Emotional arousal and recognition memory are differentially reflected in pupil diameter responses during emotional memory for negative events in younger and older adults

Hämmerer, D * [1,2,3], Hopkins, A * [1], Betts, MJ.[3,4], Maaß, A. [3,5], Dolan, RJ [2,6] and Düzel, E. [1,3,4]

1. Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London, WC1N3AZ
2. The Wellcome Trust Centre for Neuroimaging, University College London, 12 Queen Square, London, WC1N 3BG, UK
3. German Center for Neurodegenerative Diseases (DZNE), Magdeburg, Germany
4. Institute of Cognitive Neurology and Dementia Research, Otto-von-Guericke University Magdeburg, Leipziger Str. 44, 39120 Magdeburg, Germany
5. Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA, United States
6. Max Planck – UCL Centre for Computational Psychiatry and Ageing Research, Russell Square House, 10-12 Russell Square, London, WC1B 5EH, UK

* contributed equally

corresponding author:

d.hammerer@ucl.ac.uk

Dorothea Hammerer

17 Queen Square

WC1N3AR London, UK

Abstract

A better memory for negative emotional events is often attributed to a conjoint impact of increased arousal and noradrenergic modulation. A decline in noradrenergic modulation (NA) during ageing is well documented but its impact on memory function during ageing is unclear. Using pupil diameter (PD) as a proxy for noradrenergic modulation, we examined age differences in memory for negative events in younger (18-30 years) and older (62-83 years) adults based upon a segregation of early arousal to negative events, and later retrieval related PD responses. In keeping with the hypothesis of reduced age-related NA influences, older adults showed attenuated induced PD responses to negative emotional events. The findings highlight a likely contribution of noradrenergic modulation to negative emotional memory, mediated via arousal that may be compromised with ageing.

Keywords:

emotional memory, pupillometry, aging, noradrenaline

1. Introduction

Better memory for emotionally arousing events is well documented. Emotional stimuli ranging from stories (Heuer and Reisberg, 1990), to words (Phelps et al., 1998), film clips (Cahill et al., 1996) and pictures (Bradley et al., 1992) are all related to improvements in long-term memory recall. This emotional enhancement was suggested to be due to an effect of arousal, as this emotional bias extends to both positively and negatively valenced stimuli (Bradley et al., 2008; Garavan et al., 2001; Mather, 2007). Negative emotional arousal is accompanied by activation of the locus coeruleus (LC) and concomitant release of the neurotransmitter noradrenaline (NA) that serves to improve memory for the negative event by mediating long-lasting synaptic plasticity in the medial temporal lobe (Klukowski and Harley, 1994; Sara, 2009). In line with this role of noradrenaline in negative emotional memory, memory enhancement for negative emotional items has been found to be abolished with administration of a beta-adrenergic antagonist, propanolol, during early retrieval or encoding (Kroes et al., 2010; Strange and Dolan, 2004).

By the age of 60, older adults have lost between 20-40% of the neurons in the LC (Mann, 1983; Vijayashankar and Brody, 1979). This cell loss might occur as a correlate of healthy ageing or might reflect a presymptomatic reduction in LC integrity related to tau pathology (Mather and Harley, 2016). The age-related decline in noradrenergic modulation is expected to contribute to cognitive decline during healthy ageing (Arnsten and Golman-Rakic, 1985; Arnsten and Goldman-Rakic, 1985). Although this has been a prevalent hypothesis in the field for several decades, very little progress has been made in addressing this question. In particular the encoding and consolidation of declarative long-term memory is affected in

healthy ageing (Nyberg et al., 2012). This is the case for more complex episodic memory contents as well as negative emotional memory events (Jacques et al., 2009; Naveh-Benjamin et al., 2003). A recent study in rats showed that an age-related deficit in negative emotional memory formation is accompanied by reduced levels of extracellular noradrenaline and can be attenuated by administrating noradrenaline or blocking noradrenaline reuptake (Luo et al., 2015). Here we investigated whether physiological indicators of LC activation can inform age differences in emotional memory for negative events.

Given the LC's small size and its location deep within the brainstem, it is difficult to obtain noninvasive recordings of its activity using approaches such as functional neuroimaging. However, (Samuels and Szabadi, 2008) observed that when monkey LC neurons were stimulated, pupil diameter (PD) increased in parallel to its firing rate. Similarly, Joshi et al. (2016) used electrical microstimulation of LC in monkeys to show that phasic LC activation produces robust changes in PD. PD can therefore be taken as an indicator of LC activation and serve as a proxy measure for LC firing to salient, arousing (Chen and Sara, 2007) or task-relevant events (Aston-Jones et al., 1994).

In line with the above, larger PD have been consistently observed in response to emotional events such as the viewing of pleasant and unpleasant stimuli, relative to neutral stimuli (Bradley et al., 2008). This effect is seen across valences (and modalities), and is therefore analogous to the noradrenergically mediated modulation driven by bottom-up stimulus properties as seen in emotional memory findings (Partala and Surakka, 2003; Sara, 2009).

Studies investigating PD during memory encoding or recognition tasks have furthermore consistently shown larger PD responses to old as opposed to novel stimuli (recognition effect, also known as familiarity effect) (Heaver and Hutton, 2011; Otero et al., 2011; Võ et al., 2008). These larger PD responses to old compared to new stimuli have been attributed to increased effort necessary for memory retrieval (Võ et al., 2008). However, an alternative strength-of-memory trace account (Otero et al., 2011) provides contradictory evidence as the PD to old stimuli was larger for deeply encoded items which would seem at odds with an increased retrieval effort account. Given LC-NA activation in occurrence with goal-oriented target stimuli (such as old stimuli in a recognition test), the old/new effect could therefore be a combination of task-goals and also memory strength which reflect the saliency of old stimuli in a recognition task.

The present study aimed to examine age differences in PD responses during emotional memory while disentangling opposing views on the old/new recognition effect in PD responses. An important goal for our study was to separate PD responses associated with retrieval success from PD responses associated with different levels of emotional arousal and to explore age differences in both processes attributed to noradrenergic modulation.

Finally, some studies predict that PD at an encoding stage should predict subsequent memory accuracy (Papesh et al., 2012), yet findings remain inconsistent with others reporting no relationship between PD responses at encoding and subsequent remembering (Võ et al., 2008), or that constriction, as opposed to dilation predicts higher memory recall (Kafkas and Montaldi, 2011). Thus, an

additional aim was to investigate subsequent memory in the presence or absence of arousal-based modulation.

2. Methods

2.1. Participants

44 participants took part in the study comprising 22 healthy younger adults (15 female, aged between 18 and 30, mean 24 years) and 22 healthy older adults (11 female, aged 62-83, mean 71 years). Younger adults were recruited using the Institute of Cognitive Neuroscience (ICN) subject database, older adults were recruited using advertisements in local newspapers and via flyers. All participants had normal or corrected-to-normal vision and no history of any psychiatric disorders. Informed written consent was gained from each participant and reimbursement was set at £8 per hour. The study was approved by the local ethics committee (UCL Research Ethics Committee reference 5506/001). Three older adults had to be excluded from the analyses. One due to poor performance in the recognition tests, one due to eyesight problems that prevented proper engagement with the task, and one due to unwillingness to complete the task. The final sample therefore comprised 41 participants, 22 of which were younger and 19 older adults. Moreover, 4 participants' (2 younger adults and 2 older adults) performance on the first recognition test was more than 2 standard deviations lower (Hit-False Alarm on first recognition close to 0%) than that of the rest of the participants due to difficulties in understanding the instruction on the early recognition test. Note that this low performance was not due to general difficulties of understanding the task or overall lower recognition memory, as their performance was in a normal range on the delayed recognition test. We therefore replaced behavioral as well as pupil data on this first test with the group mean of their respective age group for these 4 participants. Replacing the data in these 4 participants did not affect analyses as

control analyses showed that all statistically reliable results were robust to excluding these four participants.

2.2 Materials and stimuli

The stimuli consisted of 120 indoor and outdoor pictures containing negative emotional or neutral scenes partly taken from the International Affective Picture System (IAPS) database (Lang and Bradley, 2007; $N_{negative}$=48, $N_{neutral}$=27) and partly taken from an image set collected from the internet ($N_{negative}$=72, $N_{neutral}$=93). The internet-based image set was built as part of a different study and was rated on valence and arousal by a sample of 60 young adults (mean age = 28±2 yrs, 50% female) Therefore, the Self-Assessment Manikin (SAM) was used, which is an affective rating system devised by (Lang, 1980) that also underlies IAPS ratings. For each image to be rated, subjects could select from a 9-point rating scale with 9 representing a high rating on each dimension (i.e., high pleasure, high arousal), and 1 represents a low rating on each dimension (i.e., low pleasure, low arousal). The rating was performed on a total of 387 novel images collected from the internet as well as 45 IAPS images. Ratings on the IAPS images derived from our sample of young adults did not differ from the established ratings available in the IAPS database ($p>0.18$), such that both databases could be merged. For the current study, we chose negative emotional pictures with low valence (M=2.85, SD=0.45) and moderately high arousal (M=5.87, SD=0.73), and neutral pictures with neutral valence (M=5.26, SD=0.45) and low arousal (M=3.19, SD=0.55).

Stimuli were displayed on a 22inch monitor and viewed from a distance of 80 cm. The stimuli, fixation crosses and grey patterned background were luminance adjusted in order to control for trivial luminance-related effects on PD responses.

Stimuli and fixation crosses were displayed in the centre of the screen (cf. Figure 1). The text was presented in size 50 Arial font and coloured white. Participants sat on a comfortable chair, with their head position stabilized with a desktop-mounted chin and headrest. Two four-choice button boxes were used to record responses during the task.

