germany

²Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

³Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

Correspondence

Lars Schwabe, Department of Cognitive Psychology, Institute of Psychology, University of Hamburg, Von-Melle-Park 5, 20146 Hamburg, Germany.
Email: lars.schwabe@uni-hamburg.de

Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Number: DO 1240/2-1, DO 1240/3-1

Abstract

At any time, we are processing thousands of stimuli, but only few of them will be remembered hours or days later. Is there any way to predict which ones? Here, we tested whether the pupil response to ongoing stimuli, an indicator of physiological arousal known to be relevant for memory formation, is a reliable predictor of long-term memory for these stimuli, over at least 1 day. Pupil dilation was tracked while participants performed visual and auditory encoding tasks. Memory was tested immediately after encoding and 24 hr later. Irrespective of the encoding modality, trial-by-trial variations in pupil dilation predicted reliably which stimuli were recalled in the immediate and 24 hr-delayed tests, in particular for emotionally arousing stimuli. These results show that our eyes may provide a window into the formation of long-term memories. Furthermore, our findings underline the important role of central arousal systems in the rapid formation of memories in the brain, possibly by gating synaptic plasticity mechanisms in the neocortex.

KEYWORDS

arousal, emotional memory, long-term memory, memory encoding, pupil dilation

1 | INTRODUCTION

Our memories define to a large extent who we are. They help us to adapt to current and future realities and without memory any form of education would be unthinkable. Moreover, both the failure to form memories (e.g., in Alzheimer's disease) and overly strong (emotional) memories (e.g., in posttraumatic stress disorder, PTSD) are clinically highly relevant. Predicting whether, and which, ongoing events will be remembered later on is thus a significant goal of memory

research, with considerable implications for various applied contexts.

Recent years have witnessed a revitalized interest in pupil dilation as indicator of phasic changes in central arousal state during cognitive processes (De Gee, Knapen, & Donner, 2014; Eldar, Niv, & Cohen, 2016; Nassar et al., 2012; Urai, Braun, & Donner, 2017; Wierda, van Rijn, Taatgen, & Martens, 2012). Task-evoked pupil responses reflect phasic activity of neuromodulatory brainstem nuclei controlling arousal. One of these is the locus coeruleus (Aston-Jones & Cohen, 2005; Joshi, Li, Kalwani, & Gold, 2016), a brainstem nucleus that provides the major noradrenergic input to the limbic system and neocortex, which is crucial for memory formation (Cahill & McGaugh, 1998). Thus, the pupil response might provide a window into the making of memories. Accordingly, some earlier studies have examined the link of pupil dilation and memory (Ariel & Castel, 2014; Clewett, Huang, Velasco, Lee, & Mather, 2018; Goldinger & Papesh, 2012; Hoffing & Seitz, 2015). In these studies, however, the association of pupil dilation and memory was assessed immediately after

Abbreviations: EEG, electroencephalography; fMRI, functional magnetic resonance imaging; IAPS, International Affective Picture System; PTSD, post-traumatic stress disorder.

*These authors contributed equally

Edited by Prof. John Foxe. Reviewed by Edward Freedman, University of Rochester, USA and Mathias Weymar, University of Potsdam, Germany

All peer review communications can be found with the online version of the article.

encoding and mainly at the group level, which precluded an item-specific prediction. Whether the pupil response may indeed forecast the long-term retrieval of individual stimuli remained unknown.

We tested here the hypothesis that the pupil response during encoding of stimuli—and, by inference, phasic elevation of central arousal—predicts trial-by-trial long-term memory formation. In two independent tasks, a visual picture encoding task and an auditory word encoding task, participants encoded either pictures or words while their pupillary responses were tracked with an eye tracker. The use of two tasks allowed us to assess the robustness of the hypothesized predictive value of the pupil response during encoding for subsequent memory. Moreover, because visual stimulation per se leads already to a pupil response, the use of an additional auditory task enabled us to examine the association between pupil dilation during encoding and later memory when any visual artifacts could be ruled out. Memory for the stimuli was tested both immediately after encoding and 24 hr later. Because pupil dilation reflects also emotional arousal (Goldinger & Papesh, 2012; Lempert, Glimcher, & Phelps, 2015) and the well-known memory enhancement for emotional relative to neutral events (Christianson, 1992) is crucial for several psychopathologies, including PTSD (Pitman, 1989; de Quervain, Schwabe, & Roozendaal, 2017), we included neutral and emotionally arousing stimuli to further examine whether pupil dilation may have particular predictive value for emotional memory formation.

2 | METHODS

2.1 | Participants

In all, 54 healthy native speakers of German (age: 18–35 years, $M = 25.35$ years; 27 women, 27 men) without a history of any neurologic or mental disorders participated in this study. All of them reported normal or corrected-to-normal visual acuity and were naïve to the purpose of the experiment. The sample size was based on an a priori sample size calculation using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), showing that a sample of 54 participants is required to detect a medium-sized effect of $f = 0.25$ with a power of 0.95, given an α of 0.05. Due to technical failure, pupil data were missing for three participants in the picture encoding task and for six participants in the word encoding task, thus leaving a sample of 51 and 48 participants, respectively, in the corresponding analyses. All participants provided written informed consent and were paid a moderate monetary compensation for participation. The study protocol conforms to the Declaration of Helsinki and was approved by the ethics committee of the Faculty of Psychology and Human Movement at the University of Hamburg (approval no. 2016_79).

2.2 | Apparatus

The experiment was programmed and presented in MATLAB (The MathWorks, Natick, MA) using the Psychophysics Toolbox (Brainard, 1997), in combination with the eye-tracking software BeGaze 3.0 (SensoMotoric Systems, SMI). Stimuli were presented on a 24-inch Dell monitor with a resolution of $1,920 \times 1,200$ pixels and a refresh rate of 60 Hz. Participants sat in a dimly lit, sound-attenuated room with their head in a chin rest at a distance of 60 cm from the screen. Pupil size was monitored in both eyes using a RED250mobile (SMI; sampling rate: 250 Hz). The eye tracker was calibrated applying the 9-point calibration and validation procedure before each of the two encoding tasks.

2.3 | Experimental tasks, stimuli, and procedure

After their arrival at the laboratory, participants first completed standard questionnaires to assess their depressive mood, chronic stress level, as well as state and trait anxiety, all of which may affect (emotional) memory processes (see Appendix S1). Next, the eye tracker was calibrated and participants performed two encoding tasks: a picture encoding task and a word encoding task. Task order was counterbalanced across participants.

2.3.1 | Picture encoding task

The stimulus set for the picture encoding task consisted of 150 emotionally neutral and 150 negative pictures taken from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 1997) and other open online sources (all images are available at <https://doi.org/10.5281/zenodo.1246100>). Pictures were presented in greyscale and modified in MATLAB so that they all had the same average luminance. During encoding, 75 neutral and 75 negative pictures were randomly chosen from the picture pool and presented in randomized order for 3 s at the center of the screen, against a grey background that was equiluminant to the pictures. While encoding the pictures, participants were requested to memorize the pictures (intentional encoding) and to evaluate the emotionality of the shown picture on a 4-point scale from 0 (“neutral”) to 3 (“very negative”). Between pictures, we presented a grey fixation cross for 3–6 s.

2.3.2 | Word encoding task

The stimulus set for the word encoding task consisted of 120 emotionally neutral and 120 negative German nouns. Words were taken from standardized German word data sets (Böcker, Gruber, & Gauggel, 2014; Schwibbe, Räder, Schwibbe, & Geiken-Pophanken, 1981). We created audio

files for these words with the help of the software Audacity®. All audio files and a list of used words are available at <https://doi.org/10.5281/zenodo.1246100>. During encoding, 60 neutral and 60 negative words were randomly chosen from the word pool and presented in randomized order via headphones. While listening to the words, participants looked at a fixation cross shown at the center of the screen with the same grey background as in the picture encoding task and were requested to memorize the words (intentional encoding). The inter-trial interval between the presentations of words varied between 3 and 6 s.

2.3.3 | Immediate and delayed memory testing

Immediately after each of the two encoding tasks as well as 24 hr after the encoding session, participants performed a free recall test, in which they were asked to report verbally as many of the presented pictures and words, respectively, as possible. The experimenter noted the recalled items on a check list, while standing in the back of the participant, thus preventing any direct feedback. If it was not entirely clear to which picture a participant was referring to in the free recall test for the picture, he/she was asked to provide more details until the recalled picture could be clearly identified. There was no time limit for the free recall tests. To assess the predictive value of the pupil response during encoding for long-term memory, participants completed a second free recall test 24 hr after encoding. The procedure of this delayed memory test was exactly the same as in the immediate free recall test.

After the 24 hr-delayed free recall test, participants completed also recognition tests for the pictures and words. In these tests, participants saw all pictures and words, respectively, that were presented on the first day and an equal number of novel neutral and negative items in randomized order on a computer screen. Participants were requested to indicate for each item whether it had been presented on day 1 (“old”) or not (“new”). For items that were identified as “old,” participants were further asked to rate on a scale from 1 (“not certain”) to 4 (“very certain”) how confident they were that the item was indeed “old.” Because free recall reflects participants’ ability to actually retrieve information better than recognition, free recall appears to be more sensitive to arousal effects than recognition (Bradley, Greenwald, Petry, & Lang, 1992) and recall and recognition appear to rely on distinct encoding mechanisms (Staresina & Davachi, 2006), our analyses focused primarily on the free recall tests. Data for the recognition test are presented in the Appendix S1 and Figure S1.

2.4 | Pupil data preprocessing and analyses

The pupil data were preprocessed as described in Urai et al. (2017). Missing data and blinks, as detected by the SMI

software, were linearly interpolated. We estimated the effect of blinks and saccades on the pupil response through deconvolution, and removed these responses from the data using linear regression. The residual pupil time series were *z*-scored per run, and resampled to 50 Hz. We segmented the continuous pupil data into epochs corresponding to experimental trials, and baseline corrected the single-trial data by subtracting the average pupil size in the 2 s before stimulus onset. We then defined pupil responses as the average in 1–3 s after stimulus onset; this window was chosen to take into account the delay of the pupil response (De Gee et al., 2014) and encompass the full presentation duration of the pictures. Statistics on the pupil time course were corrected using cluster-based permutation testing.

2.5 | Behavioral data analysis

Data from the picture and word recall tasks on day 1 and 2 were quantified as the fraction of recalled stimuli relative to the number of stimuli presented during encoding. Recall performance for neutral and negative stimuli was subjected to paired *t*-tests. To assess the predictive value of the pupil size during encoding for subsequent memory, we first subjected the data to a subsequent memory analysis, in which we asked whether the pupil size during encoding differed for subsequently remembered and forgotten items. To this end, we subjected the pupil data to an ANOVA with the factors subsequent memory (remembered vs forgotten) and stimulus emotionality (neutral vs. negative).

To analyze on a trial-by-trial basis whether subsequent memory for each individual item can be predicted by pupil dilation during encoding, we employed a logistic regression approach. More specifically, we performed for all participants individual logic regressions estimating the predictive value of the pupil response to an individual item for the subsequent recall of this item. The logistic regression was performed both separately for neutral and negative items and for all items together. The beta values from these individual logistic regressions were then subjected to *t*-tests at the group level to assess whether the beta values were reliably different from zero and different for neutral and negative items. All reported *p* values are two-tailed.

3 | RESULTS

3.1 | Pupil response predicts emotional memory formation for pictures

Participants’ emotionality ratings during picture encoding confirmed the classification into neutral and negative pictures (mean rating (*SEM*) for neutral pictures: 0.15 (0.02), for negative pictures: 1.98 (0.05); $t(53) = 38.19$, $p < 0.001$, $d = 7.35$). As expected (Christianson, 1992), negative

pictures were significantly better remembered than neutral pictures, both in the immediate free recall test ($t(53) = 12.89$, $p < 0.001$, $d = 2.47$; Figure 1a) and in the 24 hr-delayed recall ($t(53) = 12.39$, $p < 0.001$, $d = 2.38$; Figure 1b).

This emotional memory enhancement was reflected in participants' pupil responses (Figure 1c). The pupil initially constricted during stimulus presentation, an effect only evident for pictures, not for words (compare with Figure 2c), which is due to the pupil response to the presentation of high-contrast images. This constriction was followed by an evoked dilation, the amplitude of which was modulated by emotional content: Pupil responses (defined as the average baseline-corrected pupil size from 1 to 3 s after stimulus onset, see Methods) were significantly stronger in response to negative as compared to neutral pictures ($t(50) = 11.16$, $p < 0.001$, $d = 2.15$; Figure 1c).

Importantly, pupil responses during encoding were significantly stronger for items that were subsequently remembered in the immediate free recall test (Figure 1d; main effect subsequent memory: $F_{1,50} = 9.28$, $p = 0.004$, $\eta_p^2 = 0.10$) and the 24 hr-delayed test (Figure 1e; $F_{1,50} = 6.62$, $p = 0.013$, $\eta_p^2 = 0.09$). This effect was driven by the emotionally negative stimuli: Significant

interactions of stimulus emotionality and subsequent memory in the immediate ($F_{1,50} = 10.58$, $p = 0.002$, $\eta_p^2 = 0.21$) and delayed ($F_{1,50} = 4.21$, $p = 0.045$, $\eta_p^2 = 0.08$) recall tests revealed that the pupil response during encoding was larger for remembered than forgotten pictures, when pictures were negative (immediate recall: $t(50) = 4.83$, $p < 0.001$, $d = 0.97$; 24 hr-delayed recall: $t(50) = 3.81$, $p < 0.001$, $d = 0.76$) but not when pictures were neutral (immediate recall: $t(50) = 0.51$, $p = 0.610$, $d = 0.10$; 24 hr-delayed recall: $t(50) = 0.08$, $p = 0.939$, $d = 0.02$). Adding the factor retention delay (immediate vs 24 hr-delayed) explicitly to the model confirmed that there was, across recall tests, a significant difference in pupil dilation during encoding between subsequently remembered and forgotten pictures ($F_{1,50} = 9.00$, $p = 0.004$, $\eta_p^2 = 0.15$), between neutral and negative pictures ($F_{1,50} = 92.41$, $p < 0.001$, $\eta_p^2 = 0.65$), a significant interaction of stimulus emotionality and subsequent memory ($F_{1,50} = 7.33$, $p = 0.009$, $\eta_p^2 = 0.13$), but no effect of retention delay (main effect: $F_{1,50} = 0.64$, $p = 0.428$, $\eta_p^2 = 0.01$; all interaction effects including the factor retention delay: all $F < 2.73$, all $p > 0.106$, all $\eta_p^2 < 0.06$). The absence of any effects of retention delay