2.3 Experimental Procedure

Participants were invited to attend for a double testing session. Incidental Encoding test and Early Recognition test were performed in the morning and a Delayed Recognition test was performed 6 hours later on the afternoon of the same day. Participants were familiarized with the use of the button boxes and encouraged to reduce blinking while a picture was displayed on the screen. Prior to every eye tracking recording, the eye-tracker was calibrated to the participant's right eye using a standard 9-point or 6-point calibration and validated using the same calibration setup.

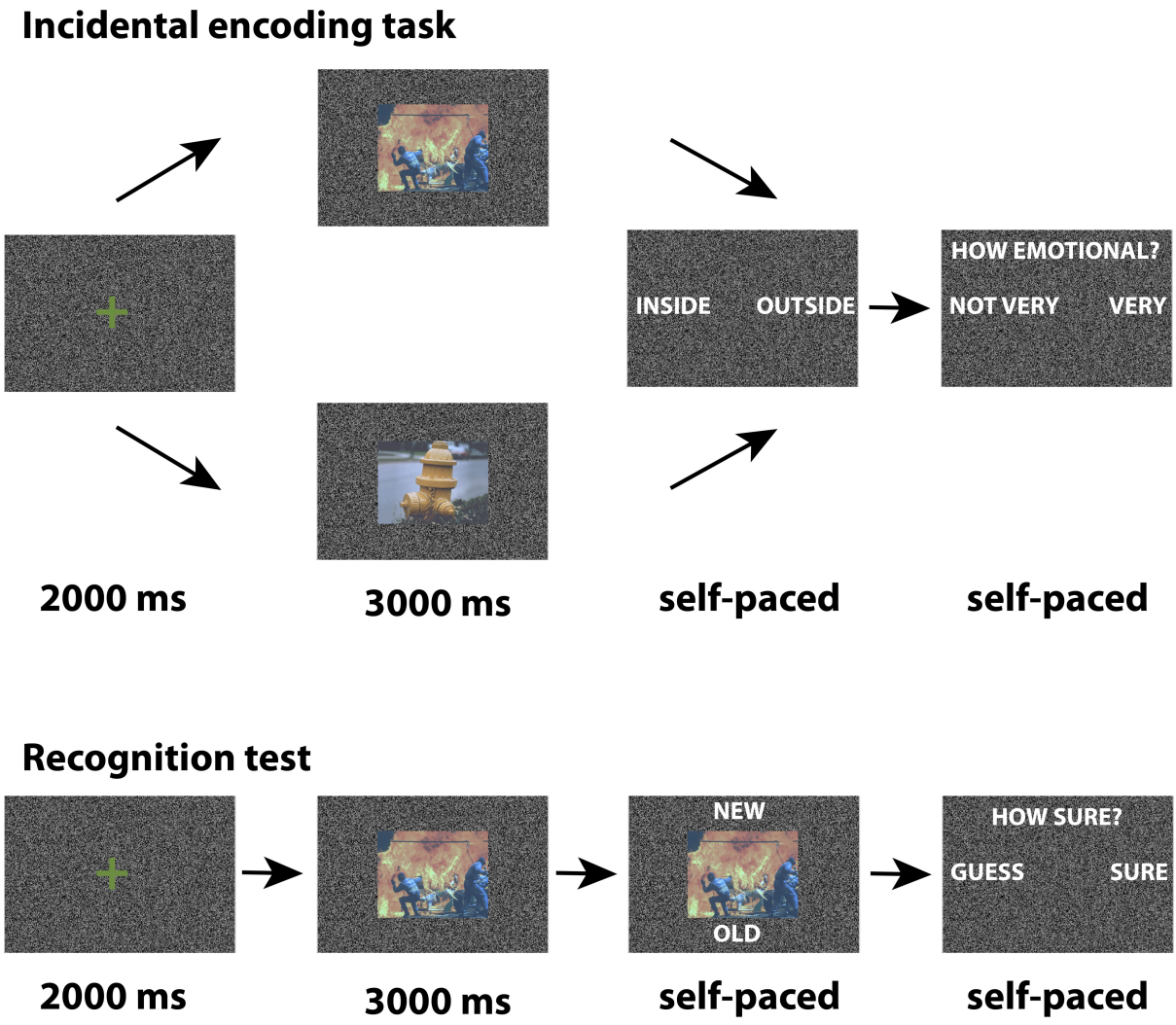


Figure 1. Incidental emotional memory task. During incidental encoding (upper plot), participants classified neutral and negative emotional stimuli as indoor or outdoor scenes and indicated how emotional the picture made them feel on every trial. During early and delayed recognition tests (lower plot), participants classified pictures as new or old and indicated how certain they were in their memory judgement on a four-point scale (four buttons, with certainty increasing from left to right). Responses were withheld during the initial 3 seconds of stimulus presentation. Background, fixation cross and scene stimuli were adjusted in luminance in order to avoid changes of pupil diameter due to stimulus luminance.

2.3.1 Incidental Encoding Task

The incidental encoding task consisted of 120 trials, with a short break after 60 trials. As a cover task, participants were asked to classify scenes as indoor or outdoor. Each trial began with a grey fixation cross, followed by a picture showing a neutral or negative inside or outside scene. The amount of inside, outside, negative and neutral scenes were balanced yielding 30 trials per emotional and stimulus category.

After the picture had disappeared, participants were prompted to classify the picture as indoor or outdoor and then indicate using a 4-point scale (four buttons) how emotional the picture made them feel ('not very emotional, somewhat emotional, quite emotional or very emotional'). This provided a subjective measure of negative emotionality of the pictures for manipulation evaluation as well as assessment of interindividual differences in perceived emotionality. Button positions were matched to the display to facilitate responding for the older adults in particular.

Inside/outside responses, and emotional ratings, were self-paced and could only be given after the picture disappeared and the respective question was displayed in order to avoid confounds of PD responses to stimuli and response execution.

2.3.2 Early and Delayed Recognition Test

Each test consisted of 120 trials, with a short break after 60 trials. Trials for each test were separated into 60 old (from the incidental encoding task) and 60 new pictures balanced for inside and outside as well as negative and neutral scenes. Trial

structure and timing was kept similar to the encoding task. However, in this phase the pictures did not disappear after 3000ms and stayed on the screen when participants were prompted to classify pictures as old or new as well as how certain they were in the old or new classification on a 4-point scale (four buttons).

2.4 Acquisition and preprocessing of pupillometric data, statistical analyses

Pupillometric and eye-tracking data were continuously recorded from the right eye using an infrared EyeLink 1000 Desktop Mount (SR Research), with a sampling rate of 1000Hz and luminance levels of 100%. Pupil diameter (PD) was acquired using the centroid measure in order to provide more accurate estimates of changes in PD over time. Both eye-tracking and pupil data were analysed using custom-made scripts in MATLAB 2015b and the software toolbox FieldTrip (Version 2, http://www.fieldtriptoolbox.org/), implemented in MATLAB.

PD data were segmented in a time window of 500ms before and 3000ms after picture onsets. A custom-made filter was used to detect eye-blinks in segmented data based on changes in pupil diameter. Periods of missing data due to blink-related eye artifacts were cut out in time windows of 200ms and 30ms around large and small artifacts, respectively, and replaced by linear interpolation. Trials with excessively noisy or missing data were excluded (on average 15 trials in younger and 24 trials in older adults). Pupil data were then baseline corrected in a time window -200 to 0ms before picture onset and z-scored per individual to allow comparing task conditions independent of individual differences in PD size (Nassar et al. 2012). RTs +/- 3.5 standard deviations from the mean were excluded from RT analyses.

Negative emotionality and recognition effects were first examined in repeated measures ANOVAs with age groups as a between subjects factor. For these analyses, PD responses were averaged in a time window of 1 to 2 seconds after stimulus onset. Prior studies have shown that an old-new recognition effect on pupil data is independent of correct and incorrect answers (i.e. hits and misses or correct recognitions and false alarms) (Kafkas and Montaldi, 2015), suggesting that it reflects unconscious or implicit components of episodic memory. We therefore included test trials with correct as well as incorrect (recognition) responses in our analyses, but additionally assessed whether results were robust to excluding incorrect trials.

We then used GLM analyses to examine a concurrent effect of negative emotionality and recognition on PD, as well as to test for the time course of these effects while controlling for interindividual differences in the intercepts. We included interactions between negative emotionality and recognition as well as recognition and certainty of responses as repeated measures ANOVAs showed a modulation of negative emotionality and recognition effect with old/new stimuli and certainty of response, respectively (see paragraph above as well as Results below). Thus, the entered trial-wise regressors were negative emotionality of the stimuli, the old/new status of stimuli, certainty of old/new responses, retention interval (early or delayed), old/new and negative emotionality interaction and old/new and certainty interaction (as six concurrent logistic regressors in a GLM per person, predicting PD per trial). GLMs were calculated separately for every sampling point. At the individual level, group-level GLMs tested for differences between age groups in trial-wise effects and for additional variation of the trial-wise effects with other behavioral variables of interest (e.g. modulation of old/new effect on pupil data by mean memory accuracy

per person). Time points of significant differences in PD between conditions as well as between age groups were assessed with permutation analyses (compared against time series of effects resulting from randomly shuffled condition labels across trials for the individual logistic regressors (100 repetitions) or participant IDs for GLMs across individuals (1000 repetitions), respectively). On every permutation, the largest beta in a time window of 0.5 to 2.5 seconds after stimulus presentation was saved. Effects on unpermuted data were considered significant if the largest beta exceeded the 95% or 90% percentile of the distribution of maximal betas on permuted data. In this manner, we provided significance tests corrected for multiple comparisons against sequences of PD responses which did not differ in the extent of autocorrelation (Cohen, 2014). Time periods of significant condition or group-level effects are indicated as lines below pupil effects (cf Figure 3). To assess whether negative emotionality effects manifested earlier than recognition effects, we extracted peaks of the respective betas per individual in a time window 0.75 to 2.0 seconds after stimulus onsets. Differences in the timing of negative emotionality and recognition peak effects were assessed using non-parametric tests, given non-normality of the data.