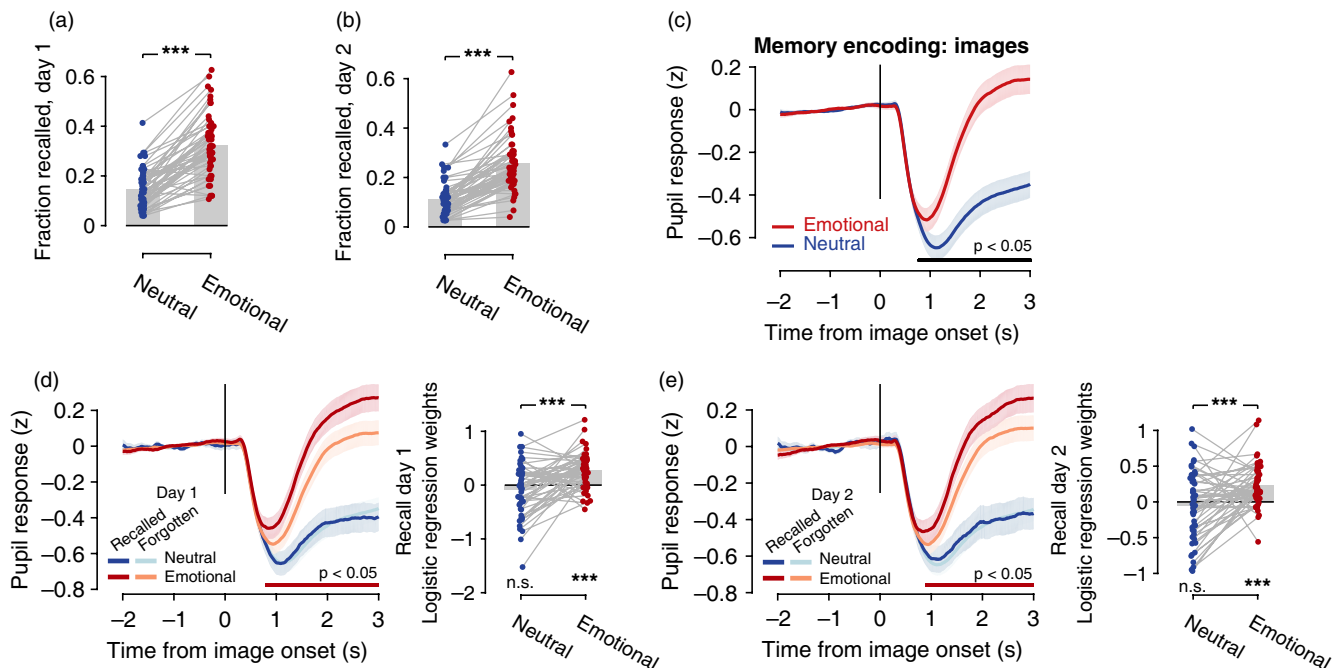


FIGURE 1 Predictive value of pupil dilation for memory formation in the picture task. (a) Recall performance in the immediate and (b) 24 hr-delayed memory, expressed as fraction of recalled pictures, was better for negative than for neutral items. (c) Timecourse of baseline-corrected pupil size in response to picture presentation. After picture onset, the pupil initially constricts in response to the higher contrast images. From 0.78 s after pictures onset, emotion pictures elicit a larger pupil response than neutral pictures. Black line indicates significant timepoints, obtained from a cluster-corrected permutation test. Lines and shaded error regions indicate mean \pm SEM. (d) Left: Pupil timecourses, separately for emotional and neutral pictures that were recalled or forgotten in the recall test on day 1. Horizontal lines indicate timepoints at which the pupil dilation is significantly different between forgotten and recalled stimuli, as obtained from cluster-corrected permutation test (separately for emotional and neutral stimuli). Right: Individual beta values from logistic regression analyses, indicating that the pupil response during encoding predicts immediate recall of negative but not neutral pictures. (e) Same as panel (d) but for 24 hr-delayed recall test. *** two-tailed $p < 0.001$

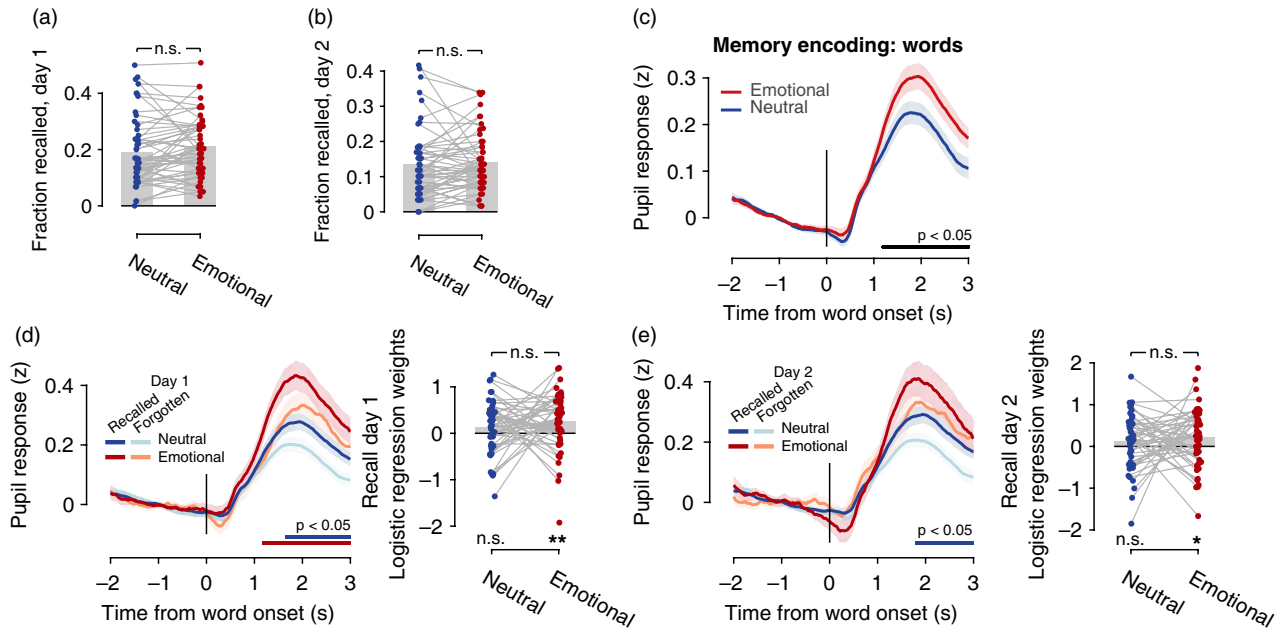


FIGURE 2 Predictive value of pupil dilation for memory formation in the word task. (a) Recall performance in the immediate and (b) 24 hr-delayed memory, expressed as fraction of recalled words. (c) Timecourse of baseline-corrected pupil size in response to auditory word presentation. Negative words elicit a significantly larger pupil response than neutral words. Black line indicates significant timepoints, obtained from a cluster-corrected permutation test. Lines and shaded error regions indicate mean \pm SEM. (d) Left: Pupil timecourses, separately for emotional and neutral words that were recalled or forgotten in the recall test on day 1. Horizontal lines indicate timepoints at which the pupil dilation is significantly different between forgotten and recalled stimuli, as obtained from cluster-corrected permutation test (separately for emotional and neutral stimuli). Right: Individual beta values from logistic regression analyses, indicating that the pupil response during encoding predicts immediate recall of negative but not neutral pictures. (e) Same as panel (d) but for 24 hr-delayed recall test. ** two-tailed $p < 0.01$, * two-tailed $p < 0.05$

shows that the pattern of results was indeed comparable for the immediate and delayed tests.

These results so far established that pupil responses were larger for negative compared to neutral pictures, and that this pupil response to emotional stimuli was, on average, larger for pictures that were subsequently recalled relative to those that were not. We then set out to determine the predictive power of pupil response for subsequent memory on a trial-by-trial basis, using logistic regression (see Method). The pupil response during encoding, across items of different emotionality, was a reliable predictor of trial-by-trial memory in the immediate free recall test (average beta value (SEM): 0.27 (0.04), t test against 0: $t(50) = 6.78$, $p < 0.001$, $d = 1.36$) and in the 24 hr-delayed test (average beta value (SEM): 0.26 (0.04), t test against 0: $t(50) = 6.77$, $p < 0.001$, $d = 1.35$). When taking the emotionality of the pictures into account, pupil dilation predicted subsequent memory for negative pictures (average beta value (SEM) for immediate recall: 0.24 (0.05), t test against 0: $t(50) = 5.75$, $p < 0.001$, $d = 1.15$; for 24-delayed recall: 0.20 (0.04), $t(50) = 5.42$, $p < 0.001$, $d = 1.08$) but not for neutral ones (average beta value (SEM) for immediate recall: -0.09 (0.07), t test against 0: $t(50) = -1.29$, $p = 0.211$, $d = -0.26$; for 24-delayed recall: -0.06 (0.06), $t(50) = -0.82$, $p = 0.418$, $d = -0.16$; Figure 1d and e). These differences in the predictive value of

the pupil response for subsequent memory between neutral and negative pictures were statistically significant (immediate recall: $t(50) = 4.46$, $p < 0.001$, $d = 0.88$; 24 hr-delayed recall: $t(50) = 3.69$, $p = 0.001$, $d = 0.73$). Accordingly, an emotionality \times retention delay ANOVA on the beta values of the regression analysis showed a main effect of picture emotionality ($F_{1,50} = 18.62$, $p < 0.001$, $\eta_p^2 = 0.27$) but no influence of the retention delay on the predictive value of pupil dilation during encoding for subsequent memory (main effect: $F_{1,50} = 0.18$, $p = 0.893$, $\eta_p^2 < 0.01$; emotionality \times retention delay: $F_{1,50} = 1.95$, $p = 0.169$, $\eta_p^2 = 0.04$).

3.2 | Pupil response during encoding predicts subsequent memory for words

The findings from the picture encoding task show that the pupil response was a reliable predictor of trial-by-trial long-term memory, in particular for emotionally arousing pictures. We replicated these findings in a different stimulus modality, that is, auditory encoding of words. Recall performance was not reliably different for neutral and negative words, neither immediately after encoding ($t(52) = 1.60$, $p = 0.115$, $d = 0.31$; Figure 2a), nor after 24 hr ($t(52) = 0.74$, $p = 0.462$, $d = 0.08$; Figure 2b). The finding that the emotional modulation of memory was overall lower (if present at all) for the

words than for the pictures may be due to fact that the learned words were more abstract than the pictures, making emotional words less arousing and less salient than emotional pictures, in combination with the well-known inferior memory for words relative to pictures (Paivio & Csapo, 1973); see also Figures 1a and b, 2a and b). However, also during the encoding of words the pupil response was significantly stronger for negative compared to neutral words ($t(48) = 2.72, p = 0.009, d = 0.56$; Figure 2c). Note that this emotion-related pupil dilation could not be explained by any differences in visual stimulation as items were presented auditorily.

Again, the pupil response was overall stronger for words that were remembered in the immediate ($F_{1,48} = 23.59, p < 0.001, \eta_p^2 = 0.32$) and delayed free recall test ($F_{1,46} = 10.02, p = 0.003, \eta_p^2 = 0.15$) compared to those that were not remembered. This subsequent memory effect was not influenced by the emotionality of the stimuli (subsequent memory \times stimulus emotionality for the immediate recall: $F_{1,48} = 0.09, p = 0.763, \eta_p^2 = 0.00$; for the 24 hr-delayed recall: $F_{1,46} = 0.04, p = 0.840, \eta_p^2 = 0.00$), suggesting that memory for both neutral and negative words was predicted equally well by the pupil response during encoding, as displayed in Figure 2d and e (left panels). Including the factor retention delay (immediate vs 24 hr-delayed) explicitly into the model confirmed that there was, across recall tests, a significant difference in pupil dilation during encoding between subsequently remembered and forgotten pictures ($F_{1,46} = 16.77, p < 0.001, \eta_p^2 = 0.27$), a trend for a difference between neutral and negative pictures ($F_{1,46} = 3.42, p = 0.071, \eta_p^2 = 0.07$), but neither an interaction of stimulus emotionality and subsequent memory ($F_{1,46} = 0.01, p = 0.919, \eta_p^2 < 0.01$), nor an effect of retention delay (main effect: $F_{1,46} = 0.02, p = 0.968, \eta_p^2 < 0.01$; all interaction effects including the factor retention delay: all $F < 1.05$, all $p > 0.311$, all $\eta_p^2 < 0.03$). The absence of any effects of retention delay shows that the pattern of results was indeed comparable for the immediate and delayed tests.

Again, we further used logistic regression to assess the predictive value of the pupil response during encoding for subsequent memory for words on a trial-by-trial basis. Across neutral and negative items, pupil response was a reliable predictor of recall performance in both the immediate (average beta value (SEM): 0.25 (0.05), t test against 0: $t(48) = 5.00, p < 0.001, d = 1.01$) and in the 24 hr-delayed test (average beta value (SEM): 0.22 (0.06), t test against 0: $t(47) = 4.01, p < 0.001, d = 0.82$). When looking at the neutral and negative words, separately, we obtained that the pupil response predicted the immediate (t test against 0: $t(48) = 2.82, p = 0.007, d = 0.58$) and delayed recall (t test against 0: $t(48) = 2.12, p = 0.039, d = 0.43$) of negative words. For neutral words, the pupil response did not significantly predict immediate or delayed recall (immediate recall: t test against 0: $t(47) = 1.66, p = 0.103, d = 0.34$; 24 hr-delayed recall: t test against 0: $t(47) = 1.20, p = 0.235, d = 0.25$). These

differences between neutral and negative words, however, were not statistically reliable (immediate recall: $t(47) = 1.20, p = 0.237, d = 0.25$; 24 hr-delayed recall: $t(45) = 0.49, p = 0.627, d = 0.10$; Figure 2h and i). Accordingly, an emotionality \times retention delay ANOVA on the beta values of the regression analysis showed no main effect of picture emotionality ($F_{1,45} = 0.56, p = 0.459, \eta_p^2 = 0.01$) or retention delay on the predictive value of pupil dilation during encoding for subsequent memory (main effect: $F_{1,45} = 2.59, p = 0.114, \eta_p^2 = 0.05$), nor an interaction of these factors (emotionality \times retention delay: $F_{1,45} = 0.23, p = 0.638, \eta_p^2 = 0.01$).

4 | DISCUSSION

Our results demonstrate that task-evoked pupil responses predict on a trial-by-trial basis which information will be remembered in the long-run, for at least 24 hr and in particular for emotionally arousing information. This effect generalized across visual and auditory encoding tasks, thus allowing us to establish the memory-predictive value of the pupil response across sensory modalities.

The pupil response to task is thought to reflect phasic activity of brainstem neuromodulatory nuclei, including the noradrenergic locus coeruleus (Aston-Jones & Cohen, 2005; Joshi et al., 2016). We propose that the predictive value of the pupil response during encoding for subsequent memory is mainly owing to this link of the pupil dilation to locus coeruleus noradrenergic activity. The role of noradrenaline in memory processes is very well established (Cahill & McGaugh, 1998; Cahill, Prins, Weber, & McGaugh, 1994; Mather, Clewett, Sakaki, & Harley, 2016; McGaugh, 2000, 2015; Sara, 2009; Schwabe, Nader, Wolf, Beaudry, & Pruessner, 2012; Strange & Dolan, 2004). At the neural level, noradrenaline is known to stimulate activity in the amygdala, which then modulates memory processes in other areas, such as the hippocampus (Cahill, Babinsky, Markowitsch, & McGaugh, 1995; Cahill & McGaugh, 1998; Cahill et al., 1996; McGaugh, 2000). Neurophysiological data further indicate that noradrenaline has a critical impact on long-term potentiation and long-term depression (Harley, 2007; Huang, Haganir, & Kirkwood, 2013), key processes of synaptic plasticity underlying memory formation (Bliss & Collinridge, 1993; Ito, 1989). Moreover, noradrenaline facilitates protein synthesis processes promoting long-lasting memory (Cirelli & Tononi, 2000; Gelinis & Nguyen, 2005). Together, these data show the crucial role of noradrenaline in memory formation and we assume that the task-related pupil response may provide a window into the initiation of the noradrenaline-related memory machinery.