In analyses of tonic PD responses for subsequent memory effects, data were segmented and cleaned as for the phasic data and were then z-scored across all concatenated trials (Nassar et al., 2012). In this manner, trial differences in the mean (or tonic) pupil response could be analyzed independently of individual differences in overall pupil size. The analysis interval was -0.2 to 2 seconds around stimulus onsets. Five younger adults and 2 older adults had too few not remembered trials and had to be excluded from analyses, resulting in a final sample for the analyses of tonic data of 17 younger adults and 17 older adults.

3. Results

3.1 Behavioural Results: Performance on incidental encoding task

Both age groups attended to the cover task (in-out classification), with the younger (M = 96, SD = 0.03) and older adults (M = 96, SD = 0.03) both showing high accuracy ratings. There was no significant difference between the two age groups $t(39) = 0.67$, $p = 0.42$, indicating both age groups followed the instructions relevant for incidental encoding. Subjective emotionality ratings (4-point scale) of the stimuli reflected negative and neutral stimulus types. Both age groups rated the negative pictures (M = 2.45, SD = 0.59) higher than the neutral pictures (M = 1.22,SD = 0.20; $t(40) = 16.862$, $p < 0.05$, $r = 0.51$) with no age difference in emotionality ratings for either type of stimulus (ratings for negative pictures $t(39) = .71$, $p = 0.49$, neutral pictures $t(39) = 0.29$, $p = 0.77$). Age groups were hence comparable in terms of elicited emotional responses, as evident in self-reports.

Table 1

	Younger Adults			***Older Adults***		
	Hits	*False Alarms*	*Hits – False*	*Hits*	*False Alarms*	*Hits – False*
	Mean(SD)	*Mean(SD)*	*Alarms*	*Mean(SD)*	*Mean(SD)*	*Alarms*
			Mean(SD)			*Mean(SD)*
Early Test	.96(.05)	.06(.05)	.91(.08)	.95(.07)	.06(.05)	.88(.10)
Negative Stimuli						
Early Test	.91(.08)	.03(.04)	.88(.09)	.91(.06)	.03(.06)	.87(.08)
Neutral Stimuli						
Delayed Test	.90(.07)	.05(.06)	.84(.09)	.86(.11)	.14(.13)	.72(.16)
Neutral Stimuli						
Delayed Test	.84(.13)	.07(.07)	.77(.14)	.75(.18)	.10(.10)	.65(.18)
Neutral Stimuli						

Table 1. Descriptive Statistics for Hits and False alarms (FA) on early and delayed recognition tests for both age groups.

3.2 Behavioural Results: Early and Delayed Recognition Tests

3.2.1 Overall accuracy Hit-FA

In line with the hypothesis that negative events improve memory accuracy (assessed as hits-false alarms), we found a main effect of negative emotionality on memory accuracy ($F(1,39) = 13.64$, $p < 0.05$, $rICC = 0.51$), with higher accuracy to negative stimuli ($M = .84$, $SD = .09$) than to neutral stimuli ($M = .79$, SD = .11). Furthermore, a main effect of test interval; ($F(1,39) = 71.72$, $p < 0.05$, $rICC = 0.81$) showed an expected reduced memory performance with delayed recognition testing.

In line with negative stimuli improving long-term memory accuracy in particular, we found an negative emotionality by test interval (early/delayed) interaction ($F(1,39) = 71.72$, $p < 0.05$, $rICC = 0.80$). We also found evidence for a reduced memory performance in older adults ($F(1,39) = 5.22$, $p < .05$, $rICC = 0.34$), as well as a test interval by age interaction ($F(1,39) = 10.28$, $p < 0.05$, $rICC = 0.46$) indicating older adults' memory was more compromised in a delayed recognition test. This confirms reduced long-term memory performance in older compared to younger adults. We did not observe an age X negative emotionality interaction. However, when recognition trials with high certainty alone were considered (certainty rating > 3), we observed a trend for younger adults to profit more from negative emotional content in their recognition memory (age x negative emotionality interaction $F(1,38) = 3.02$, $p = 0.09$, $rICC = 0.27$). We did not observe a significant 3-way interaction between negative emotionality, test interval and age group indicating that both age groups were affected similarly by negative emotional memory content. This might reflect insufficient task difficulty needed to allow for a differential modulation of memory performance by negative stimulus emotionality in the two age groups (see Discussion).

Finally, research on time of day effects in testing suggest that in particular older adults' recognition performance can be affected by being tested at suboptimal times (Hasher et al., 2005; May et al., 1993). In our study, this effect might have contributed to the worse recognition performance of older adults on delayed memory tests in the afternoon. To gauge the extent of this effect, we compared recognition memory as assessed via hits or false alarms. The latter have been shown to be particularly sensitive for time of day effects (May et al., 1993). Indeed, we did

observe a somewhat weaker age x test interval effect for hits ($F(1,39) = 3.84$, $p = 0.06$, $rICC = 0.30$) as compared to false alarms ($F(1,39) = 4.91$, $p < 0.05$, $rICC = 0.33$). However this difference was not reliable (age x hit versus false alarm x test interval $F(1,39) = 0.34$, $p = 0.86$).

In summary, these results show expected effects of negative emotion increasing memory (for both age groups, and marginally more so in younger adults) and a reduction in long-term memory performance in older adults, although the latter was not modulated by negative emotion. These effects were observed in the absence of age differences in cover task performance or emotionality ratings (Windmann and Kutas, 2001).

3.2.2 Certainty ratings

Certainty ratings (on a scale from on 1 to 4) were obtained after categorizing old/new stimuli showed both age groups were more certain in the early ($M = 3.68$, $SD = .32$) compared to the delayed test ($M = 3.43$, $SD = .44$, $F(1,39) = 36.73$, $p < 0.05$, $rICC = 0.70$), as expected given the lower memory performance on delayed recognition. There was no difference in certainty ratings for negative and neutral stimuli and no difference in certainty ratings between the younger and older age groups for both the early and delayed recognition tests. Certainty ratings and memory performance (hits-false alarms) were overall positively correlated across participants (part1: younger adults: $r = .27$, $p = .23$; older adults: $r = .42$, $p = .07$; part2: younger adults: $r = .43$, $p < .05$; older adults: $r = .52$, $p < .05$; the correlations between certainty ratings and memory performance did not differ between age groups).

3.2.3 Reaction times for old or new and negative or neutral stimuli

We observed longer RTs for negative stimuli compared to neutral stimuli, independent of the question responded to $F(1,39) = 52.76$, $p < .05$, $rICC = 0.76$.This was seen in the encoding task (indoor – outdoor classification: $F(1,39) = 33.34$, $p < .05$, $rICC = 0.68$, emotionality ratings: $F(1,39) = 19.02$, $p < .05$, $rICC = 0.57$) as well as early ($F(1,39) = 8.07$, $p < .05$, $rICC = 0.41$) and late ($F(1,39) = 8.24$, $p < .05$, $rICC = 0.42$) recognition tests. Older adults also showed slower reaction times overall ($F(1,39) = 55.51$, $p < .01$, $rICC = 0.77$) and showed more slowing to negative stimuli than younger adults $F(1,39) = 8.33$, $p < .05$. This consistent effect of longer reaction times to negative stimuli, independent of whether the negative emotionality of the stimuli is relevant for the response or not, is consistent with the effect of an attentional capture by negative stimuli (Schwartz et al, 2003).

There was also an overall main effect of old/new stimuli on RTs across tasks, $F(1,39) = 7.04$, $p < .05$, $rICC = 0.39$, with longer RTs for new stimuli ($M = .77$, $SD = .26$) than for old ($M = .71$, $SD = .26$). A similar effect has been previously reported in a recognition memory task (Kafkas and Montaldi, 2015). Longer RTs to new as compared to old stimuli suggest that participants focused in particular on old stimuli when tasked to classify stimuli as old or new (see below for a more detailed discussion of this effect).

Table 2

	Younger Adults Neutral	**Younger Adults Negative**	**Older Adults Neutral**	**Older Adults Negative**
	Mean(SD)	*Mean(SD)*	*Mean(SD)*	*Mean(SD)*
Task (indoor / outdoor)	.80(.33)[a]	1.03(.41)[a]	.92(.23)[a]	1.11(.30)[a]
Task (emotionality rating)	.63(.37)[ab]	.78(.52)[ab]	1.40(.52)[ab]	2.02(1.01)[ab]
Early Test (old/new classification)	.51(.14)[b]	.55(.19)[b]	1.02(.61)[ab]	1.2(.87)[ab]
Delayed Test (old/new classification)	.55(.23)[b]	.59(.23)[b]	1.02(.44)[ab]	1.14(.39)[ab]
Early Test Certainty Rating	.42(.14)[b]	.44(.19)[b]	.88(.26)[b]	.88(.26)[b]
Delayed Test Certainty Rating	.40(.19)[b]	.40(.19)[b]	.94(.31)[b]	.98(.35)[b]

Table 2. Reaction times (in seconds) following negative or neutral stimuli in younger and older adults (mean across individuals, median per individual). Overall, reaction times were longer after negative stimuli. (a) indicates a significant difference within age groups between reaction times after negative and neutral stimuli and (b) indicates a significant difference between age groups in reaction times to neutral or negative stimuli.

3.3 Pupil Results

3.3.1 *Comparability of PD measures between age groups*

Younger and older adults had similar trial numbers for all tasks following artefact correction ((*M* = 105.89, *SD* = *10.22*, *M* = 96.07, *SD* = 19.61, respectively, $F(2, 78) = 2.56$, $p = 0.12$), suggesting a comparable quality of pupil measurements in younger and older adults. Similarly, younger and older adults had similar levels of baseline noise, assessed as the mean across trials of the standard deviation in the

first 500ms of pupil data taken on individual trials, $F(1, 39) = 3.06$, $p = .09$). On this basis we would not expect to see age differences in pupil data solely arising out of age differences in PD measurement noise.