Some previous studies have examined the link between pupil dilation and memory (Clewett et al., 2018; Goldinger & Papesh, 2012; Hoffing & Seitz, 2015; Kafkas & Montaldi,

2011; Papesh, Goldinger, & Hout, 2011). Our present findings extend these previous studies in several important ways. First, previous studies have shown that pupil responses, averaged across many trials, differ between memorized and forgotten items. By contrast, we tested here the predictive value of the pupil response at the single-trial level. Doing so is critical for the prediction of specific memories, as well as for evaluating the utility of pupil responses as an easily measurable physiological marker of memory formation.

Second, the current study is, to the best of our knowledge, one of the first to show that the pupil response predicts whether stimuli will be retained in the long run (for, at least, 24 hr). The delays between encoding phase and memory test were confined to less than 30 min in previous work, when memory consolidation, known to take hours (McGaugh, 2000), had (at best) just started. Assessing the stability of pupil-linked memory effects over several delays is important to determine whether the pupil predicts, beyond encoding, also consolidation processes and actual long-term memory and thus to evaluate their real-life behavioral significance. Doing so for two delays a day apart in the current study revealed that pupil responses predicted the immediate and delayed recall equally well. This, in turn, showed that pupil responses are reliable predictors of long-term memories and indicates that pupil-linked arousal mechanisms appear to specifically facilitate the encoding of new memories rather than the memory consolidation processes. This is in line with the idea that the phasic release of modulatory neurotransmitters reflected in pupil dilations (De Gee et al., 2017; Joshi et al., 2016) help memorize information by gating synaptic plasticity mechanisms in the cerebral cortex (Cooke, Komorowski, Kaplan, Gavornik, & Bear, 2015; Roelfsema, van Ooyen, & Watanabe, 2010). It should be noted, however, that the immediate free recall test may well have affected performance in the 24 hr-dealyed test. In particular, there is considerable evidence that the retrieval of an item may foster its subsequent memory (Karpicke & Roediger, 2008; Roediger & Karpicke, 2006). Thus, future studies on the predictive value of the pupil response during encoding for subsequent memory should include both a group that recalls the learned material immediately after encoding and at a later time point and a control group that omits the immediate recall.

Third, while previous studies used mainly recognition tests to assess memory, we here assessed both free recall and recognition performance. Free recall provides more insight into the search in, or retrieval from, memory than recognition, while the latter requires merely the comparison of present information to representations in memory. In fact, in our data, pupil dilation significantly predicted only free recall performance, but not recognition performance (see Appendix S1). This pattern of results might suggest that the arousal reflected in the pupil response aids particularly the search process in memory and less the comparison of information to the internal representation, in line with previous evidence

suggesting that free recall is more sensitive to arousal effects than recognition (Bradley et al., 1992). Alternatively, the absence of an effect on recognition may also be owing to the excellent (near-ceiling) performance in the recognition test.

Beyond spontaneous or cognitive task-evoked variation in pupil size, the pupil dilates in response to emotionally arousing events (Goldinger & Papesh, 2012; Lempert et al., 2015). Indeed, we found here a modulation of the task-evoked pupil dilations by emotional content, in particular in the picture encoding task. In the picture encoding task, pupil dilation predicted memory formation only for stimuli with emotional value, which is generally in line with the view that locus coeruleus noradrenergic activity, thought to be reflected in the pupil response, facilitates in particular high-priority information (Mather et al., 2016). For neutral pictures, however, there was no differential pupil dilation for subsequently remembered vs. forgotten items. This latter finding may point to the role of other factors than arousal in memory formation, such as the level of processing, which may be less well captured by the pupil response. In the word encoding task, however, the pupil response predicted subsequent memory for both, neutral and negative items. This difference between tasks might be due to the fact that words were presented auditorily which triggered already a pupil response, corroborating findings showing that tones may lead to pupil dilation and hence promote subsequent memory (Hoffing & Seitz, 2015). Moreover, pictures are thought to be more salient than words and to elicit emotional arousal more easily than words (Carr, McCauley, Sperber, & Parmlee, 1982). In line with this view, it has been suggested that for pictures emotion may be evoked more rapidly and that regions such as the amygdala respond faster to emotional pictures than they do to emotional words (Gianotti et al., 2008; Kensinger & Schacter, 2006; Kim, Yoon, & Park, 2004). The different potency of emotional pictures and words to elicit emotional arousal and saliency may well have contributed to the reduced (to absent) emotional modulation of memory for words, whereas there was a strong emotional memory enhancement for pictures, as well as to the slightly different findings with respect to the emotionality-specific predictive value of the pupil response for memory.

In conclusion, our findings demonstrate that our eyes may indeed provide a window into the making of (emotional) long-term memories. So far, subsequent memory paradigms have been used in combination with electroencephalography (EEG) or functional magnetic resonance imaging (fMRI) to identify neural predictors of later memory (Paller & Wagner, 2002). Compared to these complex neuroimaging techniques, pupilometry provides an easily accessible and much cheaper index of memory formation, in particular in the face of recently developed mobile eye-tracking devices. Our data suggest that such devices may be used, for instance in therapeutic or educational settings, to achieve a key goal of memory research, to predict which information will be remembered in the future.

ACKNOWLEDGEMENTS

This work was supported by funding from the University of Hamburg (to LS), the German Research Foundation (DFG), DO 1240/2-1, DO 1240/3-1, and SFB 936 project A7 (to THD) and the German Academic Exchange Service (DAAD, to AEU). LS is a member of the FENS-Kavli Network of Excellence.

COMPETING INTERESTS

None of the authors has any conflicts of interest.

DATA ACCESSIBILITY

All raw data, stimulus materials, and Matlab code for data analyses are available at <https://doi.org/10.5281/zenodo.1246100>.

AUTHOR CONTRIBUTIONS

L.S. and T.H.D. designed research; A.B. performed research; A.B., A.E.U., and L. S. analyzed data; A.E.U. and L.S. drafted the manuscript, T.H.D. and A.B. provided critical revisions.