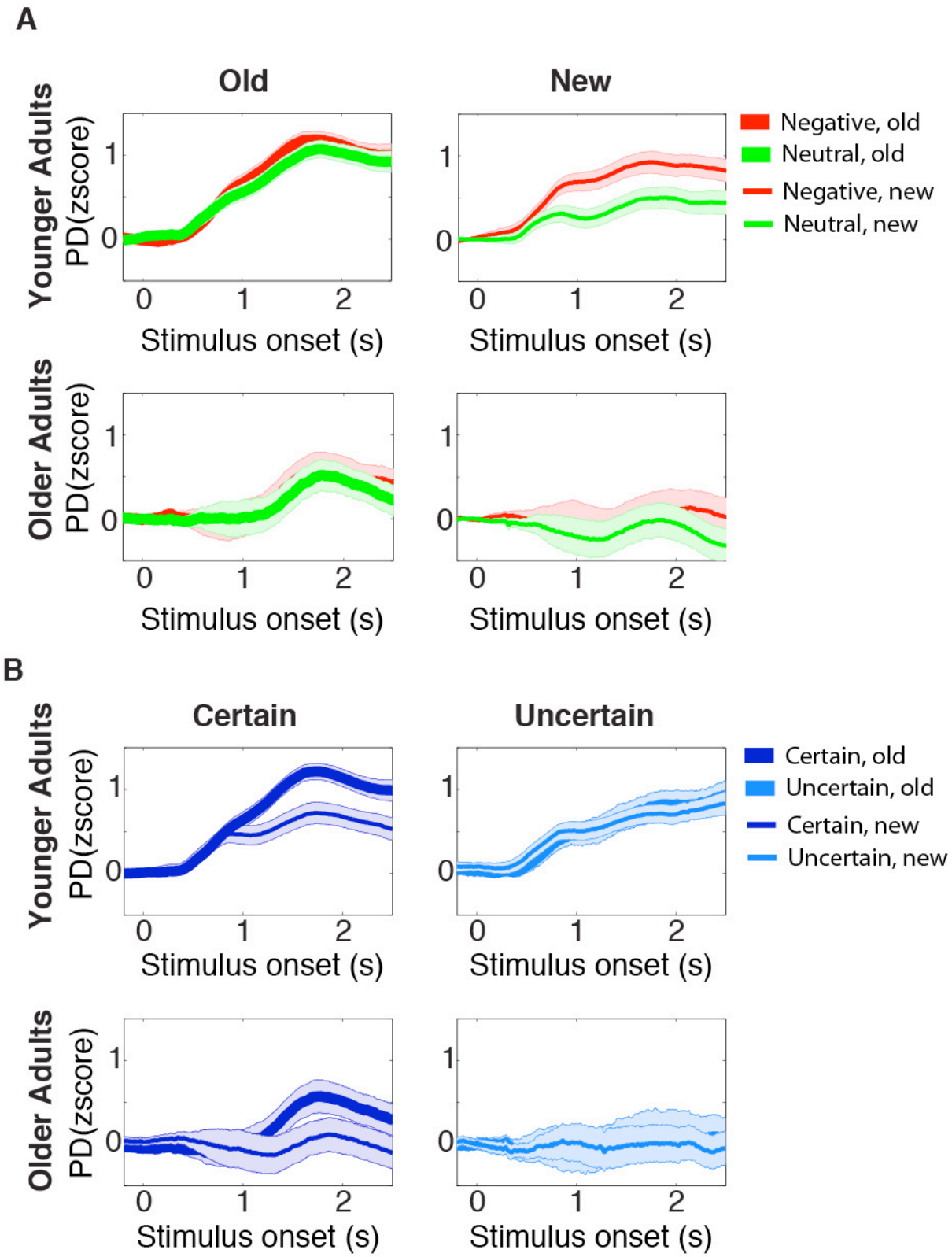


Figure 2. Mean PD responses related to negative emotionality and recognition of stimuli in recognition tests (collapsed across early and delayed recognition). A) PD was increased for negative stimuli (red). The effect of negative emotionality on PD was particularly pronounced for new stimuli and in younger adults (thin line, e.g. upper right plot). B) Pupil diameter was increased for old stimuli (thick line). Memory certainty (dark blue) modulated PD further, but only for old stimuli (thick line, e.g. upper left plot). A) and B) Older adults showed an overall reduced responsiveness in PD. Shaded error bars represent standard error (across individuals within age groups).

3.3.2 Repeated measures ANOVAs of negative emotionality and recognition effects in pupil data

Cleaned pupillometric data were analysed with respect to condition and age differences in mean PD responses in a time window 1 to 2 seconds after stimulus onsets across the tasks in two separate repeated measures ANOVAs. The first analysis examined a four-way interaction of negative emotionality and recognition effects across the two recognition delays as well as age differences therein (Figure 2A). Younger adults had overall larger PD responses than older adults ($F(1,39) = 9.57$, $p < 0.05$, $rICC = 0.44$). As predicted given higher levels of noradrenergic modulation during negative emotional events, there was a main effect of negative emotionality on mean PD ($F(1,39) = 23.49$, $p < 0.05$, $rICC = 0.61$), with negative stimuli eliciting a larger mean PD ($M = .55$, $SD = .70$) than neutral stimuli ($M = .36$, $SD = .64$) (Figure 2A). This was modulated by age group ($F(1,39) = 5.67$, $p < 0.05$, $rICC = 0.36$), with younger adults showing a larger reaction to negative stimuli than older adults (Figure 2A, upper right plot). There was also the expected main effect of old/new stimuli on PD ($F(1,39) = 65.89$, $p < 0.01$, $rICC = 0.79$), with larger PD responses observed to old stimuli ($M = .64$, $SD = .70$) than to new ($M = .27$, $SD = .83$), which was not modulated by age (Figure 2A, left column). Interestingly, we observed an negative emotionality X old/new interaction ($F(1,39) = 4.21$, $p < 0.05$,

rICC = 0.31) with negative stimuli eliciting more pronounced PD responses in particular for new stimuli during recognition tests (cf. Figure 2A, right column). There was no main effect of recognition delay (early or delayed tests), but a trend for an interaction of older adults showing less differentiation of negative and neutral stimuli in their PD responses in particular on delayed tests ($F(1,39) = 3.27$, $p = .08$, *ICC* = 0.28). All effects were robust to excluding error trials, except for the trend of older adults showing less differentiation of negative and neutral stimuli on the delayed test. This suggests that this somewhat weaker effect might be more sensitive to a reduction in trial numbers due to the exclusion of incorrect trials. Finally, negative and neutral stimuli on the encoding task showed an effect of negative emotionality similar to the new stimuli on the recognition tests (Supplementary material and Supplementary Figure 1).

The second four-way, repeated measures ANOVA examined the interaction between the old/new effect and the certainty of the old/new response as indicated on a rating scale (separated into high and low certainty trials, cf. Figure 2B) across the two recognition tests and age groups. Again, there was a main effect of recognition ($F(1,39) = 16.92$, $p < 0.05$, *rICC* = 0.55), with old stimuli eliciting larger PD responses (Figure 2B, thick lines). We also observed higher PD responses if participants were more certain in their old-new responses ($F(1,39) = 4.42$, $p < 0.05$, *rICC* = 0.32). Interestingly, there was also an old/new X certainty interaction ($F(1,39) = 6.87$, $p < 0.05$, *rICC* = 0.39), with a higher PD response for old stimuli detected with high certainty (Figure 2B, left column). This suggests that PD was modulated by the certainty with which stimuli could be classified as old and might therefore reflect a combination of task focus and strength of the memory trace. There were no age interactions or influence of early/delayed recognition tests. Note however, that

certainty ratings were overall higher on the first test, a difference in pupil diameter between early and delayed tests might thus have been captured in the main effect of higher PD with higher certainty. Again, all effects were robust to excluding error trials.

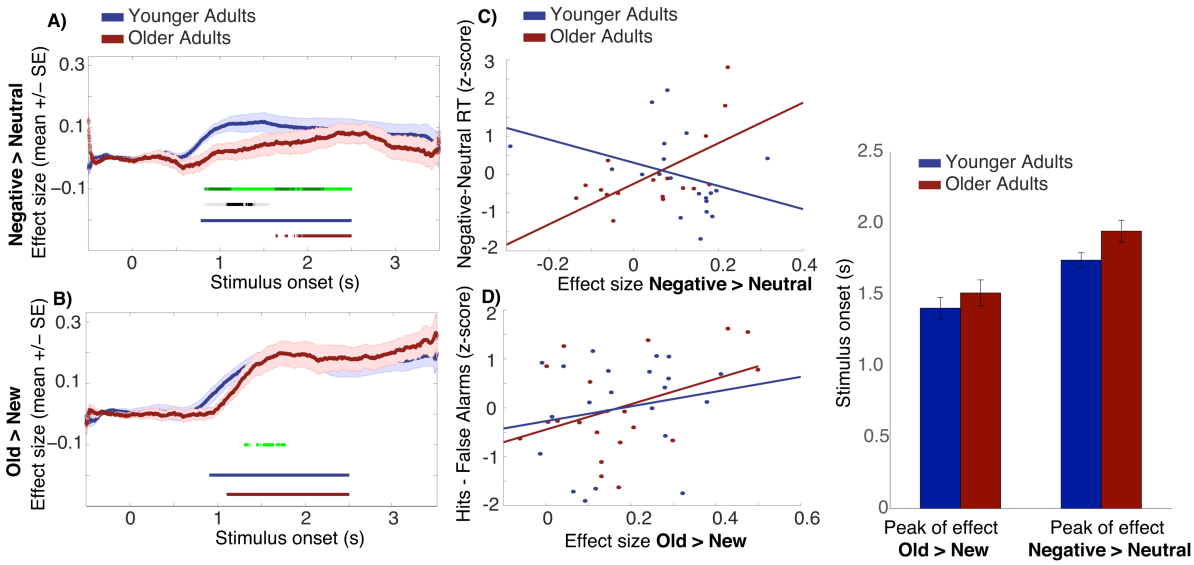


Figure 3. GLM results. A) Beta weights (effect sizes) for negative > neutral and B) old > new stimuli on PD data. Values above 0 indicate a positive effect of the regressor (e.g. PD response to negative stimuli larger than to neutral stimuli in 3A). A) and B) Lines indicate time points of significant effects based on permutation analyses within age groups (blue: within younger adults, red: within older adults). Light green lines indicate timepoints of significant relationship at $p < 0.10$ of PD to negative-neutral RTs for older adults (A) and to memory accuracy (Hit-False Alarms) for all participants (B) and dark green lines indicate significance at $p < 0.05$. In A) Grey line indicates age group difference of negative > neutral effect at $p < 0.10$ and black indicates age group difference at $p < 0.05$. C) Scatterplot shows positive correlation between effect size of the PD response to negative > neutral stimuli and negative – neutral RTs in time window indicated in A). D) Scatterplot shows positive correlation between effect size of PD response to old > new stimuli and memory accuracy in time window indicated in B). E) Bar chart showing the timing of the maximum value for negative emotionality and recognition effect in the time window 0.75 to 2s.

3.3.3 GLM analyses of negative emotionality and recognition effects in pupil data

To examine differences in the time course of negative emotionality and recognition effects, we used GLM analyses which allowed to test for the time course of the effects while controlling for inter-individual and age differences in the intercepts of PD responses. As outlined in the Methods section, analyses focused on the old/new and negative/neutral effects, while controlling for interaction effects as evident in repeated measures analyses detailed above. Furthermore, group-level GLMs tested for differences between age groups in these trial wise GLM effects. Also, inter-individual differences of the trial-wise old/new or negative/neutral effects were predicted by inter-individual differences in behavioral variables of interest to further validate the cognitive correlates of effects observed in PD (e.g. is there a larger PD recognition effect in individuals with better memory performance?).

3.3.3.1 Effect of negative emotionality on PD

As can be seen in Figure 3A, there is a clear positive effect of negative emotionality for both age groups, indicating that PD responses were larger for emotional than neutral stimuli, with the peak dilation occurring at around 1.46 seconds (cf. Figure 3E). As expected from the repeated MANOVA results, this effect was slightly larger in younger adults. Furthermore, when emotional RTs were added as a behavioural variable at the individual level, we observed a modulation of the increase in PD with emotional capture as evident in prolonged reaction times to negative stimuli (Figure 3C). Specifically, those older adults who showed a stronger modulation of PD with negative emotional stimuli show a greater slowing of RTs to negative stimuli (younger adults: r = -.36, p = .11, older adults: r = .63, p < 0.05, age difference in correlation coefficients, z = -3.09, p <0.05). Inter-individual differences in neuromodulation can be expressed in larger interindividual differences in cognitive

performance in a population which expresses reduced levels of neuromodulation (Hämmerer et al., 2013; Nagel et al., 2008). This has been attributed to a non-linear, u-shaped association between levels of neuromodulation and cognitive functions which would result in increasingly large effects on cognition as neuromodulation recedes from higher to medium levels (Lindenberger, 2008). Given the weaker PD response to emotional stimuli observed in older adults, one might speculate whether the stronger relationship between negative emotionality and RTs in older adults is indicative of an increased relevance of altered emotional reactivity to negative events in older adults.

3.3.3.2 Effect of recognition on PD

There is also a clear reliable effect of old stimuli eliciting larger PD in both age groups as evident in Figure 3B, with the peak dilation occurring at around 1.84 seconds (cf. Figure 3E). When memory accuracy was added as an individual-level variable we observed that the recognition PD effect was largest in individuals with higher memory accuracy (Figure 3D). This indicates that individuals with larger PD response to old stimuli are also more likely to correctly classify those stimuli as old (yas: r =.19, p = .39, oas: r = .41, p = .08, no reliable age difference in correlation coefficients, correlation collapsed across younger and older adults' z-scored data r = .31, p = .05). Finally, as can be seen in Figure 3E, the maximal peak of the negative emotionality effect was reliably earlier than the maximal peak of the recognition effect (Z = -2.58, p < .05), suggesting separable processes contributing to PD. All effects were robust to including memory errors as a separate logistic regressor (Kafkas and Montaldi, 2015).

3.3.4 Subsequent memory effects

We found no evidence for a difference in phasic PD during the encoding task for later remembered versus later forgotten stimuli ($F(1,32) = 0.91$, $p = 0.35$, cf Supplementary Figure 2). It has been argued that subsequent memory effects during incidental encoding tasks should be evident in changes in tonic PD that precede stimulus presentation and last throughout the trial as compared to phasic PD responses to stimulus properties. This is because memory is held to be guided by local fluctuations in PD rather than phasic PD responses to stimulus properties for incidental encoding. An examination of tonic PD at encoding showed greater tonic PD for later forgotten stimuli (Supplementary Figure 3, $F(1,32) = 5.80$, $p < 0.05$, *rICC* = 0.39), an effect that was greater for neutral stimuli (interaction negative emotionality X subsequent memory $F(1,32) = 5.62$, $p < 0.05$, , *rICC* = 0.39). Note this effect was robust to lowering trial numbers in the remembered condition to a level present for non-remembered stimuli and hence cannot be attributed to a higher likelihood for more extreme values in the later forgotten category.

4. Discussion

We present the first study, to our knowledge, investigating age-related differences in pupil diameter (PD) in an emotional memory task with negative events. Our aim was to assess age-related differences in processes attributed to noradrenergic modulation, namely arousal to negative stimuli, recognition memory and encoding (subsequent memory).

We found a reliable increase in PD to negative compared to neutral stimuli in both age groups, with the older adults showing a smaller PD response to negative stimuli (Figure 2A, 3A). An increased PD response to negative stimuli is a consistent

effect in the literature (Bradley et al., 2008; Bradley and Lang, 2015; Partala and Surakka, 2003; van Stegeren, 2008) and thought to reflect phasic bottom-up influences on the LC-NA system elicited by salient stimulus properties (Sara and Bouret, 2012). Furthermore, we add to this literature by showing that an increase in PD to negative stimuli is especially pronounced in older adults who are more susceptible to emotional stimulation, as evident in longer RTs to negative stimuli. The slowing in response to the negative stimuli in our task is in line with the idea of an 'emotional capture' effect, whereby the task-irrelevant emotional event diverts attention away from the relevant task (Hodsoll et al., 2011) of in-out classification. Finally, in further support that the emotional PD effect observed was driven by the negative emotional impact of the presented pictures, we observed an increased effect of stimulus emotionality on PD for novel pictures in particular (Bradley and Lang, 2015).

It should be noted that the observed reduced PD in older adults to negative emotional stimuli can not be necessarily assumed to generalize to positive stimuli. In contrast to their reduced reactions to negative events, previous studies demonstrated that older adults show generally slightly increased memory for positive events or increased startle responses to positive pictures, as compared to younger adults (Feng et al., 2011; Leigland et al., 2004; Mather, 2016). This 'positivity effect' in older adults has been suggested to reflect an adaptive emotion regulation strategy in old age which emphasizes positive events and dampens the impact of negative events (Mather and Carstensen, 2005). One might thus rather expect to see a slight increase in PD to positive stimuli in older adults.

Similarly, the current state of research does not allow to judge to what extent a reduction in PD to negative events is due to age differences in noradrenergic

modulation or age differences in emotional regulation. Studies in humans, rats as well as primates suggest that a reduction of noradrenergic modulation during ageing can be relevant for altered cognitive functions (Arnsten and Golman-Rakic, 1985; Clewett et al., 2016; Mann, 1983), including memory for negative emotional events (Luo et al., 2015). On the other hand, imaging studies which show increased prefrontal and decreased amygdala activation during negative events in older adults suggest that older adults actively downregulate their arousal response to negative events (see Mather, 2016 for a review). Of course, these two views do not have to be mutually exclusive. It is well conceivable that a tendency towards dampening negative exeriences is met or even facilitated by a reduced initial arousal response to negative stimuli. However, answering this question conclusively requires measures of LC integrity and function during a concurrent assessment of control areas and ideally also a pharmacological manipulation of noradrenergic levels in younger and older adults when examining age differences in memory. Finally, valenced information which is not arousing can also lead to improved memory formation (Kensinger, 2004), suggesting multiple cognitive as well as physiological pathways via which valenced material can affect memory.

We observed larger PD responses to old stimuli compared to new during recognition tests, (Heaver and Hutton, 2011; Otero et al., 2011; Võ et al., 2008) in both younger and older adults. Moreover, PD during recognition was larger in individuals with higher memory accuracy (Figure 3D). This profile is supportive evidence for the idea that increased PD to old stimuli is an index of memory strength (Papesh et al., 2012). Our observation that higher certainty ratings modulated this PD effect, such that PD responses to old stimuli were enhanced the more certain the

stimulus was classified as old (Figure 2B), is further evidence in support of the memory strength account.

Our results therefore speak against the idea that an old-new PD effect is related to increased effort in memory retrieval (Beatty and Kahneman, 1966). We can only speculate how an old-new PD response may relate to the physiology of LC neurons. One parsimonious account is that the phasic PD response to old, but not to new stimuli is akin to a consistently observed increased phasic responses of LC neurons to target stimuli in oddball tasks (Aston-Jones et al., 1994). This aspect of noradrenergic modulation has been shown to be independent of stimulus properties and stimulus frequency and driven by a combination of task set focus and stimulus type (Aston-Jones and Cohen, 2005). Unlike noradrenergic responses to arousing events, these phasic LC responses are considered to reflect attentional top-down modulation of salient events, as determined by a given task set (Sara and Bouret, 2012). In a recognition test, as in ours, the task set might require looking for old stimuli, which would then become the focus of evaluation. Old stimuli on a recognition test could therefore be understood to function similarly to targets in oddball tasks. The fact that we see faster response times to old compared to new stimuli, when classifying stimuli as old or new, might be seen as in line with a 'targetness' interpretation. If PD responses are driven by task focus and response certainty, one would expect faster reactions to stimuli in line with the current task focus.