ORCID

Anne E. Urai  <http://orcid.org/0000-0001-5270-6513>

Tobias H. Donner  <http://orcid.org/0000-0002-7559-6019>

Lars Schwabe  <http://orcid.org/0000-0003-4429-4373>

REFERENCES

- Ariel, R., & Castel, A. D. (2014). Eyes wide open: Enhanced pupil dilation when selectively studying important information. *Experimental Brain Research*, *232*, 337–344. <https://doi.org/10.1007/s00221-013-3744-5>
- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–450. <https://doi.org/10.1146/annurev.neuro.28.061604.135709>
- Bliss, T. V. P., & Collinridge, G. L. (1993). A synaptic model of memory: Long-term potentiation in the hippocampus. *Nature*, *361*, 31–39. <https://doi.org/10.1038/361031a0>
- Böcker, M., Gruber, S., & Gauggel, S. (2014). *Aachener Emotionale Wortliste*. Retrieved from https://www.ukaachen.de/fileadmin/files/institute/medizinische-psychologiesoziologie/Aachener_Emotionale_Wortliste.pdf
- Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering pictures: Pleasure and arousal in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 379–390.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. <https://doi.org/10.1163/156856897X00357>
- Cahill, L., Babinsky, R., Markowitsch, H. J., & McGaugh, J. L. (1995). The amygdala and emotional memory. *Nature*, *377*, 295–296. <https://doi.org/10.1038/377295a0>
- Cahill, L., Haier, R. J., Fallon, J., Alkire, M. T., Tang, C., Keator, D., ... McGaugh, J. L. (1996). Amygdala activity at encoding correlated with long-term, free recall of emotional information. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 8016–8021. <https://doi.org/10.1073/pnas.93.15.8016>
- Cahill, L., & McGaugh, J. L. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends in Neurosciences*, *21*, 294–299. [https://doi.org/10.1016/S0166-2236\(97\)01214-9](https://doi.org/10.1016/S0166-2236(97)01214-9)
- Cahill, L., Prins, B., Weber, M., & McGaugh, J. L. (1994). Beta-adrenergic activation and memory for emotional events. *Nature*, *371*, 702–704. <https://doi.org/10.1038/371702a0>
- Carr, T. H., McCauley, C., Sperber, R. D., & Parmlee, C. M. (1982). Words, pictures, and priming: On semantic activation, conscious identification, and the automaticity of information processing. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 757–777.
- Christianson, S. A. (1992). *The handbook of emotion and memory*. Hillsdale, NJ: L. Erlbaum Associates.
- Cirelli, C., & Tononi, G. (2000). Differential expression of plasticity-related genes in waking and sleep and their regulation by the noradrenergic system. *Journal of Neuroscience*, *20*, 9187–9194. <https://doi.org/10.1523/JNEUROSCI.20-24-09187.2000>
- Clewett, D., Huang, R., Velasco, R., Lee, T.-H., & Mather, M. (2018). Locus coeruleus activity strengthens prioritized memories under arousal. *Journal of Neuroscience*, *38*, 1558–1574. <https://doi.org/10.1523/JNEUROSCI.2097-17.2017>
- Cooke, S. F., Komorowski, R. W., Kaplan, E. S., Gavornik, J. P., & Bear, M. F. (2015). Visual recognition memory, manifested as long-term habituation, requires synaptic plasticity in V1. *Nature Neuroscience*, *18*, 262–271. <https://doi.org/10.1038/nn.3920>
- De Gee, J. W., Colizoli, O., Kloosterman, N. A., Knapen, T., Nieuwenhuis, S., & Donner, T. H. (2017). Dynamic modulation of decision biases by brainstem arousal systems. *Elife*, *6*, e23232.
- De Gee, J. W., Knapen, T., & Donner, T. H. (2014). Decision-related pupil dilation reflects upcoming choice and individual bias. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 618–625. <https://doi.org/10.1073/pnas.1317557111>
- de Quervain, D., Schwabe, L., & Roozendaal, B. (2017). Stress, glucocorticoids and memory: Implications for treating fear-related disorders. *Nature Reviews Neuroscience*, *18*, 7–19. <https://doi.org/10.1038/nrn.2016.155>
- 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. *Psychological Science*, *27*, 1632–1643. <https://doi.org/10.1177/0956797616665578>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. <https://doi.org/10.3758/BF03193146>
- Gelinas, J. N., & Nguyen, P. V. (2005). Beta-adrenergic receptor activation facilitates induction of a protein synthesis-dependent late phase of long-term potentiation. *Journal of Neuroscience*, *25*, 3294–3303. <https://doi.org/10.1523/JNEUROSCI.4175-04.2005>
- Gianotti, L. R., Faber, P. L., Schuler, M., Pascual-Margui, R. D., Kochi, K., & Lehmann, D. (2008). First valence, then arousal: The temporal dynamics of brain electric activity evoked by emotional stimuli. *Brain Topography*, *20*, 143–156. <https://doi.org/10.1007/s10548-007-0041-2>

- Goldinger, S. D., & Papesh, M. H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science*, 21, 90–95. <https://doi.org/10.1177/0963721412436811>
- Harley, C. W. (2007). Norepinephrine and the dentate gyrus. *Progress in Brain Research*, 163, 299–318. [https://doi.org/10.1016/S0079-6123\(07\)63018-0](https://doi.org/10.1016/S0079-6123(07)63018-0)
- Hoffing, R. C., & Seitz, A. R. (2015). Pupillometry as a glimpse into the neurochemical basis of human memory encoding. *Journal of Cognitive Neuroscience*, 27, 765–774. https://doi.org/10.1162/jocn_a_00749
- Huang, S., Huganir, R. L., & Kirkwood, A. (2013). Adrenergic gating of Hebbian spike-timing-dependent plasticity in cortical interneurons. *Journal of Neuroscience*, 33, 13171–13178. <https://doi.org/10.1523/JNEUROSCI.5741-12.2013>
- Ito, M. (1989). Long-term depression. *Annual Review of Neuroscience*, 12, 85–102. <https://doi.org/10.1146/annurev.ne.12.030189.000505>
- 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. <https://doi.org/10.1016/j.neuron.2015.11.028>
- Kafkas, A., & Montaldi, D. (2011). Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *Quarterly Journal of Experimental Psychology*, 64, 1971–1989. <https://doi.org/10.1080/17470218.2011.588335>
- Karpicke, J. D., & Roediger, H. L. III (2008). The critical importance of retrieval for learning. *Science*, 319, 966–968. <https://doi.org/10.1126/science.1152408>
- Kensinger, E. A., & Schacter, D. L. (2006). Processing emotional pictures and words: Effects of valence and arousal. *Cognitive, Affective, & Behavioral Neuroscience*, 6, 110–126. <https://doi.org/10.3758/CABN.6.2.110>
- Kim, K. H., Yoon, H. W., & Park, H. W. (2004). Spatiotemporal brain activation pattern during word/picture perception by native Koreans. *NeuroReport*, 15, 1099–1103. <https://doi.org/10.1097/00001756-200405190-00003>
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). *International Affective Picture System (IAPS): Technical manual and affective ratings*. Gainesville, FL: National Institute of Mental Health Center for the Study of Emotion and Attention.
- Lempert, K. M., Glimcher, P. W., & Phelps, E. A. (2015). Emotional arousal and discount rate in intertemporal choice are reference dependent. *Journal of Experimental Psychology: General*, 144, 366–373. <https://doi.org/10.1037/xge0000047>
- Mather, M., Clewett, D., Sakaki, M., & Harley, C. W. (2016). Norepinephrine ignites local hotspots of neuronal excitation: How arousal amplifies selectivity in perception and memory. *Behavioral and Brain Sciences*, 39, e200. <https://doi.org/10.1017/S0140525X15000667>
- McGaugh, J. L. (2000). Memory—a century of consolidation. *Science*, 287, 248–251. <https://doi.org/10.1126/science.287.5451.248>
- McGaugh, J. L. (2015). Consolidating memories. *Annual Review of Psychology*, 66, 1–24. <https://doi.org/10.1146/annurev-psych-010814-014954>
- Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasley, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, 15, 1040–1046. <https://doi.org/10.1038/nn.3130>
- Paivoy, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, 5, 176–206. [https://doi.org/10.1016/0010-0285\(73\)90032-7](https://doi.org/10.1016/0010-0285(73)90032-7)
- Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, 6, 93–102. [https://doi.org/10.1016/S1364-6613\(00\)01845-3](https://doi.org/10.1016/S1364-6613(00)01845-3)
- Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2011). Memory strength and specificity revealed by pupillometry. *International Journal of Psychophysiology*, 83, 56–64.
- Pitman, R. K. (1989). Post-traumatic stress disorder, hormones, and memory. *Biological Psychiatry*, 26, 221–223. [https://doi.org/10.1016/0006-3223\(89\)90033-4](https://doi.org/10.1016/0006-3223(89)90033-4)
- Roediger, H. L. III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Roelfsema, P. R., van Ooyen, A., & Watanabe, T. (2010). Perceptual learning rules based on reinforcers and attention. *Trends in Cognitive Sciences*, 14, 64–71. <https://doi.org/10.1016/j.tics.2009.11.005>
- Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience*, 10, 211–223. <https://doi.org/10.1038/nrn2573>
- Schwabe, L., Nader, K., Wolf, O. T., Beaudry, T., & Pruessner, J. C. (2012). Neural signature of reconsolidation impairments by propranolol in humans. *Biological Psychiatry*, 71, 380–386. <https://doi.org/10.1016/j.biopsych.2011.10.028>
- Schwibbe, M., Räder, K., Schwibbe, M., & Geiken-Pophanken, G. (1981). Zum emotionalen gehalt von substantiven, adjektiven und verben. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 28, 486–501.
- Staresina, B. P., & Davachi, L. (2006). Differential encoding mechanisms for subsequent associative recognition and free recall. *Journal of Neuroscience*, 26, 9162–9172. <https://doi.org/10.1523/JNEUROSCI.2877-06.2006>
- Strange, B. A., & Dolan, R. J. (2004). β -adrenergic modulation of emotional memory-evoked human amygdala and hippocampus responses. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 11454–11458. <https://doi.org/10.1073/pnas.0404282101>
- Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*, 8, 14637. <https://doi.org/10.1038/ncomms14637>
- Wierda, S. M., van Rijn, H., Taatgen, N. A., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 8456–8460. <https://doi.org/10.1073/pnas.1201858109>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Bergt A, Urai AE, Donner TH, Schwabe L. Reading memory formation from the eyes. *Eur J Neurosci*. 2018;00:1–10. <https://doi.org/10.1111/ejn.13984>