In a second set of analyses, we used GLM analyses to disentangle the time course of PD effects related to emotional arousal and recognition independent of individual and age differences in baseline PD. Here emotional effects on PD emerged earlier than a PD recognition effect, a temporal dissociation that supports a

bottom-up/top-down distinction in these two effects. Specifically, emotional arousal to negative events would seem to be a faster and more stimulus driven process compared to a slower more top-down driven old/new recognition effect. Moreover, we also observed specific, but inversely expressed, reaction time effects on these two different aspects of PD modulation. In the case of emotional arousal, we observed an ensuing slowing in reaction times indicative of increased emotional capture by stimulus properties. Instead, stronger recognition PD effects were followed by faster reaction times for old responses, suggesting an effect of task focus on old-new classifications. These two aspects of PD responses most likely reflect different processes attributed to noradrenergic modulation, namely bottom-up emotional arousal and top-down 'target detection' (of old stimuli) based on memory traces.

In support of evidence for a reduction in noradrenergic modulation with ageing we found a reduced PD in older adults. Moreover, we observed a reduced emotional arousal PD effect in older adults to negative events while a recognition memory PD effect appeared comparable to that in younger adults. Subjective ratings of emotional involvement or memory certainty did not differ between age groups and thus did not account for observed age differences in these effects. We can only speculate as to why we find different age effects for negative emotionality and recognition-related PD effects. Under an assumed differentiation of bottom-up driven arousal PD effects and top-down driven recognition PD effects, it is conceivable that older adults compensate more easily for a lower baseline level of LC responses for the top-down driven effects than stimulus-driven arousal effect. We note here that a stronger emphasis of task set in easier task conditions is a well-known finding in older adults (Hämmerer et al., 2014).

We observed an expected decrease in memory performance in older adults on a delayed compared to immediate recognition test, confirming a well-established long-term memory deficit in older adults (Nyberg et al., 2012). Although we observed an altogether reduced PD modulation related to emotional arousal in older adults, and a trend for better memory for negative stimuli in younger adults on high certainty recognition trials, we did not observe a decrease in long-term memory in older adults for negative stimuli or a link between PD to negative stimuli and emotional memory. This might have been due to the fact that we used a comparatively easy memory paradigm, which could have prevented a stronger modulatory effect on memory performance. Indeed, age differences in memory for negative emotional events were larger in a task that put more emphasis on more difficult memory retrieval by asking participants to recall rather than recognise stimuli (Charles et al., 2003). However, given we were interested in assessing pupil diameter recordings during the memory tests, which required the use of a chinrest, a recall test could not be included easily in our task.

The present study did not explicitly address to what extent age differences in memory performance may also be related to time of day effects in cognitive capacities as well as changes in circadian patterns in noradrenergic modulation. Both, encoding in the morning and delayed recognition in the afternoon should provide more conservative estimates of performance as both, older as well as younger adults are not tested on their peak time of day (Hasher et al., 2005; May et al., 1993). However, older adults can be comparatively more affected by having to perform during non-optimal test times. It is thus possible that age differences in particular in memory performance on the delayed test might be in part due to effects of test time. Time of day effects in older adults have been shown to manifest

particular in increased false alarms during non-optimal test times (May et al., 1993). We did not observe a stronger age effect on false alarms for the delayed test. However, our study sample might have been too small to capture the effect reliably. Future studies should include short benchmark tests at the respective test time points to assess time of day effects in cases where test times can not be optimally adjusted for age comparative testing. In this way, performance differences due to test times can be assessed as covariates. Circadian patterns in noradrenergic modulation in humans as well as age differences therein are still relatively unexplored. Diurnal peaks in NA levels have been observed in the afternoon as well as late morning, in CSF NA levels and plasma NA levels, respectively (Prinz et al., 1979; Wood et al., 1976). There is some evidence for elevated plasma NA levels in older adults (Prinz et al., 1979; Ziegler et al., 1976). However, it currently remains unclear how these changes may relate to recognition memory, although more generally, elevated NA in CSF, independent of age, has been shown to negatively correlate with cognition (Wang et al., 2013).

Finally, we did not observe a subsequent memory effect on phasic PD at encoding. The existing literature on subsequent memory effects on PD is sparse and provides a mixed picture (see Goldinger and Papesh, 2012 for a review) where one view is that *phasic* PD effects are present on memory tasks with explicit instructions for encoding whereas incidental memory, as in our task, is better reflected in *tonic* PD differences during encoding (Kafkas and Montaldi, 2011; Papesh et al., 2012). Accordingly, phasic increases are thought to reflect the strength of an attentional focus or encoding effort in instructed conditions (Papesh et al., 2012). Whereas in the case of incidental memory, subsequent memory is assumed to be influenced by fluctuations in *tonic* PD preceding or concurrent to stimulus presentation. However,

to complicate matters further, theories on the influence of tonic noradrenergic modulation on cognitive performance offer contradicting predictions for incidental memory effects related to tonic PD levels. Some accounts associate increased tonic levels of noradrenaline with greater distractibility (Aston-Jones and Cohen, 2005), predicting poorer subsequent memory with *high* tonic PD levels during encoding. Accounts that associate increased levels of noradrenaline with increased neuronal gain, would predict poorer subsequent memory with *lower* tonic PD levels at encoding (Eldar et al., 2013). We observed higher tonic PD levels for later forgotten stimuli which supports an assumption of greater participant distractibility during encoding of later forgotten stimuli. The effect was considerably weaker for negative stimuli, suggesting that on trials with greater stimulus capture, subsequent memory is less driven by temporal fluctuations in attentional states. A recent theory suggests that the effects of noradrenergic modulation on incidental memory should be examined in a context of an interaction of attentional focus on the one hand and temporal modulations of tonic levels on the other hand (Mather et al., 2015; Sakaki et al., 2014). Future studies should address this assumption in more detail by introducing an additional modulation in attentional focus during stimulus encoding combined with PD measurements in the emotional memory task (Sakaki et al., 2014).

5. Conclusion

In summary, we found overall reduced PD responses in older adults as well as a reduced PD response reflecting emotional arousal to negative events. An intriguing possibility is that these reduced PD responses in older adults indicate age-related LC-NA system degeneration. Further research should investigate if this PD

effect is consistent across cognitive paradigms and can be altered by pharmacological manipulation. Furthermore, we segregated emotional and cognitive processes reflected in PD and provide evidence for a link to dissociable bottom-up and top-down modulation of LC function, respectively. However, despite the existence of causal evidence for a link between PD and LC activity (Joshi et al., 2016), PD measures only provide very indirect physiologic window into noradrenergic modulation during emotional memory for negative events. For instance, it is unclear whether the observed general decrease in PD and PD reactivity to altered luminance levels with age might be due to an altered muscular or neurochemical reactivity (Winn et al., 1994). Pharmacological studies (Kroes et al., 2010; Strange and Dolan, 2004) are needed to firmly establish causal neuromodulatory influences on PD and to explore whether these effects are exclusively due to the LC-NA system, or whether there are other neurotransmitters involved.

Funding sources

This work was supported by the German Academic Exchange Service (ID 91515956).

Disclosure statement

All authors report no actual or potential conflicts of interest.

References

Arnsten, A.F., Golman-Rakic, P.S., 1985. Catecholamines and cognitive decline in aged nonhuman primates. Ann. N. Y. Acad. Sci. 444, 218–234.

Arnsten, A.F.T., Goldman-Rakic, P.S., 1985. Alpha 2-adrenergic mechanisms in prefrontal cortex associated with cognitive decline in aged nonhuman primates. Science 230, 1273–1276.

Aston-Jones, G., Cohen, J.D., 2005. Adaptive gain and the role of the locus coeruleus–norepinephrine system in optimal performance. J. Comp. Neurol. 493, 99–110.

Aston-Jones, G., Rajkowski, J., Kubiak, P., Alexinsky, T., 1994. Locus coeruleus neurons in monkey are selectively activated by attended cues in a vigilance task. J. Neurosci. 14, 4467–4480.

Beatty, J., Kahneman, D., 1966. Pupillary changes in two memory tasks. Psychon. Sci. 5, 371–372.

Bradley, M.M., Greenwald, M.K., Petry, M.C., Lang, P.J., 1992. Remembering pictures: pleasure and arousal in memory. J. Exp. Psychol. Learn. Mem. Cogn. 18, 379.

Bradley, M.M., Lang, P.J., 2015. Memory, emotion, and pupil diameter: Repetition of natural scenes. Psychophysiology 52, 1186–1193. doi:10.1111/psyp.12442

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

Cahill, L., Haier, R.J., Fallon, J., Alkire, M.T., Tang, C., Keator, D., Wu, J., Mcgaugh, J.L., 1996. Amygdala activity at encoding correlated with long-term, free recall of emotional information. Proc. Natl. Acad. Sci. 93, 8016–8021.

Charles, S.T., Mather, M., Carstensen, L.L., 2003. Aging and emotional memory: the forgettable nature of negative images for older adults. J. Exp. Psychol. Gen. 132, 310.

Chen, Sara, S.J., 2007. Locus coeruleus activation by foot shock or electrical stimulation inhibits amygdala neurons. Neuroscience 144, 472–481.

Clewett, D.V., Lee, T.-H., Greening, S., Ponzio, A., Margalit, E., Mather, M., 2016. Neuromelanin marks the spot: identifying a locus coeruleus biomarker of cognitive reserve in healthy aging. Neurobiol. Aging 37, 117–126.

Cohen, M.X., 2014. Analyzing Neural Time Series Data. MIT Press.

Eldar, E., Cohen, J.D., Niv, Y., 2013. The effects of neural gain on attention and learning. Nat. Neurosci. 16, 1146–1153. doi:10.1038/nn.3428

Feng, M.C., Courtney, C.G., Mather, M., Dawson, M.E., Davison, G.C., 2011. Age-related affective modulation of the startle eyeblink response: older adults startle most when viewing positive pictures. Psychol. Aging 26, 752.

Garavan, H., Pendergrass, J.C., Ross, T.J., Stein, E.A., Risinger, R.C., 2001. Amygdala response to both positively and negatively valenced stimuli. Neuroreport 12, 2779–2783.

Goldinger, S.D., Papesh, M.H., 2012. Pupil Dilation Reflects the Creation and Retrieval of Memories. Curr. Dir. Psychol. Sci. 21, 90–95. doi:10.1177/0963721412436811

Hämmerer, D., Biele, G., Müller, V., Thiele, H., Nürnberg, P., Heekeren, H.R., Li, S.-C., 2013. Effects of PPP1R1B (DARPP-32) Polymorphism on Feedback-Related Brain Potentials Across the Life Span. Front. Psychol. 4. doi:10.3389/fpsyg.2013.00089

Hämmerer, D., Müller, V., Li, S.-C., 2014. Performance monitoring across the lifespan: Still maturing post-conflict regulation in children and declining task-set monitoring in

older adults. Neurosci. Biobehav. Rev. 46, 105–123. doi:10.1016/j.neubiorev.2014.06.008
Hasher, L., Goldstein, D., May, C.P., 2005. It's About Time: Circadian Rhythms, Memory, and Aging.
Heaver, B., Hutton, S.B., 2011. Keeping an eye on the truth? Pupil size changes associated with recognition memory. Memory 19, 398–405. doi:10.1080/09658211.2011.575788
Heuer, F., Reisberg, D., 1990. Vivid memories of emotional events: The accuracy of remembered minutiae. Mem. Cognit. 18, 496–506.
Hodsoll, S., Viding, E., Lavie, N., 2011. Attentional capture by irrelevant emotional distractor faces. Emotion 11, 346.
Jacques, P.L.S., Dolcos, F., Cabeza, R., 2009. Effects of aging on functional connectivity of the amygdala for subsequent memory of negative pictures a network analysis of functional magnetic resonance imaging data. Psychol. Sci. 20, 74–84.
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.
Kafkas, A., Montaldi, D., 2015. The pupillary response discriminates between subjective and objective familiarity and novelty. Psychophysiology 52, 1305–1316.
Kafkas, A., Montaldi, D., 2011. Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. Q. J. Exp. Psychol. 64, 1971–1989. doi:10.1080/17470218.2011.588335
Kensinger, E.A., 2004. Remembering emotional experiences: The contribution of valence and arousal. Rev. Neurosci. 15, 241–252.
Klukowski, G., Harley, C.W., 1994. Locus coeruleus activation induces perforant path-evoked population spike potentiation in the dentate gyrus of awake rat. Exp. Brain Res. 102, 165–170.
Kroes, M.C.W., Strange, B.A., Dolan, R.J., 2010. -Adrenergic Blockade during Memory Retrieval in Humans Evokes a Sustained Reduction of Declarative Emotional Memory Enhancement. J. Neurosci. 30, 3959–3963. doi:10.1523/JNEUROSCI.5469-09.2010
Lang, P., Bradley, M.M., 2007. The International Affective Picture System (IAPS) in the study of emotion and attention. Handb. Emot. Elicitation Assess. 29.
Lang, P.J., 1980. Behavioral treatment and bio-behavioral assessment: Computer applications.
Leigland, L.A., Schulz, L.E., Janowsky, J.S., 2004. Age related changes in emotional memory. Neurobiol. Aging 25, 1117–1124.
Lindenberger, U., 2008. Age-related decline in brain resources magnifies genetic effects on cognitive functioning. Front. Neurosci. 2, 234–244. doi:10.3389/neuro.01.039.2008
Luo, Y., Zhou, J., Li, M.-X., Wu, P.-F., Hu, Z.-L., Ni, L., Jin, Y., Chen, J.-G., Wang, F., 2015. Reversal of aging-related emotional memory deficits by norepinephrine via regulating the stability of surface AMPA receptors. Aging Cell 14, 170–179.
Mander, B.A., Rao, V., Lu, B., Saletin, J.M., Lindquist, J.R., Ancoli-Israel, S., Jagust, W., Walker, M.P., 2013. Prefrontal atrophy, disrupted NREM slow waves and impaired hippocampal-dependent memory in aging. Nat. Neurosci. 16, 357–364.
Mann, D.M., 1983. The locus coeruleus and its possible role in ageing and degenerative disease of the human central nervous system. Mech. Ageing Dev. 23, 73–94.
Mather, M., 2016. The affective neuroscience of aging. Annu. Rev. Psychol. 67, 213–238.

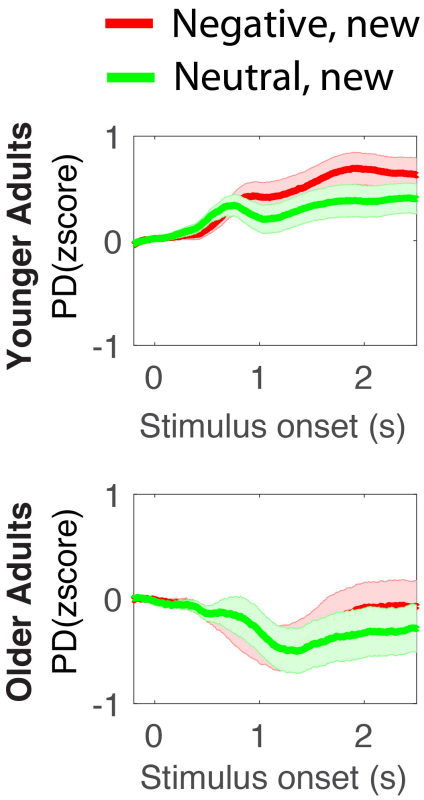
Mather, M., 2007. Emotional arousal and memory binding: An object-based framework. Perspect. Psychol. Sci. 2, 33–52.
Mather, M., Carstensen, L.L., 2005. Aging and motivated cognition: the positivity effect in attention and memory. Trends Cogn. Sci. 9, 496–502. doi:10.1016/j.tics.2005.08.005
Mather, M., Clewett, D., Sakaki, M., Harley, C.W., 2015. Norepinephrine ignites local hot spots of neuronal excitation: How arousal amplifies selectivity in perception and memory. Behav Brain Sci 1, 1–100.
Mather, M., Harley, C.W., 2016. The locus coeruleus: Essential for maintaining cognitive function and the aging brain. Trends Cogn. Sci. 20, 214–226.
May, C.P., Hasher, L., Stoltzfus, E.R., 1993. Optimal time of day and the magnitude of age differences in memory. Psychol. Sci. 4, 326–330.
Nagel, I.E., Chicherio, C., Li, S.-C., Von Oertzen, T., Sander, T., Villringer, A., Heekeren, H.R., Bäckman, L., Lindenberger, U., Nagel, I.E., Chicherio, C., Li, S.-C., Oertzen, T. von, Sander, T., Villringer, A., Heekeren, H.R., Bäckman, L., Lindenberger, U., 2008. Human aging magnifies genetic effects on executive functioning and working memory. Front. Hum. Neurosci. 2, 1. doi:10.3389/neuro.09.001.2008
Nassar, M.R., Rumsey, K.M., Wilson, R.C., Parikh, K., Heasly, B., Gold, J.I., 2012. Rational regulation of learning dynamics by pupil-linked arousal systems. Nat. Neurosci. 15, 1040–1046. doi:10.1038/nn.3130
Naveh-Benjamin, M., Hussain, Z., Guez, J., Bar-On, M., 2003. Adult age differences in episodic memory: Further support for an associative-deficit hypothesis. J. Exp. Psychol. Learn. Mem. Cogn. 29, 826–837. doi:10.1037/0278-7393.29.5.826
Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U., Bäckman, L., 2012. Memory aging and brain maintenance. Trends Cogn. Sci. 16, 292–305.
Otero, S.C., Weekes, B.S., Hutton, S.B., 2011. Pupil size changes during recognition memory. Psychophysiology 48, 1346–1353. doi:10.1111/j.1469-8986.2011.01217.x
Papesh, M.H., Goldinger, S.D., Hout, M.C., 2012. Memory strength and specificity revealed by pupillometry. Int. J. Psychophysiol. 83, 56–64. doi:10.1016/j.ijpsycho.2011.10.002
Partala, T., Surakka, V., 2003. Pupil size variation as an indication of affective processing. Int. J. Hum.-Comput. Stud., Applications of Affective Computing in Human-Computer Interaction 59, 185–198. doi:10.1016/S1071-5819(03)00017-X
Phelps, E.A., LaBar, K.S., Anderson, A.K., O'connor, K.J., Fulbright, R.K., Spencer, D.D., 1998. Specifying the contributions of the human amygdala to emotional memory: A case study. Neurocase 4, 527–540.
Prinz, P.N., Halter, J., Benedetti, C., Raskind, M., 1979. Circadian variation of plasma catecholamines in young and old men: relation to rapid eye movement and slow wave sleep. J. Clin. Endocrinol. Metab. 49, 300–304.
Sakaki, M., Fryer, K., Mather, M., 2014. Emotion strengthens high-priority memory traces but weakens low-priority memory traces. Psychol. Sci. 25, 387–395.
Samuels, E.R., Szabadi, E., 2008. 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, 235–253.
Sara, S.J., 2009. The locus coeruleus and noradrenergic modulation of cognition. Nat. Rev. Neurosci. 10, 211–223. doi:10.1038/nrn2573
Sara, S.J., Bouret, S., 2012. Orienting and Reorienting: The Locus Coeruleus Mediates Cognition through Arousal. Neuron 76, 130–141. doi:10.1016/j.neuron.2012.09.011