SUPPLEMENTARY MATERIAL

Reading memory formation from the eyes

Anne Bergt*, Anne E. Urai*, Tobias H. Donner, and Lars Schwabe

Supplementary methods

Control variables

After their arrival at the lab, participants first completed German versions of the Beck Depression Inventory (BDI; Hautzinger, Bailer, Worall, & Keller, 1994), the Trier Inventory of Chronic Stress (TICS; Schulz, Schlotz, & Becker, 2004), and the State-Trait Anxiety Inventory (STAI; Laux & Spielberger, 1981) to assess their depressive mood, chronic stress level, as well as state and trait anxiety, all of which may affect (emotional) memory processes.

Analysis of recognition data

Data from the recognition experiment on day 2 was quantified as the relative fraction of hits and misses, computed on only the stimuli that were previously presented. Additionally, using all the presented (old and new) stimuli, we quantified recognition performance in terms of signal detection-theoretic d' (Green and Swets, 1966):

$$d' = \Phi^{-1}(H) - \Phi^{-1}(FA) \quad (1)$$

where Φ was the normal cumulative distribution function, H was the fraction of hits and FA the fraction of false alarms. Both H and FA were bounded between 0.001 and 0.999 to allow for computation of d' in case of near-perfect performance (Stanislaw & Todorov, 1999). This measure gives an unbiased estimate of memory performance. Lastly, we used the average confidence rating for correctly recognized stimuli as a measure of memory strength.

The predictive value of the pupil response during encoding for subsequent recognition memory was analyzed in a comparable manner as for the free recall data (see main text). In brief, pupil data were subjected to an ANOVA to test whether pupil size differed for subsequently correctly detected ('hit') and missed neutral and negative items. In addition, logistic regressions testing the predictive value of the pupil response for subsequent recognition performance (hit vs. miss) was performed for all participants and the beta-values from these analyses were analyzed at the group-level.

Supplementary results

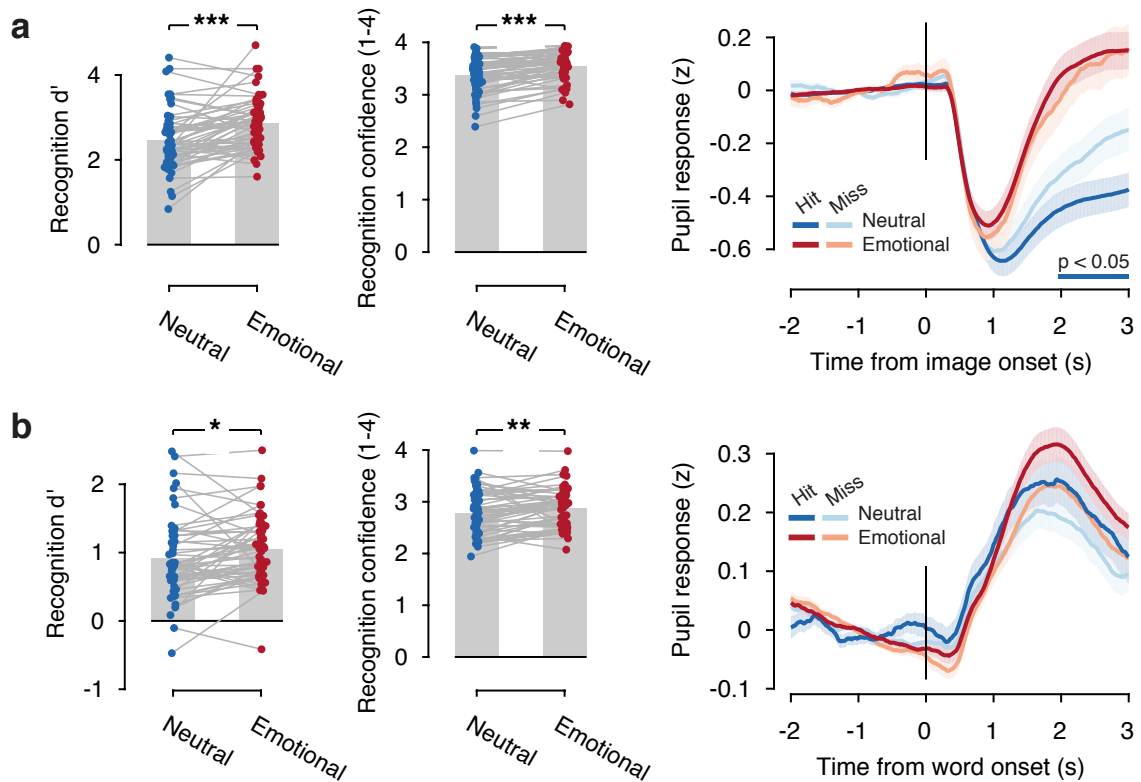
Recognition memory performance for pictures

Recognition performance was significantly better for negative than for neutral pictures (Figure S1a and b), as reflected in a higher d' -score ($t(53) = 4.97$, $p < 0.001$, $d = 0.96$) and higher confidence in correctly identified old pictures ($t(53) = 6.27$, $p < 0.001$, $d = 1.21$).

The subsequent memory analysis showed no differences in pupil size during encoding for those pictures that were correctly recognized and those that were not (hits and misses, respectively) 24 hours later ($F(1,49) = 0.88$, $p = 0.352$, $\eta_p^2 = 0.02$; Figure S1c). Whereas there was no difference in pupil dilation for subsequently identified and missed negative pictures ($t(1,49) = 0.20$, $p = 0.839$, $d = 0.04$), there was even a trend for a larger pupil size for subsequently identified vs missed neutral items ($t(1,49) = -1.95$, $p = 0.057$, $d = -0.39$). In line with these findings, the logistic regression analysis showed no evidence for a prediction of subsequent recognition performance by pupil size, neither overall (average beta-value (SEM): 0.02 (0.06), t-test against 0: $t(50) = 0.38$, $p = .704$, $d = 0.08$) nor for negative pictures alone (-0.07 (0.11), t-test against 0: $t(49) = -0.66$, $p = 0.510$, $d = -0.13$). For neutral pictures, increased pupil size during encoding was even linked to reduced recognition performance 24 hours later (-0.20 (0.09), t-test against 0: $t(50) = -2.20$, $p = 0.032$, $d = -0.44$).

Recognition memory performance for words

The pattern of results for the word recognition was very similar to the pattern observed for picture recognition. Again, negative words were significantly better recognized than neutral ones, as indicated by a higher d' -score ($t(52) = 2.42$, $p = 0.019$, $d = 0.47$) and a higher confidence for correctly identified negative relative to neutral words ($t(52) = 2.22$, $p = 0.031$, $d = 0.43$; Figure S1d and e). Pupil dilation tended to be higher during encoding of subsequent hits compared to subsequent misses, both overall ($F(1,47) = 2.78$, $p = 0.102$, $\eta_p^2 = 0.056$) and for negative pictures, selectively ($t(47) = 1.92$, $p = 0.06$, $d = 0.30$), whereas there was no such effect for neutral words ($t(47) = 0.41$, $p = 0.683$, $d = 0.073$; (Figure S1f). The logistic regression analysis indicated that the pupil size during encoding was a significant predictor of subsequent word recognition (average beta-value (SEM): 0.15 (0.06), t-test against 0: $t(47) = 2.52$, $p = .015$, $d = 0.51$). When analyzed separately, however, there was a trend for a prediction of negative word recognition by the pupil response (0.14 (0.08), t-test against 0: $t(47) = 1.81$, $p = .077$, $d = 0.37$), whereas there was no such effect for recognition of neutral words (0.06 (0.08), t-test against 0: $t(47) = 0.82$, $p = .414$, $d = 0.17$).