Scullin, M.K., 2013. Sleep, memory, and aging: the link between slow-wave sleep and episodic memory changes from younger to older adults. Psychol. Aging 28, 105.
Strange, B.A., Dolan, R.J., 2004. -Adrenergic modulation of emotional memory-evoked human amygdala and hippocampal responses. Proc. Natl. Acad. Sci. 101, 11454–11458. doi:10.1073/pnas.0404282101
van Stegeren, A.H., 2008. The role of the noradrenergic system in emotional memory. Acta Psychol. (Amst.), Learning, memory and psychopathology 127, 532–541. doi:10.1016/j.actpsy.2007.10.004
Vijayashankar, N., Brody, H., 1979. A quantitative study of the pigmented neurons in the nuclei locus coeruleus and subcoeruleus in man as related to aging. J. Neuropathol. Exp. Neurol. 38, 490–497.
Võ, M.L.-H., Jacobs, A.M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., Hutzler, F., 2008. The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect. Psychophysiology 45, 130–140. doi:10.1111/j.1469-8986.2007.00606.x
Wang, L.Y., Murphy, R.R., Hanscom, B., Li, G., Millard, S.P., Petrie, E.C., Galasko, D.R., Sikkema, C., Raskind, M.A., Wilkinson, C.W., others, 2013. Cerebrospinal fluid norepinephrine and cognition in subjects across the adult age span. Neurobiol. Aging 34, 2287–2292.
Windmann, S., Kutas, M., 2001. Electrophysiological correlates of emotion-induced recognition bias. J. Cogn. Neurosci. 13, 577–592.
Winn, B., Whitaker, D., Elliott, D.B., Phillips, N.J., 1994. Factors affecting light-adapted pupil size in normal human subjects. Invest. Ophthalmol. Vis. Sci. 35, 1132–1137.
Wood, J.H., ZIEGLER, M., LAKE, C., Ebert, M.H., 1976. Circadian rhythm in cerebrospinal fluid noradrenaline of man and monkey. Nature 264, 656–658.
Ziegler, M.G., Lake, C.R., Kopin, I.J., 1976. Plasma noradrenaline increases with age. Nature 261, 333–335. doi:10.1038/261333a0

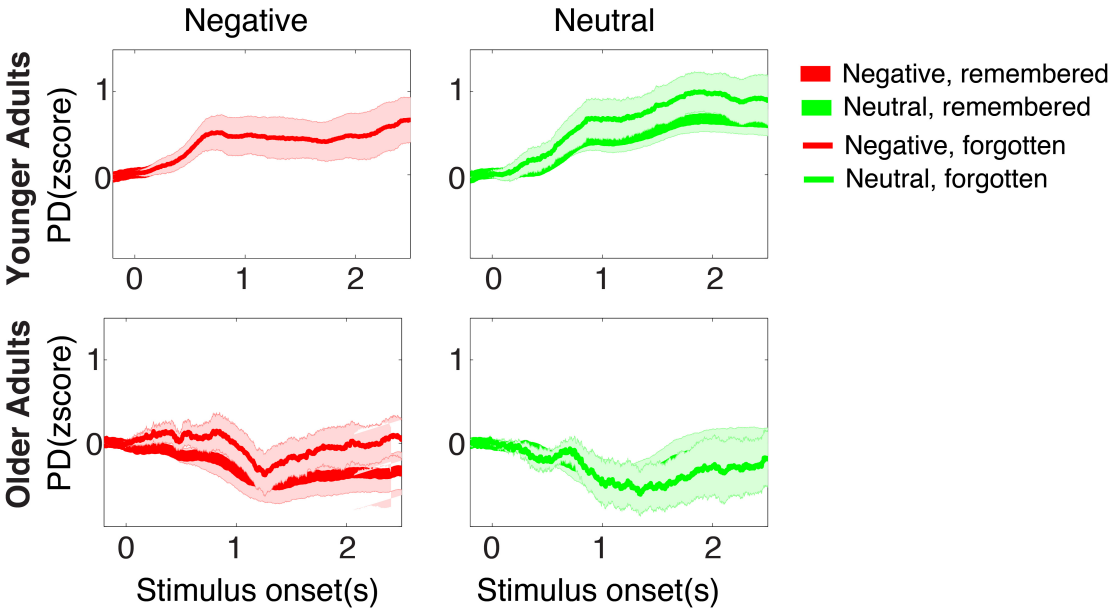
Supplementary Material

Effects of negative and neutral stimuli on PD in encoding task

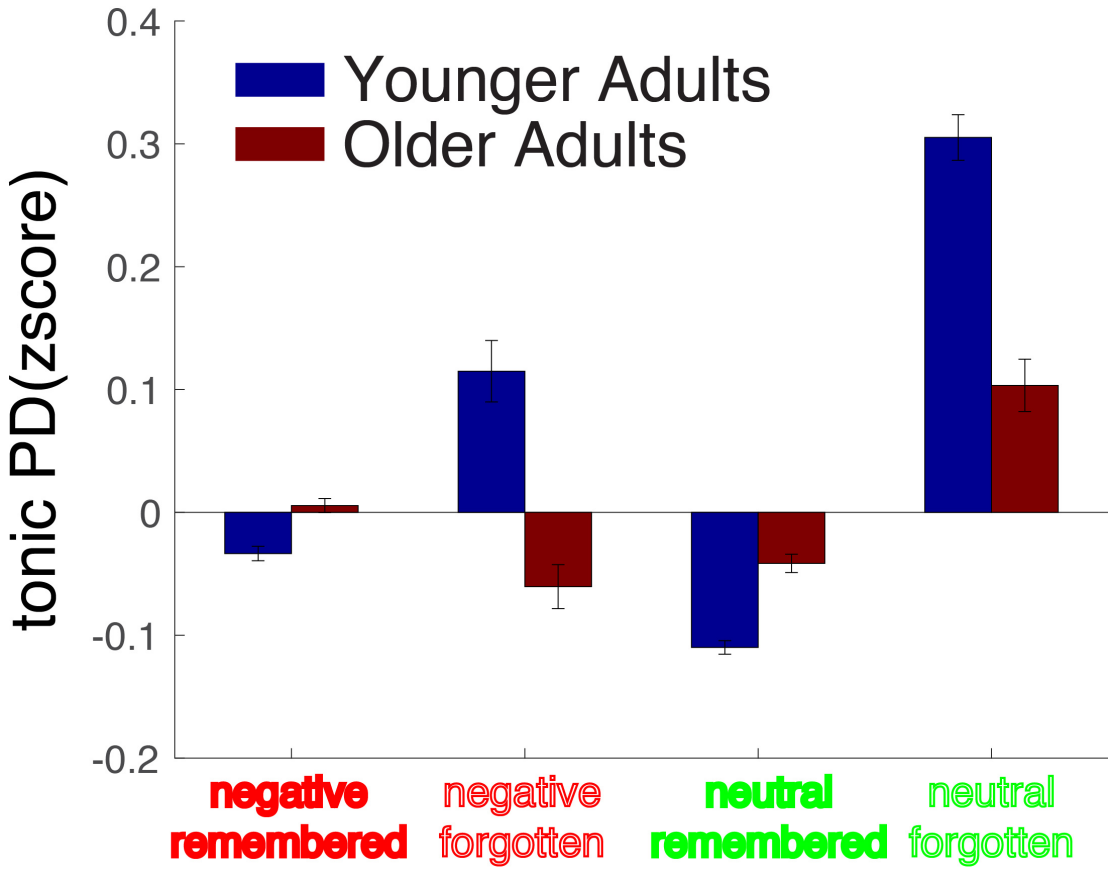
Negative and neutral stimuli in the encoding task should be expected to show similar PD effects as new negative and new neutral stimuli in the recognition tests. We compared negative and neutral stimuli on the encoding task with the new negative and neutral stimuli on the recognition tests (repeated measures ANOVA task-test x negative-neutral x agegroup). Inded, we did not observe reliable task-test interaction effects with stimulus emotionality or agegroups. Rather, main effects of negative emotionality ($F(1,39) = 22.82$, $p < 0.01$, $rICC = 0.61$) and the interaction agegroup x negative emotionality remained unaltered ($F(1,39) = 5.44$, $p < 0.05$, $rICC = 0.35$) by including the task PD data. There was however, an overall effect of PD being smaller during the encoding as compared to the recognition tasks ($F(1,39) = 15.29$, $p < 0.01$, $rICC = 0.53$), which might be due to the fact that no elevations of PD due to old stimuli could be expected on the encoding task.



Supplementary Figure 1. Mean PD responses related to negative emotionality during the incidental encoding task. As for the recognition tests, PD was increased for negative stimuli (red). Shaded error bars represent standard error (across individuals within age groups).



Supplementary Figure 2. Mean PD changes to negative/neutral effects on subsequent memory for both age groups. No reliable effects of subsequent memory are apparent in *phasic* changes in PD. Red indicates negative, green indicates neutral stimuli. Thin lines indicate forgotten stimuli, thick lines indicate lateron remembered stimuli. Shaded error bars represent standard errors.



Supplementary Figure 3. Mean *tonic* PD levels during lateron forgotten or remembered stimuli in the incidental encoding task in time interval -.2 to 2s during stimulus presentation. Lateron forgotten stimuli are associated with overall increased tonic PD levels. This effect is larger for neutral stimuli. Error bars represent standard errors.

Subsequent memory effects

As evident in supplementary Figures 2 and 3, we observed subsequent memory effects during the incidental encoding task in tonic pupil diameters only. Specifically, tonic pupil diameters were larger during lateron forgotten stimuli ($F(1,32) = 5.80$, $p < .05$, $rICC = .39$), and in particular larger for lateron forgotten neutral stimuli ($F(1,32) = 5.62$, $p < .05$, $rICC = .39$). This might suggest increased distractability and hence worse subsequent memory in particular on neutral trials. There were no main or interaction effects of age group differences (ANOVA agegroup x negative-neutral x old-new stimuli).