Supplementary Figure S1. Recognition memory performance in the picture and word tasks. (a) Recognition memory expressed as d' was better for negative than for neutral pictures. (b) Furthermore, participants were more confident in their responses for negative compared to neutral pictures that were correctly identified as 'old' (hits). (c) For negative pictures, the pupil dilation during encoding did not differ for later recognized vs. not recognized pictures, whereas for neutral pictures the dilation was even larger for subsequently not recognized pictures. (d-f) Same data but for word encoding task. *** $P < .001$, ** $P < .01$, * $P < .05$.

Control variables

Participants' levels of chronic stress, depressive mood, trait and state anxiety were in the normal range of healthy individuals (supplemental Table S1). In order to test whether these parameters were associated with the pupil response during encoding, memory performance and the predictive value of the pupil response for memory, we performed explorative correlational analyses. We obtained negative correlations between both trait anxiety and chronic stress with the predictive value of the pupil dilation for immediate and delayed recall of neutral pictures (all $r < -.28$; all $p < .044$; supplemental Table S2). Furthermore, immediate free recall of pictures and both the immediate and delayed free recall of words tended to be negatively correlated with chronic stress level (all $r < -.25$, all $p < .07$; supplemental Tables S2 and S3). Although these findings dovetail with earlier reports suggesting impairing effects of chronic stress and trait anxiety on memory (Lupien, McEwen, Gunnar, & Heim, 2009; Pajkossy, Keresztes, & Racsmany, 2017), they need to be interpreted with caution because (i) these correlations would clearly not survive a correction for the number of correlations performed and (ii) it remains unclear why these correlations were observed only for some of the stimuli but not for others.

Supplementary references

Green DM, Swets JA (1966) Signal Detection Theory and Psychophysics. John Wiley and Sons.

Hautzinger, M., Bailer, M., Worall, H., & Keller, F. (1994). *Beck-Depressions-Inventar (BDI). Testhandbuch*. Bern: Hans Huber.

Laux, L., & Spielberger, C. D. (1981). *Das State-Trait-Angstinventar: STAI*. Weinheim: Beltz.

Lupien, S. J., McEwen, B. S., Gunnar, M. R., & Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behavior and cognition. *Nature Reviews Neuroscience*, 10(6), 434-445.

Pajkossy, P., Keresztes, A., & Racsmany, M. (2017). The interplay of trait worry and trait anxiety in determining episodic retrieval: The role of cognitive control. *Quarterly Journal of Experimental Psychology*, 70, 2234-2250.

Schulz, P., Schlotz, W., & Becker, P. (2004). *Das Trierer Inventar zum Chronischen Stress*. Göttingen: Hogrefe Verlag.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31, 137-149.

Supplementary table S1. Measures of chronic stress, depressive mood, state and trait anxiety.

<i>Scale</i>	<i>Mean</i>	<i>SEM</i>
Beck Depression Inventory	4.94	0.81
Trier Inventory of Chronic Stress, screening scale (T-score)	46.57	1.68
State-Trait Anxiety Inventory, Trait (T-score)	50.62	1.34
State-Trait Anxiety Inventory, State (T-score)	34.67	0.90

Supplementary table S2. Correlations of chronic stress, depressive, state and trait anxiety with memory, pupil response and predictive value of pupil response for subsequent recall in the picture encoding task.

	Depressive mood	Chronic Stress	Trait anxiety	State anxiety
Immediate recall, neutral	-0.014 <i>0.920</i>	-0.268 <i>0.050</i>	-0.140 <i>0.313</i>	-0.046 <i>0.741</i>
Immediate recall, negative	-0.094 <i>0.500</i>	-0.159 <i>0.251</i>	-0.134 <i>0.336</i>	-0.022 <i>0.873</i>
Delayed recall, neutral	-0.022 <i>0.875</i>	-0.164 <i>0.237</i>	-0.179 <i>0.194</i>	-0.070 <i>0.614</i>
Delayed recall, negative	-0.032 <i>0.819</i>	-0.182 <i>0.188</i>	-0.052 <i>0.709</i>	0.033 <i>0.811</i>
Pupil dilation, neutral pictures	0.025 <i>0.863</i>	0.164 <i>0.249</i>	0.019 <i>0.894</i>	0.010 <i>0.943</i>
Pupil dilation, negative pictures	-0.144 <i>0.313</i>	-0.005 <i>0.973</i>	-0.076 <i>0.596</i>	-0.106 <i>0.460</i>
Prediction immediate recall, negative pictures	-0.80 <i>0.576</i>	-0.055 <i>0.703</i>	-0.111 <i>0.439</i>	0.022 <i>0.876</i>
Prediction immediate recall, neutral pictures	-0.232 <i>0.102</i>	-0.283 <i>0.044</i>	-0.291 <i>0.038</i>	-0.096 <i>0.501</i>
Prediction delayed recall, negative pictures	0.044 <i>0.758</i>	0.072 <i>0.614</i>	0.093 <i>0.515</i>	0.170 <i>0.233</i>
Prediction delayed recall, neutral pictures	-0.243 <i>0.086</i>	-0.363 <i>0.009</i>	-0.294 <i>0.036</i>	-0.168 <i>0.238</i>

Data show Pearson correlations, two-tailed p-values in italics.

Supplementary table S3. Correlations of chronic stress, depressive, state and trait anxiety with memory, pupil response and predictive value of pupil response for subsequent recall in the word encoding task.

	Depressive mood	Chronic Stress	Trait anxiety	State anxiety
Immediate recall, neutral	-0.116 <i>0.409</i>	-0.262 <i>0.058</i>	-0.033 <i>0.817</i>	-0.175 <i>0.210</i>
Immediate recall, negative	-0.180 <i>0.196</i>	-0.285 <i>0.039</i>	-0.196 <i>0.161</i>	-0.261 <i>0.059</i>
Delayed recall, neutral	-0.189 <i>0.175</i>	-0.439 <i>0.001</i>	-0.155 <i>0.266</i>	-0.249 <i>0.072</i>
Delayed recall, negative	-0.074 <i>0.600</i>	-0.252 <i>0.069</i>	-0.087 <i>0.534</i>	-0.132 <i>0.348</i>
Pupil dilation, neutral pictures	-0.128 <i>0.388</i>	-0.284 <i>0.050</i>	-0.185 <i>0.207</i>	-0.093 <i>0.529</i>
Pupil dilation, negative pictures	0.002 <i>0.992</i>	-0.075 <i>0.611</i>	0.126 <i>0.392</i>	0.038 <i>0.798</i>
Prediction immediate recall, negative pictures	-0.014 <i>0.922</i>	0.045 <i>0.761</i>	0.224 <i>0.126</i>	0.202 <i>0.168</i>
Prediction immediate recall, neutral pictures	-0.068 <i>0.645</i>	0.115 <i>0.435</i>	-0.128 <i>0.385</i>	-0.107 <i>0.470</i>
Prediction delayed recall, negative pictures	0.040 <i>0.787</i>	0.074 <i>0.618</i>	0.116 <i>0.430</i>	0.239 <i>0.102</i>
Prediction delayed recall, neutral pictures	-0.108 <i>0.473</i>	0.115 <i>0.446</i>	-0.191 <i>0.205</i>	-0.168 <i>0.264</i>

Data show Pearson correlations, two-tailed p-values in italics.